

**IJCSIS Vol. 14 No. 7, July 2016**  
**ISSN 1947-5500**

# **International Journal of Computer Science & Information Security**

**© IJCSIS PUBLICATION 2016**  
**Pennsylvania, USA**

*Indexed and technically co-sponsored by :*



Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

## CALL FOR PAPERS

### International Journal of Computer Science and Information Security (IJCSIS) January-December 2016 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

**Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.**

**Deadline:** see web site

**Notification:** see web site

**Revision:** see web site

**Publication:** see web site

Context-aware systems

Networking technologies

Security in network, systems, and applications

Evolutionary computation

Industrial systems

Evolutionary computation

Autonomic and autonomous systems

Bio-technologies

Knowledge data systems

Mobile and distance education

Intelligent techniques, logics and systems

Knowledge processing

Information technologies

Internet and web technologies

Digital information processing

Cognitive science and knowledge

Agent-based systems

Mobility and multimedia systems

Systems performance

Networking and telecommunications

Software development and deployment

Knowledge virtualization

Systems and networks on the chip

Knowledge for global defense

Information Systems [IS]

IPv6 Today - Technology and deployment

Modeling

Software Engineering

Optimization

Complexity

Natural Language Processing

Speech Synthesis

Data Mining

For more topics, please see web site <https://sites.google.com/site/ijcsis/>

arXiv.org Google scholar

SCIRUS  
search engine for science

ScientificCommons

Scribd

.docstoc  
find and share professional documents

BASE  
Bielefeld Academic Search Engine

CiteSeerX beta

uni-trier.de  
Computer Science  
Bibliography

DOAJ  
DIRECTORY OF  
OPEN ACCESS  
JOURNALS

EBSCO  
HOST

ProQuest

## Editorial Message from Editorial Board

*It is our great pleasure to present the July 2016 issue (Volume 14 Number 7 Part I, II & III) of the International Journal of Computer Science and Information Security (IJCSIS). High quality research, survey & review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and technology. It especially provides a platform for high-caliber academics, practitioners and PhD/Doctoral graduates to publish completed work and latest research outcomes. According to Google Scholar, up to now papers published in IJCSIS have been cited over 6611 times and the number is quickly increasing. This statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. There have been many improvements to the processing of papers; we have also witnessed a significant growth in interest through a higher number of submissions as well as through the breadth and quality of those submissions. IJCSIS is indexed in major academic/scientific databases and important repositories, such as: Google Scholar, Thomson Reuters, ArXiv, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, ResearchGate, Academia.edu and EBSCO among others.*

*On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers for their outstanding efforts to review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status.*

*"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." For further questions or other suggestions please do not hesitate to contact us at [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com).*

*A complete list of journals can be found at:  
<http://sites.google.com/site/ijcsis/>*

*IJCSIS Vol. 14, No. 7, July 2016 Edition*

*ISSN 1947-5500 © IJCSIS, USA.*

*Journal Indexed by (among others):*



**Open Access** This Journal is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source.



**Bibliographic Information**

ISSN: 1947-5500

Monthly publication (Regular Special Issues)

Commenced Publication since May 2009

**Editorial / Paper Submissions:**

**IJCSIS Managing Editor**

[\(ijcsiseditor@gmail.com\)](mailto:ijcsiseditor@gmail.com)

**Pennsylvania, USA**

**Tel: +1 412 390 5159**

# IJCSIS EDITORIAL BOARD

IJCSIS Editorial Board	IJCSIS Guest Editors / Associate Editors
<b>Dr. Shimon K. Modi</b> <a href="#">[Profile]</a> Director of Research BSPA Labs, Purdue University, USA	<b>Dr Riktesh Srivastava</b> <a href="#">[Profile]</a> Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE
<b>Professor Ying Yang, PhD.</b> <a href="#">[Profile]</a> Computer Science Department, Yale University, USA	<b>Dr. Jianguo Ding</b> <a href="#">[Profile]</a> Norwegian University of Science and Technology (NTNU), Norway
<b>Professor Hamid Reza Naji, PhD.</b> <a href="#">[Profile]</a> Department of Computer Engineering, Shahid Beheshti University, Tehran, Iran	<b>Dr. Naseer Alquraishi</b> <a href="#">[Profile]</a> University of Wasit, Iraq
<b>Professor Yong Li, PhD.</b> <a href="#">[Profile]</a> School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China	<b>Dr. Kai Cong</b> <a href="#">[Profile]</a> Intel Corporation, & Computer Science Department, Portland State University, USA
<b>Professor Mokhtar Beldjehem, PhD.</b> <a href="#">[Profile]</a> Sainte-Anne University, Halifax, NS, Canada	<b>Dr. Omar A. Alzubi</b> <a href="#">[Profile]</a> Al-Balqa Applied University (BAU), Jordan
<b>Professor Yousef Farhaoui, PhD.</b> Department of Computer Science, Moulay Ismail University, Morocco	<b>Dr. Jorge A. Ruiz-Vanoye</b> <a href="#">[Profile]</a> Universidad Autónoma del Estado de Morelos, Mexico
<b>Dr. Alex Pappachen James</b> <a href="#">[Profile]</a> Queensland Micro-nanotechnology center, Griffith University, Australia	<b>Prof. Ning Xu,</b> Wuhan University of Technology, China
<b>Professor Sanjay Jasola</b> <a href="#">[Profile]</a> Gautam Buddha University	<b>Dr . Bilal Alatas</b> <a href="#">[Profile]</a> Department of Software Engineering, Firat University, Turkey
<b>Dr. Siddhivinayak Kulkarni</b> <a href="#">[Profile]</a> University of Ballarat, Ballarat, Victoria, Australia	<b>Dr. Ioannis V. Koskosas,</b> University of Western Macedonia, Greece
<b>Dr. Reza Ebrahimi Atani</b> <a href="#">[Profile]</a> University of Guilan, Iran	<b>Dr Venu Kuthadi</b> <a href="#">[Profile]</a> University of Johannesburg, Johannesburg, RSA
<b>Dr. Dong Zhang</b> <a href="#">[Profile]</a> University of Central Florida, USA	<b>Dr. Zhihan Lv</b> <a href="#">[Profile]</a> Chinese Academy of Science, China
<b>Dr. Vahid Esmaeilzadeh</b> <a href="#">[Profile]</a> Iran University of Science and Technology	<b>Prof. Ghulam Qasim</b> <a href="#">[Profile]</a> University of Engineering and Technology, Peshawar, Pakistan
<b>Dr. Jiliang Zhang</b> <a href="#">[Profile]</a> Northeastern University, China	<b>Prof. Dr. Maqbool Uddin Shaikh</b> <a href="#">[Profile]</a> Preston University, Islamabad, Pakistan
<b>Dr. Jacek M. Czerniak</b> <a href="#">[Profile]</a> Casimir the Great University in Bydgoszcz, Poland	<b>Dr. Musa Peker</b> <a href="#">[Profile]</a> Faculty of Technology, Mugla Sitki Kocman University, Turkey
<b>Dr. Binh P. Nguyen</b> <a href="#">[Profile]</a> National University of Singapore	<b>Dr. Wencan Luo</b> <a href="#">[Profile]</a> University of Pittsburgh, US
<b>Professor Seifeidne Kadry</b> <a href="#">[Profile]</a> American University of the Middle East, Kuwait	<b>Dr. Ijaz Ali Shoukat</b> <a href="#">[Profile]</a> King Saud University, Saudi Arabia
<b>Dr. Riccardo Colella</b> <a href="#">[Profile]</a> University of Salento, Italy	<b>Dr. Yilun Shang</b> <a href="#">[Profile]</a> Tongji University, Shanghai, China
<b>Dr. Sedat Akylek</b> <a href="#">[Profile]</a> Ondokuz Mayis University, Turkey	<b>Dr. Sachin Kumar</b> <a href="#">[Profile]</a> Indian Institute of Technology (IIT) Roorkee

<b>Dr Basit Shahzad</b> [ <a href="#">Profile</a> ] King Saud University, Riyadh - Saudi Arabia	
<b>Dr. Sherzod Turaev</b> [ <a href="#">Profile</a> ] International Islamic University Malaysia	

ISSN 1947 5500 Copyright © IJCSIS, USA.

## TABLE OF CONTENTS

### 1. PaperID 30061602: HITH: A Novel Hybrid IP Traceback Approach for Heterogeneous Wireless Networks (pp. 1-11)

*Ikbel Daly, Faouzi Zarai, Lotfi Kamoun  
LETI laboratory, University of Sfax, Tunisia*

**Full Text:** PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

*Abstract* — Among the most critical attacks in wireless networks, there is the Denial of Service (DoS) attack. This threat is becoming increasingly vulnerable with heterogeneous wireless networks. To remedy this attack, it is fundamental to identify the source of attack by exploiting an IP traceback technique. There are two major categories of approaches; packet marking and packet logging. In packet marking, it is characterized by adding supplementary information to mark packets. This method moderates the problem of overhead but requires a large amount of packets to reconstruct the attack path. In packet logging, it is based on saving packets in digest tables. This approach enables the identification of attack source through a single packet but necessitates a huge storage space. In this paper, we propose a novel Hybrid IP Traceback for Heterogeneous wireless networks, which is called HITH (Hybrid IP Traceback for Heterogeneous wireless network). Our solution presents a precise IP traceback method with low overhead storage and improved accuracy. To evaluate the effectiveness and the feasibility of HITH approach, we use mathematical analysis and simulations. The results of a comparison with an existing solution in literature prove the capacity to trace a single IP packet while reducing storage overhead and data access time.

*Keywords*—Heterogeneous Wireless Network; Security, Hybrid IP traceback; Marking packet; Logging packet; Denial of Service attack.

### 2. PaperID 30061603: Developer Companion: A Framework to Produce Secure Web Applications (pp. 12-16)

*Mamdouh Alenezi, College of Computer & Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia*

*Yasir Javed, College of Computer & Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia*

**Full Text:** PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

*Abstract* — Software engineering and development is a very complex endeavor that contends with limited resources, potentially causing software to behave in an unexpected manner. Software developers often lack secure coding skills and it's a major reason behind development of insecure web applications. In this work, we propose a developer companion as an integrated framework that can be integrated to any IDE to educate and help developers produce more secure code. This framework can be adopted and can be made more intelligent by focusing on historical security flaws in the development team. Expert developers practices to overcome the security vulnerabilities.

*Keywords*—web applications, source code, security, static analysis

### 3. PaperID 30061608: A Review on Influential Factors of Information Privacy Concerns in the Use of Electronic Medical Record (pp. 17-27)

*Fiza Abdul Rahim, Department of Systems and Networking, College of Computer Science and Information Technology, Universiti Tenaga Nasional, Kajang, Malaysia*

*Zuraini Ismail and Ganthan Narayana Samy, Advanced Informatics School, Universiti Teknologi Malaysia Kuala Lumpur, Malaysia*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Healthcare organisations process massive amount of electronic medical records (EMR) utilised by their employees in supporting the organisation's services. Having given privileged access to sensitive and valuable patient information in the EMR, healthcare employees may cause privacy breaches, which may lead to detrimental consequences. Therefore, it is paramount to impose particular attention to healthcare employees' concerns on privacy in the use of EMR. The aim of this study is to identify the factors that influence information privacy concerns (IPC) in the use of EMR from healthcare employees' perspective. Systematic literature review (SLR) was conducted to identify articles pertinent to IPC. EBSCOhost, IEEE Explore, SAGE, MEDLINE, ScienceDirect, SpringerLink, Wiley Online Library and Taylor & Francis Online database were searched for reviews relevance articles. A total of 38 full articles were reviewed to extract the factors that influence the IPC. From the review, it revealed three influential factors, namely privacy risk, privacy awareness, and privacy policy. Furthermore, preliminary qualitative study has been done in this study helps in understanding the privacy practices, to validate the identified factors and relationships with IPC. This study may be of significance in providing useful information for healthcare organisations to understand IPC from their employees' perspective in ensuring the compliance towards privacy regulations.

*Keywords-information privacy concerns; electronic medical records; healthcare information system*

#### **4. PaperID 30061609: Moment Based Copy Move Forgery Detection Methods (pp. 28-35)**

*Khaled W. Mahmoud, Arwa Husien Abu Al-Rukab  
Computer Science Department, Zarqa University, Zarqa, Jordan*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Copy-Move forgery is one important type of image forgery. In this type of forgery, part of the image is copied and pasted into another position in the same image. This is done in order to hide an object inside the image by covering it. To detect this type of forgery, many methods (algorithms) were published. Each method has its own strong points and drawbacks. One of the most important aspects in detecting copy-move forgery is how to read the image; the features which used to represent the image. It is important to realize that having invariant features, will support the robustness of the detection method against different attacks that the copied parts may affected by. Different studies show that moment invariants are one of the best choices in image processing. In this paper, a brief introduction to moment is given and detection methods that are based on moments are illustrated and analyzed.

*Keywords- Forgery; Forensics; Moments; Zernike; Hu*

#### **5. PaperID 30061611: A New Approach to Predict Stock Big Data by combination of Neural Networks and Harmony Search Algorithm (pp. 36-44)**

*Kiarash Aghakhani, Young Researchers and Elite Club, Arak Branch, Islamic Azad University, Arak, Iran  
Abbas Karimi, Department of Computer Engineering, Arak Branch, Islamic Azad University, Arak, Iran*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Nowadays, due to the vast volume and complicated interrelation of daily stock data, the prediction of the stock price is very crucial in order to earn the highest profit of the shareholder's investment is the main target. For these purposes, data mining techniques such as correlation analysis and prediction, and likewise data modeling and pattern recognition are utilized. Since the stock market is a chaotic and nonlinear system, the exact prediction of the massive data exchange, requires intelligent and advanced tools such as neural networks and meta-heuristic algorithms. This purpose method is conducted on the stock data of IBM, Apple and Dell companies and gold price in the global

market. Moreover, the prediction error is compared with results of ARIMA, ANN, HMM ANN-ICA, ANN-GA, ANN-PSO, HMM-Fuzzy, HMM-ANN-GA methods. The comparison indicates that the purposed method provides remarkable improvement in the prediction performance.

*Keywords- Data mining, Big Data, Predict Stock Price, Artificial Neural Network, Harmony Search Algorithm*

**6. PaperID 30061622: Effective Techniques for Reduction of Impulse, Gaussian and Speckle Noises (pp. 45-51)**

*Md. Golam Moazzam, Tanzila Rahman, Mohammad Shorif Uddin*

*Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Noise is a common phenomenon and usually introduce during acquisition and transmission of images. Reduction and removal of noise from digital images is a prerequisite for subsequent analysis and recognition. Hence, nowadays it becomes an active area of research. Different types of noise can be added with digital images, such as impulse noise, Gaussian noise, speckle noise and so on. Impulse noise can be defined by replacing the intensity of an image point with random value of either higher-end or lower-end. Gaussian noise can be described by randomly adding values with zero mean maintaining Gaussian distribution to the intensities of image points. With a view to eradicate of these noises in this paper we briefly describe some important noise reduction methods. On the other hand speckle is a multiplicative noise that usually occurs in SAR and ultrasound images. For effective reduction of this noise here we have modified an existing technique and perform experimentation to confirm its superiority.

*Keywords- Gaussian noise, median filter, fuzzy filter, frost filter, mean square error.*

**7. PaperID 30061623: Performance Evaluation of Femtocell Based LTE Network under the Concept of Cross-layer Optimization (pp. 52-60)**

*Jesmin Akhter, Associate Professor, Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh*

*Md. Imdadul Islam, Professor, Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh*

*M. R. Amin, Member IEEE, Professor, Electronics and Communications Engineering, East West University, Dhaka, Bangladesh*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — To achieve end-to-end maximum throughput, the wireless Internet access requires (a) sufficient SNR at evolved Node-Bs and UEs (user equipment) at physical layer and (b) congestion control algorithm in determining appropriate window size at transport layer. Considering above, the paper deals with both the layers for Femto cellular LTE network and relates the fading parameters of physical layer and congestion parameters of transport layer. One of the promising approaches of 4G mobile cellular network is to incorporate Femto cell inside macro cells to get access of a MS (Mobile Station) within few meters. This approach is adopted to combat the small scale fading of wireless link so that a MS can achieve optimum throughput, otherwise huge capacity of a mobile cellular network is lost under fading environment. This paper deals with the relation among outage probability, density of Femto cell, threshold link capacity, threshold SNR (signal to noise ratio) and mean congestion window size under fading environment. We found that Nakagami-m environment provides better result compare to Rayleigh case (because of several direct link in Nakagami-m environment) at the same time path loss exponent is a vital factor for such network. Next we analyze the performance of end-to-end TCP (Transmission Control Protocol) link under the concept of congestion window control with newly developed state transition chain. The impact of fading parameter on outage probability, mean transmission rates, mean window size and throughput are analyzed explicitly for such network.

*Keywords-component; 4G mobile, small scale fading, optimum throughput, outage probability, Nakagami-m fading and moment generation function.*

**8. PaperID 30061625: Data Mining Approach to Extract the Interdependency among Different Attributes of Cardiac Patients (pp. 61-68)**

*Rao Muzamal Liaqat, Bilal Mehboob, Nazar Abbas Saqib, Muazzam A Khan  
College of E&ME, National University of Sciences and Technology, Islamabad, Pakistan*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract -* Nowadays we are surrounding with large data related to patient history, test results and reports. Usually, doctors diagnose the disease on the basis of recommended tests. A final recommendation about patient health may involve a lot of factors including patients test results and doctor experience. In this paper, we will use the data mining approach to extract the dependency among different tests recommended by practitioners, as well as relations of important parameters in cardiac patient's dataset. In this paper, we have used ID3, CHAID, Random Tree, Random Forest, Decision Tree and Decision Stump to extract the interdependency among different attributes in cardiac patients. We have performed the comparative analysis of these algorithms; according to analysis, ID3 give the best result. In this paper we have used the dataset provided by AFIC (Armed Force Institute of Cardiology), our dataset consists of 1500 records along with 36 attributes.

*Keywords:* *Data Mining; Cardiac Patients; Supervised Learning; DT (Decision Tree), ID3*

**9. PaperID 30061626: Predicting Student Performance and Risk Analysis by Using Data Mining Approach (pp. 69-76)**

*Bilal Mehboob (1,2), Rao Muzamal Liaqat (1), Nazar Abbas Saqib (1)  
(1) Department Of Computer Engineering, College of EME, National University of Sciences and Technology (NUST), H-12 Islamabad, Pakistan;  
(2) se@ts global, 54 Orchard, Paragon city Lahore, Pakistan*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract -* Today we are surrounding with large data related to student performance (class participation, attendance, pre student history, quiz result, subject dependency, student CGPA till to final semester). In this paper we will evaluate the reason of student failure basis on the previous data, predict the risk of failure for next course so that students may be mentally prepare for offered course as well dependency level of the course. In engineering it is common practice if a student doesn't know about the basic course he/she can't perform well in advance courses of same scopes. In this paper we will back trace the failure cause with the help of six algorithms. This work will also help out to estimate the risk in early phase, which can help the teachers to design an effective planning for the students who are at risk. We have used the six algorithms for prediction and risk analysis and ID3 algorithm gives the best results as compared to other five algorithms. In this paper we have used the data set of CEME, NUST. Our dataset consists of 450 records extracted from five degrees (DE-29, DE-30, DE-31, DE-32, and DE-33).

*Keywords:* *Data Mining, ID3, Risk, Performance Prediction*

**10. PaperID 30061629: Android-Based Health Care Management System (pp. 77-87)**

*Fazal Masud Kundi, Institute of Computing and Information Technology, Gomal University, D.I.Khan, Pakistan  
Ammara Habib, Institute of Computing and Information Technology, Gomal University, D.I. Khan, Pakistan  
Anam Habib, Institute of Computing and Information Technology, Gomal University, D.I. Khan, Pakistan*

*Muhammad Zubair Asghar, Institute of Computing and Information Technology, Gomal University, D.I. Khan, Pakistan*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract** — Objective: The primary goal of this study is to develop an android-based healthcare application, which can assist the users to monitor their health-related conditions for improving their health. Methods: The application is developed using android operating system environment. A Visual block programming language, namely MIT App Inventor is used to develop the system. The modification is presented as: (1) integration of different modules and their offline usage, (2) history facility, (3) user friendly. The qualitative method is used to study the objective. Findings: The research paper depicts a brief study of existing systems and the new development that has made in the application and also it is better in the manner that it works as a guide to control risk factors. The descriptive analysis point outs that the application is effective to deal with health related issue. Applications/Improvement: Integration of modules is performed on the android platform of different applications that are located on different websites, the storage facility is added by using Tiny DB, guidance in the form of charts and text is provided to the users. Such features are not provided in the previous work.

**Keywords**—*Health Care; App Inventor; Android; Diabetes; Target Heart Rate.*

**11. PaperID 30061631: Genetic Algorithm Based Novel Approach for Load Balancing Problem in Cloud Environment (pp. 88-93)**

*Dr. Surjeet Dalal, Shilpa Kukreja*

*Department of Computer Science & Engineering, SRM University, Haryana, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract** — Cloud computing has come up as one of the most promising & reliable technologies in the IT sector. However presently there exists a major issue of load balancing in the cloud computing environment. This paper consists of a solution for optimizing the load using genetic algorithm. Genetic algorithm which follows the evolutionary mechanism is able to develop a solution close to optimal solution. The proposed algorithm is developed by merging two existing algorithms by considering cost value as the fitness function. The workload is balanced by the considering the combination of both the load percentage and cost value of the resources. Allocation of resources is performed by taking the best fit value and reducing the response time and overall cost. Simulation results are shown using the cloud analyst simulator.

**Keywords**- *Cloud computing; genetic algorithm; load balancing; fitness value; load percentage;*

**12. PaperID 30061632: Real Time Algorithm for the Smart Home Automation Based on the Internet of Things (pp. 94-99)**

*Salman Ali Khan (1), Arhsad Farhad (2), Muhammad Ibrar (1), Muhammad Arif (1)*

*(1) Department of computer science, City University of Science & Information Technology, Peshawar, Pakistan*

*(2) Department of computer science, COMSATS Institute of Information Technology, Sahiwal, Pakistan*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract**—Internet of Things (IoT) is enabled by the advancements in the latest technologies including sensors, Radio Frequency Identification (RFIDS), internet protocols and communication technologies. The most important premises of IoT is to connect devices and sensors without human intervention. The proposed smart home automation system differs from other systems by allowing the user to access and operate the system from anywhere around the world through internet along with decision controls according to the needs. In this paper, we propose an algorithm for smart home automation system based on IoT using sensor nodes which are directly connected to Arduino Nano. The

algorithm perform some basic local functions such as; Turning ON/OFF the lights based on the motion sensor and generating the alarm based on the gas sensor. In the proposed algorithm the Arduino Mega is connected to the internet using Wi-Fi module to monitor the power consumption of different home appliances and can be controlled from anywhere on the internet. The objective of the proposed system is to provide a low cost and efficient solution for home automation system by using IoT. Results show that the proposed system is able handle all controlling and monitoring function of home.

*Keywords* — Smart home system, Internet of Things, Motion sensor, Gas sensor, Alarm system

### **13. PaperID 30061633: Security of Dynamic and Multipoint Virtual Private Network (pp. 100-106)**

*Ayoub BAHNASSE, Faculty of Sciences, University Chouaib DOUKALI, El Jadida, Morocco  
Najib EL KAMOUN, Faculty of Sciences, University Chouaib DOUKALI, El Jadida, Morocco*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — Nowadays, the solutions of virtualizing network infrastructure have become one of the most preoccupations of small, medium and large enterprises. These solutions make the extension of companies' sites possible and easier with a transparent and flexible manner. These solutions allow also the remote access to personal data, stored on several distributed sites, securely. Dynamic and Multipoint Virtual Private Network, stands for DMVPN, is considered as a main component of these solutions, this technology involves a suite of protocols for a smooth functioning, such as : IPsec, mGRE and NHRP. Nonetheless, even the considerable security and modularity level of DMVPN solution, this latter suffers from several security issues linked to each components' protocol, which might threaten availability, confidentiality, authentication and integrity of communications. In this article, we will discuss the key vulnerabilities related to DMVPN technology and the possible countermeasures.

*Index Terms* — DMVPN, Security, Vulnerability, IPsec, NHRP, mGRE.

### **14. PaperID 30061635: Predominant Factors Influencing Software Effort Estimation (pp. 107-110)**

*Sumeet Kaur Sehra (1), Yadwinder Singh Brar (2), Navdeep Kaur (3)  
(1) Research Scholar, I.K.G. Punjab Technical University, Jalandhar, Punjab, India  
(2) Professor, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India  
(3) Associate Professor, Shri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — Software effort estimation is a crucial task affecting the success of a software project. Inaccurate estimates can lead to incomplete, over-budgeted and failed projects. Accurate estimate of software development effort, which has always been a challenge for both the software industry and academia. Many models have been developed and validated by researchers to estimate the effort. But none of the models are successful for all types of projects and every type of environment. The reason is the prevalence of some fundamental issues which have a negative influence on the effort estimation process. In this paper, some of the issues affecting software effort estimation have been discussed.

*Index Terms* — Software Effort Estimation, Estimator, Factors, Environment, Dataset

### **15. PaperID 30061640: Development of an Autopsy Forensics Module for Cortana Artifacts Analysis (pp. 111-121)**

*Bernard Allen Sabernick III  
Department of Computing Security, Rochester Institute of Technology, Rochester, NY USA*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Forensic tools are a critical component of a forensic investigators job. As new features are added in operating systems, these tools need to adapt and be updated to analyze these new features. Microsoft recently released its Windows 10 operating system with a new voice activated personal digital assistant called Cortana. Cortana is capable of storing information about a user which could be used as evidence in criminal cases. Using the open source forensic tool Autopsy, this information is currently not being gathered in an effective manner. In order to address this problem, this paper proposes enhancements to the Autopsy tool to allow forensic investigators to collect the needed information about Cortana and analyze it more quickly.

*Keywords:* Digital Forensics, Windows 10, Cortana, Autopsy, Development

**16. PaperID 30061644: An Effort Estimation Approach for Agile Software Development using Fireworks Algorithm Optimized Neural Network (pp. 122-130)**

*Thanh Tung Khuat, My Hanh Le*

*DATIC Laboratory, IT Faculty, University of Science and Technology – The University of Danang, Vietnam*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Software effort estimation is one of the most critical steps in the software development process. The success or failure of projects relies greatly on the accuracy of effort estimation and schedule results. Agile software development process has become prevalent in the industry and replacing the conventional approaches of software development. Nevertheless, the question of accurate estimation of effort for this novel method has still been a challenging problem with regard to researchers and practitioners. This study aims to propose a novel method to ameliorate the accuracy of agile software effort prediction process using Artificial Neural Network (ANN) optimized by Fireworks Algorithm (FWA). The performance of the proposed approach is compared to the various types of neural networks and the regression model. In addition, the role of Fireworks Algorithm in optimizing the weights and biases of the ANN is also compared with other optimization algorithms.

*Index Terms*— Software Effort Estimation, Agile Software Development, User Story, Artificial Neural Network, Fireworks Algorithm, Levenberg-Marquardt.

**17. PaperID 30061645: Critical Evaluation of Maintainability Parameters using Code Metrics (pp. 131-136)**

*Bhawana Mathur, Manju Kaushik*

*Dept. of CS&E, JECRC University, Jaipur, India*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Software maintenance is a noteworthy feature of software development life cycle, hence earlier approximation of work for maintainability plays a vibrant role. The C sharp small programs are programmed in console applications like a reverse number & check if it a palindrome, check whether given string is a palindrome or not, and so many. The 40 programs on Visual Studio 2012 are compiled and analyze the code metrics. After analysis code metrics parameters like MI, DIT, LOC, class coupling and cyclomatic complexity results are found. Existing approaches for maintainability estimation are the correlation between code metrics like maintainability index, cyclomatic complexity, Depth in Inheritance, class coupling, Line of Code in the experiments. On the off chance that the coefficient quality is in the negative shift, before which it implies the relationship between the variables is adversely associated, or as one worth increases, the different declines ,like Depth of Inheritance between cyclomatic complexity , Depth of Inheritance between class coupling, Lines of Code between Depth of Inheritance, Maintainability Index between Lines of Code. This paper likewise gives various understanding to the viable utilization of Maintainability Index. To reiterate, stay to remark the source codes yet don't put a lot of trust in remarks to enhance maintainability.

*Keywords- Code Metrics; Maintenance; Maintainability Index; Lines of Code; Halstead Volume; Cyclomatic Complexity; Depth of Inheritance; Class coupling; smells; Lines of Code (LOC).*

**18. PaperID 30061646: An efficient (n,n) - Threshold Secret Sharing Scheme using on FAPKC4 and Hash Function (pp. 137-140)**

*Ali Saeidi Rashkolia, Department of Mathematics, Graduate University of Advanced Technology, Kerman, Iran  
Mohammad Mahdi Zahedi, Department of Mathematics, Graduate University of Advanced Technology, Kerman, Iran  
Masoud Hadian Dehkordi, School of Mathematics, Iran University of Science and Technology, Tehran, Iran*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* —The main purpose of this paper is to give a (n,n) - thresholdsecret sharing scheme based on the inversion of weakly invertible finite automata. It is varifiable, practical it does not face with time-spending computation such as "discrete logarithm" moreover both combiner and participants can investigate the validity of exchanged data. Security can be reduced to the generalization of finite automata public key cryptosystem FAPKC4, because the secret is encrypted by using it. This is a strong property since the FAPKC4 is believed to be secure.

*Keywords - component; finite automaton, secret sharing, weakly invertible, weak inverse, hash function, public key crypyosystem.*

**19. PaperID 30061652: Trust and Risk Based Approach for the Management of Dynamic Break-Glass Access in the Cloud Federation Environments (pp. 141-152)**

*Manoj V. Thomas, K. Chandrasekaran  
Department of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal, Mangalore, Karnataka, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — Personal Health Records (PHRs) are highly sensitive; and hence proper access control mechanisms need to be enforced in dealing with access requests involving such data. With the emergence of the inter-cloud computing, the PHR service providers can combine different services from multiple Cloud Service Providers (CSPs) into a single service or application for advantages such as better quality of health care and reduced health care cost. In this combined service delivery model, patients' data are stored in the CSPs in cloud federation, and hence the effective access control mechanism should be enforced by the CSPs. During emergency situations, availability of the healthcare data is more important than confidentiality, and hence relevant medical data should be made available to the concerned people irrespective of the employed access control model. But, how to identify the legitimate access request is an issue to be solved in this domain. In this paper, we are proposing a trust and risk-based mechanism for finding the legitimacy of the emergency access requests in the cloud federation environment. The proposed mechanism calculates the risk involved in the access request and takes a suitable access decision by calculating the trust value of the user. The workflow of the proposed approach is also discussed. We have implemented the proposed approach using the CloudSim toolkit, and the analysis of the results is also given. The analysis shows that the proposed approach is efficient in dealing with the break-glass access requests in the cloud federation environment.

*Index Terms — authorization; break-glass; cloud federation; emergency; PHR; risk; trust.*

**20. PaperID 30061653: Cross Slot Microstrip Patch Antenna with Dual Polarization (pp. 153-157)**

*Nazia Hasan, ECE Deptt, UTU Dehradun, Uttarakhand Technical University Dehradun, Dehradun, India  
Dr. S. C. Gupta, ECE Deptt, DIT Dehradun, Dehradun Institute of Technology Dehradun, Dehradun, India*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — A single-feed circular microstrip patch antenna having reconfigurable polarization capability is proposed. This proposed antenna has a very simple structure; two slots are created at an angle of 45 degree and 135 degree in the shape of X at the centre of patch antenna, and one Micro Electromechanical switch is inserted at the centre of the created slot to alter the polarization of antenna. When switch is in ON position, the polarization will be linear and if switch is OFF, polarization will be circular. Polarization will be confirmed with the help of axial ratio plot. Microstrip feed line is used in this structure.

*Keywords*—Circular Polarization, microstrip patch, Xshape slot, MEM switch.

**21. PaperID 30061660: Layer Based Log Analysis for Enhancing Security of Enterprise Datacenter (pp. 158-164)**

*Samuel Getachew Tadesse, Department of Computer Science, Haramaya University, Haramaya, Ethiopia  
Dejene Ejigu Dedefa, Department of Computer Science, Addis Ababa University, Addis Ababa, Ethiopia*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — The paper explores how log analysis is key for enhancing network security of enterprises. Now a days the issues of security becomes great concern because of the interconnection among organizations with WWW. Routine log analysis is beneficial for identifying security incidents, policy violations, fraudulent activity, and operational problems. Security is a means for assuring health and help to identify attacks. Enterprises must perform log analysis to discover different attacks by considering heterogeneous log records. We used multilevel log analysis to identify attacks found at different layers of data center through scrutinizing log events of various network devices, applications and others. Thus, to discover different attacks considering heterogeneous log records are basis for analysis. In our work log records were organized together into common format and analyzed based on their features. In central engine clustering and correlation are core of log analyzer that work together with attack knowledge base to identify attacks. Clustering algorithms such as Expectation Maximization, K-means were used to determine the number of clusters and filter events based on filtering threshold respectively. On the other hands, correlation finds a relationship or association among log events and generates new attack definitions. Finally, we evaluated log analyzer prototype of the proposed system and obtained an encouraging result with average precision of SOM#34 and AAU is 84.37 and 90.01 respectively. Further study and implementation of log analysis can significantly enhance data center security of enterprises. Generally, this paper demonstrates the application of log analysis for enhancing security of enterprise data center and our proposed solution will be discussed.

*Keywords*—Log File; Log Analysis; Layered approach; Attack Identification; Data Center; Network Security

**22. PaperID 30061662: Solving Bi-objective Two-Dimensional Rectangle Packing Problem using Binary Cuckoo Search (pp. 165-169)**

*Amandeep Kaur Virk, Dr. Kawaljeet Singh  
SGGSWU, Fatehgarh Sahib*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* - The work presented here optimizes the rectangular packing problem in which rectangular items are packed on a rectangular stock sheet. Our objective is to maximize the utilization of the rectangular sheet and to minimize the number of non-guillotine cuts required to cut various pieces. Binary version of cuckoo search algorithm has been used to solve this discrete problem. A series of computational experiments have been conducted to evaluate the performance of the new cuckoo search metaheuristic technique. It appears from the computational analysis that the cuckoo search algorithm is able to give good solutions.

*Keywords:* Nesting Problem, Cuckoo Search, Multi-objective optimization, Non-guillotine cutting.

**23. PaperID 30061663: Optimising Mobile Adhoc Energy demands with Probabilistic Max Drift and Longevity Scheme for realizing Green Campus status in Higher Education Institutions (pp. 170-175)**

*Kesava Rao Alla, Soong Der Chen*

*Department of Graphics and Multimedia, College of Information Technology, University Tenaga Nasional, Malaysia*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Green computing is an approach of optimizing the usage of Computer systems without compromising on system output and performance. As each hardware and software component contributes to the overall system energy requirement, this research is about presenting an investigation on minimizing the energy demand over a network in a Higher Education Institution (HEI) which can contribute towards achieving and maintaining a green campus environment. The principal contribution of this research is an amplification method that uses Probabilistic Max Drift in reducing Mobile Ad hoc energy demands and improving the durability. Comparison of performance of amplified networks was simulated using Java with their initial layouts. Furthermore, extended probabilistic method is added to Max Drift Scheme, and the effects are assessed by comparing on network lifetime when combined with network amplification. This system uses bi-connectivity directly to improve network lifetime, and also it introduces the network maintenance improvement to promote green environment. The results show that the energy consumption was reduced to a significant level of 17% when tested for one of the HEI, which thus plays a key role in fulfilling the green computing requirements and provides a pathway to realising the green campus. With these findings, it is envisaged that this system with less network resource usage could very well be applicable for any other HEI or any other environment with a demand for higher volumes of network communication resources.

*Index Terms*—Adhoc, Mobile, Energy Demands, Green Computing, Institution or University

**24. PaperID 30061665: NeTMids: Neighbor Node Trust Management Based Anomaly Intrusion Detection System for Wireless Sensor Networks (pp. 176-183)**

*Syed Muhammad Sajjad, Muhammad Yousaf*

*Riphah Institute of Systems Engineering, Riphah International University, Islamabad, Pakistan*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Timely detection of anomalous activities in Wireless Sensor Network is critical for the smooth working of the network. This paper presents an intrusion detection technique based on the calculation of trust of the neighboring nodes. In the proposed IDS, each node observes the trust level of its neighboring nodes. Based on these trust values, neighboring nodes may be declared as trustworthy, risky or malicious. Trustworthy nodes are recommended to the forwarding engine for packet forwarding purposes. The proposed scheme successfully detects Hello Flood Attack, Jamming Attack and Selective Forwarding Attack by analyzing the network statistics and malicious node behavior. The simulation results show that network performs better when neighbor node trust management based anomaly detection technique is in place.

*Index Terms*—Wireless Sensor Network, Intrusion Detection System, Trust management, Risk, Trusted Node

**25. PaperID 30061668: Novel Hybrid Image Encryption (64-Bit) Based On Rubik Cube and Block Cipher (pp. 184-192)**

*Jasdeep Singh Chauhan, Amanpal Singh Rayat*

*Dept. of CSE, Rayat Bahra Campus, Ropar*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract** - Cryptographic Encryption is a method, for the protection of useful information so that only those for whom it is intended can read and process it. Numerous applications are there which require the rapid and strong security against the unauthorized users. For example, securing Military related information, securing sensitive online transactions, securing online transmission of data for real time applications like stock market apps, electronic mails or data transmission of social applications and online personal photograph albums like applications demand for the high security as these are stored and transmitted throughout the internet. The image Encryption is one of the techniques used for alteration of the images into faint form so that the image cannot be seen by the prohibited person. In this paper we explore the Novel Hybrid technique to encrypt image by following the concept of Rubik Cube encryption phenomenon (stream cipher) and combine it with block cipher.

**Index Terms:** - *Novel Hybrid, Encryption, Decryption, Rubik Cube, Symmetric key cryptography, Secure Force Algorithm, secret key.*

## **26. PaperID 30061670: E-Learning Systems Risks and their Security (pp. 193-200)**

*Kassid Asmaa, El kamoun Najib  
STIC Laboratory, Chouaib Doukkali University*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract** — The security of Information Systems is a major challenge for all organizations today, because people can only use a system if they trust it. Especially when they are using open and distributed environment like E-learning platforms, as e-learning increases in popularity and reach, the need to understand security concepts will also increase. The goal of this research is to identify some key security issues that must be taken into consideration in developing and using an E-learning platform. In order to do it, this paper examines the basic concepts of security in computing, and some characteristics of E-learning platforms that introduce new threats and ways to attack, we will also discuss some security aspects of one of the most popular E-learning systems: Moodle.

**Index Terms**—*Security requirements, E-learning platform, Security in E-learning platform.*

## **27. PaperID 30061671: Classification of households after a traumatic shock, with the aid of Bayesian Networks: example of the post- electoral crisis in Côte D'Ivoire (pp. 201-207)**

*SAHA Kouassi Bernard, BROU Konan Marcellin, BABRI Michel, OUMTANAGA Souleymane  
Institut National Polytechnique, Félix Houphouët-Boigny de Yamoussoukro, Côte d'Ivoire*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract**-: Classification is a branch of multidimensional descriptive statistical analysis. This field of study has been the subject of several publication works. For the last couple's years, it is facing a renewal and a remarkable development with the multiplication of data .This situation requires, a deep analysis before the adoption of probabilistic model as suggested by the results. In this paper, we intend to study the social resilience and the vulnerability of urban populations'. Owing to the high concentration rate of population in big cities and the subsequent increase of modern plagues like rural exodus, galloping and blind urbanization with such corollaries as the creation of precarious districts and at times upper-crust in high-risks zones. So, within the framework of this study , we propose a deep analysis of data in general , the classification of Ivorian households according to their income , dwelling place after the shock of the social, political and the military crisis .This classification study should confirm or invalidate the opinion according to which the crisis was salutary to some people and a disaster for others, by causing a delay in the

development of the country. Also through a modelling of the data collected on households made vulnerable by the post electoral crisis, in the form of Bayesian multidimensional models.

*Index Terms:- Bayesians networks, HIV-AIDS, Household, Resilience, Traumatic Shock, Post electoral crisis, Vulnerability.*

**28. PaperID 30061673: Secure Approach for Net Banking by Using Fingerprint Authentication in Distributed J2EE Technology (pp. 208-213)**

*Rachid ALAOUI (1), Khalid ABBAD (2), Ahmad EL ALLAOUI (3), Moulay Abdellah KASSIMI (4)*

*(1) Laboratory of Systems Engineering and Information Technology (LISTI), ENSA, Ibn Zohr University Agadir, Morocco*

*(2) SIA Laboratory, FST, FSDM University, FEZ, Morocco*

*(3) ENSA AL Hoceima and Labo MATSI Mohammed I University OUJDA, Morocco*

*(4) LGEMS Laboratory, ENSA, Ibn Zohr University Agadir, Morocco*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract -* Today, Net Banking or Internet Banking System is popular technology typically used by individuals to carry out a variety of personal and business financial transactions and banking functions by using mobile technology. Net Banking is used to describe banking transactions through internet application. But there are many security problems like fraudulent websites, fake emails from banks, capturing user IDs and passwords, hacking personal bank accounts and ATM card etc. Security and Authentication of individuals is necessary for our daily lives especially in net Banking. It has been improved by using biometric verification techniques like fingerprints. This research paper gives a security solution mobile through a new model with biometric recognition and SMS service.

*Keywords:* *Secure Internet banking, Smartphone, Fingerprint, Banking transaction.*

**29. PaperID 30061675: Towards the Design of Fault Tolerant Binary Comparator by Parity Preserving Reversible Logic based Multi Layer Multiplexer (pp. 214-221)**

*Biswajit Das, Murshidabad College of Engineering & Technology, Berhampore, India*

*Shefali Mamataj, Murshidabad College of Engineering & Technology, Berhampore, India*

*Saravanan Chandran, National Institute of Technology, Durgapur, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract —* Reversible circuits which are Parity-preserving are nowadays getting more weight towards the progress of designing systems having fault-tolerance in the field of nanotechnology. The reversible circuit which preserves parity must have the parity preserving property means the input vector parity must be the same to the output vector parity. It contributes a expansive category of finding faults in the circuit which can be detect at the circuit outputs. Thus in a single word reversible logic circuits which preserves parity will be more beneficial towards the progress of fault free circuit realization. In this paper we have proposed three new fault tolerant reversible gates FTM, FTC and FATOC for optimizing the circuit in terms of the gate number, garbage outputs, hardware complexity and constant inputs. This work targets implementation of reversible Fault Tolerent Comparator (FTCom) by Reversible Logic-based Multi Layer Multiplexer of proposed FTM. Furthermore the design is also presented by the obtainable fault tolerant reversible gates and the proposed gates FTC & FATOC. We have also presented three lemmas to verify the fault tolerance or parity preserving property of these proposed FTM, FTC and FATOC gate respectively.

*Keywords-* *Fault Tolerance, Parity-Preserving Reversible Gate, Reversible Logic, Comparator*

**30. PaperID 30061678: Sentiment Analysis of Twitter data using Hybrid Method of Support Vector Machine and Ant Colony Optimization (pp. 222-225)**

*Jasleen Kaur, Sukhjit Singh Sehra, Sumeet Kaur Sehra  
Guru Nanak Dev Engineering College, Ludhiana, India, 141006*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Sentiment analysis is the process of elicitation, comprehension, classification and illustration of opinions or sentiments expressed by various users concerning a topic or object. It has become prominent due to the increase in crowdsourced information on social media. Social media has bestowed users with much more power than they possessed before its advent. Presently, Twitter is a prominent micro-blogging platform which empowers its users to post their opinions in form of “tweets”. These can be utilised to gain insights into opinions and sentiments of people for better decision making and marketing. This research aims to use Twitter data to inspect sentiments of the crowd regarding a particular subject. Retrieved tweets are classified into two opinion classes: Positive or Negative. This classification is performed by using a hybrid strategy of Machine Learning algorithm Support Vector Machine (SVM) and Ant Colony Optimization (ACO). Unigrams are employed for feature extraction with term frequency-inverse document frequency as feature weighting criteria. The average accuracy of classification enhances from 75.54% (using SVM) to 86.74% (using SVMACO).

*Index Terms*—*Crowdsourced data, Machine Learning Techniques, Sentiment Analysis, Twitter*

**31. PaperID 30061679: Defending Against Attacks from the Dark Web Using Neural Networks and Automated Malware Analysis (pp. 226-237)**

*Eng. Mihai-Gabriel IONITA, Prof. Victor-Valeriu PATRICIU  
Computer Sciences and Information Technology Doctoral School, Military Technical Academy, Bucharest, Romania*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — In an Internet connected world, cyber security assurance is critical for protecting an organization’s critical infrastructures. For this task, we propose a connected infrastructure that offers various types of malware analysis capabilities. This infrastructure’s architecture is based on customized open-source projects. This proposed implementation has been integrated into an already built platform that aims to protect an organization’s geographically distributed network. Our proposed implementation is based on software defined network components, and it uses artificial neural networks for protecting these critical infrastructures. The malware analysis component is based upon three sub-components that perform static and behavioural analysis against suspected pieces of code, documents or traffic. In addition, when attacks that involve zombie computers come from the Dark Web, the proposed platform tries to uncover their true source, so it can inform the unsuspecting users or defer them to justice. As detecting Tor traffic is not a trivial task, the platform includes a dedicated module for scanning and making a risk assessment of inbound and outbound connections. An intelligent firewall separates the protected infrastructure from malicious internet traffic by telling apart malevolent Tor traffic from other benign traffic flows. The platform also offers added protection against 0-day vulnerabilities and APT attacks by using its behavioural analysis techniques.

*Keywords*- *cyber security, artificial neural networks, automated malware analysis, Tor, dark-web*

**32. PaperID 30061680: Reduce Collisions and Increase the Efficiency of the RFID Network System by using Manchester Encoding (pp.238-243)**

*Fahimeh Afshamnia, Mohammad Reza Soltan Aghaei  
Department of Computer Eng., Faculty of Eng., Isfahan (khorasan) branch, Islamic Azad University, Isfahan, Iran.*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — RFID networks represent a system that uses radio waves to transmit information. This network plays a key role in a wide range of applications such as traffic control, transportation, military and medical use. In such networks, data collision is inevitable. The thing that made it difficult and seriously affected the desire to progress in the field of practical applications of radio networks is the problem of collision. Collision as a key problem in the RFID system, can waste energy consumption and bandwidth and leading to an increasing the time requirement for the process of tags identification. In this article, we review some adversaries to consider anticollision algorithm first of all, and then present a method that use Manchester encoding to reduce collision, which aims to increase the system efficiency, reducing the amount of energy consumption and collision. Finally, evaluate of the proposed algorithm in system efficiency parameters such as the number of collision. The result of the comparison shows that the performance of the proposed algorithm will reduce energy consumption and increase the system efficiency.

*Keywords* - Data collision, Radio networks, System efficiency, Slot, Manchester encoding Commas

**33. PaperID 30061682: Numerical Solution of Nonlinear Optimality Problem by PSO & GA (pp. 244-250)**

*H. Hossein zadeh, E. Salehpour*

*Department of Mathematics, University of Mazandaran, Babolsar, Iran*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — The present research aims to introduce a combined method for solving optimization problems namely PSO-GA. In this algorithm, particle swarm optimization (PSO) is operated in order to improve vector while genetic algorithm is used in order to improve decision vectors through genetic algorithms. A balance between exploration capabilities and exploitation is improved in PSO algorithm through genetic operators namely cut and swap. Defined limitations are used in the problem through penalty function without parameter. Empirical results of optimization problems are compared to different kinds of methods in the published paper. The obtained solution compared with better suggested method of the solution existing in this paper and published texts. Moreover, empirical results show that the proposed method is the best solution for engineering problems.

*Keywords*—Particle swarm optimization, Genetic algorithm, Constraint optimization, PSO -GA.

**34. PaperID 30061688: GPASS: A Graphical Password Scheme Using Alphanumeric Characters and Pictures (pp. 251-258)**

*Shah Zaman Nizamani, Department of Information Technology, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Pakistan*

*Syed Raheel Hassan, Department of Computer Systems Engineering, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Pakistan*

*Muhammad Mubashir Khan, Department of Computer Science & IT, NED University of Engineering & Technology, Karachi, Pakistan*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Authentication is very important for secure use of any computerized system. Textual password is serving to authentication since long time, but it is vulnerable to different kinds of attacks. To make authentication process more secure and easy to memorize, graphical password authentication has been introduced. This approach solved most of the problems present in textual passwords. However shoulder surfing attack is common in graphical password schemes. Anyone monitoring the process of login, through camera or some kind of recording software can recognize the password easily. To overcome this issue researchers developed different graphical password schemes but most of them suffer from usability and memorability issues. Therefore a graphical password scheme is required, which is resistant to shoulder surfing and similar attacks along with better usability and memorability. In this paper a combined textual and graphical password scheme (GPASS) is proposed with its implementation and usability results. In the

GPASS scheme users select password by clicking on a group of four password elements which help to improve the authentication process. Security analysis of GPASS scheme is also presented along with comparison of other recognition based graphical password schemes.

*Index Terms — Graphical Authentication, Security, Usability, Alphanumeric password*

**35. PaperID 300616100: Concept Based Text Document Clustering with Vector Suffix Tree Document Model (pp. 259-264)**

*Dr. N. Sandhya (1), Dr. A. Govardhan (2) , Dr. G. Rameshchandra (3)*

*(1) Professor, CSE Department, VNRVJIET, Hyderabad*

*(2) Professor & Principal, CSE Department, JNTUH, Hyderabad*

*(3) Professor, CSE Department, VNRVJIET, Hyderabad*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract —* The most popular way for representing documents is the vector space model, because of its speed and versatility. The vector space model has some drawbacks. To overcome the bag of words problems, text documents are treated as a sequence of words and documents are retrieved based on sharing of frequent word sequences from text databases. The sequential relationship between the words and documents is preserved using a suffix tree data structure. Syntax based disambiguation is attempted by enriching the text document representations by background knowledge provided in a core ontology. Word Net is used for this purpose in our model. This work aims to extend a document representation model which is elegant by combining the versatility of the vector space model, the increased relevance of the suffix tree document model and also retains the relationship between words like synonyms. The effectiveness and the relevance of this concept based model compared to the existing models is evaluated by a partitioning clustering technique and then a systematic comparative study of the impact of similarity measures in conjunction with different types of vector space representation on cluster quality is performed. This document model will be called the Concept Based Vector Suffix Tree Document Model (CBVSTDM).

*Keywords- Text Clustering, Similarity Measures, Suffix tree WordNet, Cluster Accuracy*

**36. PaperID 300616101: Trajectory Planning for a Four-Wheel Robot using Decomposition Coordination Principle (pp. 265-274)**

*Hala El Ouarrak #, Mostafa Rachik #, Mohammed Mestari \**

*#Department of Mathematics, Faculty of science Ben M'Sik, Université Hassan II Mohammedia, Av Driss El Harti B.P 7955, Sidi Othmane, Casablanca, Morocco.*

*\*Department of Computer science, ENSET Mohammedia, Hassan II Mohammedia, Morocco*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract —* In this paper we treat the path planning problem with a new approach based on the Decomposition Coordination Method. This method allows the resolution of a complex non-linear model of a four-wheel robot, while integrating its kinematic and dynamic constraints. The method consists of the decomposition principle, which treats non-linearity of the system on a local level. The coordination is then achieved by use of Lagrange multipliers. One of the best features of this method is the fast reactivity and its flexibility to adapt to even the most complicated of systems. A numerical application is presented to highlight the advantages of the approach we use in this paper.

*Keywords— Path planning, Autonomous navigation, Robotic, Control theory, Nonlinear control systems.*

**37. PaperID 300616107: FractAntBee Algorithm for Data Clustering Problem (pp. 275-283)**

*Amira Hamdi (1,2), Nicolas Monmarché (2), Mohamed Slimane (2), Adel M Alimi (1)*  
*(1) REGIM-Lab.: REsearch Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Sfax, Tunisia*  
*(2) Polytech Tours, University of Tours, France Tours, France*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — We present in this paper a new swarm based algorithm named FractAntBee for data clustering problem. This algorithm uses the stochastic principles of ant colonies in conjunction with the geometric characteristics of the bee's honeycomb, the basic principles of stigmergy and the main characteristics of fractals theory. Experimental results show that the proposed approach is significantly better than both K-means and Ant Clustering Algorithms in terms of the number of clusters and relevant clustering indices.

*Index Terms*— *swarm intelligence; artificial ants; data clustering problem; clustering validity indices*

**38. PaperID 300616108: Asymptotically Almost Automorphic Solution of High Order Recurrent Neural Networks with Mixed Delays (pp. 284-295)**

*Hajer Brahmi (1)\*, Boudour Ammar (1), Farouk Cherif (2) Adel M. Alimi (1), Ajith Abraham (3)*  
*(1) Research Group on Intelligent Machines, Department of Electrical and Computer Engineering, National Engineering School of Sfax, University of Sfax, Sfax 3038, Tunisia,*  
*(2) ISSATS, Laboratory of Math Physics; Specials Functions and Applications, LR11ES35, Ecole Supérieure des Sciences et de Technologie, 4002- Sousse- Tunisia*  
*(3) Machine Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, Auburn, Washington 98071, USA*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — This work aims to investigate a class of high-order recurrent neural networks. Various criteria are established for the existence and uniqueness of asymptotic almost automorphic solutions in a given convex domain. Besides, several approaches are applied to derive sufficient condition for the globally exponential stability of the considered model. Our method is based on finding suitable Lyapunov functional and the well-known Banach's fixed point principle. Lastly, two numerical examples are given to illustrate the effectiveness of the analytical findings.

*Index Terms*— *high order recurrent neural network, exponential stability, asymptotically almost automorphic functions.*

**39. PaperID 300616110: Dynamic Topology Control for Reliable Group Communication over MANETs (pp. 296-303)**

*Amit Chopra, Dr. Rajneesh Kumar*  
*CSE dept., MMEC, M. M. University, Ambala, Haryana, India*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Group communication based applications can be used for education, defense, medical emergencies, traffic management and conferencing. The end user experience depends upon the reliable multicast routes which are formed under various constraints, i.e. dynamic topology, high mobility, weak wireless links, short transmission range and low power backup etc. Packet drop due to the dynamic change in network topology can reduce the network performance. So multicast routing should be able to operate in different situations such as mobile environment, heavy traffic load and uncertain network topology, in order to achieve satisfactory performance. In this paper, the impact of mobility models over the different multicast routing protocols under the different constraints will be explored to provide a solution for reliable multicast communication over ad hoc networks.

*Keywords-MANET, Multicast, Reliability, Dynamic Topology, Mobility, MAODV, PUMA, MZRP*

**40. PaperID 300616113: Logistics Reverse Chain Optimization Based on Genetic Algorithms (pp. 304-313)**

*Walid Ellili, Mounir SAMET, Abdennaceur KACHOURI*

*Laboratoire d'Electronique et des Technologies de l'Information 'LETI', University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — The introduction of reverse logistics has been remarkably enhanced by the relevant legislative acts passed in several industrial countries, with the aim of preserving the environment. Given the diverse activities and definitions attributed to this particular type of logistics, the relevant design and control turn out to be too sophisticated. It is in this particular context that the present paper can be set, with the major objective of implementing a mixed integer nonlinear programming model for the design of an integrated distribution network, which appears to be dynamic once the integrated nature of reverse logistics network optimization is being considered. As demonstrated, the genetic algorithms' effectiveness in achieving the most optimum solutions within a reasonable time framework has also been highlighted. In a last stage, these methods' limitations have been underlined, along with some suggested research perspectives. An analysis of some mathematical reverse logistics models has also been undertaken with respect to five related application areas, namely: location, life cycle assessment, production planning, inventory management, along with the establishment of the most appropriate product-collection routes. In addition, the reverse logistics' case studies, as treated in the pertinent literature, have also been thoroughly analyzed and grouped according to industries.

*Keywords-* Reverse logistics, Genetic algorithm, Optimization, models, End of life.

**41. PaperID 300616114: IoT Operating Systems and Security Challenges (pp. 314-318)**

*Muhammad Asim, Waseem Iqbal*

*National University of Science and Technology (NUST), Islamabad, Pakistan*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — The emerging trend of pervasive computing aims embedded devices such as smart phones, autofocus cameras, musical instruments, home video systems etc with microprocessor and wireless communication capability. This type of computing paradigm is known as IoT (Internet of Things). IoT connects myriad of things for providing service to machines and humans. In 2020 it is expected billions of things in IoT will be deployed worldwide. Centralized computing approach does not provide sustainable model, so a new architecture is needed as trusted platform for expansion of Internet of Things (IoT). Data gather with IoT are often unstructured and noisy, so more computation power require for analysis and getting efficient results and also needed efficient mechanism for authentication in lightweight devices like IoT where less computation power, limited resources, low memory and low battery life.

This paper is about operating systems of IoT and current security challenges in IoT using RPL and 6LoWPAN (IPv6 over low-power WPAN) protocols and also we will discuss possible solutions related to IoT Security challenges.

*Keywords-- Wireless Sensor Network, Low power Wireless Personal Area Networks, Software Define Network.*

**42. PaperID 300616117: A Spiking Neural Network Model for Complex Handwriting Movements Generation (pp. 319-327)**

*Mahmoud Ltaief, Hala Bezine, Adel M. Alimi*

*REGIM-Lab.: REsearch Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — In this paper a spiking neural network model for online complex handwriting movement generation is proposed. Online handwriting is described as the superposition of strokes with the elliptical shape which is the result of the algebraic sum of the beta profiles. Handwriting can be partitioned into simple strokes. Each one is fully modeled by a set of ten parameters which characterize the handwriting in both the kinematics and the static fields. The network is composed of an input layer which uses a set of Beta-elliptic parameters as input, a hidden layer and an output layer dealing with the estimation of the script coordinates X(t) and Y (t). An additional input is used as a timing network to prepare the input parameters. This later, acts as starting pulse of each stroke belonging to a given handwriting script. The simulation results showed that the spiking neural network model could generate both Latin and Arabic handwriting scripts. Similarity degree is measured between original scripts and generated scripts to evaluate our model. The proposed spiking neural network model can be applied in new ways such as: signature verification and shape recognition.

*Index Terms* — Spiking neural network, Beta-elliptic model, Handwriting generation, Similarity degree.

**43. PaperID 300616125: Making PIN and Password Entry Secure Against Shoulder Surfing Using Camouflage Characters (pp. 328-335)**

*Suliman A. Alsuhibany, Computer Science Department, College of Computer, Qassim University, Buridah, Saudi Arabia*

*Saad G. Almutairi, Department of English & Translation, College of Arabic & Social Studies, Qassim University, Buridah, Saudi Arabia*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Authentication in phones as well as in public spaces or even in shared spaces such as digital tabletops is inherently vulnerable to attacks and has the weakness of being susceptible to shoulder surfing attack. Shoulder surfing attack is a type of attack that uses direct observation techniques such as looking over someone's shoulder to get information. This paper introduces a novel way of using the simple PIN (Personal Identification Number) entry technique to conceal the actual password within contingent randomly selected entries. In particular, the traditional password concept where what you input is what you get is redefined. That is, the distinction between the actual password and the act of entering a password is achieved using two master keys. The proposed approach allows the entry of very long passwords and thus prevents unwanted access even with exact copying of the entered password. Furthermore, it allows also to the entry of very short password. The prototype of the proposed approach is implemented. A user study has been conducted to evaluate both security and usability perspectives of this technique. The results showed that proposed approach is strength against observing the password and usable for participants to have a good control over the different parts of the entry.

*Keywords- passwords; tabletops; security; usability; shared space; authentication; shoulder surfing attack*

**44. PaperID 300616126: Support Vector Machine, Multilayer Perceptron Neural Network, Bayes Net and k-Nearest Neighbor in Classifying Gender using Fingerprint Global Features (pp. 336-340)**

*S. F. Abdullah, A.F.N.A. Rahman, Z. A. Abas*

*Optimisation, Modelling, Analysis, Simulation and Schedulling (OptiMASS) Research Group, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia.*

*W. H. M. Saad, Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka,Malaysia.*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — A scientific study of fingerprints, lines, mounts and shapes of hands are called dermatoglyphics. Dermatoglyphics features from fingerprint are statistically differ between the gender, ethnic groups, region and age categories. From the previous study of gender classification in forensic area, the process of feature extraction is done manually and classify using a statistical approach. The features extracted were; ridge count (RC), ridge density (RD), ridge thickness to valley thickness ratio (RTVTR) and white lines count (WLC). The sample use consists of 300 respondents where each respondent gives 10 different fingerprints. Four classifiers which are Bayes Net, Multilayer Perceptron Neural Network (MLPNN), k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) are used in order to evaluate the performance of the proposed algorithm. The overall performance of the classifier is 95% of the classification rate. From all classifiers, SVM emerges as the best classifier for proposed algorithm.

*Keywords*—fingerprint, gender classification, SVM, MLPNN, k-NN, Bayes Net

**45. PaperID 300616129: An Implementation of Segmentation in Citrus Canker Disease Detection Using Patches and Labels (pp. 341-345)**

*K. Padmavathi, Research and Development Centre, Bharathiar University, Coimbatore-641 046*

*Dr. K. Thangadurai, Assistant Professor and Head P.G & Research Department of Computer Science, Government Arts College (Autonomous), Karur-639 005*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — The proposed work employs the segmentation method using patches and labels to segment the citrus canker leaf diseased portion. The patches and labels method is based on the region merging, color mapping and clustering techniques with statistical tests to determine the merging of regions. The method utilizes the color feature of the leaf images, where the leaf image can be segmented into multiple parts by its colors. The color intensity feature of the leaf image is used as basis for grouping the pixels into patches. Range of colors are considered for process and grouping of respective pixels within the color range to form patches which is based on color threshold levels (Q values). The leaf image is represented at 9 different color threshold levels (Q), where the nth level of threshold applies  $2^{n-1}$  number of colors in color space for further color mapping to form patches. The patched image divides and represents different regions of the leaf image as segmented output. The patched image forms the grouping pixels within neighborhood connectivity, is represented as collection of clustered color patches with labels. The boundary information of each labeled patch is achieved. The labeling of the clustered color patches aids in segregation between region of interest and other uninterested region.

*Keywords*- Density based clustering, Patches and Labels, Citrus Canker Disease Detection, Region merging, Segmentation, Threshold levels (Q)

**46. PaperID 300616130: Hybrid (OCDMA/WDM) System with DPSK Modulation Using Different Detection Technique at Bit Rate 1Gbps for Local Area Network (pp. 346-351)**

*Monirul Islam, Nasim Ahmed, Mijanur Rahman, S.A. Aljunid, R. B. Ahmad*

*School of Computer and Communication Engineering, University Malaysia Perlis (UniMAP), 02600 Arau, Perlis, Malaysia.*

*Shohel Sayeed*

*Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia.*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Design of a hybrid (OCDMA/WDM) system with advance modulation technique DPSK is proposed in this paper. In hybrid system, two different technologies OCDMA and WDM combines with one system to increase the transmission capacity, security and potentiality of the network. To develop OCDMA/WDM system, multiple access interference (MAI) is one of the main reasons that degradation the performance. According to the MAI

degradation, Modified double weight (MDW) codes are used as signature address code. OptiSystem ver.12 has been used for simulation to measuring the performance of the system. In this paper, compare the simulation results between AND subtraction detection with complementary subtraction detection techniques at data bit rate 1Gbps. The simulation results revealed that complementary subtraction technique is better than direct detection technique as referred to bit error rate (BER) 1Gbps where the targeted BER  $\leq 10^{-9}$ .

*Keywords-optical code division multiple access; wavelength division multiplexing; bit error rate; on-off keying; differential phase shift keying*

**47. PaperID 300616132: Analytical Approach for Quality Assessment of Dynamic Web Environment through Metrics (pp. 352-356)**

*Rohini, Dr. Indu Chhabra*

*Department of Computer Science and Applications, Panjab University, Sector- 14, Chandigarh, India*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Quality assessment of dynamic web environment is continuous process. Websites are used as important communication channels and information delivery tools with potential for reaching a wide audience. In the current scenario, websites are designed and developed according to ISO based quality models and customized further as per specific requirements. Therefore, evaluating websites at runtime in their dynamic environment has become a key issue warranting attention. In this regard, quality analytics are used to work on this issue and evaluate the relevant website on the basis of web analytics captured from dynamic web learning environment by exercising key performance metrics. This paper covers assessment of the website for its dynamic environment using web analytics captured for specific metrics. These metric directly related to the sessions, users, page views and bounce rate analysis for website. An analytical approach for the same has been discussed and the results are analysed for single sample website form business sector.

**48. PaperID 300616133: A Novel Approach for Multi-modality Image Fusion with Conjugation of DWT and RT using Region Consistency (pp. 357-361)**

*Keyur N. Brahmbhatt, Department of Computer Engineering, Changa, India*

*Dr. Ramji M. Makwana, Department of Information Technology, VVP College, Rajkot, India*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Complementary multi-modality image information of more than one image when joined together and create single new informative image, this is known as image fusion. Main purpose of it is to reduce ambiguity and idleness in a resultant image by enhancing relevant details specific to any task or any application. In medical imaging images comes from various input sources which has different detailed information. Thus there will be an interesting task to perform merging operation on registered multimodality images. The image fusion is very valuable in medical analysis. Here in our research paper, the fusion process has been done in two transform named discrete wavelet transform (DWT) & Ripplet transform (RT). Region consistency check and Maximum selection fusion rules has been used. Implementation task has been performed on Computed Tomography and Magnetic Resonance Imaging images. Evaluating and doing comparative study of fusing methods, measuring parameters are used. Implementation of method shows that, our suggested method exhibits a fine results in area of medical imaging, because our method provides arbitary scaling due to RT and good local features due to DWT.

*Keywords:- Image Fusion, DWT (Discrete Wavelet Transform), Ripplet Transform*

**49. PaperID 300616138: Design of Pixel Neighborhood Based Offline Handwritten Thinning Framework for Devnagri Numeral Script using Elman Neural Network (pp. 362-368)**

*Gulshan Goyal, I. K. Gujral Punjab Technical University (Punjab), Faculty of Engineering, CCET (Chandigarh), India  
Dr. Maitreyee Dutta, Professor & Head, NITTTR (Chandigarh), India*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Processing of Handwritten script has been an active area of research due to non-uniformity of handwritten data. Thinning is an important pre processing stage in processing of handwritten script. A large number of traditional thinning algorithms and techniques available so far involve trade-off between one or more of the topological and geometrical requirements. This paper proposes a neural network based framework for thinning of handwritten Devnagri numeral script. The pre-thinning steps including resizing, gaussian blurring and binarization are applied on dataset. Training rules are formed based on Zhang and Suen thinning algorithm with the inclusion of unit pixel width templates. An Elman neural Network is trained and applied on test images. Experimental results show that improvements in proposed framework over traditional approach. Multiple sub iterations of a conventional thinning algorithm are reduced to single one.

*Index Terms* — *Handwritten devnagri numerals, Elman neural network, skeleton, Thinning.*

#### **50. PaperID 300616139: Fault Tolerant System with Maximum Efficiency (pp. 369-375)**

*Minakshi Memoria, Computer Engineering, Suresh Gyan Vihar University, Jaipur, India  
Dr. Ripu Ranjan Sinha, Research and Development, Suresh Gyan Vihar University, Jaipur, India*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Today highly secure systems in networks are very demanding in which they can share available resources in network and complete their execution even in the occurrence of fault. In this paper we proposed a fault tolerance technique to improve resource utilization with maximum throughput and great efficiency.

*Keywords* - *Replication, Check pointing, Scheduling*

#### **51. PaperID 300616145: Rule Extraction Algorithm for Deep Neural Networks: A Review (pp. 376-380)**

*Tameru Hailesilassie  
Department of Computer Science and Engineering, National University of Science and Technology, Moscow, Russia*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Despite the highest classification accuracy in wide varieties of application areas, artificial neural network has one disadvantage. The way this Network comes to a decision is not easily comprehensible. The lack of explanation ability reduces the acceptability of neural network in data mining and decision system. This drawback is the reason why researchers have proposed many rule extraction algorithms to solve the problem. Recently, Deep Neural Network (DNN) is achieving a profound result over the standard neural network for classification and recognition problems. It is a hot machine learning area proven both useful and innovative. This paper has thoroughly reviewed various rule extraction algorithms, considering the classification scheme: decompositional, pedagogical, and eclectics. It also presents the evaluation of these algorithms based on the neural network structure with which the algorithm is intended to work. The main contribution of this review is to show that there is a limited study of rule extraction algorithm from DNN.

*Keywords*- *Artificial neural network; Deep neural network; Rule extraction; Decompositional; Pedagogical; Eclectic.*

**52. PaperID 300616147: Verification of Pipelined Microprocessors using Maude LTL Model Checker (pp. 381-388)**

*Mustapha Bourahla, Computer Science Department, University of M'sila, BP 166 Ichebilia, M'sila, 28000, Algeria*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — This paper presents an approach for the verification of a pipelined microprocessor using the Rewriting Logic. To express many machine-relevant properties, we have modeled the stream of instructions with the system Maude which is based on Rewriting Logic. It is used to run and debug the pipelined machine specification. The Maude LTL model-checker is also used to verify the pipelined machine properties and eventually to verify a complete pipelined machine design, whose correctness is defined using the idea of pipeline flushing.

*Index Terms* — *Rewriting Logic, Maude LTL Model Checking, Pipelined Microprocessors*

**53. PaperID 300616150: A Novel Multi-Stage Authentication System for Mobile Applications (pp. 389-396)**

*Monther Aldwairi #\*, Rima Masri \*, Haneen Hassan, May ElBarachi*

*# Department of Network Engineering and Security, Jordan University of Science and Technology, PO. BOX. 3030, Irbid, 22110, Jordan*

*\* College of Technological Innovation, Zayed University, Abu Dhabi, P.O. Box 144534, U.A.E*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Mobile devices and applications are quickly becoming the main platform to access the Internet and web applications. Successful repetitive attacks on conventional authentication systems made it necessary to reinvent the wheel and come up with new authentication systems that increase security while maintaining convenience. Mobile devices and applications require specific authentication systems that combine, security, simplicity and speed. In this paper we propose, design, implement and evaluate a multi-stage authentication system, more specifically, a three-stage authentication system. It takes the user into a series of secure but simple challenge-response stages, before allowing access to the system. The first stage is the identification where the user chooses a username and the device serial number is automatically logged and tied to the username. The second stage presents the user with a large grid of n independent squares and he must highlight the correct m squares he previously selected during the registration process. In the final stage, the user must select s out of i images in the same order he picked them during registration. We logically argue the advantages and disadvantages of the proposed system and present a formal and probabilistic analysis to gauge the systems security. To better quantify the convenience and simplicity of use, we deploy the system and survey the opinions of regular users and security professionals. The results of our analysis and survey show very low probability of guessing attacks and high user satisfaction. The probability of successful brute force attack is as low as 1:22314E  $\square$  28 for selecting 21 out of 25 squares/images.

**54. PaperID 300616152: Energy Efficient Clustering Multipath Routing Protocol (EECMRP) with Strategic Route Discovery for Heterogeneous Wireless Sensor Network (pp. 397-402)**

*Kakelli Anil Kumar\* (1), Dr. Addepalli VN Krishna (2), Dr. K. Shahu Chatrapati (3)*

*(1) Associate Professor, Dept. of CSE, IIST, RGPV, Indore, MP, India*

*(2) Professor, Dept. of CSE, FOE, Christ University, Bangalore, KA, India*

*(3) Head, Dept. of CSE, COE, Manthani, JNTUH, Hyderabad, TS, India.*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — Heterogeneous wireless sensor networks (HTWSNs) are the emerging field to monitor the unattended and unmanned areas efficiently. Heterogeneous wireless sensor networks are having the vast number of applications in other important major fields also like military, environmental, industrial and agriculture. The routing protocol

performs a key role for energy-efficient quality data transmission in sensor network. However, the existing routing protocols like TEEN, MACS, MP and MRP are not achieved the desired throughput due to traditional route discovery and data transmission methods. These methods are resulting higher overhead within the network and result the low network throughput. To achieve the better network lifetime and performance, strategic routing discovery mechanisms are highly essential for resource contained HTWSNs. The proposed energy efficient clustering multipath routing protocol (EECMRP) with strategic routing discovery mechanism has discovered the energy-efficient multipath for quality data transmission for better network lifetime in HTWSNs. The performance evolution shows that, EECMRP protocol has given higher performance in terms of throughput 14%, 10%, 10% and 4 %, energy efficiency of 78%, 58%, 28% and 13%, packet delivery ratio 13%, 10% and 8%, low latency 75%, 83%, 63% and 53 % and network lifetime 26%, 30%, 33.5% and 10% as compared with TEEN, MACS, MP and MRP.

**55. PaperID 300616160: Time-Frequency Analysis of Epileptic EEG for Seizure Detection (pp. 403-411)**

*Tessy E, Muhammed Shanir P. P, Shaleena Manafuddin*

*Department of Electrical and Electronics Engineering, TKM College of Engineering, Karicode, Kerala, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — Epilepsy is a serious disorder characterized by abnormal firing of electrical impulses in the brain, producing repeated, involuntary seizure activities. EEG measures the electrical potentials in the brain through electrodes and is the most important diagnostic tool to detect any abnormalities associated with this normal electrical activity. An automatic novel method based on wavelet transform was used for analyzing the EEG signal and for detecting epileptic seizure activity. The proposed method was tested on a dataset, consisting of five sets of EEG data, recorded on healthy and epileptic subjects. Using daubechies wavelet db4, the data was decomposed into five sub signals in different frequency bands and features namely line length, difference absolute standard deviation value (DASDV), mean absolute value (MAV), median absolute deviation (MAD), and variance were extracted for each sub signal. Classification algorithms- 1) Quadratic discriminant analysis (QDA), 2) K-Nearest Neighbor (KNN), 3) Linear discriminant analysis (LDA) were used for classifying the EEG signals into normal and seizure class, and their performance was determined in terms of sensitivity, specificity and accuracy. The three classifiers obtained pretty good performance for the different combinations of EEG data. The performances obtained show that the features were able to classify epileptic and non-epileptic EEG segments with less complexity and low cost.

*Keywords*— EEG, Epilepsy, Wavelet Transform, Feature Extraction, Classifiers, Seizure Detection.

**56. PaperID 300616161: A Review of Retrieval Algorithms of Indexing Techniques on Learning Material (pp. 412-418)**

*Hamid Paygozar, Department of Computer Engineering, Khomain Branch, Islamic Azad University, Khomain, Iran  
Athena Samadi, Department of Computer Engineering, Khomain Branch, Islamic Azad University, Khomain, Iran*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* - This paper presents a survey to develop retrieval algorithms of indexing techniques framework on learning material. Analysis of the framework is based on surveys on literature review and experiment on online campus Learning Materials. Data indexing problem of online learning material occurs where online data become larger on the system daily which comprised of many types, formats and words of documents. Thus, searching capability for corrects information become slower. Learning materials comprised of words, image and video documents which makes it is more difficult to get the correct information. Objective of this research is to analyse the existing indexing technique in modeling new retrieval indexing algorithms framework mainly for data mining. Four existing indexing technique for a learning material are reviewed. It is identified that the best used technique are Inverted File, Suffix Array, Suffix Tree and Signature File. Based on the four techniques, characterizations and parameters to enhance a new indexing technique (NIT) is identified and five User Acceptance Test (UAT) are performed. A framework for NIT is designed and experiments are done on a Campus Learning Material. Identified parameters are successfully inserted in five test

experiments. The conceptual framework is continuously applied to develop NIT for retrieval algorithms on learning material. This research is significant for fast accessing on real live campus learning material system that benefits users and fast retrieval of needed information.

*Keywords:* Indexing Technique, Data Mining, Retrieval Algorithms, Learning Material, Text, Graphic, Video, Framework.

**57. PaperID 300616163: Implementation of Library Management System Using Radio Frequency Identification Technology in Sindh Libraries (pp. 419-422)**

*Hafeezullah Abdul Rehman (1), Sattar Ahmed Soomro (2), Farhan Ali Surahio (3), Awais Khan Jumani (3),*

*(1) Govt: Department of Education & Literacy, Karachi, Pakistan*

*(2) Department of Mathematic & Computer Science, Sindh University of Jamshoro, Hyderabad, Pakistan*

*(3) Department of Computer Science, Shah Abdul Latif University, Khairpur Mirs, Pakistan*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract*— Shifting of manual records to computerize has highlighted the need to provide an easier, faster, convenient, User-friendly environment for storing, updating and retrieving books information rapidly in Sindh Libraries. The aim of this paper is to propose Radio Frequency Identification (RFID) technology based Library Management System for Sindh Province. Traditionally, Systems are based on two tier application and these are being considered time consume systems as well as less secure. RFID is auto Identification book recognition and retrieving information technology contains on four components, it allows for handling quick Book issuance, to theft detection and returning transactions. It is used readers and passive tags that are able to store information into SQL Server Database for reading and fetching the required record. The experimental approach consist on C# .Net Framework using Object Oriented Programming Methods and an application is presented. By implementing of this new system in to existing environment, it will be beneficial for people of Sindh Province.

*Keywords*— Radio Frequency Identification (RFID) Technology; RFID Components; reader; RFID Tags

**58. PaperID 300616166: Speaker Identification: A Novel Fusion Samples Approach (pp. 423-427)**

*Qeethara Kadhim Al-Shayea, MIS Department, Al-Zaytoonah University of Jordan, Amman, Jordan*

*Muzhir Shaban Al-Ani, Computer Science Department, Al-Anbar University, Anbar, Iraq*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract*— Speaker identification systems are an important part of the biometric techniques. Many speaker identification systems were designed and implemented during the last few years and these systems depend on different techniques. This paper presented a simple speaker identification approach based on fusion via samples and statistical approach to generate the adequate features. This approach describes a simple method that employs statistical approach to generate feature vectors that were defined each speaker.

*Keywords-component; speaker identification; speaker recognition; feature extraction; windows and fusion approach.*

**59. PaperID 300616167: Approximation Algorithm for N-distance Minimum Vertex Cover Problem (pp. 428-433)**

*Tarun Yadav, Scientist, Scientific Analysis Group, Defence R & D Organisation, India*

*Koustav Sadhukhan, Rao Arvind Mallari, Scientist, Defence Research and Development Organisation, India*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Evolution of large scale networks demand for efficient way of communication in the networks. One way to propagate information in the network is to find vertex cover. In this paper we describe a variant of vertex cover problem N-distance Vertex Minimal Cover (N-MVC) Problem to optimize information propagation throughout the network. This problem is equivalent to finding r-Dominating set but a new approach for approximation solution using vertex cover domain is presented in this paper. A minimum subset of vertices of a unweighted and undirected graph  $G = (V;E)$  is called N-MVC if  $\forall v \in V$ ,  $v$  is at distance  $N$  from at least one of the vertices in N-MVC. In the following paper, this problem is defined, formulated and an approximation algorithm is proposed with discussion on its correctness and upper bound.

*Index Terms* — Minimal Vertex Cover, Approximation, N-Trail, N-distance, Maximal Matching, Graph Reduction, Extended Graph

#### **60. PaperID 300616168: CDA-Clone Detection Algorithms in Wireless Sensor Networks (pp. 434-442)**

*Gulista Khan, Computer Science and Engineering Department, Teerthanker Mahaveer University, Moradabad, India*

*Dr. R. K. Dwivedi, College of Computing Sciences and Information Technology, Teerthanker Mahaveer University Moradabad, India*

*Kamal Kumar Gola, Computer Science and Engineering Department, Teerthanker Mahaveer University, Moradabad, India*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Wireless Sensor Networks (WSN) are used in almost every area of life. Due to its advancements wireless sensor networks are exposed to a number of threats. Among which Clonation is one of the most harmful threats. In this threat adversary capture the original node from the network. Then creates its number of clones and implement back them into the network. Now by using these clone's adversary gets all the information and control over the network activities and insider attacks could also launched. Many Clone detection schemes are proposed in literature among them some are based on single node, where only centralized node is responsible to detect Clone, and some other are location dependent in which some nodes are assigned the responsibility to detect Clone. In this paper we have proposed three types of Clone detection algorithm namely CCDA (Cluster based Clone Detection Algorithm), RECDA (Residual energy based Clone Detection Algorithm) and DSCDA (Digital Signature based Clone Detection Algorithm). First scheme is based on cluster, in which a network is divided into clusters and cluster head is responsible to detect Clone, this algorithm showing efficiency in inter cluster Clone detection as well as intra cluster Clone detection, second algorithm is non centralized in which all the nodes are responsible to detect Clone. This scheme is based on residual energy of the nodes. In third algorithm we have used public key cryptosystem and digital signature to detect Clones in WSN.

*Keywords-* Clone, Wireless sensor networks, Public key cryptosystem, Digital Signature, Residual energy.

#### **61. PaperID 300616169: A New Possibilistic Classifier for Heart Disease Detection from Heterogeneous Medical Data (pp. 443-450)**

*Karim Baati (#\*), Tarek M. Hamdani (#+), Adel M. Alimi (#), Ajith Abraham (&*

*(#) REGIM-Lab.: REsearch Groups on Intelligent Machines, University of Sfax, National Engineering School of Sfax, (ENIS), BP 1173, Sfax, 3038, Tunisia*

*(\*) Esprit School of Engineering, Tunis, Tunisia*

*(+) Taibah University, College Of Science And arts at Al-Ula, Al-Madinah al-Munawwarah, KSA*

*(&) Machines Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, P.O. Box 2259, Auburn, WA 98071, USA*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — In this paper, we propose a new Hybrid Naïve Possibilistic Classifier (HNPC) for heart disease detection from the heterogeneous data (numerical and categorical) of the Cleveland dataset. The proposed classifier stands for an extension of two versions of HNPC which have been already proposed to deal with the same problem. As the two former HNPC, the proposed classifier separates data into two subsets (numerical and categorical) and then estimates possibility beliefs using the two versions of the probability-possibility transformation method of Dubois et al. for numerical and categorical data, respectively. Moreover, like the recent version of HNPC, our new classifier performs a common fusion to combine the obtained beliefs. Nevertheless, instead of using the product and the minimum as combination operators during the fusion step, the proposed classifier calls a Generalized Minimum-based algorithm (G-Min) as an improvement of the minimum operator when making decision from possibilistic beliefs. Experimental evaluations on the Cleveland dataset show that the proposed G-Min-based HNPC outperforms the two former versions of HNPC as well as the main classification techniques which have been used in related work.

*Index Terms*—*Naïve possibilistic classifier, G-Min algorithm, heterogeneous data, subjective data, Cleveland dataset, heart disease.*

**62. PaperID 300616170: Biorthogonal Compactly Supported Wavelet with Vanishing Moments for Musical Instrument Sounds (pp. 451-456)**

*A. Raghavendra Sharma, Anand Engineering College, Agra, India,  
B. V Prem Pyara, Dayalbagh Educational Institute, Agra, India*

**Full Text:** PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

*Abstract* — In this paper biorthogonal compactly supported wavelet from the sounds of musical instrument is identified. This paper is based on maximizing projection of the sound signal onto successive scaling subspace, which results in minimization of energy of sound signal in the wavelet subspace. First, 2-band FIR biorthogonal perfect reconstruction filter bank is identified from the given sound signal, which leads to the design of biorthogonal compactly supported wavelet. Second, the wavelet with desired support as well as desired number of vanishing moments is also identified.

*Index Terms*—*Bi-orthogonal wavelet, Vanishing moments, Compact support, Optimal wavelet, Scaling function, Wavelet function.*

**63. PaperID 300616176: Effect of PSO Algorithm on a ECG Data Fusion System (pp. 457-461)**

*Elhoucine BEN BOUSSADA, Computer Engineering and Applied Mathematics Department, National Engineering School, Sfax, Tunisia  
Mounir BEN AYED, Computer Science and Communication department, Faculty of Sciences, Sfax, Tunisia  
Adel M. ALIMI, Computer Engineering and Applied Mathematics Department, National Engineering School, Sfax, Tunisia*

**Full Text:** PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

*Abstract* — With the study and analysis of an intelligent data fusion agent-based model for electrocardiogram (ECG) classification, a contribution of (Particle Swarm Optimization) PSO algorithm on this ECG data fusion system is presented. This data fusion system is based on agents, group of agents, and Swarm. In the context of assistance to medical decision system, the proposed technique helps doctors to quickly and precisely diagnose a heart disease by examining the ECG beats class. In data fusion we discuss the architecture of the proposed system and correlated result without using PSO algorithm. On the other hand, PSO is presented and we discuss the effect of PSO algorithm on the result of data fusion ECG classification. All the results presented in this work are tested on the MIT-BIH database.

*Keywords*-PSO; data fusion; classification; ECG; agent; Swarm; MIT-BIH; Multi-agent system;

**64. PaperID 300616185: The Reasoning in the Description Logic with Vague Concepts (pp. 462-471)**

*Mohamed Gasmi, Mustapha Bourahla  
Computer Science Department, University of M'sila, Algeria*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — The Description Logic languages are considered the core of the knowledge representation systems, given both the structure of a DL knowledge base and its associated reasoning services. “concept” is used to refer to the expressions of a DL language, denoting sets of individuals; however DL becomes less suitable in domains in which the concepts to be represented have not precise definition. We will face the problem of vague concepts. This paper discusses a vagueness theory to express the vague concepts in OWL2 and propose reasoning technique for reasoning tasks of extended Vagueness DL by introduce new expansion rules in Tableau algorithm for reasoning over vague DL.

*Index Terms* — *Vagueness, Ontologies, Description Logics, Automatic Reasoning.*

**65. PaperID 300616186: Knowledge Based Reduction Technique for Virtual Machine Provisioning in Cloud Computing (pp. 472-475)**

*Bhaskar R., Dept. of Computer Science and Engg, Don Bosco Institute of Technology, Bangalore, India.  
Dr. Shylaja B. S., Dept. of Information Science and Engg, Dr. Ambedkar Institute of Technology, Bangalore, India.*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — The recent growth of Cloud computing, service providers major challenging problem is designing efficient mechanism for managing the limited resources shared by different applications. Resource management mechanism has to do efficient sharing of resources for virtual machines by ensuring optimal resource utilization of available physical machines. Resource management mechanism allows Cloud users and also Service providers to effective utilization of their available resources. This paper proposes application of Rough Set model for provisioning of virtual machines. The proposed method uses Knowledge/Attribute based reduction technique, it generates the rules to reduce unnecessary attribute of the virtual machines. These rules help virtual machine manager for making efficient selection of virtual machine.

*Keywords* — *Cloud computing, Virtual Machines, Rough set model.*

**66. PaperID 30041685: An Unsupervised Stemming: A Review (pp. 476-489)**

*Miral Patel, G H. Patel college of Engineering and Technology, GJ, India  
Apurva Shah, Maharaja Sayajirao University, Computer science department GJ, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* - The present article reviews work on morphological analysis, a subfield of computational linguistics. Special focus is given on statistical approach for morpheme segmentation. We delineate morphological analysis as a problem of persuading a narrative of some kind. However, this paper considered the problem of morpheme segmentation. Morpheme segmentation is also referred to as stemming; Stemming is a minimum unit of language that carries a meaning as a root word. In this paper, we concisely discuss the history and motivation of this problem. By referring numbers of papers from reputed journals and conferences, we here present a brief classification of the work done and unfavorably discuss the most significant viewpoints of major unsupervised techniques in the field. While summarizing the achievements so far, we give a clear direction for future work in the related field with justified points of view.

*Keywords: stemming, application of morphological analysis, morpheme segmentation, suffix striping, suffix removal, word segmentation, unsupervised morphology.*

**67. PaperID 31051680: A Survey on Congestion Handling Techniques in Opportunistic Networks (pp. 490-502)**

*Ahthasham Sajid, Department of Computing, Shaheed Zulfiqar Ali Bhutto Institute, Science and Technology, Islamabad, Pakistan*

*Khalid Hussain, Department of Computer Science, Muslim Youth university, Islamabad, Pakistan*

*Imran Baig, Department of Electrical & Computer Engineering, College of Engineering, Dhofar University Salalah, Sultanate of Oman*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract - Opportunistic Networks an emerging research area that encompasses Mobile Adhoc Networks (MANETs) and a subset of Delay Tolerant Network (DTNs). In such networks one of the most dynamic and challenging tasks is to detect congestion timely and effectively. Congestion both at link and node level can occur due to Irregular Connectivity, Long and Variable Length Delays, and Heterogeneous Networks Infrastructure. Intermediate nodes between source and destination are seems to be critical for detecting a congestion issue as they have to store data in their respective buffer until a next best forwarding opportunistic node is found. The prime objective of our research is to evaluate different congestion handling techniques that addressed issues of storage/node level congestion in opportunistic networks. Consequently it may help to detect congestion issue at node level in a Pre-active manner rather than pro-active manner.*

*Keywords: Estimation/ probability theory, DTN, MANETs, ICN*

**68. PaperID 310516117: Optimizing Frequency-Based Thread Pool System By Non-Blocking Queues And Automated Timers (pp. 503-513)**

*Ghazala Muhammad Ashraf (#), \*Faisal Bahadur (#), Mohammad Abrar Khan (+), Arif Iqbal Umar (#)  
(# Department of Information Technology, Hazara University, Mansehra, Pakistan*

*(+) Institute of Information Technology, Kohat University of Science and Technology, Kohat, Pakistan*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract — Thread Pool based server's face a challenge of dynamic optimization and Frequency Based Optimization Strategy (FBOS) was one of such attempts that dynamically adjusts the size of pool on the basis of request frequency if and only if the Turnaround time of client's request is degrading. Whereas FBOS strategy is developed by synchronization primitive known as Locks in java, which slows down its performance due to context switch and synchronization overheads. This paper presents a dynamic thread pool optimization system namely, Non-blocking Frequency Based Optimization Strategy with Automated Timers (NBFBOS with Automated Timers) which makes use of non-blocking synchronization primitives offering advantages of substantial scalability and liveness. We also automate timers in Non-Blocking FBOS which previously remained constant in FBOS. The results of our analysis show that NBFBOS outperforms previous FBOS strategy. Simulation results show that NBFBOS with Automated Timers outperforms existing FBOS scheme by decreasing wait time of clients by 98%. For 90th percentile response times, NBFBOS with automated timers outperforms FBOS by 100ms. Reducing pool size of NBFBOS with automated timers to 11% of FBOS resulted in less memory utilization.*

*Keywords- multithreading, thread pool models, multithreading approaches, Non-Blocking Queues, Non-Blocking Algorithms, Frequency Based Optimization Strategy, Synchronization primitive, Automated Timers.*

**69. PaperID 310516150: A Secure Protocol for Vanet Using Proxy Blind Signature Based on Elliptic Curve (pp. 514-517)**

*Hizbulah Khattak, Arif Iqbal Umar, Insaf Ullah Khattak*  
*Department of Information Technology, Hazara University Mansehra*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — Vehicular ad-hoc networks play important role in modern transportation system. This network faces different un-social and immoral events because of its adhoc nature. The message sender requires authenticity, unforgeability, message integrity for secure communication in vehicular networks environments. This paper presents a new protocol for vehicular ad-hoc networks using proxy blind signature based on elliptic curve cryptosystem. The proposed protocol ensures sender authenticity and message integrity with less computational and communication cost due to shorter key size of elliptic curve.

*Keywords:* *VANET, Proxy Blind Signature, Elliptic Curve*

**70. PaperID 30061618: Evaluation of Best Suitable Scenario for Vehicular Ad Hoc Network (pp. 518-524)**

*Kaushika Patel, J M Rathod*  
*B V M Engineering College, Vallabh Vidya Nagar-388120, Gujarat, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* - This paper is oriented towards performance of various existing reliable transport layer protocol implementations over Vehicular Ad-hoc Network. This includes the comprehensive survey of different Transmission Control Protocol (TCP) implementations using traffic and mobility models developed for Vehicular Ad-hoc Network (VANET). VANET considers car as moving node so the topology of it changes repeatedly. In VANET, vehicular to vehicular (V2V) and vehicular to infrastructure (V2I) communication are possible. Transmission control protocol provides end-to-end, reliable and congestion controlled connections over the Internet. It is required to understand congestion control algorithm for TCP, as it is heart of TCP. Out of all TCP implementations, TCP WestwoodNR has dynamic adjustment of congestion window and has control of congestion window with the help of end-to-end estimation of bandwidth. TCP WestwoodNR continuously estimates the available end-to-end bandwidth based on rate of Acknowledgement reception. Efforts are made to evaluate different mobility models for TCPs on VANET and as part of it different mobility models were implemented and evaluated. Routing protocols plays vital role in performance of any network, hence three of the most commonly used routing protocols DSR, DSDV and AODV were evaluated. In case of random drops due to bit error rate, which are frequent in case of wireless networks, WestwoodNR's bandwidth estimation algorithm gives best results amongst all.

*Keywords:* *Vehicular Adhoc Network, TCP WestwoodNR, Network Simulator 2, BonnMotion, MANET, cwnd, RTT, Mobile Adhoc Network.*

**71. PaperID 30061620: A Fuzzy Based Approach for the Elicitation of Attributed Values of Goal Models (pp. 525-536)**

*Sameena Naaz, Farhana Mariyam*  
*Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jamia Hamdard, New Delhi, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* - Requirements elicitation is an important process of requirements engineering. Different requirements elicitation methods have been proposed in literature to identify the need of the requirements of the software. For example, goal oriented requirements elicitation, group requirements elicitation process etc. Among various requirements elicitation methods, goal oriented requirements elicitation method is a promising method in which the

need of the stakeholders are identified using goal models. In goal models, the high level objective of the stakeholders are decomposed and refined into sub-goal. These sub-goals are further refined and decomposed until the responsibility of the sub-goal is assigned to some stakeholder or software. After refining and decomposing the goals, we will get the set of requirements. Therefore, it is an important research issue that how to select and prioritize the requirements for the next release of software development. In real life application, several stakeholders participate in requirements elicitation and they may have different opinion for the same requirement. Therefore, the objective of our work would be to propose a method to elicit the attributed values of goal models under fuzzy environment.

*Keywords:* Fuzzy set, Goal Model, Group Decision Making, Goal oriented requirements elicitation process, Requirement Engineering.

**72. PaperID 30061621: Real Time Lane Departure Warning System for Drivers (pp. 537-547)**

*Kadir İLERİ (1), Salih GÖRGÜNOĞLU (2)*

(1) Department of Electrical & Electronics Engineering, Karabuk University, Karabuk, Turkey

(2) Department of Computer Engineering, Karabuk University, Karabuk, Turkey

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* - Recently, real time driver warning systems have received a new increased interest for safety driving in the automobile industry. In this paper, we present real time vision based lane departure warning system. First, the captured image is divided into two parts as a road part and a non-road part by using the camera geometry information. Then, inverse perspective mapping is applied to avoid disadvantage of perspective effect. Next, gradient method is used to filter lane marks and Canny edge detection is applied. Additionally, Hough transform method is used for lane marks detection. Finally, driver is warned according to right or left lane departure by using detected lane marks' angles. The system works accurately in real time with various weather conditions and different road types. Additionally, this system has implemented on different embedded systems and their performances have been compared.

*Keywords:* Lane Detection, Inverse Perspective Mapping, Hough Transform, Canny Edge Detection, Real Time

**73. PaperID 30061638: Building and Evaluating an Adaptive User Interface using a Bayesian Network Approach (pp. 548-565)**

*Rebai Rim, Mohamed Amin Maalej, Mahfoudhi Adel, Abid Mohamed*

*ENIS, Embedded Computer System, University of Sfax, Tunisia*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* - The World Wide Web is a popular platform for providing adaptive user interfaces. In the Web context, user modeling has been the subject matter of many studies. Several techniques can be exploited to the user's preferences. Bayesian Networks (BNs), in particular, provide an effective approach for constructing probabilistic models. This paper presented an adaptive user interface (Social Network) for web applications. First, our Bayesian user model was constructed. Learning algorithms were compared in order to train the Bayesian structure. Evidence, in a Bayesian network, is a point of inference methods and originates from information based on the variables observation. Two types of evidence were clearly defined: hard evidence and probabilistic evidence. In this paper, we were interested in updating a probabilistic evidence distribution represented by a Bayesian user model. Then, inference algorithms were used so that the user model could predict the user's preferences. The Bayesian Network was confirmed to be effectively able to predict the user's preferences by evaluating the inferred results of the necessary variables based on several scenarios. Finally, the adaptive user interface was confirmed to be more comfortable for use than the fixed user interface.

*Keywords:* user model, Bayesian network, soft evidence, adaptation, web interface, evaluation.

**74. PaperID 30061643: A New Approach in Steganography of Digital Images using Saliency Map Algorithm and Pixel Neighbors (pp. 566-571)**

*Meysam Ghasemi (1), Mehran Emadi (2)*

*(1) Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran*

*(2) Faculty of Electrical Engineering, Mobarakeh Branch, Islamic Azad University, Mobarakeh, Iran*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract:* In current years, called the age of communication, communication can be different from a simple interaction. Sometimes due to security issues, the information that is going to be communicated between a sender and a receiver should not have its real appearance. This is where the discussion of steganography arises. Steganography includes a large set of communication methods that hide the presence of the message. Some of these methods include micro points, digital signature, and steganography channels. Despite great progress in methods available, the carrier file content has not received much attention. In fact, all the existing algorithms work based on its designer's policy, not the file content. Thus, the purpose of this study is to focus on the content of the media file. In the proposed procedure, after selecting the media image, the image prominences will be selected by AIM algorithm and accurately separated from the surrounding area using active contour. Next, the input chosen for steganography is embedded in the new image without the prominent part, applying pixel neighbors method and LSB. In this study, color and black and white images will be used for steganography employing stated method. Based on the results of the tests, the proposed method has an average of 75 for PSNR peak signal noise rate that in some approaches has more than 15 percent improvement over other methods.

*Keywords:* *Steganography, Content-Based Steganography, Saliency Map, Pixel Neighbors Steganography*

**75. PaperID 30061647: Performance Improvement in MapReduce via Overlapping of Mapper and Reducer (pp. 572-588)**

*Saurabh Gupta, Manish Pandey*

*Maulana Azad National Institute of Technology, Bhopal, 462003, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* — The MapReduce model supports big data processing on cluster by specifying mapper and reducer function. User defines a mapper function to process input key-value pairs and produces intermediate key-value pairs. Reducer function merges all the values for the same key and produces output key-value pairs. Hadoop is one of the popular framework which supports MapReduce model. In the traditional MapReduce model, Hadoop forces reducer to start its execution after all mappers finishes its execution. In turn, this causes an inefficient utilization of system resources and also impacts the performance. To overcome the limitation of traditional Hadoop, this article proposes two approaches which together solves the above mentioned limitation. The first solution, overlapping of mapper and reducer i.e. starts reducer task as soon as a predefined number of map tasks completed. The second solution, hierarchical reduction, in which there are several stages of reducer task. When reducer task completed its processing on the data that is generated by corresponding mapper task, another stage of reduce task is started. By combining both the solutions, three algorithms i.e. PageRank, Kmeans and WordCount are implemented in this article. The experimental results have shown that the speedup can be achieved by 6.5%, 7.02% and 10.38% over the traditional Hadoop for WordCount, Kmeans and PageRank applications respectively.

*Keywords* - distributed computing, MapReduce, Hadoop, cloud computing.

**76. PaperID 30061659: An Optimized Hybrid Multi-Digit BCD Adder Using QCA (pp. 589-598)**

*D. Ajitha, Dept. of E.C.E, Research Scholar, JNTUA, Ananthapuramu, India.*

*K. V. Ramanaiah, Dept. of E.C.E, Y.S.R Engineering College of Yogi Vemana University, Proddatur, India.*

*V. Sumalatha, Dept. of E.C.E, JNTUA College of Engineering, Ananthapuramu, India*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract** - Quantum-dot Cellular Automata (QCA), a possible alternative to contemporary CMOS technology. QCA is gaining its prominence in the digital circuit due to its high device density and clocking speed. In this paper, a Hybrid Multi-digit BCD adder (HMDBA) design is presented to perform the decimal addition with the optimized area and reduced delay. The HMDBA is constructed with the cascading of an enhanced architecture using Single digit BCD adder design. The HMDBA accomplishes the two 4-digit, 8-digit Decimal addition with 36% higher speed compared to the CFA-based BCD adder with the slight overhead of the area. The HMDBA occupies 38% and 29% decrease in the cell count, 23% and 14% less in the area compared to the CLA-based adder. Furthermore, the overall cost of the HMDBA is decreased by 50% compared to the CLA-based BCD adder.

**Keywords:** *Quantum-dot Cellular Automata (QCA), Decimal Adder/Binary Coded Decimal (BCD) Adder, CFA Type-II Adder, Multiplexer, Nanocomputing.*

**77. PaperID 30061661: Optimization and Traffic Management in IEEE 802.16 Multi-hop Relay Stations using Genetic and Priority Algorithms (pp. 599-616)**

*Jawwad Ibrahim (1), A. Rehman (1), M. Saad Bin Ilyas (1), Mohsin Shehzad (2), Maryum Ashraf (1)*

*(1) Department of Computer Science & Information Technology, The University of Lahore, Gujrat, Pakistan*

*(2) Huawei Technologies Co., Ltd. Islamabad, Pakistan*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract** - Wireless networks have become a widely accepted technology for rapid access to network infrastructure by remote locations. The Cooperative relaying strategy is considered as one of the most effective solutions for maximum utilization of a wireless network. Paper provides the study on the placement of Relay Stations to achieve an efficient and scalable design in wireless networks. A framework will be developed to optimize the capacity and to meet the minimum traffic demand of subscribers. To utilize the benefits of relaying, relay station placement problems is formulated and bandwidth allocation into an integer linear program that can be easily solved by any Artificial intelligence tool.

**Index Terms :** *WiMAX, Multi-hop Relay Stations, Relaying Techniques, Genetic Algorithm, Priority Algorithm*

**78. PaperID 30061672: A Novel Approach for Ranking Authors in an Academic Network (pp. 617-623)**

*Muhammad Farooq, Department of Computer Science, Government College, Rehmatabad, Rawalpindi, Pakistan.*

*Hikmat Ullah Khan, Department of Computer Science, COMSATS Institute of Information Technology, Wah, Pakistan.*

*Tahir Afzal Malik, Department of Management Information Systems, King Khalid University, Abha, Kingdom of Saudi Arabia.*

*Syed Muhammad Saqlain Shah, Department of Computer Science and Software Engineering, International Islamic University, Islamabad*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

**Abstract** - Ranking the authors in an academic network is a significant research domain to find the top authors in various domains. We find various links based ranking algorithms and index based approaches to measure the productivity and impact of an author in a social network of authors. The research problem to rank the experts has vast applications such as advisor finding, domain expert identification. In this paper, we propose a novel approach to rank the scholars in the academic network of DBLP, a well-known computer science bibliography website. A huge data set is prepared covering the publications of more than 70 years. We propose AuthorRank and Weighted AuthorRank

algorithms based on the state of the art ranking algorithms of PageRank and weighted PageRank algorithms respectively. For weighted algorithms, existing methods lack to provide diverse weights. We introduce the novel weights of h-index, gindex and R-index and elaborate their impact to identify the top authors in the scholarly network. The results confirm that the proposed algorithms find the top authors in an effective manner.

*Keywords:* Ranking, Social Network, PageRank, Academic Network.

**79. PaperID 30061686: A Semantic Multi-Agent Architecture for Multilingual Machine Translation (pp. 624-635)**

*Salam Fraihat, Qusai Shambour, Mou'ath Hourani*

*Software Engineering Department, Faculty of Information Technology, Al-Ahliyya Amman University, PO Box 19328, Amman, Jordan*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* - Machine Translation is the use of computerized methods to automate all or part of the translation process from one natural language into another. Machine Translation systems used to overcome the language barriers, for example, by making digital information understandable to people across the world in minimum amount of time. A Multiagent system is a software system that consists of multiple active, task-oriented and autonomous intelligent agents. Such agents can communicate and coordinate between each other in order to produce high quality solutions to complex problems in different domains. The semantic web is realized by adding semantics to the web in which it gives well-defined semantic meaning of information. It makes it possible to facilitate the representation, interpretation, sharing, searching, and reusing of information. This paper proposes a Semantic Multi-Agent Architecture for Multilingual Machine Translation system. In the proposed architecture, the multi-agent technology and ontologies will be integrated to produce collaborative working environment for multilingual machine translation. The automatic reasoning capacity of agents and their collaboration will improve the quality of the translation process. While, the incorporation of semantic features of languages, using ontologies, can be effective in increasing the quality of translations as such features focus more on the intended meaning of words rather than their syntactical structure.

**80. PaperID 30061694: Language and Security for None English Speakers (pp. 636-644)**

*Ibrahim Mohammed Alseadoon, Rabie A. Ramadan, Ahmed Y. Khedr*

*College of Computer Science and Engineering, Hail University, Hail, KSA*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* - Security is one of the major issues especially when dealing with bank accounts and money related business. Since most of the world are now using websites, e-mails and SMS messages to handle business, the security issues becomes a critical issue. For instance, we may receive at least a couple of phishing e-mails every day. These messages most of the time are written in English language. For none English speakers, it may be hard not to fall in the phishing trap due to their language illiteracy. Throughout this paper, we try to study the impact of English language on the security of none English speakers. The conclusion will be drawn based on the results obtained from our experiments.

**81. PaperID 30061695: A Proposed Technique for Preventing Criminal Attack (pp. 645-649)**

*Gamal H. Eladl, PhD*

*Mansoura University, Faculty of Computers and Information Sciences, Egypt*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* - Today, visitors of bank systems are growing and securing its important buildings and their staffs are become veryessential. Although, all banks have different rigid physical security systems with the support of many men's security bodyguards with monitoring cameras but there are large numbers of attacks, crimes, and victims. Moreover, crime's execution doesn't need more time to happen. So, this paper introduces a proposed technique in order to prevent any bank teller from suddenly dead. The proposed technique converts the bank teller's predefined face emotion (secret key) into a silent alert and sending a warning/alert message such as an SMS message to the bank security staff. The proposed technique is based on Artificial Neural Network (ANN) that will be used to detect the known bank teller's face emotion (secret key) and convert it into a silent alert with the assistance of high capabilities of the smart cameras that will be used as a pattern recognition system. The proposed technique will be helpful, more secured rather than the existing model. It will be used to prevent crime execution in all bank systems and securing their bank staffs.

*Keywords:* *Criminal Attack, Face emotion, ANNs, Smart Camera, Bank system.*

**82. PaperID 30061697: Modeling and Optimization the Four-Level Integrated Supply Chain: Sequential Quadratic Programming (pp. 650-669)**

*Abolfazl. Gharaei, Seyed Hamid Reza. Pasandideh,*

*Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract*- In this paper, we modeled a four-level integrated supply chain, contains a supplier, a producer, a wholesaler and multiple retailers. These four levels interacted and agreed with each other on having the same period length and same number of stockpile for each product in order to make an integrated chain to minimize total cost of supply chain. Products in this model are multi stage and there is limitation on production time capacity for producing the products. Other constraints of this model include: limitation on "total procurement cost, production cost, number of orders, space cost, number of stockpile for each level and setup cost". Objectives are to find both the number of agreed optimum stockpile and the agreed optimum period length for products that levels agree to minimize total inventory cost of chain while the constraints are satisfied. Problem model is nonlinear and large, so sequential quadratic programming (SQP) as one of the best exact optimization methods for solving nonlinear and large problems is used to solve this model. Three numerical examples are solved in order to demonstrate the applicability of this model and to evaluate SQP optimum performance. The results illustrate that SQP method has high efficiency in terms of optimum solutions, number of iterations to achieve the optimum solution, infeasibility, optimality error and complementarity for solving research nonlinear and large model. At the end, a sensitivity analysis is performed on the change rate of the obtained integrated objective function based on the change rate of the number of stockpile.

*Keywords-* *Four-level integrated supply chain; Multi Stage Products; Stockpile; Nonlinear Programming; Sequential Quadratic Programming; Period Length.*

**83. PaperID 30061698: Different Languages Classification Engine (pp. 670-681)**

*Khalaf F. Khatatneh, Samer S. Khanfar*

*Faculty of Graduate Studies, Al-Balqa'a Applied University, Assalt, Jordan*

**Full Text:** [PDF](#) | [Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)

*Abstract* - The Multi-Lingual Classification Engine (MLCE) is an automatic text classification system. Concerned data here are unstructured data; they are in a text fashion. The MLCE had been applied on Arabic and English languages as a model can be used for any other language. Initially, this work reviewed applications of MLCE, and then it reviewed related works and previous tries that aim to create automatic classification system. It has listed some complexities, difficulties and complications in both languages Arabic and English. It described principles of the design of (MLCE); these principles are components of the life cycle of MLCE. Finally, Experiments have been done by using MLCE and results have been registered and discussed.

*Keywords:* Automatic Classification; Machine Learning; Classification Engine; Classification System; Preprocessing; Naïve Bayes; supervised Learning.

**84. PaperID 300616105: Level Set Segmentation of Oil Spill Images with Non-Separable Wavelet Transform (pp. 682-693)**

*Ganta Raghatham Reddy (1), Ramudu Kama (2), R. Srikanth (3), Rameshwar Rao (4)*

*(1) Research Scholar, Department of ECE, OUCE, Osmania University, Hyderabad,*

*(2-3) Dept.of ECE, KITS-Warangal-15,*

*(4) Rtd. Professor, Department of ECE, Osmania University, Hyderabad,*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* - We propose, a new region based method for segmentation of the oil spills in the SAR satellite images with a fast level set model using non-separable quincunx wavelets. In the Synthetic Aperture Radar (SAR) images of oil spills, due to heavy leakages of oils on sea surface, which is in form of capillary waves some areas appear brighter which amounts to presence of glitter. Segmentation of these images is still a tedious task and cumbersome. The main reason is the large amount of inhomogeneity present in the background and foreground of image. The automatic segmentation of such images is very difficult due to glitter presence in the SAR images. The conventional methods like C-V model leads to improper segmentation with unconvincing results. We proposed an efficient segmentation method on oil spills images with level set approach using non-separable quincunx wavelets. The accuracy of segmentation greatly depends on the coefficients of the quincunx wavelet transform. We modified the Signed Pressure function (SPF) by combining it with quincunx wavelet domain. This new approach is very helpful for detection of the oil spill regions accurately. Satisfactory and convincing results are obtained when compared with conventional methods.

*Keywords*— SAR, Oil spills, wavelet toolbox, Quincunx wavelet transform, Level sets, Image Segmentation and SPF.

**85. PaperID 300616106: A Fast Priority-Flood Algorithm with Pruning for Depression Filling in Hydrologic Analysis (pp. 694-705)**

*Ruchi Lawaniya, Manish Pandey*

*Maulana Azad National Institute of Technology, Bhopal, 462003, India*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract*— The Digital elevation models are widely used spatial data source to incorporate the topographic information within geographical and hydrological applications. Depressions in DEM are lower areas surrounded by surface without any outlet. They interrupt or disconnect the flow path and create inaccurate drainage pattern. Subsequently, recognizing and removing the depression is a vital necessity for any hydrological study, which is commonly done prior to the use of DEM to conduct the hydrologic analysis. Usually, handling the depressions is a time consuming task for applications of huge terrain dataset with high resolution. This paper presents an improvement on priority-flood algorithm for recognizing and processing the depressions based on gridded digital elevation model in digital terrain analysis. The improvement on previous method is done by introducing a novel concept of pruning the dead cells from the priority queue. The priority queue cells that will never be used for further computation are considered as dead cells. Pruning of the dead cells can reduce the number of cells in the priority queue. Thus, the overall running time of Insertion and Deletion operations within the priority queue is asymptotically decreased. The proposed Priority Flood Pruning algorithm runs in  $O(K \log K)$  time, where  $K$  is the number of cells present in the priority queue after pruning. The proposed Priority Flood Pruning algorithm shows 1.13x to 1.25x speedup.

*Keywords* - Digital Elevation Model, depression filling, priority-flood, hydrologic analysis

**86. PaperID 300616109: Content-Based Video Browsing: Semantic Similarity and Personalization (pp. 706-724)**

*Jamel SLIMI, Anis Ben AMMAR, Adel M. ALIMI*

*REGIM-Lab.: REsearch Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* - In this paper, we present an intelligent video browsing system covering all tasks in video data visualization process. Visualization process is composed by categorization step followed by a representation of video collection step. The specificity of our work resides in the integration of personalization module allowing an appropriate interface to the user preferences. Our tool is based on multimodal video indexing (video text extraction, audio features and visual features). Video Indexing allows the construction of video data descriptor vectors. Based on these vectors, we calculate semantic similarity distance between documents composing video collection. This task permits a semantic classification of video corpus. Obtained classes will be projected in the visualization space. Video data visualization graph is in the form of a network. This network is composed by nodes (keyframes extracted from video shot) and color edges representing the similarity distance between data collection. Visualization interface components comportment is inspired from biological neuron comportment. By clicking on keyframe representing document; all the documents which are strongly connected to this one will be posted in the visualization space. An important step in our tool is dedicated to integrating personalization module in the video data visualization system. Personalization is based on user preferences collection. These preferences are collected via user interaction with the system. User profile is based on static indicators, dynamic indicators and navigation history. Compared to existing video browsing; our system includes a personalization module allowing appropriate interface to the user preferences. Network form of visualization representation permits easier navigation in large video corpus.

*Keywords:* *data visualization, video semantic similarity, personalization, video indexing, content-based video browsing.*

**87. PaperID 300616137: Creating and Configuring Cloud Computing Environment (pp. 725-738)**

*Vivek Thapar, Ph.D. Research Scholar, I.K.G. P.T.U. , Kapurthala, Punjab, India*

*Assistant Professor, Department of Computer Science and Engineering, G.N.D.E.C., Ludhiana, Punjab, India*

*Dr. O.P. Gupta, Associate Professor and Head, School of Electrical Engineering. & Information Technology, PAU, Ludhiana, Punjab, India*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* - Cloud computing is gaining popularity in delivering services to users in efficient and cost effective manner. Various services are offered to users in pay as you go model. The basic cloud services include Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Apart from this, various other services like Business Process as a Service (BPaaS), Testing as a Service (TaaS), Integration as a Service (InaaS), and Governance as a Service (GaaS) etc. are some other emerging cloud services. SaaS service model is the most popular service model of cloud .Under this model, the user is offered software or application on a subscription basis. The user uses the application (App) of the provider to interact with the resources of the cloud. The App needs to be loaded at various datacenters of the provider to improve the performance and response time. As Cloud environment is a complex and dynamic environment, testing the performance of the App on the real cloud environment is a very difficult task. Simulation Tools and techniques can be used to test the performance of App before being actually deployed in the real environment. .In this article first, we explore the CloudAnalyst simulation tool to simulate complex cloud environment. We explore the various packages and classes of the simulation tool. Then we use the simulation tool to create and configure a virtual cloud environment to test the performance of App on the cloud.

*Keywords:* *CloudAnalyst, Simulation of Cloud, SaaS, Cloud Environment, Facebook.*

## **88. PaperID 300616141: Goal Modeling Techniques for Requirements Engineering (pp. 739-746)**

*Nagy Ramadan Darwish (1), Bassem S.M. Zohdy (2)*

*(1) Information Systems Department, Cairo University, Cairo*

*(2) BTEC Department, Canadian International College, Cairo*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — This research aims to introduce the goal oriented requirements engineering GORE, also defining what is meant by goal, the main differences between goal and requirement, also the types of goals and the sources of extracting these goals, in addition, the birth of goal modeling techniques and the reason behind using goal modeling, at last the goal oriented approaches, early and late requirements goal modeling techniques, this research tries to get out with the result of how goal modeling is very important in requirements engineering, in order to extract the goals and requirements in correspondence to business context, which in turn will aid in better analyses and extract the functions and processes in any organization or business.

*Keywords* — *Goal Oriented Requirements Engineering GORE, Goal Modeling Techniques, Requirements Engineering RE.*

## **89. PaperID 300616142: Using Multiple Criteria Decision Making Approaches to Assess the Quality of Web Sites (pp. 747-761)**

*Rim Rekik, REGIM-Lab: Research Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia*

*Ilhem Kallel, REGIM-Lab: Research Groups in Intelligent Machines, University of Sfax, Tunisia*

*ISIMS: Higher Institute of Computer Science and Multimedia of Sfax, University of Sfax, Tunisia*

*Jorge Casillas, Department of Computer Science and Artificial Intelligence, University of Granada and the Research Center on Information and Communications Technology (CITIC-UGR) Granada, E-18071, Spain*

*Adel M. Alimi, REGIM-Lab: Research Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* - Multiple Criteria Decision Making (MCDM) is a widely used discipline in everyday life especially to make decisions about conflict and multiple criteria that need to be evaluated and analyzed. In this paper, the aim is to explore the known MCDM techniques to assess web sites information in specific domains or identify the current developments in on-line literature. Based on applying a Systematic Literature Review (SLR) process, this paper identifies MCDM methodology and provides a comparison of existing research. Further, the analysis highlights the features and limitations of MCDM methods. In order to assess the quality of web sites, it requires a list of criteria and sub-criteria. The metrics depend on web site category that generally the decision makers choose the suitable ones. So, weighing criteria in MCDM problems are usually used to determine their importance. The evaluation with crisp MCDM methods is not largely used. The trend is to make hybridization among them or a combination with fuzzy reasoning.

*Keywords:* *Quality assessment, Multiple Criteria Decision Making, Preferences, Fuzzy numbers*

## **90. PaperID 300616157: Analysis of Rank Aggregation Techniques for Metasearch: A Case Study (pp. 762-774)**

*Parneet Kaur (1), Manpreet Singh (2), Gurpreet Singh Josan (3)*

*(1) Department of Computer Science and Engineering, I.K.G Punjab Technical University, Jalandhar, Punjab, India*

*(2) Department of IT and Computer Science and Engineering, GNDEC, Ludhiana, Punjab, India*

*(3) Department of IT and Computer Science and Engineering, Punjabi University, Patiala, Punjab, India*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* - For surfing the internet many users rely on search engines but results are not fully effective. This gave birth to the invention of Meta-search Engines (MSEs), which merge and aggregate results from multiple search engines to derive user preferred and efficacious results. MSE takes the query from users and supply it to different search engines which in turn provide the various decisions as well as ranking of query. Hence, the cornerstone of all these processes used by MSE is directly or indirectly depends upon the merging techniques of ranking which uses Rank aggregation methods. Rank Aggregation prominence on combining of non-identical rank ordering which is applied on similar type of data set or candidates to refine the rank order. Rank Aggregation techniques are applied for numerous applications like voting, social network, metasearch under search engine performance check and selection. This paper focuses on various Rank Aggregation methods with implementation on real world dataset.

*Keywords* Meta Search Engine, Rank Aggregation, Rating, Metasearch , Rapid Miner.

**91. PaperID 300616164: Construction of a Jacobi matrix by given n Eigenpairs (pp. 775-784)**

*Seyed Abolfazl Shahzadeh Fazeli, Soodeh Kakuei Nejad, Maryam Babaie  
Parallel Processing Laboratory, Faculty of Mathematics, Yazd University, Yazd, Iran*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract*- In this paper, an algorithm for construction of a Jacobi matrix is proposed by given some eigenvalues and corresponding eigenvectors. Then we discuss about solvability of this problem with n eigenvalue, and some sufficient conditions for existence of the solution are proposed. Finally, a generalized method for this problem by given n eigenpairs is proposed. At the end a numerical algorithm and some examples are presented.

*Keywords:* Inverse problem, Jacobi matrix, Eigenpair

**92. PaperID 300616165: Enhancing Genetic Algorithms using Multi Mutations: Experimental Results on the Travelling Salesman Problem (pp. 785-801)**

*Ahmad B. A. Hassanat (1)\*, Esra'a Alkafaween (2), Nedal A. Al-Nawaiseh (3), Mohammad A. Abbadi (4),  
Mouhammd Alkasassbeh (5), and Mahmoud B. Alhasanat (6)  
(1,2,4,5) IT Department, Mutah University, Mutah, Karak, Jordan  
(3) Department of Public Health and Community Medicine, Mutah University, Mutah, Karak, Jordan  
(6) Department of Civil Engineering, Al-Hussein Bin Talal University, Maan, Maan, Jordan*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* — Mutation is one of the most important stages of genetic algorithms because of its impact on the exploration of the search space, and in overcoming premature convergence. Since there are many types of mutations one common problem lies in selecting the appropriate type. The decision then becomes more difficult and needs more trial and error to find the best mutation to be used. This paper investigates the use of more than one mutation operator to enhance the performance of genetic algorithms. New mutation operators are proposed, in addition to two election strategies for the mutation operators. One is based on selecting the best mutation operator and the other randomly selects any operator. Several experiments were conducted on the Travelling Salesman Problem (TSP) to evaluate the proposed methods. These were compared to the well-known exchange mutation and rearrangement mutation. The results show the importance of some of the proposed methods, in addition to the significant enhancement of the genetic algorithms' performance, particularly when using more than one mutation operator.

*Index Terms*— Mutation operator, Nearest Neighbor, Multi Mutations, TSP, GA, AI, Evolutionary Computation.

**93. PaperID 300616177: Hybrid Metaheuristic Optimization based on ACO and Standard PSO applied to Traveling Salesman Problem (pp. 802-823)**

*Sonia Kefi (#), Nizar Rokbani (\*), Adel Mohamed Alimi (#)*

*(#) REGIM-Lab: Research Groups in Intelligent Machines, University of Sfax, ENIS, Sfax, Tunisia*

*(\* ) Higher Institute of Applied Sciences and Technology of Sousse, University of Sousse, Sousse, Tunisia*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* - Hybrid Metaheuristics Optimization have emerged along the paradigm itself. Now, they are very famous because the hybrid metaheuristics methods give best results for combinatorial optimization problems compared to exact methods. In this paper, we will propose Metaheuristic method which are applied to difficult problems. This method is based on hybridization between population based solution methods like Ant colony optimization (ACO) and standard Particle swarm optimization (SPSO) algorithms and single based solution methods like 2-Opt algorithm. Our developed approach is called "Standard Ant Supervised by PSO" (SAS-PSO-2Opt) applied to routing problem like Traveling Salesman Problem (TSP), which is considered as NP-complete problem. Therefore, the ACO algorithm can explore the search space, PSO algorithm is used to optimize the ACO parameters and the 2-Opt algorithm improves the obtained solution and reduce the probability of falling into a local minimum. To evaluate our proposed hybrid approach, we have used several standard tests benches from TSPLIB and we have compared the results with other hybrid metaheuristics approaches from litterature.

**94. PaperID 300616178: Generative Software Development Techniques of User Interface: Survey and Open Issues (pp. 824-842)**

*Thouraya SBOUI (1), Mounir BEN AYED (2)*

*(1) Department of Computer Science, National Engineering School of Sfax, Sfax, Tunisia*

*(2) Department of Computer Science and Communication, Faculty of Science of Sfax, Sfax, Tunisia*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* - the multiplication of digital devices and multimedia development has led to the amplification and the variety of User Interface (UI). Much research of late has focused on the User Interface development process from the task analysis stage up to the code generation stage. Due to the complexity of User Interface construction, most approach now uses a generative software development approach which recommends abstraction and reusability to achieve more efficient software, resulting in quicker results at a lower cost. Additionally, Software Product Line is a software engineering paradigm which transposes the industrial product line into a software development process in order to create a collection of similar software systems. The development in Software Product Line is based on the management of a set of features that satisfy the specific needs of a particular market segment or mission developed from a common set of core assets in a prescribed way. In this regard and to construct User Interfaces, this paper will present, a small survey of generative processes dedicated to develop User Interfaces with a special focus on Software Product Line approaches.

*Keywords—UI development; generative software process; MDE; MBUID; SPL.*

**95. PaperID 300616181: Energy Aware Resource Management for Cloud Data Centers (pp. 844-853)**

*Santosh Kumar, Manish Pandey*

*Maulana Azad National Institute of Technology, Bhopal, India*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract* — In modern age a huge number of different kinds of applications are processed by data centers. These data centers establishment incur high cost in purchasing IT resources and their maintenance. Cloud computing model

facilitates creation of extensive scale virtualized data centers with the goal that clients can utilize them on interest on a compensation as-you-go premise. These data centers consume unprecedeted amount of electrical energy which increases the overall operating cost and carbon dioxide emission. Energy consumption of cloud data centers can be reduced by using dynamic consolidation of virtual machines (VMs) which optimizes their resource usage. In dynamic consolidation of VMs based on lower threshold and upper threshold of utilization, VMs migrate live from one host to other and idle nodes are switched to sleep mode which results optimized resource usage and less energy consumption. However, providing high quality of service to the customers brings issue of energy-performance tradeoff. Since workloads experienced by applications are variable, VM placement need to be optimized online on a regular basis. This paper proposes an adaptive VM consolidation approach which determines upper threshold to detect if a host is overloaded or not based on an analysis of historical information of resource usage. The proposed strategy significantly lessens the energy consumption while fulfilling the Service Level Agreement (SLA) to a high level of adherence. This article shows simulation results of proposed strategy using real-world workload traces of PlanetLab.

*Keywords:* Cloud computing, Dynamic VM consolidation, Resource management.

**96. PaperID 300616191: A Survey on Association Rule Hiding Methods (pp. 854-860)**

*Mahsa Roghanian, Mohammad Naderi Dehkordi*

*Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* - Rapid growth of information technology has led to creation of huge volumes of data which will be useless if they are not efficiently analyzed. Therefore, various techniques have been provided for retrieving valuable information from huge amounts of data, one of the most common of which is mining association rules. As much as data mining can be important for extracting hidden knowledge from data, it can also reveal sensitive information, which has created some concerns for data owners. Thus, the issue of hiding sensitive knowledge and preserving privacy was raised in data mining. In this paper, different methods for preserving privacy was studied and by mentioning advantages and disadvantages of each method, a suitable platform was provided for researchers to be able to implement the best technique for sanitizing the considered database.

*Keywords:* Data Mining; Association Rule mining; Privacy Preserving; Hiding Sensitive Knowledge

**97. PaperID 31031656: A Light Weight Secure Protocol for Data Transmission in Vanet (pp. 861-870)**

*Shawar Gul, Noor Ul Amin, Faisal Bahadur, Insafullah, Moazzam Ali Khan*

*Department of Information Technology, Hazara University, Mansehra, Pakistan, NUST, Pakistan*

**Full Text:** PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

*Abstract* - Vehicular Ad-hoc Network (VANET) is becoming the next generation networking technology. It provides the communication among Vehicle to Vehicle (V2V) or vehicle to Road Side Unit (RSU) using wireless communication. However, vehicular ad-hoc network faces different security issues because of open air communication of information which needs to be resolved. This paper presents a light weight secure protocol for data transmission in VANET. The protocol is based on the hardness of hyper elliptic curve cryptography using authenticated key exchange with road side unit aiming to secure VANET communication. So, the proposed protocol meets the security properties such as authenticity, confidentiality, non-repudiation, unforgeability, Integrity. The protocol also reduces the computational cost 48.11% as compared to the existing scheme. Our scheme is best suited for the vehicle communication system.

*Keywords:* VANET; RSU; HECDM; DLP; ECDLP

## **98. PaperID 310516198: Dynamic Edge Detection in a Digital Video Stream Using Kirsch Filters (pp. 871-878)**

*S. Aparna, Department of Computer Science Asst. Professor GITAM University Research Scholar JNTUH  
Dr. M. Ekambaram Naidu, Principal, ARJUN College of Engineering and Technology, Hyderabad*

**Full Text:** PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

*Abstract:* Edge detection plays a vital role in various image processing applications. Basically the term ‘edge’ refers to sharp edges of discontinuity in images and edge detection allows one to locate boundaries of various regions, each region having certain uniform pixel values. One of the applications of edge detection in streaming video is an area of recent research. One can use any of the many algorithms that are available in the standard literature. Kirsch filter seems to be computationally efficient as well as in detecting edges without losing the image content. This paper presents the results of an intensive research carried out in locating moving boundaries regions present in a streaming video using Kirsch directional filters.

*Keywords:* Edge detection, Kirsch compass kernel, Video image processing

## **99. PaperID 300616143: An Approach for Scheduling Problem on Single Machine (pp. 879-883)**

*Omar Selt (1), Tarak Benslimane (2) and Thameur Abdelkrim (3)*

*(1) Laboratory of pure and applied mathematics, Department of Mathematics, University of M'sila, Algeria*

*(2) Department of Electrical Engineering, University of M'sila, Algeria*

*(3) Unité de Recherche Appliquée en Energies Renouvelables, URAER, Centre de Développement des Energies*

**Full Text:** PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

*Abstract* — This paper considers the elaboration of tabu search approach to solve a scheduling problem of n tasks on single machine. This problem is strongly NP-difficult, which makes finding an optimal solution looks impossible. To improve the performance of this approach, we used, on one hand, different diversification strategies (T1 and T2) with the aim of exploring unvisited regions of the solution space, and on the other hand, we proposed three types of neighborhoods (neighborhood by swapping, neighborhood by insertion and neighborhood by blocs). It must be noted that tasks movement can be within one period or between different periods. Besides that, all data in this problem are supposed to be integer and deterministic. The weighted sum of the end dates of tasks constitutes the optimization performance criterion for the problem treated in this paper.

*Keywords:* Scheduling, single machine, NP-difficult, Tabu search

## **100. PaperID 300616118: RICA: Reform based Imperialist Competitive Algorithm for Mapping Applications to Network on Chip based, Many-core architectures (pp. 884-890)**

*Alireza Mahini, Computer engineering department, Islamic Azad University, Gorgan, IRAN*

*Hossein Pedram & Seyedeh Fateme Hosseini, Computer engineering department, Amir Kabir University, Tehran, IRAN*

**Full Text:** PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

*Abstract* — One of the most important problems in designing the many-core architectures on Network on Chip (NoC) platform is task mapping. In this article, we are concerned with proposing a method for mapping aimed to reduce the consumed energy and utilize the Imperialist competitive algorithm which is called RICA. Reform policy has been proposed instead of revolution in algorithm and results proved reason of this selection. Implementation of RICA in MATLAB, and comparison of it with the previous methods, shows that reduction of energy consumption, and in similar conditions, it reaches better results in less iteration than genetic-based algorithms.

*Keywords - Imperialist competitive algorithm, manycore processor, task mapping*

**101. PaperID 300616120: An Overview of Service Oriented Architecture, Cloud Computing and Azure Platform (pp. 891-896)**

*Kamran Shaukat, Muhammad Umair Hassan, Haider Ali, Muhammad Shah Zaib, Muhammad Muhibb Ullah  
Department of Information Technology, University of the Punjab, Jhelum Campus, Jhelum, Pakistan*

**Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]**

*Abstract —* This research paper is about the cloud computing, and benefits of service oriented architecture. Its objective is to make a uniform cloud computing model that will permit the general population to move starting with one cloud supplier then onto the next easily. This paper discusses service oriented architecture and why it is so imperative in the consequent objective of a unified architecture. There are a number of areas described in which cloud applications are being used whether it is being used in healthcare organizations or in cloud technology platforms. Manjrasoft Aneka describes the rapid creation of scalable applications, and their development on various types of clouds in a seamless and elastic manner. At the end we have described an overview of five Microsoft Azure scenarios by using which we can make such an environment through which a user can switch to different clouds using the service oriented architecture.

*Keywords—Service Oriented Architecture (SOA), Organization for Advancement Structured Information Standards (OASIS), Application Service Providers (ASPs)*

# HITH: a novel Hybrid IP Traceback approach for Heterogeneous wireless networks

Ikbel Daly, Faouzi Zarai, and Lotfi Kamoun

LETI laboratory  
University of Sfax, Tunisia

**Abstract**—Among the most critical attacks in wireless networks, there is the Denial of Service (DoS) attack. This threat is becoming increasingly vulnerable with heterogeneous wireless networks. To remedy this attack, it is fundamental to identify the source of attack by exploiting an IP traceback technique. There are two major categories of approaches; packet marking and packet logging. In packet marking, it is characterized by adding supplementary information to mark packets. This method moderates the problem of overhead but requires a large amount of packets to reconstruct the attack path. In packet logging, it is based on saving packets in digest tables. This approach enables the identification of attack source through a single packet but necessitates a huge storage space. In this paper, we propose a novel Hybrid IP Traceback for Heterogeneous wireless networks, which is called HITH (Hybrid IP Traceback for Heterogeneous wireless network). Our solution presents a precise IP traceback method with low overhead storage and improved accuracy. To evaluate the effectiveness and the feasibility of HITH approach, we use mathematical analysis and simulations. The results of a comparison with an existing solution in literature prove the capacity to trace a single IP packet while reducing storage overhead and data access time.

**Keywords**-Heterogeneous Wireless Network; Security, Hybrid IP traceback; Marking packet; Logging packet; Denial of Service attack.

## I. INTRODUCTION

The infrastructure of communication becomes increasingly heterogeneous following the integration of several technologies in order to meet the growing needs of the users' community. This heterogeneity has several mechanisms and techniques that are characterized by a specific composition and precise services and features.

To ensure this connectivity, we use several methods and strategies. Indeed, interoperability depends on the network topology, traffic pattern, interference, etc. According to [1], the connectivity of low-priority network component depends on the characteristics of high-priority component, in order to ensure the diversity of techniques and to reduce infrastructure complexity.

The heterogeneity notion is adopted through the variety of benefits guaranteed by several types of wireless networks and its integration with numerous technologies. This combination takes advantage of assets provided by each existing technology and method in its cover space. First, it expands the capacity and functionality offered by the diversity network types. In

addition, the request of higher bit rate, the efficiency of networks utilization and the flexibility can be relatively assured.

In order to ensure the networks' heterogeneity, primordially we must procure the interworking between existing technologies from the third generation (3G) to the fifth generation (5G). Fig. 1 illustrates an example of Heterogeneous Wireless Networks (HWN) which integrates the two technologies; LTE (Long Term Evolution) and Wireless Mesh Network (WMN). Referring to [2] and [3], there is an analysis of the different interworking solutions in HWN. Furthermore, [4] presents an overview of the various networking implementations within interworking architectures.

The primary concern for such a wireless network remains the problem of security. Indeed, the freedom of users' mobility and the rapid development of data transmission technology may intensify the network vulnerability. Especially in heterogeneous wireless network, characterized by the integration of multiple network types and wireless technologies, the problem of security is becoming more critical [5]. Consequently, this issue may deserve urgent and effective solutions to ensure secure communications and maintain the confidence between network equipments.

Among the most vulnerable attacks, we find DoS (Denial of Service) [6] and DDoS (Distributed DoS) attacks [7]. Their purpose is to make such a machine, server or network unreachable. This type of attack presents the most spread vulnerability in the internet [8] which can affect multiple targets, even critical, in a very short duration.

To apply these attacks, attackers conduct the IP Spoofing technique to hide their real identities. This can complicate the task of tracking and monitoring the real source of attack. However, in wireless networks, the mission of traceability, known by IP Traceback, becomes increasingly difficult due to limited resources and energy in the various equipments and the lack of infrastructure and unpredictable routing behaviors in particular with ad-hoc wireless networks [9].

The remaining part of this article is organized as follows. In section II, we introduce some related works which treat the existing IP traceback techniques. Section III describes our proposed hybrid IP traceback approach for Heterogeneous wireless Networks in details. Section IV evaluates the performance of our approach through mathematical analysis and simulation.

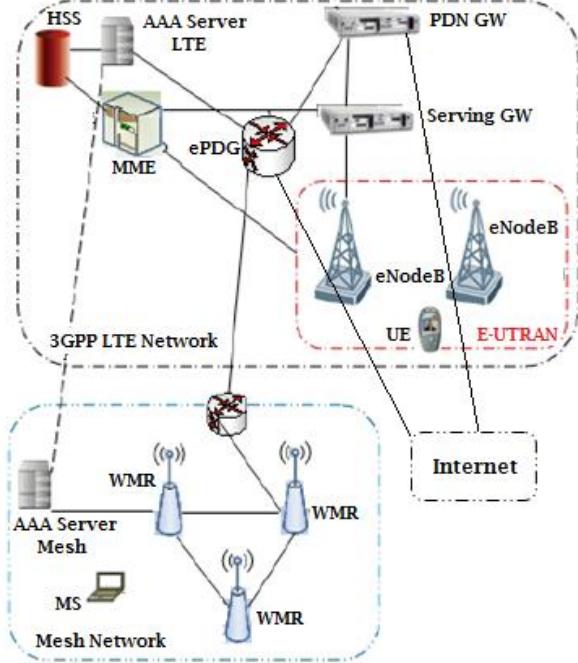


Figure 1. Example of Heterogeneous Wireless Networks topology (LTE-Mesh).

Section V presents a discussion on the issues of mobility, Handover, Splitting and Merging. Finally, Section VI summarizes the article.

## II. RELATED WORKS

In this work, we treat the reactive defense methods and in particular techniques for source attack identifying. In reality, the knowledge of the origin of such vulnerability serves to protect a network from different types of attacks and even those which may occur in the future. For this reason, researchers do not stop treating this topic by proposing new solutions and techniques [10] for IP traceback.

### A. IP Traceback Techniques

Therefore, there is a variety of IP traceback methods including primarily:

- Probabilistic Packet Marking (PPM) [11]
- ICMP traceback (ITrace) [12]
- Hash-based IP traceback (Packet logging) [13]
- Hybrid IP traceback [14]

1) *Probabilistic Packet Marking (PPM)*: This solution is based on the idea of adding supplementary information to packets IP header traversing the network routers. This information is used to identify the routers, which participate in the packets traffic, and then build the IP packets traversed path from source (attacker) to destination (victim).

The marking action can be performed either by using some bits in the IP packet header, or by generating a new packet based on router's address or a part of its address. This idea appeared with [11] and has been improved in several works

such as [15], [16] and [17]. The choice of marking packet depends on a probability value in the order of 0.04 [11].

The major constraint to this method is the need to collect a large amount of packets in order to identify the source of attack. But thanks to works done in [18] the number was reduced to 1000 packets. In addition, this approach uses two additional functions; marking and path reconstruction. The first must be implemented at all network routers.

This addition, which requires the handling of packets, brings additional work with processor and therefore causes a processing overhead. Although the first implemented function necessitates the execution of supplementary operations in routers, it is still less than the requirements to apply reconstruct the true attack path.

On the other hand, the addition of software at routers can be independent to the other network devices. This allows a better network scalability. Furthermore, the packet can be transformed or modified in order to falsify the traceback results. For this reason, PPM supports different packets transformations within an environment without reflectors to ensure the sureness of IP traceback procedure results. Thanks to its efficiency, PPM can identify the source of majority of DoS and DDoS attacks.

A variety of packet marking techniques have been proposed; we quote as examples the technical Probabilistic Packet Marking (PPM) [18] Proactive Signaling Architecture for IP Traceback (PSAT) [19] and Hybrid messaging-based scheme for IP traceback [20] as an improvement of PSAT for wired network.

2) *ICMP traceback (ITrace)*: This method is based on the concept of adding a message named "ICMP traceback message" or "ITrace" [12]. First, the router selects a packet from 20000 forwarding packets. Then, it generates a packet containing the ITrace message. Finally, this information will be oriented to the same destination. This message includes additional data about previous and next Hop (router), a timer (timestamp) and the exploitation of the TTL (Time To Live) field to reconstruct the IP packets traversed path.

In order to handle this method, we need to implement two functions; ITrace and path reconstruction. Therefore, routers must be improved with the aim of designing new services. This addition does not affect the operation of other equipments that compose the communication network. This independency ensures a better scalability.

With the intention of building the suitable path towards the source of attack, we need thousands of packets. As with PPM, the addition of new functions requires more processor requirements in particular with ITrace which generates a new packet. However, those changes do not deserve a wide range of additional memory. Similarly, this method supports the transformation and modification of packets except with reflectors. Then, a second solution was proposed as an improvement of ITrace, known by Intention-Driven ICMP traceback [21]. This new technique processes the DDoS attacks.

*3) Hash-based IP Traceback (Packet Logging):* This technique is known as SPIE (Source Path Isolation Engine). The idea of this method, presented in [13], is based on the store of information related to packets traversed network. This data can be a digest or a packet signature, which will be saved in intermediate routers. Foremost, these routers are asked to extract fields from the IP header and the first 8 bytes of the field "payload" of each packet. Then, this information undergoes a hash function in order to produce digests.

Indeed, approaches based on the hash technical such as [22] and [23] have appeared in order to reduce the storage space required for packet logging operations by exploiting the concept of packet digests and hash function instead of the utilization of the whole packets. The main drawback for this IP traceback solution is the huge amount of resources reserved for storing packet digests. In order to reduce the required storage space, routers operate a space-efficient data structure technique called Bloom filter [24].

Concerning the network infrastructure, the packet logging approach requires the addition of new equipments to ensure the storage packets procedure. Therefore, the router software must be upgraded to implement three additional functions which are:

- Data Generation Agent (DGA)
- SPIE Collection and Reduction Agent (SCAR)
- SPIE Traceback Manager (STM)

In addition, this evolution requires a modification in various network components configuration. For this reason, packet logging approach does not provide the scalability needed in heterogeneous wireless network. Also, the network will be overloaded by the repetitive action of saving digests in each router.

On the other hand, the primary advantage of this technique is the ability to establish the attack path through a single packet. Moreover, even with packet modification or transformation, packet logging is able to provide reliable and effective results for most DoS and DDoS attacks.

*4) Hybrid IP Traceback (HIPT):* This method is based on two techniques; Packet marking and Packet logging. This alliance benefits from the advantages provided by each approach. Indeed, hybrid IP traceback is characterized by the reduction of reserved resources for packet marking and the utilization of a small number of packets to identify the attack source by using saved digests [25], [26], [27]. Despite the multiplicity of advantages, this combination does not eliminate the drawbacks brought by the two used techniques mentioned in parts 1 and 3 of this current section.

#### B. Evaluation of IP traceback techniques

This subsection involves a comparative study between the different IP traceback techniques mentioned in the previous subsection (A). Based on [28], a representative method in each category was evaluated. Indeed, the proposed scheme in [29] was selected to represent the Probabilistic Packet Marking technique, [12] is chosen as a representative ICMP based traceback technique, SPIE [22] represents the Packet Logging

method and RIHT [25] is chosen under Hybrid Traceback scheme. The comparative study of IP traceback techniques is based on the following evaluation metrics:

- ISP Involvement: in some IP traceback schemes, in order to trace the attack route, we call for ISP (Internet Service Provider) intervention to provide some additional information aimed to identify the sources of attack.
- Number of attack packets: the number of packets required to determine the source of attack.
- Processing overhead: presents the additional processing related to the traceback scheme. It can take place in two levels, either in ISP's devices or in the part of victim.
- Protection: to address this matter, we should consider the non-belonging of equipment to its network if this device becomes subverted.
- Scalability: in some traceback schemes, we need to add some new devices in the network. Such complementary equipment may require an independent configuration or with others devices. Thus, minimizing the dependency improves the scalability of the scheme.
- Memory requirements : present the quantity of additional storage needed at the network equipments. This metric can be computed in two levels ; at the network components (routers and servers) and at the victim.
- Accuracy : this metric evaluates the precision of IP traceback method by defining the false positive and the false negative parameters.
- Knowledge of Network : precises if the IP traceback scheme requires a prior knowledge about the topology of studied network.
- Ability to handle major DDoS attacks: this metric proves the capacity of the scheme to perform the traceback of DDoS attacks under rigorous circumstances such as IP spoofing, and manipulation of reflectors [30].

Based on these selected evaluation parameters, Table I illustrates a comparison between different IP traceback techniques.

### III. PROPOSED SOLUTION

#### A. Motivation

In section II, we have mentioned the most common IP traceback approaches as well as their positive and negative points. The Hybrid IP traceback technique is presented as the most appropriate approach to heterogeneous wireless networks. Indeed, the diversity of networks and the interworking between various technologies require the incorporation of IP traceback methods in order to benefit from a variety of features and to give effective results. Regarding implementation, the hybrid method can be applicable in our studied network, characterized by the variety and the heterogeneity of technologies.

On the other hand, this IP traceback technique has proved its efficiency and feasibility. In literature, most of the proposed hybrid methods focuses on the study of one approach and disregards the other. Indeed, some researchers improve the storage overhead and neglect the reconstruction of the attack IP packets traversed path and others treat the problem of false paths keeping a huge amount of storage.

TABLE I. COMPARISON OF IP TRACEBACK TECHNIQUES

Evaluation Metrics	PPM	ITrace	Packet logging	Hybrid Scheme
<b>ISP Involvement</b>	Fair	Good	Poor (Huge memory requirement)	Fair
<b>Number of attack packets</b>	Large number of packet	Number of ICMP messages and huge number of attack packets	One packet	One packet
<b>Processing overhead (Router)</b>	Medium	Low	High	Low
<b>Protection</b>	Good	Good and practically feasible	Poor	Poor
<b>Scalability</b>	Poor	Good	Fair	Fair
<b>Memory requirement (Network)</b>	Not required	Not required	Very High	Low
<b>Memory requirement (Victim)</b>	Very High	Medium	Not required	Not required
<b>Accuracy</b>	Medium (Huge false positive rate in case of DDoS attack)	Good for less numbers of attackers	Medium with high false positive and false negative	High (less false positive and false negative rate)
<b>Knowledge of Network</b>	Not needed (Faster traceback and low false positive if known)	Not needed	Not needed	Not needed
<b>Ability to handle major DoS attacks</b>	DoS/DDoS flooding attacks	DoS/DDoS network layer attacks	DoS/DDoS flooding attacks	DoS/DDoS flooding attacks

Indeed, some researchers improve the storage overhead and neglect the reconstruction of the attack IP packets traversed path and others treat the problem of false paths keeping a huge amount of storage. In this work, we deal with this nuance by proposing a new hybrid IP traceback approach, which is more efficient and robust with minimal storage space.

### B. Main idea

In this subsection, we describe the principle of our proposed solution named HITH (Hybrid IP Traceback for

Heterogeneous wireless network). This approach needs to meet several specifications related to the studied environment and the conditions of traceback process implementation.

On one hand and as it is mentioned in the introduction part, HWN is characterized by a variety of wireless networks in which each category has its own properties. On the other hand, the proposed model allows encompassing the notion of mobility by studying the interworking between different networks and the handover procedure in layer 3. In order to ensure reliability and robustness of our proposed IP traceback solution and in particular sureness of the established attack paths, the traceback model is based on a set of assumptions:

- Network routers are trusted
- The attacker does not know in advance the traceback mechanism
- The IP header can be changeable

HITH is a hybrid IP traceback method, which owns the packet marking capabilities and the storage of some network information. This combination of two traceback techniques ensures the effectiveness of the proposed approach and the reduction of storage overhead problem.

The principle of HITH method is based on the identification of the first packet for each new connection crossing a new router. Indeed, following the receipt of a packet, the router checks a field called "LogMark". If it is set to 0, the router proceeds with the logging operation. Otherwise (i.e. the value of LogMark is different to 0), the router proceeds by packet marking method.

*1) Marking Procedure:* In this part, we describe in details the packets marking mechanism at routers for the studied network. First of all, we must point out that our proposed solution is valid for IPv4 packets. And for IPv6 packet, more researches and improvements should be procured to HITH since this new version includes new features that should be considered by our method in future work.

The majority of marking approaches use the 3 fields of IP header; "Identification", "Reserved flag" and "Fragment offset". Similarly, HITH uses these three fields to perform the marking operation and logging digests at network routers.

*a) Identification:* This field is carried on 16 bits. It is useful in the case of packets' fragmentation, since it designates the fragment number. A study [31], performed in the Internet environment, shows that less than 0.25% of packets crossing the network is fragmented. Therefore, the research community in the field of IP traceback processes the ability to reuse this field for other functionalities. Indeed, a study of feasibility of this proposal and also of incompatibility problems of that field introduces a topic of discussion in [11]. Ultimately, the obtained results confirm the hypothesis of "Identification" field reusing.

As illustrated in Fig. 2, HITH replaces the "Identification" field by two new fields; "LogMark" and "Router ID". For "LogMark", this first element is carried on 2 bits and is used to precise the operation to be performed by the router either marking or logging. A logical combination of two bits can

compose four possible values; 00, 01, 10 and 11. This field is initialized to (00) by the first router. Then, it undergoes an incrementing until reaching the last value (11). Afterwards, it returns to the first value (00) and increments once again till the attainment of destination. This value reveals the operation to perform at router level.

The second part of identification field indicates the router identifier (Router ID). This information may be fixed to a length of 14 bits based on [32], which illustrates a study performed on the Internet and also the concept of routers neighborhood. Reference [32] asserts that the allocation of 14 bits allows identifying routers uniquely.

*b) Reserved flag:* This field is carried on one bit which is not yet assigned. In the domain of IP traceback, [33] proposes the use of this field to execute the marking operation.

*c) Fragment offset:* In the IP header and following the packet fragmentation, this field is used to specify the offset of the fragment from the original datagram (64 bit words). But because of the elimination of the "Identification" field, "Fragment offset" becomes useless. This facilitates its exploitation in favor of IP traceback approaches [34] and [35]. Indeed, HITH operates these two fields ("Reserved flag" and "Fragment offset") to denote the identifier of the previous marking router (Last Marking Router ID). This identifier is generated by the network administrator for each traceback-enabled router.

*2) Logging Procedure:* To pursue and rebuild the attack path and minimize the storage space at routers, various traceback approaches apply hash functions on the data set from the IP packet to extract the digests. With hash-based approach, digests are derived from the integration of different IP header fields (except TTL, TOS "Type Of Service" and checksum) with the first 8 bytes of payload.

These fields are still operated in the hybrid traceback approach with the addition of a new field named "Logging flag" which introduces the innovation brought by this type of traceback mechanism. In our solution, HITH calculates the packet digest in the same way as PPIT (Precise and Practical IP Traceback) approach presented in [36] by integrating diverse parts of the IP header stated in the hybrid approach excluding the TTL field.

*a) TTL integration in digest:* In this part, we show through an illustrative example the utility behind the addition of TTL field in the input of digests and subsequently in the reconstruction of attack path. In Fig. 3, we present a sample network topology exploiting HIPT (Hybrid IP Traceback) approach, composed by 14 routers. The attack path is exposed through continuous red arrows and dotted arrows show the returned path to identify the source of attack.

Since this network adopts our hybrid IP traceback approach, routers (R1, R7, R10 and R14) executes packet digests storage procedure, known by logging and the other routers (R2, R3, R4, R5, R6, R8 R9, R11, R12 and R13) carried out the marking procedure.

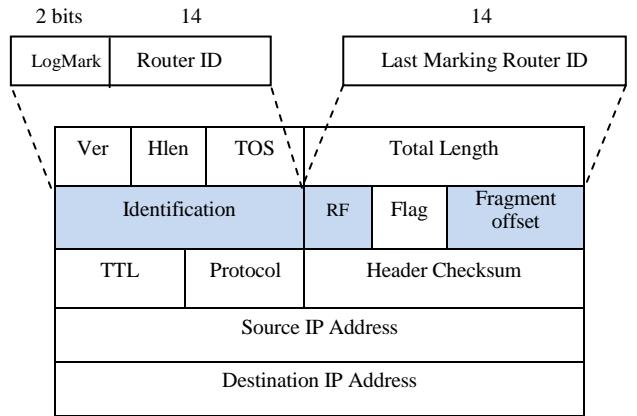


Figure 2. Structure of IP packet Header in marking procedure with HITH.

Following the detection of an attack, the traceback approach begins with the reconstruction of the attack path by identifying the concerned routers. Firstly, the hybrid approach recognizes the first router R14 since it is directly related to the victim.

Thanks to the implementation of the logging operation, we can identify the next router by referring to the router identity recorded in the packet digest. Then, the packet marking provides information on the previous router in the field "Last Marking Router ID" filled in place of "Fragment offset". On arrival at R11 router, this latter sends queries to its neighbors (R7 and R10), which are both included in the attack path. And since the responses to requests will be received randomly, it can be considered the case where the path is continued with the router R10. Consequently, that may falsify the construction of the attack path.

To remedy this problem, we must differentiate between the two responses received from R7 and R10 by a time reference which may be presented in the IP header by the TTL field. In this case, by comparing the TTL fields from digests of R7 and R10 packets, we find that the TTL value of R7 is lower than that of R10. Thus, this comparison proves that the attack packet has traversed R10 before reaching R7. This scenario can be executed only if we include the TTL field in the logging process as a part of packet digest.

*b) Digest Table:* The HITH approach saves packet digests in a digest table implemented with Bloom Filter method [24], which reduces the storage overhead and make the storage procedure more convenient. As illustrated in Fig. 4, Bloom Filter uses (k) hash functions to calculate (k) distinct packet digests, each one is composed of (n) bits. These results are indexed in a list of (2n) bits, initialized to 0. To ensure the rapidity of the reconstruction process of the attack path, we have adopted the idea of multiplying digest tables between neighbors, which is exposed in HIT (Hybrid single-packet IP Traceback) approach [37]. The routers in this approach are characterized by the management of different digest tables at the same time. Each table is associated with one or more routers identities (Router ID) used in the packet marking procedure.

After extracting the packet digest, this latter will be recorded in digest tables which are necessarily associated with the router identity supported by the concerned IP packet. Indeed, when a router decides to execute the logging operation, it looks first at the router identity on the contemplated packet (Router ID). Then, it stores the resulting digest in the corresponding table. In this context, we note that the existence of a given table depends on the router vicinity, which can find a table with its identity in each of its neighbors.

Therefore, packets from different routers can undergo the logging operation and be stored in different tables simultaneously. This may cause the reduction of access time to digest tables because this parameter is no longer proportional to the packet arrival rate but also to the maximum rate of arrival packets from the whole neighborhood router. Despite this reduction, the results of the adopted traceback mechanism remain depending on the capacity of each router essentially the memory access time. Indeed, if a router is characterized by a bad memory access time, it cannot take advantage of traceback mechanism benefits. To remedy this problem, this category of low-speed router profits from access to digest table associated with all neighboring routers, carrying all identities instead of giving to each router its own table.

In case of table saturation, packet digests will be archived for a period of time that depends on the configuration and the requirements of the studied network. Thanks to Bloom Filter procedure, the storage overhead of these tables for each router is still negligible and does not affect the network quality of service or the performance of the adopted traceback mechanism.

### C. Traceback process

In order to clarify the steps of identifying attack path by applying our IP traceback approach HITH, we adopt the same network topology mentioned in Fig. 3. The studied network has four logging routers (R1, R7, R10 and R14) and the other entities constitute marking routers. First, the HITH procedure identifies R14 since it is directly related to the victim.

Then, it joined to the attack packet the value of R14 router identity, its own TTL value and the different values of the neighboring routers identities. After calculating the digest by exploiting the same hash function as in the logging process, we proceed by comparing the obtained results with the entries in the digest table. In case of correspondence, the concerned router presents the next hop in attack path reconstruction.

After identifying the router R13, the next step is to broadcast queries to all routers' neighbors. Upon receiving this request, each router examines all digest tables referring to the time interval of packet reception to carry on with correspondence. Thanks to the exploitation of TTL value, we avoid the risk of false paths. Therefore, R12 then R11 routers are identified for the attack path reconstruction. Then, we proceed in the same manner to achieve the router R1 in order to rebuild the path leading to the source of attack.

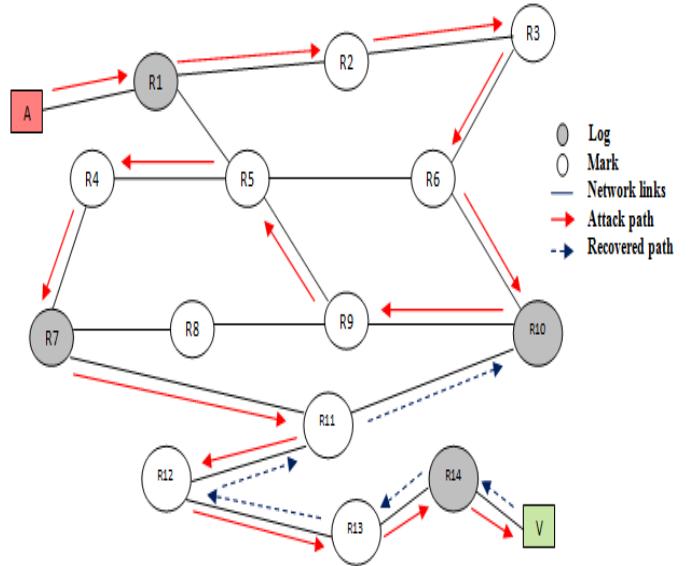


Figure 3. Example of false attack path reconstruction with HIPT.

## IV. PERFORMANCE EVALUATION

In this section, we evaluate our proposed approach HITH with analytical methods and simulation by comparing it to some existing solutions; HIT (Hybrid single-packet IP Traceback) [37] and SPIE (Source Path Isolation Engine) [22].

### A. Traceback Accuracy

Traceback accuracy presents a very important criterion in the evaluation of IP traceback mechanisms. Indeed, it defines the success rate of attack path reconstruction and subsequently the sureness of the obtained sources. However, a traceback mechanism, which does not ensure this specification, may give false results, incorrect paths, and finally the failure of the traceback procedure.

On the other hand and owing to the adoption of existing techniques in such a traceback mechanism, some imperfections can be inherited from these procedures.

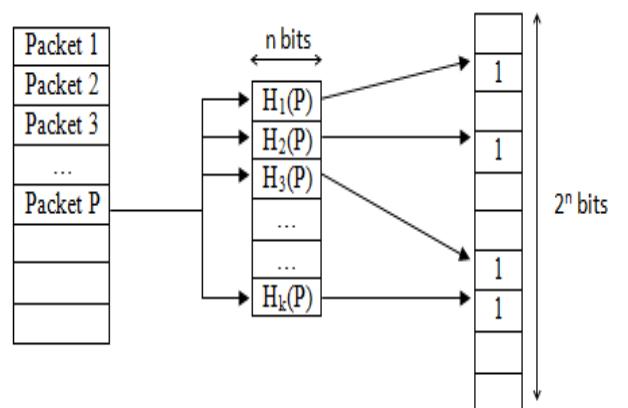


Figure 4. Bloom Filter procedure.

TABLE II. BLOOM FILTER PARAMETERS

Parameter Symbol	Description
A	Number of elements (packets)
B	Number of bits
a/b	Memory efficiency factor
P	False-positive rate
K	Number of hash functions

1) *False-positive rate*: Bloom Filter introduces a space-efficient data structure that is used to organize elements and then to check membership in this set. During the phase of membership test, Bloom Filter can produce false-positive results and never false-negative results. We suppose the set of Bloom Filter parameters which is listed in Table II.

$$P = \left(1 - \left(1 - \frac{1}{b}\right)^{ka}\right)^k \approx (1 - e^{-ka/b})^k \quad (1)$$

The false-positive rate (P) depends on the memory size of Bloom Filter as well as the size of the stored digests. Therefore, it is exponentially related to the value of memory efficiency factor (a / b) [38].

From Equation (1), the false-positive rate (P) can be managed and controlled by the right choice of the factor (a / b) [22] and the (k) hash functions. In our approach, HITH uses Bloom Filter procedure with the addition of the TTL field. This addition does not affect the rate (P) because we use (k) hash functions with the same management of digest table.

2) *IP traceback Precision*: This parameter must be taken into account since the design phase of such an IP traceback mechanism to achieve precise and accurate results. Therefore, this precision presents an essential factor in ensuring the success of traceback by tracing the attack path perfectly and identifying the true source of intrusion. According to the example mentioned in Fig. 4, we can conclude that HIT approach still suffers from some problems and precision vulnerabilities that cause the uncertainty of the obtained results. In HITH, this problem is solved by the introduction of TTL field in the packet digest in order to reduce the risk of erroneous and incorrect paths.

### B. Storage Overhead

In this part, we borrow the evaluation methods from HIT. Indeed, we evaluate the storage overhead criterion following two parameters:

- Digest Table Storage (DTS): it reveals the quantity of memory required for the registration of packet digests in a router.
- Digest Table Access time (DTA): it designates the number of packet digests stored in a table per time unit.

Similarly to SPIE, hybrid traceback approaches can resort to a single packet or transformed packet in order to identify the source of attack. During the design phase of traceback approaches, we must not neglect the case of packets transformation. Indeed, IP packets may undergo various transformations, such as fragmentation and tunneling, crossing the network. In this context and for our HITH approach, stored packets in routers can be:

- 1) IP fragments
- 2) non-fragmented packets to be logged at the router, comprising the following sub-cases:
  - a. Non-fragmented packet not logged in the two upstream routers.
  - b. Non-fragmented packet logged at the upstream router but transformed at the current router.
  - c. Non-fragmented packet logged in the upstream router two-hop away and transformed in the current router.

Similarly to HIT approach, we consider ( $P_l$ ) the percentage of packets to be logged at a router. We assume ( $\alpha$ ) the percentage of fragmented IP packets and ( $\beta$ ) the percentage of transformed packets in the router.

In addition, we set ( $Y$ ) to the percentage packet to be logged at router without fragmentation. A consolidated list of these percentages is shown in Table III. According to the parameters listed in Table III, the percentage of all of IP packets to be logged at the router is expressed by:

$$P_l = \alpha + (1 - \alpha) Y \quad (2)$$

TABLE III. PERCENTAGES OF DIFFERENT TYPES OF IP PACKETS

Type of IP packet	Percentage
1) IP fragments	$\alpha$
2) Non-fragmented packets to be logged at the router (includes 2.a, 2.b et 2.c)	$(1 - \alpha)Y$
2.a) Non-fragmented packet not logged in the two upstream routers.	$(1 - \alpha)(1 - Y)(1 - Y)$
2.b) Non-fragmented packet logged at the upstream router but transformed at the current router.	$(1 - \alpha)(1 - Y)Y\beta$
2.c) Non-fragmented packet logged in the upstream router two-hop away and transformed in the current router.	$(1 - \alpha)Y\beta$

The percentage of packets to be logged without fragmentation is:

$$Y = \frac{P_l - \alpha}{1 - \alpha} \quad (3)$$

And  $1 - Y = \frac{1 - P_l}{1 - \alpha}$  (4)

Since the second case of fragmentation includes three possible scenarios, (Y) can be expressed by the following equation:

$$Y = (1 - Y)^2 + Y\beta + Y(1 - Y)\beta \quad (5)$$

We replace the value of (Y) by (3) and  $(1 - Y)$  by (4) in (5), we obtain:

$$P_l = 1 - \frac{(1 - \alpha)(\sqrt{1 + 4(1 - \beta)^2} - 1)}{2(1 - \beta)} \quad (6)$$

Some measurement studies have proved that  $\alpha \leq 0,25\%$  [39] and  $\beta \leq 3\%$  [38], [31]. Therefore, we observe that:

$$0,382 \leq P_l \leq 0,392 \quad (7)$$

The obtained result demonstrates that 39% of packets crossing a router require the execution of logging operation. On the other side, with HIT approach and according to [37], the result is expressed by:

$$0,50 \leq P_l \leq 0,51 \quad (8)$$

This means that 50% of IP packets must be logged in the current router. In SPIE approach, all packets crossing the router require the execution of logging operation. If we consider  $DTS_H$ ,  $DTS_I$  and  $DTS_S$  the values of DTS in HITH, HIT and SPIE approaches respectively, we find:

$$DTS_H = P_l \times DTS_S \approx \frac{2}{3} DTS_I \approx \frac{2}{5} DTS_S \quad (9)$$

In addition, for our approach, logging packets can be performed in multiple neighboring routers simultaneously as detailed in digest table part. Therefore, the rate of access to a digest table may be reduced with the number of existing neighbors in network. We suppose that a router has (n) neighbors. In the ideal case where the traffic arrives to router in a balanced way from each neighbor, and:

$$DTA_H = P_l \times \frac{1}{n} DTA_S \cong \frac{2}{5} \times \frac{1}{n} DTA_S \quad (10)$$

With  $DTA_H$  and  $DTA_S$  represent the access time to digest table with HITH and SPIE approaches respectively. In the worst case, where all the traffic is derived from a single neighbor ( $n = 1$ ), we note that:

$$DTA_H = P_l \times DTA_S \cong \frac{2}{5} \times DTA_S \quad (11)$$

### C. Simulations

In order to approve the results of analytical methods, we conduct simulations by the Network Simulator NS2 [40]. For this method, we focus on the criterion of packet logging overhead. The simulation scenarios are designed based on a synthetic topology composed of 200 routers which 20 of them have more than 2 neighbors. In the remainder of this article, we identify this latter type of router with neighborhood by "headers". Furthermore, we consider the maximum number of hops by the value 20 with a total of 200,000 packets. For the achievement of the various scenarios, we consider the following assumptions:

- The packets traversing the network do not undergo any fragmentation or transformation.
- We consider two scenarios :
  - All packets are logged at the first router.
  - All packets are not logged at the first router.
- Each router is directly connected to a terminal host.
- Each host can send packets to any other host in network.

Under these conditions of simulations and with the fixed values of routers and packets, we count the number of logged packets and the number of transmitted packets by each router. Then, these values are used to calculate the logging probability of routers. Fig. 5 and Fig. 6 exhibit the various logging probability values of HIT and HITH approaches. For the different curves, the horizontal axis shows the logging probability value expressed by "x", and the vertical axis stands for the percentage of routers having lower probabilities than the value of x.

Fig. 5(a) and Fig. 6(b) reveal the results of simulations for the first scenario (i.e. all packets are logged at the first router) with HIT and HITH approaches for "headers" and all routers respectively. With the consideration of the neighboring routers, logging probability percentages of routers vary between 45% and 55% for HIT approach and between 35% and 40% for HITH approach.

These values are proportional to the percentage of the logging packets in routers obtained with analytical methods; (7) for HITH and (8) for HIT. In addition, this difference between these two approaches is even contemplated in Fig. 5(b) studying all routers in the network. This ascertainment may be justified by the exploitation of a new method with more efficient marking and logging operations for the proposed HITH approach.

On the other hand and considering the second scenario (i.e. all packets are not logged at the first router), the curves contemplated in Fig. 6(a) and Fig. 6(b) consolidate the observations made in Fig. 5(a) and Fig. 5(b). Indeed, for logging packet in routers except the first one, we note a variation of the logging probability percentages which may reach the value 70% of routers having logging probabilities less

than 50% for HIT approach and a value of 50% of routers having logging probabilities less than 39% for HITH approach. This is due to the dependence between logging probabilities of routers and the performed operation at the first router.

In summary, the simulation results obtained from different scenarios and for both approaches; HIT and HITH, show that between 45% and 55% of transmitted packets undergo logging operation for HIT approach and these percentages vary between 35% and 40% for HITH approach. This observation is in accordance with the analytical methods results for Digest Table Storage (DTS) and Digest Table Access time (DTA) criteria. The simulation results also prove that logging probability depends on the location of the router performing this operation either it is the first router on the network path or not.

## V. DISCUSSION

### A. Traceback and Mobility

This subsection highlights the impact of mobility on IP traceback mechanism in heterogeneous wireless network. Indeed, this type of network is characterized by a diversity of technologies and topologies that provides a dynamic architecture with mobile routers (particularly in Ad-hoc and Mesh networks). To ensure the effectiveness of IP traceback approach with the particular characteristics of HWN, we must minimize the execution time of traceback technique for marking or logging operation and for reconstruction of attack path. Furthermore, the storage space for the traceback procedure in routers must be minimized to avoid the storage overhead problem.

### B. IP Handover

HWN encompasses different technologies. Each type has specific network addresses. And even in the same type of network, there may be some divisions into sub-networks. Therefore, a mobile belonging to HWN can perform an intra-

domain or an inter-domain handover. In both cases and after the detection of network change, the mobile gets a new temporary address CoA (Care of Address). Then, the home network and the corresponding nodes need to be informed by the change of location and the addresses of its subscribers. Therefore, the idea of multiplying digest tables between neighbors can facilitate the smooth running of the traceback procedure. Indeed, the execution of the Handover phase and the allocation of new addresses such as CoA address allow the integration of new routers in the traceback mechanism and the cooperation with the original network components.

### C. Splitting and Merging

An Ad-hoc or Mesh network can coexist with other technologies to build a HWN. Both networks types are characterized by the routers' mobility. Therefore, IP routes can be changed following the modification of the studied network topology. Two possible cases can be performed following the routers' operation. For the first case, the architecture undergoes a partial change.

Then the traceback procedure takes into account these modifications by updating the marked routes to provide more efficient results. For the second case where we note a total change of architecture, the IP route is discovered once again. The traceback procedure will be restored from the first router that is informed by the link failure.

## VI. CONCLUSION

To ensure a more effective and precise IP traceback technique, the research community resort to combine the two existing approaches; packet marking and packet logging to establish a hybrid approach. Although this method inherits the benefits provided by each category, it still suffers from some vulnerability.

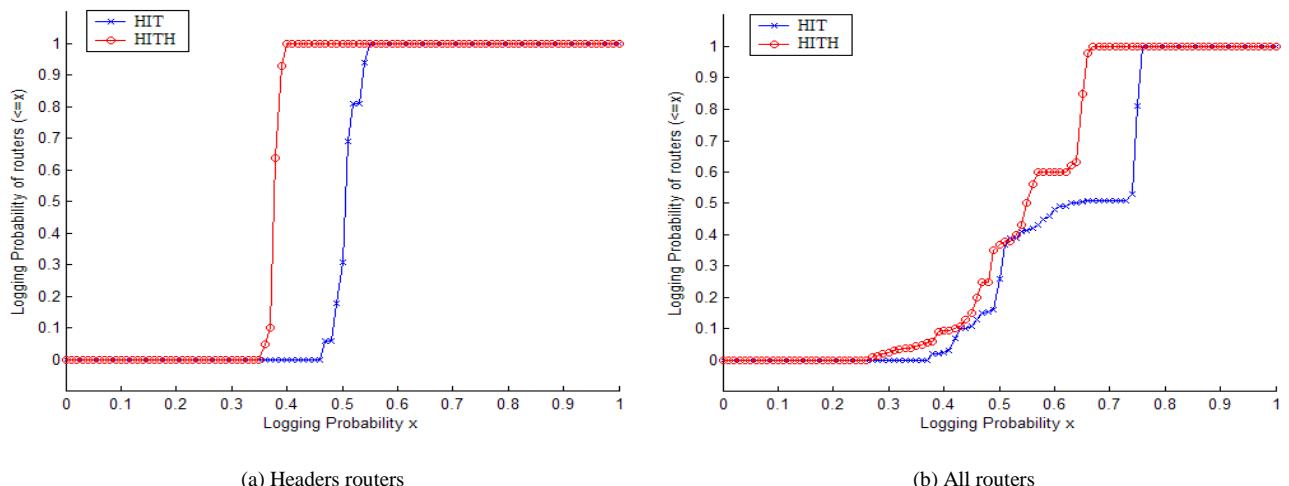


Figure 5. Simulation results of « logging probability » for scenario 1.

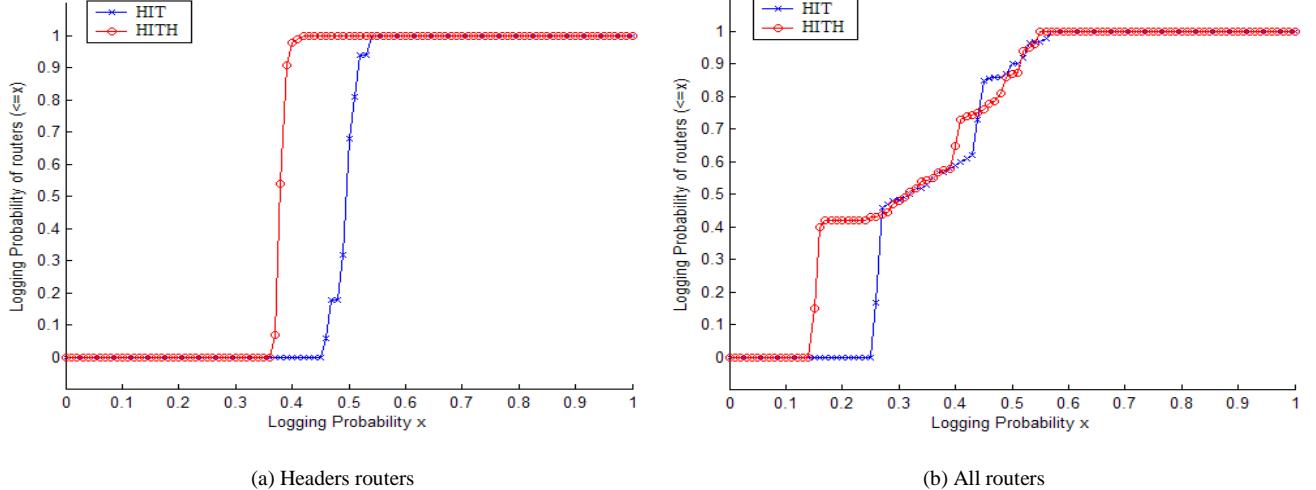


Figure 6. Simulation results of « logging probability » for scenario 2.

Indeed, the hybrid IP traceback approach can cause incorrect paths due to accuracy problems. In addition, overhead storage remains high because of the inefficient use of the marking space. In this article, we have proposed a novel Hybrid IP Traceback approach for Heterogeneous wireless networks, which is called HITH. Our solution is based on some supplementary information added in the reconstruction of the attack path to avoid incorrect results.

Moreover, HITH defines an efficient mechanism to reduce storage overhead by distributing the marking and logging roles between routers. Besides, in order to decrease the digest table access time, we have gathered the log information in multiple routers taking into account the notion of neighborhood and the limitation of some network equipments capacities. The effectiveness of the proposed IP traceback approach is proved by mathematical analysis and simulations. Indeed, HITH incurs little overhead at routers, improves accuracy and reduces overhead storage and data access time.

## REFERENCES

- [1] W. Ren, Q. Zhao and A. Swami, "Connectivity of Heterogeneous Wireless Networks". IEEE Transactions on Information Theory, vol. 57, no. 7, pp. 4315-4332, july 2011.
- [2] K. Andersson, "Interworking Techniques and Architectures for Heterogeneous Wireless Networks". Journal of Internet Services and Information Security (JISIS), vol. 2, no. 1, pp. 22-48, february 2012.
- Technologies: From Theory to Applications (ICTTA'08), pp. 1-6, Damascus, Syria, april 2008.
- [3] R. Ferrus, O. Sallent and R. Agusti, "Interworking in Heterogeneous Wireless Networks: Comprehensive Framework and Future Trends". IEEE Wireless Communications, vol. 17, no. 2, pp. 22-31, april 2010.
- [4] A. A. Atayero, E. Adegoke, A.S. Alatishe and M.K. Orya, "Heterogeneous Wireless Networks: A Survey of Interworking Architectures". International Journal of Engineering and Technology, vol. 2, no. 1, pp. 16-21, january 2012.
- [5] S. Xu, "On the security of group communication schemes". Journal of Computer Security, vol. 15, no. 1, pp. 129-169, 2007.
- [6] K.J. Houle and G.M. Weaver, "Trends in Denial of Service Attack Technology". Computer Emergency Response Team (CERT) Coordination Center, technical report v1.0, october 2001.
- [7] D. Phatak, A.T. Sherman, N. Joshi, B. Sonawane, V.G. Relan, A. Dawalbhakta, "Spread identity: A new dynamic address remapping mechanism for anonymity and DDoS defense". Journal of Computer Security, vol. 21, no. 2, pp. 233-281, 2013.
- [8] D. D. Moore, C. Shannon, D.J. Brown D.J, G.M. Voelker and S. Savage, "Inferring Internet Denial-of-Service Activity". ACM Transactions On Computer Systems (TOCS'06), vol. 24, no. 2, pp. 115-139, may 2006.
- [9] C. Zouridaki, B.L. Mark, M. Hejmo, R.K. Thomas, "Hermes: A quantitative trust establishment framework for reliable data packet delivery in MANETs". Journal of Computer Security, vol. 15, no. 1, pp. 3-38, 2007.
- [10] A. Roy, A. Datta, A. Derek J.C. Mitchell, "Inductive trace properties for computational security". Journal of Computer Security, vol. 18, no. 6, pp. 1035-1073, 2010.
- [11] S. Savage, D. Wetherall, A. Karlin and T. Anderson, "Network Support for IP Traceback". IEEE/ACM Transactions on Networking (TON), vol. 9, no. 3, pp. 226-237, june 2001.
- [12] S.M. Bellovin, "ICMP Traceback Messages", IETF draft, 2000, <http://www.research.att.com/smb/papers/draftbellovin-itrace-00.txt>.
- [13] G. Sager, "Security Fun with OCxmon and cflowd". Internet2 Working Group Meeting, november 1998, <http://www.caida.org/funding/ngi/content/security1198>.
- [14] K. H. Choi and H. K. Dai, "A marking scheme using Huffman codes for IP traceback". 7th International Symposium Parallel Architectures, Algorithms Networks (I-SPAN'04), pp. 421-428, Hong Kong, China, may 2004.
- [15] H.C. Tian, J. Bi, X. Jiang and W. Zhang, "A probabilistic marking scheme for fast traceback". 2<sup>nd</sup> International Conference on Evolving Internet (INTERNET'10), pp. 137-141 Valencia, Spain, september 2010.
- [16] P. Sattari, M. Gjoka and A. Markopoulou, "A network coding approach to IP traceback". IEEE International Symposium on Network Coding (NetCod'10), pp. 1-6, Toronto, Canada, june 2010.
- [17] A. Yaar, A. Perrig and D. Song, "FIT: Fast internet traceback". IEEE Conference on Computer Communications (INFOCOM'05), vol. 2, pp. 1395-1406, march 2005.
- [18] D.X. Song and A. Perrig, "Advanced and Authenticated Marking Schemes for IP Traceback". IEEE Conference on Computer Communications (INFOCOM'01), vol. 2, pp. 878-886, Arkansas, USA, april 2001.
- [19] A. Fadlallah and A. Serhrouchni, "PSAT: Proactive signaling architecture for IP traceback". IEEE 4<sup>th</sup> Annual Communication Networks and Services Research Conference (CNSR'06), pp. 293-299, Washington, DC, USA, may 2006.
- [20] A. Fadlallah, A. Serhrouchni, Y. Begriche and F. Naït-Abdesselam, "Hybrid messaging-based scheme for IP traceback". IEEE 3rd International Conference on Information and Communication Technologies: From Theory to Applications (ICTTA'08), pp. 1-6, Damascus, Syria, april 2008.
- [21] A. Mankin, D. Massey, C.L. Wu, S.F. Wu and L. Zhang, "On design and evaluation of intention-driven icmp traceback". IEEE International Conference on Computer Communications and Networks, pp. 159-165, october 2001.

- [22] A.C. Snoeren, C. Partridge, L. Sanchez, C. Jones, F. Tchakountio, B. Schwartz, S. Kent and W. Strayer, "Single-packet IP traceback". IEEE/ACM Transactions on Networking, vol. 10, no. 6, pp. 721-734, 2002.
- [23] A.C. Snoeren, C. Partiridge, L.A. Sanchez, C.E. Jones, F. Tchhakountio, S.T. Kent, and W.T. Strayer, "Hash-Based IP TraceBack". ACM Conference on Applications, Technologies, Architectures and Protocols for Computer Communication (SIGCOMM'01), pp. 3-14, august 2001.
- [24] B.H. Bloom, "Space/time trade-offs in hash coding with allowable errors". Communications of ACM, vol. 13, no. 7, pp. 422-426, july 1970.
- [25] M. H. yang and M. C. Yang, "RIHT: A Novel Hybrid IP Traceback Scheme". IEEE Transactions on Information Forensics and Security, vol. 7, no. 2, pp. 789-797, april 2012.
- [26] B. Sai Priyanka and N. Srihari Rao, "IP Traceback Techniques – A Selective Survey". International Journal of Computer Science and Mobile Applications, vol. 1, no. 3, pp. 40-44, september 2013.
- [27] Ch. Gong and K. Sarac, "IP Traceback based on Packet Marking and Logging", IEEE International Conference on Communications (ICC'05), vol. 2, pp. 1043-1047,16-20 may 2005.
- [28] V. Murugesan, M. Shalinie, N. Neethimani, "A Brief Survey of IP Traceback Methodologies". Acta Polytechnica Hungarica, vol. 11, no. 9, pp. 197-216, 2014.
- [29] M. T. Goodrich, "Probabilistic Packet Marking for Large Scale IP Traceback". IEEE/ACM Trans. Networking, vol. 16, no. 1, pp. 15-24, February 2008.
- [30] J. Mölsä, "Mitigating denial of service attacks: A tutorial". Journal of Computer Security, vol. 13, no. 6, pp. 807-837, 2005.
- [31] I. Stoica and H. Zhang, "Providing guaranteed services without per flow management". ACM conference on Applications, technologies, architectures, and protocols for computer communication (SIGCOMM'99), vol. 29, no. 4, pp. 81-94, Cambridge, MA, USA , october 1999.
- [32] M. Muthuprasanna, G. Manimaran, M. Manzor, and V. Kumar, "Coloring the internet: IP Traceback". 12<sup>th</sup> International Conference on Parallel and Distributed Systems (ICPADS'06), pp. 589-598, Minneapolis, USA, august 2006.
- [33] D. Dean, M. Franklin and A. Stubblefield, "An algebraic approach to IP traceback". ACM Transactions on Information and System Security, vol. 5, no. 2, pp. 119-137, 2002.
- [34] Z. Gao and N. Ansari, "Enhanced probabilistic packet marking for IP traceback". IEEE Global Telecommunications Conference (GLOBECOM'05), vol. 3, pp. 1676-1680, 28 november - 2 december 2005.
- [35] C. Gong and K. Sarac, "Toward a practical packet marking approach for IP traceback". International Journal of Network Security, vol. 8, pp. 271-281, 2009.
- [36] D. Yan, Y. Wang, S. Su and F. Yang, "A Precise and Practical IP Traceback Technique Based on Packet Marking and Logging". Journal of Information Science and Engineering, vol. 28, pp. 453-470, 2012.
- [37] Ch. Gong et K. Sarac, "A More Practical Approach for Single-Packet IP Traceback using Packet Logging and Marking". IEEE Transactions on Parallel and Distributed System, vol. 19, no. 10, pp. 1310-1324, october 2008.
- [38] A. Broder and M. Mitzenmacher, "Network applications of Bloom filters: A survey". Internet Mathematics, vol. 1, no. 4, pp. 485-509, 2005.
- [39] S. McCreary and K. Claffy, "Trends in wide area IP traffic patterns: A view from Ames internet exchange". 13<sup>th</sup> ITC Specialist Seminar on IP Traffic Modeling, Measurement and Management, pp. 1-25, Monterey, CA, USA, 2000.
- [40] Network Simulator (ns-2). <http://www.isi.edu/nsnam/ns/>.

# Developer Companion: A Framework to Produce Secure Web Applications

Mamdouh Alenezi

College of Computer & Information Sciences  
Prince Sultan University  
Riyadh 11586, Saudi Arabia  
malenezi@psu.edu.sa

Yasir Javed

College of Computer & Information Sciences  
Prince Sultan University  
Riyadh 11586, Saudi Arabia  
yjaved@psu.edu.sa

**Abstract**—Software engineering and development is a very complex endeavor that contends with limited resources, potentially causing software to behave in an unexpected manner. Software developers often lack secure coding skills and its a major reason behind development of insecure web applications. In this work, we propose a developer companion as an integrated framework that can be integrated to any IDE to educate and help developers produce more secure code. This framework can be adopted and can be made more intelligent by focusing on historical security flaws in the development team. expert developers practices to overcome the security vulnerabilities.

**Keywords**—web applications, source code, security, static analysis

## I. INTRODUCTION

Software development and engineering is a very complex endeavor that contends with limited resources [1], [2], potentially causing software to behave in an unexpected manner. Software developers often lack secure coding skills and its a major reason behind development of insecure web applications [3]. The most worrisome class of these faults can be exploited by attackers. These faults are considered security vulnerabilities [4] that are recurrent, causing companies to struggle to allocate resources for their management [5]. The practice of building secure software that functions properly under unwanted attacks is called software security.

Web applications nowadays have moved from static information about companies to a complete communication channel by providing numerous services on which clients can connect [6]. Introduction of web services and data over the web has increased the number of attacks on them thus a small security flaw in web application will have a bigger negative impact. Services that are offered by web application may range from normal purchase information to mission critical tasks or extremely sensitive information.

Development cost in terms of money and time increase whenever changes are required to be done in software at later stages [7]. Thus it is important to identify major vulnerabilities and fix them at earlier stage of development. Consequently, developers have a duty to attempt to discover weaknesses as early as possible. However, the size and complexity of the code bases and shortage of developers experience may complicate software weaknesses discoveries. Finding vulnerabilities in web applications can be done by code auditing (code

inspection or reviews), static Analysis, dynamic analysis, and security testing [8], [9].

From 1988 to 2016 total vulnerabilities recorded by National Vulnerability Database (NVD) are 72,855 out of which 6,488 are just reported in 2015. Probable security vulnerabilities can be detected using static analysis tools. These tools also provide details about each flaw like which line has flaw, type of flaw, what possible vulnerability can cause etc. Based on several studies, it was expected that 90 percent of security incidents are results of exploiting defects in the design or code of commonly used software [10]. The main purpose of static analysis tools is to find coding errors before they can be exploited. Static analysis is predominantly a good fit to security since several security issues happen in places that hard to reach and difficult to exercise by running the code. While many tools and research proposed recently have attempted to address several security exploits, we argue that a critical aspect to avoid these security problems is to target developers by educating and assisting them with developing secure code.

One of the responsibilities of software developers is to determine non-functional requirements, such as security and performance. Educating developers on security in order to help them build elasticity in applications to protect them against attacks and to prioritize security threats and handle them before designing applications. Constructing secure software needs a great deal of security education. Many software developers are not aware and equipped with enough security education. In addition, many programming books do not teach how to write secure programs [11].

Detecting vulnerabilities and finding precarious flaws in code can be classified in two main approaches: white-box analysis and black-box testing [12]. White-box analysis examines the code without the need of executing it. This can be done manually through code inspection and reviews or automatically through security static analysis [12]. Static analysis is an automated process to assess code without executing it. Code review methods, both manual and automated, try to find security issues before releasing the software. Black-box testing analyzes program execution externally. In other words, it compares the software execution outcome with expected results.

Code review needs knowledge of code as practitioners, with slight experience will not do a good job during a code review. The code review should be done by experienced senior

developers while equipping them with modern source code analysis tools. There is no silver bullet solution to ensure secure coding. However, code review provides great insights in finding security irregularities. The remainder of the paper is organized as follows: Section II discusses the related work, Section III discusses the collected data and the empirical study, Section IV explains the suggested framework, and Section V concludes the paper.

## II. RELATED WORK

Static analysis tools usually tend to produce false positive reports and it is one of the major critiques against these tools [13]. However, the vulnerabilities found by these tools are found to be reliable as reported by Walden and Doyle [14] as they reported that vulnerabilities reported by Fortify SCA tools are highly correlated to NVD vulnerabilities. Furthermore, Gegick et al. [15], [16] showed the correlation of actual vulnerabilities and warnings found by static analysis tools are highly correlated. A large scale study conducted by Zheng et al. [17] at industry showed the importance and effectiveness of using static analysis to find flaws that can lead to security vulnerabilities. We conclude from previous studies that static analysis tool can be used to give some insights about the source code problems. The analysis results should be investigated in order to educate software developers and managers.

Previous research evaluated different techniques and their capabilities in detecting vulnerabilities [18], [13]. Manual code reviews and black box testing can be both used to find the vulnerabilities where manual code reviews can find more vulnerabilities but it will take a lot of time. Both techniques complement each other as explained by Finifter and Wagner [18]. Austin and Williams [13] found that there is no single technique that can detect all vulnerabilities. After exploring the techniques like manual and automated penetration testing and static analysis, they found that automated penetration testing are better than other two techniques in terms of vulnerabilities detected. Clark et al. [19] focused on effect of legacy code on vulnerabilities and found out that it is major player in terms of vulnerabilities found in software systems.

## III. EMPIRICAL STUDY

To evaluate the current status of the security of several web applications, we conducted an empirical study on the source code of seven open source web software systems from different domains, namely, Crawler4j, Elasticsearch, WebGoat, Friki, Gestcv, Jfinal, and Jpetstore. We provide some information about these systems. Table I summarizes the collected systems. Find Security Bugs version 1.4.6 was used to find security problems. This plugin was integrated with NetBeans. It is a FindBugs plugin for security audits of Java web applications. It can detect 86 different vulnerability types with over 200 unique signatures with extensive references for each bug patterns with references to OWASP Top 10 and CWE.

Crawler4j<sup>1</sup> is an open source application for web-crawling that can crawl the web in few minutes using multi-threading. It is able to crawl almost 200 Wikipedia pages per second and waiting for 200 milliseconds between each steps. It is also possible to do resume-able crawling.

<sup>1</sup><https://github.com/yasserg/crawler4j>

Elasticsearch<sup>2</sup> is a distributed search engine built for cloud using RESTful web services. It supports multiple indexing and multiple tenant cloud. It has real time search and analytical capabilities. It can allow full text search as well as persistent where each document changes are recorded. It has JSON based document store.

WebGoat<sup>3</sup> is a deliberately designed web application for security testing maintained by OWASP. It is also designed to teach security and penetration testing system and common security flaws. It can train in cross-site scripting, access control, parameter manipulation, blind SQL injection, web services, numeric SQL injection using realistic teaching environment. It is platform independent environment that uses Java VM. When you run the webgoat it is highly probable that your machine may be hacked.

Friki<sup>4</sup> is a wiki like application built using Java and can be deployed on any modern servlet. It has some common features like wiki and its common markup tag support. It offers an easy customizable solution that can be loaded dynamically without the need of restarting the server again.

Gestcv<sup>5</sup> is a java based application used to manage Curriculum Vitae. It allows creation of CV and allows searching of its contents. It is also based on Struts, Spring and Hibernate. It is built on MVC architecture. It uses MySQL database, and allows persistent development.

JFinal<sup>6</sup> is a complete framework written in Java language and it uses RESTful web services. It allows easy development without writing large amount of code for writing RESTful web services. Its built on MVC architecture and require no configurations as uses XML. Java development and deployment doesn't need server to be restarted and is automatically loaded. Plugins can be scaled and provide struts support as well as supports multi-view.

JpetStore<sup>7</sup> is completely re-written web application pet store that was originally made by Microsoft. It is written in Java and overcomes the shortcoming of its original version. It is based on Struts with color coding conventions to ease programmer for writing codes. Presentation later is based on MVC architecture and there is HTML in database making it completely independent.

TABLE I. SUMMARY OF THE SYSTEMS

Project	Version	No. of Files	LOC
Crawler4j	4.2	43	7114
Elasticsearch	6.0.1	3865	616000
WebGoat	7.0.1	35	8474
Friki	2.1.1	21	1843
Gestcv	1.0.0	119	11524
JFinal	2	14	2379
JpetStore	6	116	25820

We report the results of running FindBugs on these applications. We report two types of bugs, namely Malicious

<sup>2</sup><https://github.com/elastic/elasticsearch>

<sup>3</sup><https://github.com/WebGoat/WebGoat>

<sup>4</sup><https://sourceforge.net/projects/friki/files/friki/>

<sup>5</sup><https://sourceforge.net/projects/gestcv/>

<sup>6</sup><http://www.jfinal.com/>

<sup>7</sup><https://sourceforge.net/projects/ibatisjpetstore/>

Code Vulnerability (MCV) and Security code. Malicious code vulnerability is a code that can be altered or exploited by other code. It can be in form of worms, viruses, Trojan horses or other programs that can exploit other security parameters. There are numerous Malicious code vulnerabilities like (1) exposing internal representation to reference object that pose a threat to security if that object is accessed through different purpose, (2) Usually the field that has last results should be declared is final but is missed and poses a threat of being used by malicious code to change the value. (3) Returning the mutable object as reference poses a serious security threat and can be used by malicious code, (4) A field is defined as static but not protected can be accessed by malicious code and can be changed.

Security code gaps means finding errors that might impact the application security by exploiting security vulnerabilities. It can be in form of malicious data injection or manipulating the applications using malicious data. There are couple of security categories that should be checked as these provide open threats to any web application. Most common security threats are (1) Carriage return and line feed or HTTP response splitting is a usual way programmers adapt to work on response returned but if hacker can plunge the response through injections it can be used to control how web functions will act. (2) Use of predictable random generator to calculate the random number may result in finding the predicted number and can be used to find the password sent or any other secret value, (3) Usually a file is opened to read or write where filename is sent as input and can result in revealing the full path of location of file (4) Usually programmer pass JDBC connection string as prepared statement unsafely can result in SQL injection attack, (5) Use of regular expression in a variable unprotected will result in plugging a big regular expression to compile and will result in Denial of service as program will get busy in parsing the variable for large amount of time.

Figures 1 and 2 show the number of security issues found in these applications. These results clearly show that these web application have a lot security problems. These security issues can be addressed at early stages of these applications development. We believe that educating developers and giving them hints while they are developing the application will result in more secure applications. Developers and test mangers dont have to wait until they finish to find out if there is a security issue or not in the code. Learning from previous security errors can be a great aid in preventing them from happening in the future. While many tools and research proposed recently have attempted to address several security exploits, we argue that a critical aspect to avoid these security problems is to target developers by educating and assisting them with developing secure code.

#### IV. SUGGESTED FRAMEWORK

Several organizations for example MITRE [20], SANS Institute [21] and OWASP [22] have highlighted the significance of educating students, developers, managers about security issues. These organizations do their part by frequently publishing common programming errors. Our study supports the intuition that web developers usually fail in securing their web applications. The outdated approach of testing applications after they are finished proved to be problematic. We

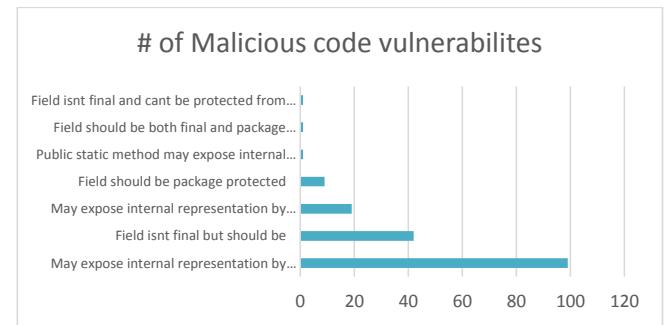


Fig. 1. MCV found in selected web applications.

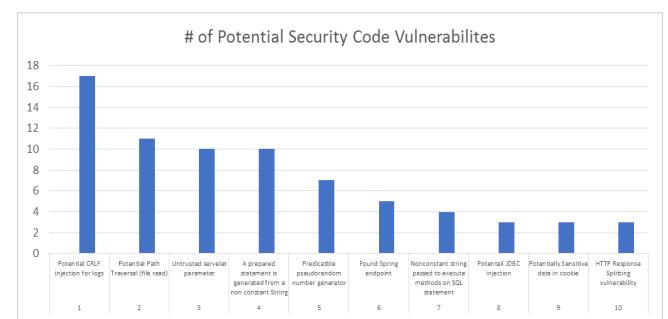


Fig. 2. Security gaps found in selected web applications.

believe that educating developers and giving them hints while they are developing the application will result in more secure applications. Developers and test mangers dont have to wait until they finish to find out if there is a security issue or not in the code. Learning from previous security errors can be a great aid in preventing them from happening in the future.

Software security researchers have measured vulnerabilities using both databases of reported vulnerabilities such as the National Vulnerability Database (NVD) and static analysis results. In our suggested framework we make use of both approaches. We utilize the valuable knowledge in vulnerability repositories such as Common Weakness Enumeration (CWE) and National Vulnerability Database (NVD), which is contributed by software security experts around the world, and is available for public use for free. In Figure 3, we explain our suggested framework. The framework can be integrated with any integrated development environment (IDE). The idea is to enable developers and testers to find security problems in the code while the system is in implementation. After a piece of code has been written, the framework will run that code on several static analysis tools, check the code in two available databases, CWE and NVD, and eventually give a recommendation based on the collected data from three different sources. This will give an instant feedback to the developer about the written code. It will make him/her confident about his code. It will also educate him/her in the go since these recommendations will help him/her learn a lot about code security problems.

For the static analysis tools, in the framework, several

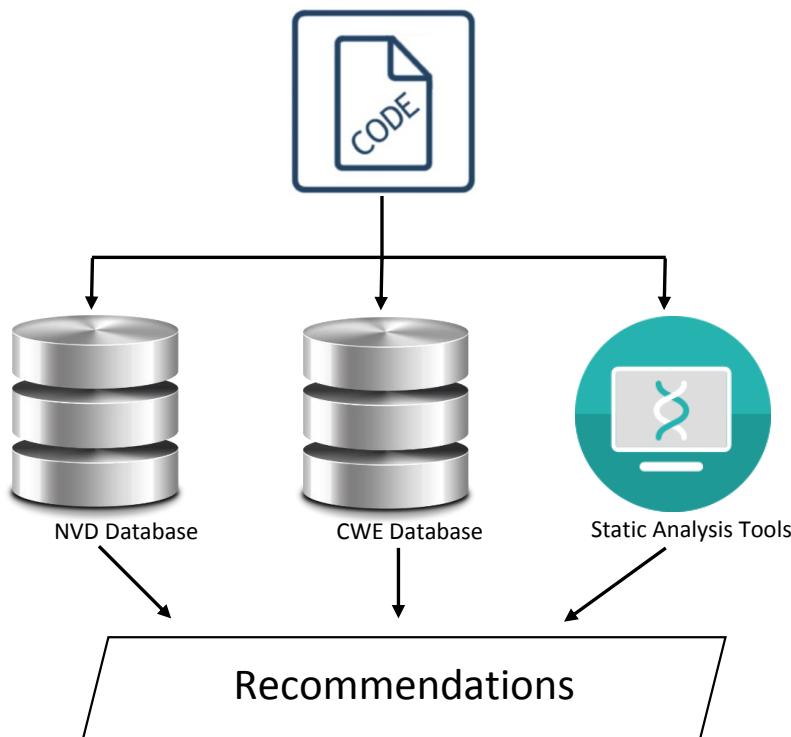


Fig. 3. The Suggested Framework.

tools can be integrated, namely FindBugs<sup>8</sup>, PMD<sup>9</sup>, Yasca<sup>10</sup>, and LAPSE<sup>11</sup>. The Static Code Analysis Module contains static analysis tool(s) which scan the given code repository to find vulnerabilities. These tools are open source and can be integrated with several available IDEs. In the suggested framework, a filtered summary of these tools can be presented as recommendations for developers to educate them and to help them avoid such problems in their source code. The proposed framework is based on the static analysis of code written by software developers. Static analysis tools report suspicious security vulnerabilities found in source code. These results can be utilized for recommending issues and best practices to the software developers who contributed to writing those class and components, hence improving their software security skills.

Numerous kinds of security vulnerabilities can take place in code, design, or architecture. The security community uses Common Weakness Enumerations (CWE) to differentiate security vulnerabilities. For the Common Weakness Enumeration (CWE), the list is available in multiple machine-readable formats including html, XML, and PDF. The CWE Schema is also provided for processing the complete CWE List. The framework can check the developer code against these common problems in source code again to educate them and to help them avoid such problems in their source code. CWE is a software community project that targets creating and maintain a catalog of software weaknesses and vulnerabilities.

The goal of the project is to identify, fix, and prevent those vulnerabilities.

NVD complete database about vulnerabilities can be downloaded from their website in unified XML format. Each XML document contains the security vulnerabilities found with a description about them. Our designed framework can check the developer code against these common problems in source code again to educate them and to help them avoid such problems in their source code.

These three source of recommendations will be aggregated and shown to the developers. Later on, some intelligence can be integrated to the framework as well. Based on the historical errors and flaws, the framework will adopt and focus on these specific security issues.

## V. CONCLUSION

It is observed that selected projects have common vulnerabilities in all types of security flaws, malicious code and security code. These vulnerabilities are mostly inserted due to developers' non awareness or bad programming practice. The vulnerabilities from selected projects also reveals that they have security and malicious code vulnerabilities making them more prone to attacks. The suggested framework will allow developers to produce more secure code and help them learn about best practices in developing web applications. The framework will also tackle these issues at early stages before even going to testing which will reduce the cost of software development.

<sup>8</sup><http://find-sec-bugs.github.io/>

<sup>9</sup><http://pmd.github.io/>

<sup>10</sup><http://scovetta.github.io/yasca/>

<sup>11</sup><https://code.google.com/archive/p/lapse-plus/>

## REFERENCES

- [1] M. Alenezi and F. Khellah, "Evolution impact on architecture stability in open-source projects," *International Journal of Cloud Applications and Computing (IJCAC)*, vol. 5, no. 4, pp. 24–35, 2015.
- [2] N. Fenton and J. Bieman, *Software metrics: a rigorous and practical approach*. CRC Press, 2014.
- [3] I. Abunadi and M. Alenezi, "An empirical investigation of security vulnerabilities within web applications," *Journal of Universal Computer Science*, vol. 22, no. 4, pp. 537–551, 2016.
- [4] M. Bishop, *Introduction to computer security*. Addison-Wesley Boston, MA, 2005.
- [5] M. Nyanchama, "Enterprise vulnerability management and its role in information security management." *Information Systems Security*, vol. 14, no. 3, pp. 29–56, 2005.
- [6] M. Alenezi and I. Abunadi, "Evaluating software metrics as predictors of software vulnerabilities," *International Journal of Security and Its Applications*, vol. 9, no. 10, pp. 231–240, 2015.
- [7] M. Alenezi and K. Magel, "Empirical evaluation of a new coupling metric: Combining structural and semantic coupling," *International Journal of Computers and Applications*, vol. 36, no. 1, pp. 34–44, 2014.
- [8] V. B. Livshits and M. S. Lam, "Finding security vulnerabilities in java applications with static analysis." in *Usenix Security*, vol. 2013, 2005.
- [9] T. Lee, G. Won, S. Cho, N. Park, and D. Won, "Detection and mitigation of web application vulnerabilities based on security testing," in *IFIP International Conference on Network and Parallel Computing*. Springer, 2012, pp. 138–144.
- [10] S. Chung, L. Hansel, Y. Bai, E. Moore, C. Taylor, M. Crosby, R. Heller, V. Popovsky, and B. Endicott-Popovsky, "What approaches work best for teaching secure coding practices," in *Proceedings of the 2014 HUIC Education and STEM Conference*, 2014.
- [11] A. S. Sodiya, S. A. Onashoga, and O. B. Ajayi, "Towards building secure software systems," *Issues in Informing Science and Information Technology*, vol. 3, pp. 635–646, 2006.
- [12] N. Antunes and M. Vieira, "Defending against web application vulnerabilities," *Computer*, vol. 45, no. 2, pp. 0066–72, 2012.
- [13] A. Austin and L. Williams, "One technique is not enough: A comparison of vulnerability discovery techniques," in *2011 International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2011, pp. 97–106.
- [14] J. Walden and M. Doyle, "Savi: Static-analysis vulnerability indicator," *IEEE Security & Privacy*, vol. 10, no. 3, pp. 32–39, 2012.
- [15] M. Gegick, L. Williams, J. Osborne, and M. Vouk, "Prioritizing software security fortification through code-level metrics," in *Proceedings of the 4th ACM workshop on Quality of protection*. ACM, 2008, pp. 31–38.
- [16] M. Gegick, P. Rotella, and L. Williams, "Predicting attack-prone components," in *2009 International Conference on Software Testing Verification and Validation*. IEEE, 2009, pp. 181–190.
- [17] J. Zheng, L. Williams, N. Nagappan, W. Snipes, J. P. Hudepohl, and M. A. Vouk, "On the value of static analysis for fault detection in software," *IEEE transactions on software engineering*, vol. 32, no. 4, pp. 240–253, 2006.
- [18] M. Finifter and D. Wagner, "Exploring the relationship between web application development tools and security," in *USENIX conference on Web application development*, 2011.
- [19] S. Clark, S. Frei, M. Blaze, and J. Smith, "Familiarity breeds contempt: The honeymoon effect and the role of legacy code in zero-day vulnerabilities," in *Proceedings of the 26th annual computer security applications conference*. ACM, 2010, pp. 251–260.
- [20] B. Martin, M. Brown, A. Paller, D. Kirby, and S. Christey, "2011 cwe/sans top 25 most dangerous software errors," *Common Weakness Enumeration*, vol. 7515, 2011.
- [21] R. Dhamankar, M. Dausin, M. Eisenbarth, J. King, W. Kandek, J. Ullrich, E. Skoudis, and R. Lee, "The top cyber security risks," *SANS Institute*, 2009.
- [22] T. OWASP, "10: Ten most critical web application security risks," 2013.

# A Review on Influential Factors of Information Privacy Concerns in the Use of Electronic Medical Records

Fiza Abdul Rahim

Department of Systems and Networking  
College of Computer Science and Information Technology  
Universiti Tenaga Nasional  
Kajang, Malaysia  
fiza@uniten.edu.my

Zuraini Ismail and Ganthan Narayana Samy

Advanced Informatics School  
Universiti Teknologi Malaysia  
Kuala Lumpur, Malaysia  
zurainiisma.kl@utm.my, ganthan.kl@utm.my

**Abstract—** Healthcare organisations process massive amount of electronic medical records (EMR) utilised by their employees in supporting the organisation's services. Having given privileged access to sensitive and valuable patient information in the EMR, healthcare employees may cause privacy breaches, which may lead to detrimental consequences. Therefore, it is paramount to impose particular attention to healthcare employees' concerns on privacy in the use of EMR. The aim of this study is to identify the factors that influence information privacy concerns (IPC) in the use of EMR from healthcare employees' perspective. Systematic literature review (SLR) was conducted to identify articles pertinent to IPC. EBSCOhost, IEEE Explore, SAGE, MEDLINE, ScienceDirect, SpringerLink, Wiley Online Library and Taylor & Francis Online database were searched for reviews relevance articles. A total of 38 full articles were reviewed to extract the factors that influence the IPC. From the review, it revealed three influential factors, namely privacy risk, privacy awareness, and privacy policy. Furthermore, preliminary qualitative study has been done in this study helps in understanding the privacy practices, to validate the identified factors and relationships with IPC. This study may be of significance in providing useful information for healthcare organisations to understand IPC from their employees' perspective in ensuring the compliance towards privacy regulations.

**Keywords-information privacy concerns; electronic medical records; healthcare information system**

## I. INTRODUCTION

The growth of information system is discernible in most organisations. In the healthcare field, a number of healthcare information systems (HIS) were developed to assist healthcare organisations in providing efficient and quality healthcare services. The massive development in healthcare technology has enabled the collection, storage, management, and sharing of electronic medical records (EMR) using HIS. EMR stored in the HIS are also known as electronic health records (EHR) which refers to the electronic patient records that are created and maintained by the healthcare organisation [1].

The use of EMR can greatly benefit healthcare organisations [2], [3]. EMR can be utilised by healthcare

employees from various employment backgrounds inside a healthcare organisation, including healthcare professionals, researchers, hospital administrators, and healthcare management personnel for specific reasons. Healthcare professionals and researchers utilise the EMR to improve diagnosis and treatment of diseases, while hospital administrators and healthcare management personnel use EMR to support their organisation's services [4].

However, healthcare employees must be fully aware of the procedures and privacy implications that may be involved in using EMR [5]. In the healthcare domain, privacy breaches include unintentional disclosure, unauthorised access, hackers attack and data theft [6]. Having given privileged access to sensitive and valuable patient information in the EMR, healthcare employees may cause privacy breaches with potentially severe concerns. It is proven based on a report by Verizon as highlighted in the 2015 Data Breach Investigations Report, in which the report claims that insider misuse is one of the attacks that harmfully affected the healthcare industry [7]. The same report also claimed that the end user is the main culprit in cases of misuse at 37.6%. Figure 1 illustrates the misuse pattern among the members of healthcare organisations.

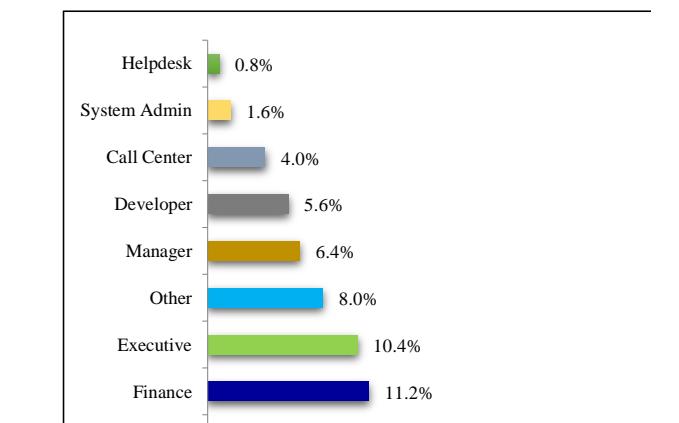


Figure 1 Misuse Pattern Among the Members of Healthcare Organisations [7]

Today, with the introduction of Personal Data Protection Act (PDPA) 2010 [8], healthcare organisations in Malaysia will have to consider the protection of personal information contained in the EMR. Appropriate security measures should also be adopted by the organisations in order to ensure their compliance to the act [9]. Hence, a better understanding of the influencing factors on information privacy concerns and their relationships may help healthcare organisations to understand their employees' perspective on privacy.

## II. RESEARCH BACKGROUND

A report from the Ponemon Institute (2012) revealed that most healthcare organisations struggle to manage with privacy risks due to the lack of technology, resources, and trained personnel needed for the job. Protecting the privacy of EMR and maintaining the confidentiality of their data have always been highlighted in previous studies [11]–[14].

From a technical perspective, specific tools or techniques to ensure the security of data such as authentication, access control, and encryption have long existed. However, the challenging work lies in the socio-technical perspective such as knowing when a particular tool or technique is needed and why. In that case, Stanton [15] stressed on the importance of understanding the perspectives and needs of those individuals whose privacy is at stake, including workers, managers, clients, customers and others whose personal information is collected, transmitted, and stored by a specific technology.

In Malaysia, no official statistic reports on privacy breaches in healthcare industry were ever published. However, few cases were reported in the local newspapers and other media regarding the misuse of EMR among healthcare organisations [16], [17].

In line with the enacted PDPA 2010 act, organisations need to take specific measures related to privacy issues in ensuring their compliance to the act [9]. With multiple healthcare employees using EMR, the organisation is responsible to ensure privacy compliance among their employees who are handling sensitive personal data contained in EMR. Hence, employees must be aware of the confidentiality of sensitive personal data and the need to protect them from any privacy threats such as the misuse of computerised information, loss of information, and identity theft.

Recent development in technology was found to have benefited the healthcare sector in reducing their operation cost and allowing for the sharing of data with other stakeholders such as government agencies, health research institutes, insurance companies, and other healthcare institutions [18], [19].

However, according to Kaletsch and Sunyaev (2011), several threats are involved when dealing with information sharing and privacy; for instance, involuntary exposure of patients' identity that should have been kept anonymous and the selling of personal information for targeted advertising. One of the major newspapers in Malaysia, Berita Harian reported on the case of thousands patients' data being sold to medicine sellers were stolen from government and private

hospitals by cyber thefts [21]. Hence, the security and privacy of personal information still remain as the main concern that must be scrutinised in an extensive manner [22].

Symantec, a global leader in providing security, storage and systems management solutions reported that for the fourth year in a row, the healthcare sector holds the record for having the largest number data breaches and disclosure; 43% in 2011 [23], 37% in 2012 [24], 44% in 2013 [25], and 37 percent in 2014 [26]. Figure 2 illustrates the top five sectors breaches by number of incidents showing that the healthcare sector as having the largest number of data breaches.

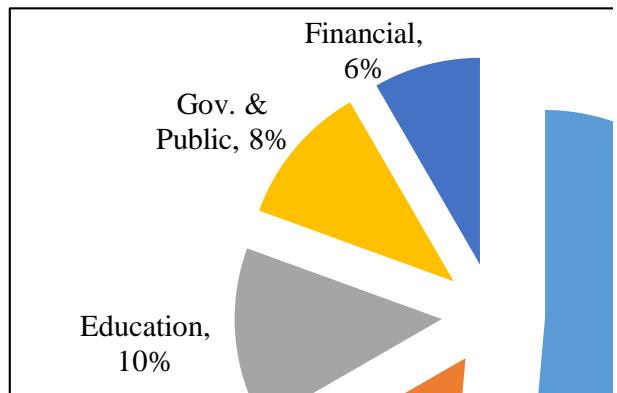


Figure 2 Top Five Sectors Breaches by Number of Incidents [26]

In the same report every single year, hackers continue to be responsible for the largest number of data breaches. Symantec also reported that insider threats remained high as employees intentionally or unintentionally leak or steal valuable data. Several possible sources of breaches were identified such as stolen laptops, misplaced memory sticks, deliberate data theft by employees, and accidental exposure of confidential data to public. This indeed shows that although the organisation is usually more concerned about vulnerabilities to external threats, they should also consider security incidents originating from inside the organisation [27].

Healthcare organisations and insurers regularly use computers, phones, and other means of technologies to record and transfer information about patients. These information consist of sensitive information related to healthcare records including personal identification, history of medical diagnoses, rendering of medical images, treatments, prescriptions, dietary habits, sexual preferences, genetic information, psychological profiles, employment history, income, and insurance information [28], [29].

In a given situation where patients' sensitive information is not properly protected, anyone who logs on to a shared computer may manipulate it for illegal purposes. This can be considered as privacy breach, whether those involving paper-based records which are susceptible to physical loss or acts of vandalism, or information stored in electronic form that could be misused in a number of ways [30]. The exposure of personal health information could result in the loss or denial of health insurance, job discrimination or personal embarrassment [31]–[33].

Although other researches related to information privacy concerns were carried out [34]–[37], limited attention was given to the aspect of employees' or users' perspective, who are the main members of the organisation that handle sensitive personal data.

In fact, the study regarding information privacy concerns related to HIS in Malaysia context is still at its infancy level. Related researches on information privacy concerns in Malaysia were done by Lallmahamood [38], [39] that focusing on e-commerce and e-government, and Mohamed and Ahmad [40], [41], who examined the social networking sites. However, there is a paucity of research done in Malaysia related to information privacy concerns in the use of EMR focusing on the employees' perspective.

Hence, this study attempts to investigate the factors that influence information privacy concerns. Consequently to ascertain the relationship of the influential and information privacy concerns among healthcare employees that are using EMR in their healthcare organisation.

#### A. Healthcare Information System

Progressive implementation of IT in the delivery of healthcare services [42], also known as healthcare information system (HIS) have been reported to provide significant benefits to healthcare organisations [43]. HIS has grown from the original hospital information systems, and has advanced additional requirements to system interoperability [44].

HIS is defined as a system that "accesses a lot of sensitive data such as personal information, physiological parameters, and medical records" [45]. Moreover, HIS is also referred as computerized programs with "a set of standards based on healthcare diagnosis, symptoms, cause, healthcare target, and measurements" [46]. In enhancing HIS, it can be integrated with the hospital system, clinical care, and administrative management [47]. Additionally, HIS is designed to assist healthcare organisation to create, store, manage, and exchange patients' medical information electronically [48].

From the above definitions, HIS deals with creating and collecting data, storing, managing, and processing information in healthcare environment. HIS is implemented to give continuous healthcare reports and assist the process of managing healthcare services [49]. HIS is used by a large number of healthcare professionals, such as healthcare practitioners (doctors, nurses, pharmacists) and a large number of services (outpatient care services, rehabilitation services) in a healthcare organisation [50]. For this reason, HIS is no longer an institutional component, but instead has become the operational backbone of the healthcare organisation [51].

This study employs the definition of HIS based on Simpson and Weaver [45], Rahman and Kreider [47], Hsu, Lee, and Su [48], Al-Sakran [50] definitions; "*Healthcare information system (HIS) is an integrated system that is used by healthcare employees in order to create, collect, store, manage, and exchange sensitive EMR electronically*".

#### B. Electronic Medical Records

In replacing the difficulty of using paper-based medical records, electronic medical records (EMR) can be considered as a solution for integrating medical records from various departments and increasing accessibility to it [52]. With the massive developments of mobile devices and web-based applications in healthcare field [53]–[55], EMR facilitate the users to improve the convenience of sharing the records throughout the organization [56].

EMR allow easier and more effective management [57] in the healthcare environment by reducing labour costs, delay, pollution, and medical errors [4]. The massive developments of mobile devices and web-based applications in this area [53]–[55] allow healthcare employees to improve the convenience of sharing the records across healthcare organisation [4], [56], [58].

Moreover, it ensures that healthcare employees are able to interact with the most recent and complete information of EMR when they need it [59]. EMR may also help healthcare professionals to identify the right medical treatment for the patients by getting such a detailed information about the patient containing identification, history of medical diagnosis, treatment and medication history, dietary habits and several other assessments from the information system [29].

In other researches, EMR may also be known as electronic health records (EHR). Indeed, an EHR which refers to patients' records stored in a computer system is often used interchangeably with EMR [60], [61]. Contrarily, Kierkegaard [4] emphasised that EMR and EHR should be defined differently, in which EMR is a record "contains the encounter information of patients in a care deliver organisation, while EHR contains information from many or all care deliver organisations where the patient has been treated or has had an encounter". Whereas, Deutsch, Duftschmid, and Dorda [62] described EMR and EHR as correlates.

Unlike Kierkegaard [4] and Deutsch, Duftschmid, and Dorda [62], Chang [63] referred EMR as legal records of events owned by healthcare organisation, while EHR is a record owned by the patient. However, several researches that have been done referred to both terms synonymously [64]–[66].

With more other variant terms used to define EMR such as Electronic Patient Record (EPR), Computerised Patient Record (CPR), and Electronic Health Care Record (EHCR) from different countries and organisations around the world, International Organisation for Standardisation (ISO) produced ISO/TR20514 in order to clarify and agree to the boundaries of EHR and to facilitate the development of international EHR standards. According to the ISO/TR20514, EHR is "a repository of information regarding the health of a subject of care, in computer-processable form" [67].

Mercuri [29] defined EMR as a "personal information containing identification, history of medical diagnosis, digital renderings of medical images, treatments, medication history, dietary habits, sexual preferences, genetic information, psychological profiles, employment history, income and physicians' subjective assessments of personality and mental

state". Notably, Sun and Fang [58] described EMR as an information stored electronically containing "information such as medical history, test results, allergy lists, radiology images, billing records, and etc.".

Hence, this study employs the term of EMR and adapts the definition of EMR based on Mercuri [29]; Sun and Fang [58]; Sandikkaya, Decker, and Naessens [56] and Kierkegaard [4] definitions; "*Electronic medical records (EMR) refer to personal information about patients containing identification, history of medical diagnosis, treatment and medication history, dietary habits and several other assessments used by healthcare employees and can be shared among individuals or groups in the healthcare organisation*".

### C. Privacy

Since most organisations collect large amounts of individual's data in their business operations, privacy becomes a critical issue [54], [68]–[71]. With a growing use of EMR in healthcare environment, information privacy (later to be described as privacy) becomes the main concern in managing EMR.

Accordingly, most scholars agreed that if the record is related to medical information, there is a need to ensure the privacy of information [6], [18], [59], [72]–[78]. Notably, Westin [79] described privacy as "the right of individuals, groups, or institutions to determine when, how, and to what extent information about them is communicated with others". This definition generates the ability for individuals, groups, or institutions to control their personal information and the ability to determine when and how that information should be processed and used.

Kurtz [73] defined privacy "as the right of an individual to control disclosure of his or her medical information". Likewise, Modaresnezhad and Chain [80] defined privacy as "a control over individual's data which includes control over what data is captured, control over the accuracy of data, control over sharing of data, and control over duration of data retention".

From law perspective, privacy is based on "the right of an individual to have control over his or her own information and to some extent, how an individual relates to and controls access to information about himself" [81]. It also involves the establishment of rules governing the collection and handling of personal data such as credit information and medical records.

In social network environment, the definition of privacy is derived from social network sites activities involving generating, collecting, processing, communicating, displaying, exchanging, and storing information. Borena, Belanger, and Ejigu [82] defined privacy as "a right of individuals, groups and organisations to define, create, preserve, communicate and control the generation, collection, manipulation, storage, communication, transformation and use of one's information but also information generated directly or indirectly from them".

At the organisational level, privacy refers to how organisation treat their customers' personal data [83]. In healthcare environment, patients can be categorised as

customers who are providing their personal data, which are then generated into EMR by the healthcare employees.

In order to protect the privacy of EMR, healthcare employees need to understand what type of information is considered as private according to the law. Based on definition of personal data in Act 709, PDPA 2010, personal data refers to any information "that relates directly or indirectly to a data subject, who is identified or identifiable from that information or from that and other information in the possession of a data user, including any sensitive personal data and expression of opinion about the data subject", while sensitive personal data means "any personal data consisting of information as to the physical or mental health or condition of a data subject".

Healthcare organisation must apply the 'need-to-know' concept to the system users when accessing EMR [84]. This kind of restriction is supposed to protect EMR belonging to specific units or departments in the organisation from being accessed by personnel from other departments without need and authorisation.

Privacy in the context of this study adapts definitions by Modaresnezhad and Chain [80] and Borena, Belanger, and Ejigu [82] to be used in the healthcare environment; "*Privacy refers to the ability of healthcare employees to control EMR during collection, maintaining the accuracy of EMR during manipulation, ensuring the confidentiality of EMR during transferring, and understanding the duration of EMR retention in the organisation*".

### D. Information Privacy Concerns

Ponemon Institute, a research centre dedicated to privacy, data protection and information security policy reported that healthcare organisations continue to struggle to comply with privacy and security regulations [85]. Moreover, failures in security and maintaining confidentiality contributing to breaches remain as headlines in the media [86], such as information leaks, non-availability, and compromise in integrity [87].

These breaches may cause tangible harm to both the organisations and individuals [88], [89] and may lead to huge losses in terms of the healthcare organisations' reputation, monetary fines, along with possible civil and criminal liabilities [18], [87], [88]. For patients, consequences of privacy breaches are tremendously serious including inappropriate and unjustified employment termination, loss of individual health insurances, and illegal use of one's identity [90].

With the enormous development of technology in the healthcare domain, privacy issues and threats have been highlighted by researchers and professionals [6], [78], [83], [89], [91]–[95]. Since security and privacy have become a central concern in the integration of EMR [96], healthcare organisations must ensure that EMR are collected and communicated with security, accessed only by authorised parties and are not disclosed to any unauthorised party, intentionally or unintentionally when disseminated.

The utilisation of EMR has raised concerns about privacy due to collection, secondary use of data, errors, and

unauthorised access [97]. Patients are worried about privacy threats and are concerned if their personal information is being released to other parties [98]. From the patients' perspective, privacy concerns refer to the extent of patients' concerns about organisational practices related to the EMR collection and usage [97], [99].

Malhotra, Kim, and Agarwal [100] defined information privacy concerns as something that "lies in fairness perceptions of an individual and likely to be generalizable across a variety of other privacy context". This definition is based on Campbell [101] who defined information privacy concerns as referring to an individual's subjective views of "fairness" within the context of information privacy, influenced by external conditions (e.g., technical knowledge, cultures, regulatory laws).

Another definition of information privacy concerns provided by Park [102] focused on information system usage, in which information privacy concerns are described "as a perception that an organisation monitors or accesses information that is disclosed when using the organisational information systems and how it is used". Tan *et al.* [103] defined information privacy concerns as "a person's awareness and assessment of risks related to privacy violations". While, in studying the specific nature of information privacy concerns among mobile users, Xu *et al.* [104] define "mobile users' information privacy concerns is a concern about the possible loss of privacy as a result of information disclosure to a specific external agent".

Most of information system researchers typically highlight the differences in privacy concerns levels or ascertain the effects of privacy concerns on several dependent variables [105]. Those researches that were done not only in the healthcare domain but also other fields, such as e-commerce [106], [107] and social network [41], [108]. Schwaig *et al.* [107] suggested that organisations need to design information practices that address consumers concerns, formulate privacy policies that highlight their information practices and make sure that the policies are well practised throughout the organisation.

In this study, the definition of information privacy concerns is adapted from Malhotra, Kim, and Agarwal (2004); I. Park (2009); Tan *et al.* (2012), which can be conceptually defined as "*healthcare employees' perception on privacy in the use of EMR*". Having defined the information privacy concerns, from hereon it will be abbreviated as IPC.

### III. EXISTING STUDIES ON INFORMATION PRIVACY CONCERNS

This section provides systematic review of existing research on IPC and reports a comprehensive taxonomy of the factors that influences IPC in various fields. The main motivation for using systematic literature review (SLR) in this study is to classify the factors that influence IPC in the use of EMR through the SLR question, "What are the factors that influence IPC in various fields?"

This review adapted eight steps outlined by Okoli and Schabram [109] in performing SLR which involves four major

stages; planning, selection, extraction, and execution. The final search string used for the searching of the literature was as follows: ("privacy concerns" OR "privacy concern" OR "information privacy concerns" OR "information privacy concern") AND ("influence" OR "impact" OR "effect" OR "affect" OR "impress").

The primary search process involved the use of 8 online databases: EBSCOhost, IEEE Explore, SAGE, MEDLINE, ScienceDirect, SpringerLink, Wiley Online Library and Taylor & Francis Online. The selection of online databases was based on the knowledge of databases that indexed the previous IPC studies that the researcher is aware of, and the list of available online databases subscribed by the Universiti Teknologi Malaysia's library.

In identification phase, the search keyword returned 686 peer-reviewed articles available from 2009 until April 2014. Four duplicate articles were removed; yielding 682 articles. After screening titles and abstract, 556 articles were removed, yielding 313 full-text articles eligible for further assessment in eligibility phase.

After detailed assessment of duplicates, abstracts, titles and research questions, 22 full-text articles were accepted for the synthesis of evidence. In addition, articles listed in the references of the 22 full-text articles were hand-searched for additional articles, resulting in 16 more articles included. Finally, 38 articles were selected for further review.

The results of the data collection methods categorization of 38 selected studies showed that majority of the data collection methods employed is survey (71%). Content analysis (16%) was the next popular data collection method followed with mixed-methods (10%) and focus group (3%). Based on the selected articles, most IPC researches were conducted in online social network field. Internet users, education and healthcare are among growing environments carrying out IPC research.

Altogether, nine similar factors were investigated in 38 studies that looked into how it affected or correlated with IPC. Table I depicts the influential factors, the studies that investigated each factor, and whether the factor had a significant positive or significant negative effect. Significant positive effects means the factor is influencing IPC, while significant negative effects means the factor is not influencing IPC.

From the SLR findings, it can be summarised that 38 selected studies acknowledged the importance of privacy risk in IPC. SLR results also showed that privacy policy and privacy awareness as significant factors for IPC. Hence, it can be generalised that privacy risk, privacy policy, and privacy awareness are needed in ensuring individuals' concern towards information privacy.

TABLE I. LIST OF FACTORS INVESTIGATED IN SELECTED STUDIES

No.	Factor	Total Studies	Significant Positive Effect	Significant Negative Effect	No Result
1.	Privacy risk	12	S5, S6, S7, S8, S17, S21, S25, S35, S36, S38	S29	S3*
2.	Privacy awareness	10	S11, S13, S14 S21**, S23, S26, S31	S21**, S30	S4*, S14*, S33
3.	Privacy policy	10	S10, S11, S17, S24, S31, S34	S22, S30	S4*, S19*
4.	Demographic	10	S2, S13, S16, S18, S25, S27, S38	S11	S19*, S33*
5.	Privacy control	7	S7, S9, S15, S36	S6, S20	S3*
6.	Privacy experience	7	S12, S28	S18, S31, S37	S19*, S33*
7.	Self-efficacy	5	S1, S25	S32, S38	S19*
8.	Culture	3	S12	-	S19*, S33*
9.	Trust	3	S8, S22	-	S14*

Legend:

\* : Reported the conceptual model / framework / proposition only.

\*\*: Both reported that privacy awareness had positive (inexperienced shoppers) & negative (experienced shoppers) effect.

Demographic was ranked the fourth factor as suggested by previous studies in considering factors that influence IPC. Next, privacy control and privacy experience were among the investigated factors in the selected studies. However, self-efficacy, culture, and trust were less investigated by the previous scholars.

#### IV. PRELIMINARY QUALITATIVE STUDY

Prior to conducting the preliminary qualitative study, three possible factors that influence IPC have been identified from SLR. Therefore, the aim of this preliminary qualitative study is to validate the identified factors and relationship with IPC. At the same time, this preliminary qualitative study is aimed to get an overview of IPC among healthcare employees from various perspectives; HIS experts, HIS users, and legal experts.

This preliminary qualitative study followed the guidelines in conducting qualitative research by Hesse-Biber & Leavy [110] and Glesne [111]. The initial step involved enlisting researchers in privacy domain with regards to EMR. The potential respondents were communicated by e-mail and telephone.

Once the respondent agreed to participate in the study, time and location for the interview was arranged based on their preferences. In any research, ethical issues relating to protection of the participants are important [112]–[115]. Therefore, informed consent has been explained to the interviewee in advance and executed at the time of the interview.

The preliminary qualitative study involves set of interviews with respondents from various backgrounds; HIS experts (HIS

developers and HIS researchers) HIS users, and legal experts as presents in Table II.

TABLE II. RESPONDENTS' PROFILE FOR PRELIMINARY QUALITATIVE STUDY

Respondent No.	Designation	Specialization	Organisation	Working Experiences (years)
R1	HIS System Analyst	HIS development	Hospital	2
R2	IT Officer	HIS development	Hospital	3
R3	Academician	HIS research (IT)	University	13
R4	Academician	HIS research (IT)	University	4
R5	Doctor	HIS user	Hospital	3
R6	Nurse	HIS user	Hospital	6
R7	Academician	PDPA (Law)	University	6
R8	Head of Department	HIS development	Government	16
R9	Head of Department	PDPA (Law)	Government	15

Interviews with nine respondents were conducted between April 2013 and May 2014. Each session took 50 minutes to 2 hours to complete. The semi-structured in-depth interviews are conducted based on an interview outline with open-ended questions for healthcare employees.

Open-ended interviews have been applied to allow the interviewees freedom in expressing their standpoint. All interviews were conducted based on respondent's available time. The findings from these interviews have been used to validate the identified factors from SLR.

#### V. FINDINGS

In healthcare domain, previous studies highlighted the importance of privacy risk, privacy awareness, and privacy policy. Thus, the top three most investigated factors found in SLR are to be discussed in this section, with the respective hypotheses and relationship.

Preliminary qualitative study helps in understanding the privacy practices and the factors that have been identified through previous IPC studies. Since few IPC studies address employees' perspective and due to limited privacy concerns studies focusing on healthcare field in Malaysia, such studies are important to expand understandings on the subjects in the local setting.

##### A. Privacy Risk

S36 described privacy risk as an individual's expectation of losses associated with personal information disclosure. It involves an evaluation of the possibility of negative consequences as well as the severity of the consequences. In relating with Communication Privacy Management (CPM) theory, when the information moves across a personal

boundary, individuals involve in an assessment about the degree of the uncertainty involved [116]. The higher the uncertainty, the higher individuals perceive the privacy risk. In healthcare setting, the individual showed their concerns towards the privacy of EMR by forming boundaries based on risk assessment. Thus, this commendation leads to the evaluation of privacy risk among healthcare employees in this study.

S6 which measuring the IPC of individuals who use the Internet, revealed that privacy risk is positively related to privacy concerns. While, S38 discovered that young adolescents perceived privacy risks to be more severe, making them less likely to disclose information to e-marketers.

S37 explored privacy concerns and privacy risk on behaviour intention during emergencies and found that privacy risk was tested as the important factor of privacy concerns which is consistent with findings by Xu *et al.* (2008). This finding is also consistent with other articles (S5, S7, S8, S17, S21, S25, S35, and S36). From this argument, this study predicts that healthcare employees will be more concerned when using EMR when they consider EMR as vulnerable to privacy risks.

Two respondents from preliminary qualitative study strongly agreed that healthcare employees who consider EMR as vulnerable to privacy are those who are concerned about the privacy of EMR. They highlighted:

*“...knowing that there is a risk in managing EMR because it contains confidential information, surely will make them know the data needs to be protected...”*

(Respondent R2)

*“...if the user knows that there is a risk if they release the protected data to the third party without consent, this shows their concern on the privacy of the data ...”*

(Respondent R7)

Adapting the privacy risk definition by Mohamed and Ahmad [41], this study defined privacy risk as the extent to which a healthcare employee believes the potential threats of losing patients' privacy in the use of EMR. Hence, the following hypothesis is developed: Privacy risk positively influences information privacy concerns in the use of EMR (Hypothesis 1).

### B. Privacy Awareness

Article S11 exhibited in their survey that privacy awareness is an accurate reflection of IPC on social networking sites. Encouraging privacy awareness would stimulate privacy concerns rather than ignorance. By providing sufficient training and knowledge, it may benefit the healthcare practitioners to improve computer-literacy towards increasing their privacy awareness.

Privacy awareness indicates the extent on how a user knows about privacy practices and policies as stated in article S14. In the context of this study, the employees should know how to

use EMR in an appropriate way guided by privacy practices and policies to ensure the privacy and security of EMR. Article S13 reported that their respondents' privacy awareness significantly influenced privacy concerns regarding unauthorized access and secondary use of EMR.

The preliminary qualitative study discovered the importance of privacy awareness initiated by top management. This is to ensure the employees are getting informed about the existing privacy policy to protect EMR. Four respondents strongly described the importance of privacy awareness. They said:

*“The top management should create awareness on how to protect patients' data. Without awareness, people tend to do what they usually do.”*

(Respondent R3)

*“Policy development and awareness program should be the main focus that organisation must looking at...”*

(Respondent R4)

*“...it is important to identify the best way to make sure that the employees are informed about the policy...”*

(Respondent R8)

*“We conducted briefings at several agencies to inform about the existence and needs of PDPA. We hope that the briefing might give awareness to the society about the importance of privacy protection.”*

(Respondent R9)

In general, the initiative from government through news and media reports regarding privacy issues and specifically from organisations regarding privacy policy may strengthen individuals' concerns about privacy. Additionally, adequate education and training about privacy can increase the basic knowledge and develop better judgment of users with regards to IPC and it can help in prevention of privacy breaches. Notably, privacy awareness which consists of comprehensive training and education on privacy, is mandatory for healthcare employees. Hence, it is hypothesised that: Privacy awareness positively influences information privacy concerns in the use of EMR (Hypothesis 2).

### C. Privacy Policy

CPM theory suggests that privacy is about opening and closing boundaries to others [116]. One of the factors in developing the privacy rule is contextual, which emphasises on boundary setting to be defined in privacy policy. Defining the context in privacy policy allow a greater control over the protected information [118]. It helps the organisations to prevent and manage any possible privacy incident effectively.

Managing the variety of policies and restriction in this study refers to privacy policies which relate to the handling of

personal data and privacy. From conjoint analysis in article S10, there is positive relationship between privacy policy and IPC. As a result of knowing about privacy policy, the users are concern towards information privacy. Article S11 highlighted that the use of multiple measures stated in privacy policy provided better understanding on privacy issues. Privacy policy is also found related with IPC in quantitative analysis by article S31. This finding is also similar with other articles (S17, S24, and S34).

Notably, two of the respondents illuminated their views on this relationship during qualitative preliminary study:

*“...Towards the preparation for PDPA 2010 enforcement, the top management must have a policy that lists all privacy mechanisms on how patients’ data in HIS should be processed and protected. Therefore, I do think that it might affect directly towards the concerns on privacy among users...”*

(Respondent R2)

*“...The organisation must have privacy policy which must tailor with PDPA 2010 requirement. To have a policy is doesn’t make people notice that privacy is a must. It should be incorporated with ‘something’, which organisation needs to think about...”*

(Respondent R8)

Organisations need to deal with a variety of policies and restrictions that emerge from different sources, such as legislations (national or international), societal expectation, business requirements, and individual preferences [119]. Managing the variety of policies and restriction in this study refers to the development of privacy policy, which describes the procedure in handling personal data and procedures in managing privacy of the protected data. Therefore, it is predicted that: Privacy policy positively influences information privacy concerns in the use of EMR (Hypothesis 3).

## VI. CONCLUSION & FUTURE WORKS

In this study, 38 articles have been reviewed to identify the factors that influence IPC in the use of EMR. 3 most investigated factors have been discussed from the articles used in the review process. Privacy risk was found to be the most frequent factor that have been deliberated in the existing studies. This study also revealed other factors, namely privacy awareness and privacy policy.

This study also showed that there is limited research on IPC from employees’ perspective which suggests the need for future research on these issues. The review and the search process are based on methodological recommendations prescribed in the literature [109], [120], [121]. However, the selection of keywords, sources, inclusion and exclusion criteria, and time frame is based on researcher own judgment, and the choice has limitations.

This in-progress study will proceed in evaluating the identified factors. The unit of analysis for this study will be from selected healthcare organisation in Malaysia. This study

may assist healthcare organisations to design or implement privacy protection mechanisms of EMR towards complying with the PDPA 2010.

## Articles Reviewed in this Study

S1	Akhter, S. H. (2014). Privacy concern and online transactions: the impact of internet self-efficacy and internet involvement. <i>Journal of Consumer Marketing</i> , 31(2), 118–125.
S2	Awwal, M. a. (2012). Influence of Age and Genders on the Relationship between Computer Self-Efficacy and Information Privacy Concerns. <i>International Journal of Technology and Human Interaction</i> , 8(1), 14–37.
S3	Bulgurcu, B. (2010). Antecedents and Outcomes of Information Privacy Concerns in Online Social Networking : A Theoretical Perspective (Vol. 10).
S4	Bulgurcu, B., Cavusoglu, H., and Benbasat, I. (2010). Understanding Emergence and Outcomes of Information Privacy Concerns: A Case of Facebook. In <i>International Conference on Information Systems (ICIS)</i> .
S5	Cho, H. (2010). Determinants of Behavioral Responses to Online Privacy: The Effects of Concern, Risk Beliefs, Self-Efficacy, and Communication Sources on Self-Protection Strategies. <i>Journal of Information Privacy &amp; Security</i> , 6(February 2015), 3–27.
S6	Dinev, T., and Hart, P. (2004). Internet privacy concerns and their antecedents - measurement validity and a regression model. <i>Behaviour &amp; Information Technology</i> , 23(6), 413–422.
S7	Dinev, T., Xu, H., Smith, J. H., and Hart, P. (2012). Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts. <i>European Journal of Information Systems</i> , 22(3), 295–316.
S8	El Emam, K., Mercer, J., Moreau, K., Grava-Gubins, I., Buckeridge, D., and Jonker, E. (2011). Physician privacy concerns when disclosing patient data for public health purposes during a pandemic influenza outbreak. <i>BMC Public Health</i> , 11, 454.
S9	Ellison, N. B., Vitak, J., Steinfield, C., Gray, R., and Lampe, C. (2011). Negotiating Privacy Concerns and Social Capital Needs in a Social Media Environment. In <i>Privacy Online</i> (pp. 19–32). Springer-Verlag Berlin Heidelberg.
S10	Hann, I., Hui, K., Lee, S., and Png, I. (2007). Analyzing Online Information Privacy Concerns: An Information Processing Theory Approach. In <i>40th Annual Hawaii International Conference on System Sciences (HICSS'07)</i> (Vol. 1, p. 210b–210b).
S11	Hazari, S., and Brown, C. (2013). An Empirical Investigation of Privacy Awareness and Concerns on Social Networking Sites. <i>Journal of Information Privacy &amp; Security</i> , 9(February 2015), 31–52.
S12	Hichang Cho, Rivera-Sanchez, M., and Sun Sun Lim. (2009). A multinational study on online privacy: global concerns and local responses. <i>New Media &amp; Society</i> , 11(3), 395–416.
S13	Hwang, H.-G., Han, H.-E., Kuo, K.-M., and Liu, C.-F. (2012). The Differing Privacy Concerns Regarding Exchanging Electronic Medical Records of Internet Users in Taiwan. <i>Journal of Medical Systems</i> , 36(6), 3783–3793.
S14	Jiang, X. (2011). Privacy concern toward using social networking services: A conceptual model. <i>2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)</i> , 3180–3183.
S15	Johnson, B. (2007). A Mixed-Methods Study of the Influence of Generational Consciousness on Information Privacy Concern. <i>Walden University</i> .
S16	King, T., Brankovic, L., and Gillard, P. (2012). Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. <i>International Journal of Medical Informatics</i> , 81(4), 279–89.
S17	Lane, J., and Schur, C. (2010). Balancing access to health data and privacy: a review of the issues and approaches for the future. <i>Health Services Research</i> , 45(5 Pt 2), 1456–67.
S18	Lankton, N., and Tripp, J. (2013). A Quantitative and Qualitative Study of Facebook Privacy using the Antecedent-Privacy Concern-Outcome Macro Model. In <i>Proceedings of the Nineteenth Americas Conference on Information Systems</i> (pp. 1–12). Chicago, Illinois.
S19	Li, Y. (2011). Empirical Studies on Online Information Privacy Concerns : Literature Review and an Integrative Framework. <i>Communications of the Association for Information Systems</i> , 28(May), 453–496.
S20	Li, Z., Lv, T., Zhang, X., and Chen, X. (2013). The effects of personal characteristics and interpersonal influence on privacy information diffusion in SNS. <i>Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics</i> , 413–418.
S21	Liao, C., Liu, C.-C., and Chen, K. (2011). Examining the impact of privacy, trust and risk perceptions beyond monetary transactions: An integrated model. <i>Electronic Commerce Research and Applications</i> , 10(6), 702–715.
S22	Lin, Y. (2005). Information Privacy Concerns in the Customer Relationship Management Context: A Comparison of Consumer Attitudes in the U.S., China, and Taiwan. <i>Golden Gate University</i> .

S23	Malandrino, D., Scarano, V., and Spinelli, R. (2013). How increased awareness can impact attitudes and behaviors toward online privacy protection. Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013, 57–62.	[8]	Laws of Malaysia, <i>Act 709: Personal Data Protection Act 2010</i> . 2010, pp. 1–95.
S24	Mekovec, R., and Vreck, N. (2011). Factors that influence Internet users' privacy perception. Proceedings of the ITI 2011, 33rd International Conference on Information Technology Interfaces, 227–232.	[9]	S. Zulhuda and A. Ibrahim, "The State of E-Government Security in Malaysia : Reassessing the Legal and Regulatory Framework on the Threat of Information Theft," in <i>ICCIT 2012</i> , 2012, pp. 810–815.
S25	Mohamed, N., and Ahmad, I. H. (2012). Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. <i>Computers in Human Behavior</i> , 28(6), 2366–2375.	[10]	Ponemon Institute, "Third Annual Benchmark Study on Patient Privacy & Data Security," 2012.
S26	Morton, A. (2013). Measuring Inherent Privacy Concern and Desire for Privacy - A Pilot Survey Study of an Instrument to Measure Dispositional Privacy Concern. 2013 International Conference on Social Computing, 468–477.	[11]	R. F. Parks, C.-H. Chu, and H. Xu, "Healthcare Information Privacy Research : Issues , Gaps and What Next?," in <i>Americas Conference on Information Systems (AMCIS) 2011 Proceedings</i> , 2011.
S27	Moscardelli, D. M., and Divine, R. (2007). Adolescents' Concern for Privacy When Using the Internet: An Empirical Analysis of Predictors and Relationships With Privacy-Protecting Behaviors. <i>Family and Consumer Sciences Research Journal</i> , 35(3), 232–252.	[12]	J. L. Fernández-Alemán, I. C. Señor, P. Á. O. Lozoya, and A. Toval, "Security and privacy in electronic health records: A systematic literature review," <i>J. Biomed. Inform.</i> , vol. 46, no. 3, pp. 541–62, Jun. 2013.
S28	Okazaki, S., Li, H., and Hirose, M. (2009). Consumer Privacy Concerns and Preference for Degree of Regulatory Control. <i>Journal of Advertising</i> , 38(4), 63–77.	[13]	S. Samsuri, Z. Ismail, and R. Ahmad, "Privacy models for protecting personal medical information: A preliminary study," in <i>2011 International Conference on Research and Innovation in Information Systems</i> , 2011, pp. 1–5.
S29	Perera, G., Holbrook, A., Thabane, L., Foster, G., and Willison, D. J. (2011). Views on health information sharing and privacy from primary care practices using electronic medical records. <i>International Journal of Medical Informatics</i> , 80(2), 94–101.	[14]	S. Samsuri, R. Ahmad, and Z. Ismail, "Towards Implementing a Privacy Policy: An Observation on Existing Practices in Hospital Information System," <i>J. e-Health Manag.</i> , vol. 2011, pp. 1–9, Jan. 2011.
S30	Pitkänen, O., and Tuunainen, V. K. (2012). Disclosing Personal Data Socially — An Empirical Study on Facebook Users' Privacy Awareness. <i>Journal of Information Privacy and Security</i> , 8(1), 3–29.	[15]	J. M. Stanton, "Information Technology and Privacy: A Boundary Management Perspective," in <i>Socio-Technical and Human Cognition Elements of Information Systems</i> , Idea Group Publishing, 2003, pp. 79–103.
S31	Samavi, R., Consens, M. P., and Chignell, M. (2014). PHR User Privacy Concerns and Behaviours. <i>Procedia Computer Science</i> , 37, 517–524.	[16]	M. Nazlina, "Court orders UKM to pay RM400 , 000 to man for revealing psychiatric medical records," <i>The Star Online</i> , 30-Oct-2013.
S32	Schwaig, K. S., Segars, A. H., Grover, V., and Fiedler, K. D. (2013). A model of consumers' perceptions of the invasion of information privacy. <i>Information &amp; Management</i> , 50(1), 1–12.	[17]	P. Dielenberg, "Court : Surgeons must get consent to take photos of intimate parts," <i>The Star Online</i> , 02-Sep-2010.
S33	Smith, H. J., Dinev, T., and Xu, H. (2011). Information Privacy Research: An Interdisciplinary Review. <i>MIS Quarterly</i> , 35(4), 989–1015.	[18]	E. AbuKhousa, N. Mohamed, and J. Al-Jaroodi, "e-Health Cloud: Opportunities and Challenges," <i>Futur. Internet</i> , vol. 4, pp. 621–645, 2012.
S34	Wirtz, J., Lwin, M. O., and Williams, J. D. (2007). Causes and consequences of consumer online privacy concern. <i>International Journal of Service Industry Management</i> , 18(4), 326–348.	[19]	R. G. Fichman, "The Role of Information Systems in Healthcare : Current Research and Future Trends," <i>Inf. Syst. Res.</i> , vol. 22, no. 3, pp. 419–428, 2011.
S35	Xu, F., Michael, K., and Chen, X. (2013). Factors affecting privacy disclosure on social network sites: an integrated model. <i>Electronic Commerce Research</i> , 13(2), 151–168.	[20]	A. Kaletsch and A. Sunyaev, "Privacy Engineering: Personal Health Records in Cloud Computing Environments," in <i>Thirty Second International Conference on Information Systems</i> , 2011, pp. 1–11.
S36	Xu, H., Dinev, T., Smith, J., and Hart, P. (2011). Information Privacy Concerns : Linking Individual Perceptions with Institutional Privacy Assurances. <i>Journal of the Association for Information Systems</i> , 12(12), 798–824.	[21]	W. N. H. Wan Alias, "Sindiket Beli, Jual Data Pesakit (The Buying And Selling of Patients' Data Syndicate)," <i>Berita Harian</i> , 18-Jan-2016.
S37	Yang, H., and Liu, H. (2014). Prior negative experience of online disclosure, privacy concerns, and regulatory support in Chinese social media. <i>Chinese Journal of Communication</i> , 7(1), 40–59.	[22]	T. Ermakova, K. Erek, and J. Huenges, "Cloud Computing in Healthcare – a Literature Review on Current State of Research," in <i>Proceedings of the Nineteenth Americas Conference on Information Systems</i> , 2013, pp. 1–9.
S38	Youn, S. (2009). Determinants of Online Privacy Concern and Its Influence on Privacy Protection Behaviors Among Young Adolescents. <i>The Journal of Consumer Affairs</i> , 43(3), 389–418.	[23]	Symantec, "Internet Security Threat Report 2011," 2012.
		[24]	Symantec, "Internet Security Threat Report 2013," 2013.
		[25]	Symantec, "Internet Security Threat Report 2014," 2014.
		[26]	Symantec, "Internet Security Threat Report 2015," 2015.
		[27]	J. M. Stanton, K. R. Stam, P. Mastrangelo, and J. Jolton, "Analysis of end user security behaviors," <i>Comput. Secur.</i> , vol. 24, no. 2, pp. 124–133, Mar. 2005.
		[28]	A. I. Antón, J. B. Earp, and A. Reese, "Analyzing Website Privacy Requirements Using a Privacy Goal Taxonomy," in <i>IEEE Joint International Requirements Engineering Conference 2002</i> , 2002.
		[29]	R. T. Mercuri, "The HIPAA-potamus in health care data security," <i>Commun. ACM</i> , vol. 47, no. 7, pp. 25–28, 2004.
		[30]	D. Grunwell, R. Gajanayake, and T. Sahama, "Demonstrating Accountable-eHealth Systems," 2014.
		[31]	N. Spector and D. M. Kappel, "Guidelines for Using Electronic and Social Media : The Regulatory Perspective," <i>Online J. Issues Nurs.</i> , vol. 17, no. 3, 2012.
		[32]	R. Cushman, M. Froomkin, A. Cava, P. Abril, and K. W. Goodman, "Ethical, legal and social issues for personal health records and applications.," <i>J. Biomed. Inform.</i> , vol. 43, no. 5 Suppl, pp. S51–5, Oct. 2010.
		[33]	S. J. Nass, L. A. Levitt, and L. O. Gostin, <i>Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research</i> . 2009.
		[34]	J. B. Earp and F. C. Payton, "Data Protection in the University Setting : Employee Perceptions of Student Privacy," in <i>Proceedings of the 34th Hawaii International Conference on System Sciences - 2001</i> , 2001, vol. 00, no. c, pp. 1–6.

## REFERENCES

- [1] Y. Tong and H.-H. Teo, "Migrating to Integrated Electronic Medical Record: An Empirical Investigation of Physicians' Use Preference," in *Proceedings of the 30th Annual International Conference on Information Systems (ICIS)*, 2009.
- [2] H. Mohd and S. M. S. Mohamad, "Acceptance Model of Electronic Medical Record," *J. Adv. Inf. Manag. Stud.* 2(1), vol. 2, no. 1, pp. 75–92, 2005.
- [3] N. I. Ismail and N. H. Abdullah, "An Overview of Hospital Information System (HIS) Implementation in Malaysia," in *3rd International Conference on Business and Economic Research Proceeding*, 2012, no. March, pp. 1176–1182.
- [4] P. Kierkegaard, "Electronic health record: Wiring Europe's healthcare," *Comput. Law Secur. Rev.*, vol. 27, no. 5, pp. 503–515, 2011.
- [5] K. Wadhwa and D. Wright, "eHealth: Legal, Ethical and Governance Challenges," in *eHealth: Legal, Ethical and Governance Challenges*, C. George, D. Whitehouse, and P. Duquenoy, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 183–210.
- [6] A. Appari and M. E. Johnson, "Information security and privacy in healthcare: current state of research," *Int. J. Internet Enterp. Manag.*, vol. 6, no. 4, pp. 279–314, 2010.
- [7] Verizon, "2015 Data Breach Investigations Report," 2015.

- [35] J. B. Earp and F. C. Payton, "Information Privacy in the Service Sector: An Exploratory Study of Health Care and Banking Professionals," *J. Organ. Comput. Electron. Commer.*, vol. 16, no. 2, pp. 105–122, Jan. 2006.
- [36] K. Ball, E. M. Daniel, and C. Stride, "Dimensions of employee privacy: an empirical study," *Inf. Technol. People*, vol. 25, no. 4, pp. 376–394, Nov. 2012.
- [37] B. Lebek and M. H. Breitner, "Investigating the Influence of Security , Privacy , and Legal Concerns on Employees ' Intention to Use BYOD Mobile Devices," no. 2008, pp. 1–8, 2013.
- [38] M. Lallmahamood, "An Examination of Individual ' s Perceived Security and Privacy of the Internet in Malaysia and the Influence of This on Their Intention to Use E-Commerce : Using An Extension of the Technology Acceptance Model," *J. Internet Bank. Commer.*, vol. 12, no. 3, pp. 1–26, 2007.
- [39] M. Lallmahamood, "Privacy over the Internet in Malaysia : A Survey of General Concerns and Preferences among Private Individuals," *Malaysian Manag. Rev.*, vol. 43, no. 1, pp. 77–108, 2008.
- [40] N. Mohamed and I. H. Ahmad, "Privacy Measures Awareness , Privacy Setting Use and Information Privacy Concern with Social Networking Sites," in *International Conference on Research and Innovation in Information Systems (ICRIS)*, 2011, 2011, no. November.
- [41] N. Mohamed and I. H. Ahmad, "Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia," *Comput. Human Behav.*, vol. 28, no. 6, pp. 2366–2375, Nov. 2012.
- [42] O. Boyinbode and G. Toriola, "CloudMR : A Cloud Based Electronic Medical Record System," *Int. J. Hybrid Inf. Technol.*, vol. 8, no. 4, pp. 201–212, 2015.
- [43] V. Mantzana, M. Themistocleous, and V. Morabito, "Healthcare information systems and older employees' training," *J. Enterp. Inf. Manag.*, vol. 23, no. 6, pp. 680–693, 2010.
- [44] L. L. Song, X. Q. Guo, and C. Wang, "Research on Model-Driven Simulation Approach for Healthcare Information System," in *International Conference on Artificial Intelligence and Industrial Engineering*, 2015, no. Aiie, pp. 592–596.
- [45] C. L. Hsu, M. R. Lee, and C. H. Su, "The role of privacy protection in healthcare information systems adoption," *J. Med. Syst.*, vol. 37, no. 5, 2013.
- [46] F.-Y. Pai and K.-I. Huang, "Applying the Technology Acceptance Model to the introduction of healthcare information systems," *Technol. Forecast. Soc. Change*, vol. 78, no. 4, pp. 650–660, 2011.
- [47] R. L. Simpson and C. A. Weaver, "Administrative application of information technology for nursing managers," in *Essentials of nursing informatics*, 4th Editio., V. K. Saba and K. A. McCormick, Eds. McGraw Hill, 2005, p. 445.
- [48] M. Rahman and C. Kreider, "Information Security Principles for Electronic Medical Record ( EMR ) Systems," in *Americas Conference on Information Systems (AMCIS) 2012 Proceedings*, 2012.
- [49] W.-S. Hsu and J.-I. Pan, "The Secure Authorization Model for Healthcare Information System," *J. Med. Syst.*, vol. 37, no. 5, 2013.
- [50] H. O. Al-Sakran, "Framework Architecture for Improving Healthcare Information Systems Using Agent Technology," *Int. J. Manag. Inf. Technol.*, vol. 7, no. 1, pp. 17–31, 2015.
- [51] K. Kaur and R. Rani, "Managing Data in Healthcare Information Systems : Many Models, One Solution," *Computer (Long. Beach. Calif.)*, vol. 48, no. 3, 2015.
- [52] H. Cripps and C. Standing, "The implementation of electronic health records: a case study of bush computing the Ngaanyatjarru lands.," *Int. J. Med. Inform.*, vol. 80, no. 12, pp. 841–8, Dec. 2011.
- [53] J. M. Haakon Bryhni and C. M. Ruland, "Secure Solution for Mobile Access to Patient's Health Care Record," *IEEE 13th International Conference on e-Health Networking, Applications and Services*. Columbia, USA, pp. 296–303, 2011.
- [54] F. Mancini, S. Gejibo, K. A. Mughal, R. A. B. Valvik, and J. Klungsøy, "Secure Mobile Data Collection Systems for Low-Budget Settings," *2012 Seventh International Conference on Availability, Reliability and Security*. IEEE, Prague, Czech Republic , pp. 196–205, 2012.
- [55] I. Ćubić, I. Markota, and I. Benc, "Application of Session Initiation Protocol in Mobile Health Systems," *33rd International Convention on Information and Communication Technology, Electronics and Microelectronics*. IEEE, Opatija, Croatia, pp. 367–371, 2010.
- [56] M. T. Sandikkaya, B. De Decker, and V. Naessens, "Privacy in Commercial Medical Storage Systems," in *Electronic Healthcare*, M. Szomszor and P. Kostkova, Eds. Springer Berlin Heidelberg, 2011, pp. 247–258.
- [57] S. V. B. Jardim, "The Electronic Health Record and its Contribution to Healthcare Information Systems Interoperability," *Procedia Technol.*, vol. 9, pp. 940–948, 2013.
- [58] J. Sun and Y. Fang, "Cross-domain data sharing in distributed electronic health record systems," *IEEE Trans. Parallel Distrib. Syst. Distrib. Syst.*, vol. 21, no. 6, pp. 754–764, 2010.
- [59] S. R. Simon, J. S. Evans, A. Benjamin, D. Delano, and D. W. Bates, "Patients' attitudes toward electronic health information exchange: qualitative study.," *J. Med. Internet Res.*, vol. 11, no. 3, p. e30, Jan. 2009.
- [60] J. A. James Baroody and S. W. Hansen, "Changing Perspective: Institutional Logics of Adoption and Use of Health Information Technology," in *Thirty Third International Conference on Information Systems*, 2012, pp. 1–18.
- [61] M. Tiggle, "Urban Alabama Physicians and the Electronic Medical Record: A Qualitative Study," Capella University, 2012.
- [62] E. Deutsch, G. Duftschmid, and W. Dorda, "Critical areas of national electronic health record programs-is our focus correct?," *Int. J. Med. Inform.*, vol. 79, no. 3, pp. 211–22, Mar. 2010.
- [63] P. H. Chang, "Modeling the Management of Electronic Health Records in Healthcare Information Systems," *2011 Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discov.*, pp. 580–584, 2011.
- [64] J. M. Holroyd-Leduc, D. Lorenzetti, S. E. Straus, L. Sykes, and H. Quan, "The impact of the electronic medical record on structure, process, and outcomes within primary care: a systematic review of the evidence.," *J. Am. Med. Inform. Assoc.*, vol. 18, no. 6, pp. 732–7, 2011.
- [65] M. Kimura, J. Nakaya, H. Watanabe, T. Shimizu, and K. Nakayasu, "A Survey Aimed at General Citizens of the US and Japan about Their Attitudes toward Electronic Medical Data Handling," *Int. J. Environ. Res. Public Health*, vol. 11, no. 5, pp. 4572–4588, 2014.
- [66] T. Mettler, "Post-Acceptance of Electronic Medical Records: Evidence from a Longitudinal Field Study," in *Thirty Third International Conference on Information Systems 2012*, 2012, pp. 1–19.
- [67] International Organisation for Standardisation, "ISO/TR 20514:2005 Health informatics -- Electronic health record -- Definition, scope and context," 2005.
- [68] J. B. Earp, A. I. Antón, L. Aiman-smith, and W. H. Stufflebeam, "Examining Internet Privacy Policies Within the Context of User Privacy Values," *IEEE Trans. Eng. Manag.*, vol. 52, no. 2, pp. 227–237, 2005.
- [69] J. Wirtz, M. O. Lwin, and J. D. Williams, "Causes and consequences of consumer online privacy concern," *Int. J. Serv. Ind. Manag.*, vol. 18, no. 4, pp. 326–348, 2007.
- [70] D. G. Taylor, D. F. Davis, and R. Jillapalli, "Privacy concern and online personalization: the moderating effects of information control and compensation," *Springer Sci. Media*, 2009.
- [71] K.-W. Wu, S. Y. Huang, D. C. Yen, and I. Popova, "The effect of online privacy policy on consumer privacy concern and trust," *Comput. Human Behav.*, vol. 28, no. 3, pp. 889–897, May 2012.
- [72] C. Randolph, J. R. Barrows, and P. D. Clayton, "Privacy, Confidentiality and Electronic Medical Records," *J. Am. Med. Informatics Assoc.*, vol. 3, no. 2, pp. 139–148, 1996.
- [73] G. Kurtz, "EMR Confidentiality and Information Security," *J. Healthc. Inf. Manag.*, vol. 7, no. 3, pp. 41–48, 2002.
- [74] M. A. Hall and K. A. Schulman, "Ownership of Medical Information," *J. Am. Med. Assoc.*, vol. 301, no. 12, pp. 1282–1284, 2009.
- [75] L. M. Lee and L. O. Gostin, "Ethical Collection, Storage, and Use of Public Health Data: A Proposal for a National Privacy Protection," *J. Am. Med. Assoc.*, vol. 302, no. 1, pp. 82–84, 2009.
- [76] The London School of Economics and Political Science, "Electronic Health Privacy and Security in Developing Countries and Humanitarian Operations," The London School of Economics and Political Science, 2010.
- [77] J. Adler-Milstein and K. J. Ashish, "Sharing Clinical Data Electronically: Critical Challenge for Fixing the Health Care System," *J. Am. Med. Assoc.*, vol. 307, no. 16, pp. 1695–1696, 2012.

- [78] I. Carrión Señor, J. L. Fernández-Alemán, and A. Toval, "Are personal health records safe? A review of free web-accessible personal health record privacy policies," *J. Med. Internet Res.*, vol. 14, no. 4, p. e114, 2012.
- [79] A. F. Westin, *Privacy and Freedom*. Atheneum. The Bodley Head Ltd, 1967.
- [80] M. Modaresnezhad and S. Chain, "The Efficacy of IS Privacy and Security Governance Structures," 2012.
- [81] A. B. Munir and S. H. M. Yasin, *Personal Data Protection in Malaysia*. Sweet & Maxwell Asia, 2010.
- [82] B. Borena, F. Belanger, and D. Ejigu, "Social Networks and Information Privacy: A Model for Low-income Countries," in *Proceedings of the Nineteenth Americas Conference on Information Systems*, 2013, pp. 1–9.
- [83] R. F. Parks, C.-H. Chu, H. Xu, and L. Adams, "Understanding the Drivers and Outcomes of Healthcare Organizational Privacy Responses," in *Thirty Second International Conference on Information Systems*, 2011, no. 2, pp. 1–20.
- [84] S. Samsuri, Z. Ismail, and R. Ahmad, "User-Centered Evaluation of Privacy Models for Protecting Personal Medical Information," in *International Conference, ICIEIS 2011, Kuala Lumpur, Malaysia*, 2011, pp. 301–309.
- [85] Ponemon Institute, "Fourth Annual Benchmark Study on Patient Privacy & Data Security," 2014.
- [86] R. Hodgkinson, L. Branz, M. Culnan, G. Dhillon, and A. MacWilson, "Information Security and Privacy: Rethinking Governance Models," in *International Conference on Information Systems (ICIS)*, 2010.
- [87] Y. K. Mittal, S. Roy, and M. Saxena, "Role of Knowledge Management in Enhancing Information Security," *IJCSI Int. J. Comput. Sci. Issues*, vol. 7, no. 6, 2010.
- [88] M. J. Culnan and C. C. Williams, "How Ethics Can Enhance Organizational Privacy: Lessons from the ChoicePoint and TJX Data Breaches," *MIS Q.*, vol. 33, no. 4, pp. 673–687, 2009.
- [89] G. N. Samy, R. Ahmad, and Z. Ismail, "Threats to Health Information Security," in *2009 Fifth International Conference on Information Assurance and Security*, 2009, pp. 540–543.
- [90] D. Wartenberg and W. D. Thompson, "Privacy versus public health: the impact of current confidentiality rules," *Am. J. Public Health*, vol. 100, no. 3, pp. 407–12, Mar. 2010.
- [91] P. Ambrose and C. Basu, "Interpreting the Impact of Perceived Privacy and Security Concerns in Patients' Use of Online Health Information Systems," *J. Inf. Priv. Secur.*, vol. 8, no. February 2015, pp. 38–50, 2015.
- [92] D. Birnbaum, E. Borycki, B. T. Karras, E. Denham, and P. Lacroix, "Addressing Public Health informatics patient privacy concerns," *Clin. Gov. An Int. J.*, vol. 20, no. 2, pp. 91–100, 2015.
- [93] W. Chung and L. Hershey, "Enhancing Information Privacy and Data Sharing in a Healthcare IT Firm: The Case of Ricerro Communications," *J. Inf. Priv. Secur.*, vol. 8, no. February 2015, pp. 56–78, 2014.
- [94] T. Ermakova, B. Fabian, and R. Zarnekow, "Security and Privacy System Requirements for Adopting Cloud Computing in Healthcare Data Sharing Scenarios," in *Proceedings of the Nineteenth Americas Conference on Information Systems*, 2013, pp. 1–9.
- [95] I. C. S. José Luis Fernández-Alemán Pedro Ángel Oliver Lozoya, Ambrosio Toval, "Security and privacy in electronic health records: A systematic literature review," *J. Biomed. Inform.*, pp. 1–22, 2013.
- [96] R. F. Parks and R. T. Wigand, "Organizational Privacy Strategy: Four Quadrants of Strategic Responses to Information Privacy and Security Threats," *J. Inf. Priv. Secur.*, vol. 10, no. February, pp. 203–224, 2015.
- [97] H. J. Smith, S. J. Milberg, and S. J. Burke, "Information Privacy: Measuring Individuals' Concerns About Organizational Practices," *MIS Q.*, vol. 20, no. 2, pp. 167–196, 1996.
- [98] J. Kolter and G. Pernul, "Generating User-Understandable Privacy Preferences," in *2009 International Conference on Availability, Reliability and Security*, 2009, pp. 299–306.
- [99] H. J. Smith, T. Dinev, and H. Xu, "Information Privacy Research: An Interdisciplinary Review," *MIS Q.*, vol. 35, no. 4, pp. 989–1015, 2011.
- [100] N. K. Malhotra, S. S. Kim, and J. Agarwal, "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model," *Inf. Syst. Res.*, vol. 15, no. 4, pp. 336–355, Dec. 2004.
- [101] A. J. Campbell, "Relationship marketing in consumer markets: A comparison of managerial and consumer attitudes about information privacy," *J. Direct Mark.*, vol. 11, no. 3, pp. 44–57, 1997.
- [102] I. Park, "The Study on The Relationship Between Privacy Concerns and Information Systems Effectiveness," in *International Conference on Information Systems (ICIS)*, 2009.
- [103] X. Tan, L. Qin, Y. Kim, and J. Hsu, "Impact of privacy concern in social networking web sites," *Internet Res.*, vol. 22, no. 2, pp. 211–233, 2012.
- [104] H. Xu, S. Gupta, M. B. Rosson, and J. M. Carroll, "Measuring Mobile Users' Concerns for Information Privacy," in *Thirty Third International Conference on Information Systems*, 2012, no. Ftc 2009, pp. 1–16.
- [105] F. Bélanger and R. E. Crossler, "Privacy in the Digital Age - A Review of Information Privacy Research in Information Systems," *MIS Q.*, vol. 35, no. 4, pp. 1017–1041, 2011.
- [106] R. Mekovec and N. Vrcek, "Factors that influence Internet users' privacy perception," in *International Conference on Information Technology Interfaces (ITI), Proceedings of the ITI 2011 33rd*, 2011, pp. 227–232.
- [107] K. S. Schwaig, A. H. Segars, V. Grover, and K. D. Fiedler, "A model of consumers' perceptions of the invasion of information privacy," *Inf. Manag.*, vol. 50, no. 1, pp. 1–12, Jan. 2013.
- [108] Y. Li, "Theories in online information privacy research: A critical review and an integrated framework," *Decis. Support Syst.*, vol. 54, no. 1, pp. 471–481, Dec. 2012.
- [109] C. Okoli and K. Schabram, "A Guide to Conducting a Systematic Literature Review of Information Systems Research," 10, 2010.
- [110] S. N. Hesse-Biber and P. Leavy, *The Practice of Qualitative Research*. SAGE Publications, Inc, 2006.
- [111] C. Glesne, *Becoming Qualitative Researchers*. Pearson, 2011.
- [112] R. K. Yin, *Qualitative Research from Start to Finish*. New York: The Guilford Press, 2010.
- [113] J. W. Creswell, *Education Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*, Fourth Edi. Pearson, 2012.
- [114] J. W. Creswell and V. L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 2nd Editio. SAGE Publications, Inc, 2011.
- [115] R. S. Barbour, *Introducing Qualitative Research*. SAGE Publications, Inc, 2008.
- [116] S. Petronio, *Boundaries of Privacy: Dialectics of Disclosure*. Albany, NY, US: State University of New York Press, 2002.
- [117] H. Xu, T. Dinev, H. J. Smith, and P. Hart, "Examining the Formation of Individual's Privacy Concerns: Toward an Integrative View," in *International Conference on Information Systems (ICIS)*, 2008.
- [118] G. C. Kane, K. S. Schwaig, and V. C. Storey, "Information Privacy: Understanding How Firms Behave Online," in *Theoretical and Practical Advances in Information Systems Development*, IGI Global, 2011, pp. 81–100.
- [119] M. C. Mont, S. Pearson, S. Creese, M. Goldsmith, and N. Papanikolaou, "A Conceptual Model for Privacy Policies with Consent and Revocation Requirements," in *Privacy and Identity Management for Life*, S. Fischer-Hübner, P. Duquenoy, M. Hansen, R. Leenes, and G. Zhang, Eds. Sweden: Springer Berlin Heidelberg, 2011, pp. 258–270.
- [120] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.
- [121] M. Petticrew and H. Roberts, *Systematic Review in the Social Sciences: A Practical Guide*. Blackwell Publishing, 2006.

# Moment Based Copy Move Forgery Detection Methods

Khaled W. Mahmoud, Arwa Husien Abu Al-Rukab

Computer Science Department

Zarqa University

Zarqa, Jordan

**Abstract**—Copy-Move forgery is one important type of image forgery. In this type of forgery, part of the image is copied and pasted into another position in the same image. This is done in order to hide an object inside the image by covering it. To detect this type of forgery, many methods (algorithms) were published. Each method has its own strong points and drawbacks. One of the most important aspects in detecting copy-move forgery is how to read the image; the features which used to represent the image. It is important to realize that having invariant features, will support the robustness of the detection method against different attacks that the copied parts may affected by. Different studies show that moment invariants are one of the best choices in image processing. In this paper, a brief introduction to moment is given and detection methods that are based on moments are illustrated and analyzed.

**Keywords-** *Forgery; Forensics; Moments; Zernike; Hu*

## I. INTRODUCTION

Images are a primary source of information. More than 95% of information that are received by human are optical [1]. On the other hand, images may subject to amendment and as a result the integrity and the authenticity of the digital images became very sensitive. This led to the appearance of a new field of science called “digital image forensic”. The aim if this science is to validating the authenticity of images by recovering information about their history [2].

Digital image forgery techniques are classified into two principle approaches; the active approach and the passive approach. In the active approach the signature that was embedded in the image during image creation is extracted. This approach is sometimes called “watermarking”. The passive approach checks the authenticity of images from an unknown and uncontrolled source (i.e. blind image forensics). This approach is categorized into three groups: copy-move, image splicing, and image retouching [3]. Copy-Move forgery or region duplication forgery is the most important type of forgery. In recent years, this forgery has become one of the most actively researched topics [4].

In Copy-Move forgery, part of the image is copied and pasted into another position of the same image. This is done in order to hide an object inside the image by covering it or just duplicating an object. Fig. 1 shows an example, where the bird

is copied and pasted into another position in the same image. Usually grass, foliage, gravel, or fabric with irregular patterns is ideal for this purpose. Since the copied areas are likely harmonious with the background and compatible with the rest of the image, human eyes cannot easily catch any suspect thing in the image [5].



Figure 1. Example of a Copy-Move Forgery: (left) is the original image, (right) is the forged image

Any forgery introduces a correlation between the copied parts of the image and the pasted ones. This can be used as a basis for a successful detection of this type of forgery. The following are very important requirements for the detection algorithm [5]:

- The detection algorithm must allow for an approximate match of small and large image parts.
- It must work in a reasonable time.
- It must introduce few false positives (i.e., detecting incorrect areas).
- The forged segment will likely be a connected component rather than a collection of individual pixels.
- The performance of the detector should also be demonstrated on several forged images.

For making the forgery undetectable, copied objects may be affected by different operations such as rotation, scaling, blurring, filtering, noise addition, JPEG compression... etc. These operations make the forgery detection process more difficult and fragile. Therefore forgery detector should be robust against different operations that may affect the image.

One of the most important aspects in detecting copy-move forgery is how to represent the image as a set of features. More

invariance these features are the more robustness the detector is. It is a very good idea to select feature insensitive to particular deformations, and provide enough discrimination power to distinguish objects belonging to different classes. This called invariant features. In this paper, detection methods that are based on moment invariants are illustrated and analyzed.

The rest of this paper is organized as follows: section 2 present the typical flow for most Copy-Move Forgery Detection (CMFD) methods. The mathematical background of the moments is given in section 3. In Section 4 and 5; the moment based methods are presented. Conclusion and comparisons are given in last section.

## II. TYPICAL IMAGE CMFD METHODS

A very rich literature in the field of CMFD focuses mainly on the robustness and the speed of the detection method. These methods can be classified into three broad categories:

- **Segment-based Methods:** Methods that subdivide the image into meaningful structure and then compute the features for each structure [3], [6].
- **Key-Point based Methods:** Methods that compute the features only on image regions with high entropy, without any image subdivision. For example, method that based on SIFT [7], and SURF [8].
- **Block-based Methods:** In these methods, image is subdivided into overlapping (rectangular or circular) blocks for feature extraction. Since moment based methods are fall into this category, more information are given next.

### A. Block-based Methods

Typically, detection process for this type of methods is shown in Fig. 2 and the main steps are explained below.

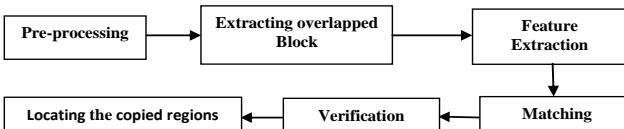


Figure 2. CMFD Pipeline

1) *Preprocessing:* This step is used to improve the computational time by preparing the image for the next step. The most popular operations in this step are: converting color images into gray scale images (gray scale image is simple to enhance and interprets), and scale down the image before going on to the remaining steps.

2) *Extracting overlapped blocks:* As shown in Fig. 3, input image with resolution  $M \times N$  is divided into  $(M - B + 1) \times (N - B + 1)$  blocks, where each block is of  $B \times B$  size.

3) *Feature extraction:* Here the feature vector of each block is computed. The robustness of this feature against different post-processing operations reduces the false match rate and gives better chance to detect the forged region.

4) *Matching:* The aim of this step is to find the duplicated blocks based on their feature vectors. This can be done by

sorting the vectors and compute the similarity between neighbor's vectors. High similarity between two vectors is interpreted as a hint for duplicated blocks. Lexicographically sorting [9], [10] and K-D tree [11], [12] are the most common sorting methods that were used.

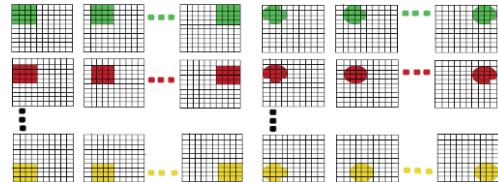


Figure 3. Overlapped blocks. (left) Square  $4 \times 4$  block (right) Circular  $4 \times 4$  block.

5) *Verification (Filtering):* This step is performed in order to group matches that jointly follow a transformation pattern and so reduce superior matches. For example matches that belong to a copied region are expected to be spatially close to each other in both source and target blocks. Furthermore, matches that originate from the same copy-move action should exhibit similar amounts of translation, scaling and rotation.

6) *Locating (highlighting) the duplicated region:* this can be done by coloring the copied blocks.

Block-based methods can be classified according to the way that used to compute the feature vector. These subcategories are illustrated next.

1) *Frequency Domain-based Methods:* These methods transform the image to the frequency domain for the purpose of using frequency components in the detection process. This transformation can be done using different operator such as:

a) *Discrete Cosine Transform (DCT):* Experiments in [13], [14] shows that features that are extracted from DCT domain can be used successfully to detect the duplicated regions even when the input image was distorted by JPEG compression, blurring or additive white Gaussian noise. However, it is not robust against rotation or scaling.

b) *Fourier Mellin Transform (FMT):* According to the experiments in [11], features extracted from FMT domain can be used to detect regions that have been rotated up to 10 degree, scaled up to 10% (started degenerating at a scale change of about 106%) and distorted by JPEG compression up to 20% quality.

c) *Discrete Wavelet Transform (DWT):* The main advantage of DWT is to yield a reduced dimension representation of the input image (i.e. use only the low-frequency sub-bands of the image). In [15], Singular Value Decomposition (SVD) is applied to each low-frequency block. The experimental results demonstrate that this approach localize the duplicated regions accurately even when the image was highly compressed.

2) *Dimensionality Reduction-based Methods:* Reducing the dimension of the tested image is the main goal of these methods. For example:

a) *Principal Component Analysis (PCA)*: PCA projects data onto axes of maximal data variance. In so doing, the dimensionality of data is reduced while minimizing the loss of information or distortion. This representation is robust to minor variations in the image due to additive noise or JPEG compression [10].

b) *Singular Value Decomposition (SVD)*: The SVD of an image  $A$  is the factorization of  $A$  into the product of three uncorrelated matrices  $A = UDV^T$ , where the columns of  $U$  and  $V$  are orthonormal and the matrix  $D$  is diagonal with positive real entries [16]. Researchers in [4] use SVD after transform the image into DCT. The experimental results demonstrate that this approach is robust to blurring, JPEG compression and their mixed operations.

3) *Moment-based Methods*: For the remainder of this paper we will concentrate on this type of methods.

Finally, the accuracy of each method can be measured by using recall rate ( $r$ ) and precision rate ( $p$ ). Recall is the ratio of the number of correctly detected region to the total number of duplicated regions, while precision is the ratio of the number of correctly detected region to the total number of detected regions. The equation of  $r$  and  $p$  is given in (1)

$$p = \frac{TP}{TP+FP}, r = \frac{TP}{TP+FN} \quad (1)$$

### III. MOMENTS BACKGROUND

Moments are widely used in many image processing applications such as pattern recognition and CMFD. Moments are a way to reduce a function (or image) down to be a set of scalar quantities. These quantities are used to characterize a function (image) in order to capture its significant features [1]. It is a global description of an image rather than local (i.e. global properties of the image are used rather than local properties) [17].

General moments  $M_{pq}$  of an image  $f(x, y)$  is defined as a projections of an image  $f$  onto a polynomial basis  $P_{pq}$  as shown in (2):

$$M_{pq} = \iint_D p_{pq}(x, y) f(x, y) dx dy \quad (2)$$

Where  $p, q$  are non-negative integer,  $r = p + q$  is called the order of the moment and  $D \subset R \times R$

Moment invariants are special functions of image moments. Invariant  $I$  is a functions that does not change its value under degradation operator  $D$  [1], i.e. satisfy the following condition

$$I(f) = I(D(f))$$

One of the main motivations when working with moments is the easy construction of rotation invariants. An efficient way to do this is to work in polar coordinates, where rotation is transformed into a shift.

Depending on the polynomial basis  $P_{pq}$  used, moments can be categorized into three types: Geometric moments, Complex

moments and Orthogonal moments. More information about these categories is given below.

#### A. Geometric Moments

In geometric moments, the most common choice is the standard power basis  $p_{kj}(x, y) = x^k y^j$ . Geometric moments are defined as in (3):

$$M_{pq} = \iint_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (3)$$

Geometric moments suffer from a high degree of information redundancy, and they are sensitive to noise for high-order moments [18]. Hu moment invariants is one example of algorithms that is defined over geometric moments

#### B. Complex Moments

Complex moments are introduced because they behave favorably under image rotation. This property can be advantageously employed when constructing invariants with respect to rotation. Here, the polynomial basis used is  $(x + iy)^p (x - iy)^q$ , where  $i$  is the imaginary unit. The complex moment is defined as in (4).

$$M_{pq} = \iint_{-\infty}^{\infty} (x + iy)^p (x - iy)^q f(x, y) dx dy \quad (4)$$

Geometric moments and complex moments carry the same amount of information. Each complex moment can be expressed in terms of geometric moments [1].

Complex moment invariants are not good features in general. They suffer from information loss, suppression, and redundancy which limit their discrimination power [17].

#### C. Orthogonal Moments

Polynomial basis is orthogonal if it is satisfying the condition of orthogonality that is given in (5): for all indexes  $p \neq m$  and  $q \neq n$  then

$$\iint_{\Omega} p_{pq}(x, y) p_{mn}(x, y) dx dy = 0 \quad (5)$$

where  $\Omega$  is the area of orthogonality.

Orthogonality here means that there is no redundancy or overlapping of information between the moments. In general, orthogonal moments are better than other types of moments in terms of information redundancy and image representation [17].

Orthogonal moments are classified into: moments orthogonal on a rectangle (e.g. Legendre moments and Chebyshev moments) and moments orthogonal on a disk (e.g. Zernike moments and Pseudo-Zernike moments). The main advantage of the moments orthogonal on a rectangle is that they preserve the orthogonality even on the sampled image. Moreover, they can be made scale-invariant but creating rotation invariants from them is very complicated. On the other hand, moments orthogonal on a disk can be easily used to construct rotation invariants. Whereas, an image must be mapped into a disk of orthogonality which creates certain resampling problems [1].

#### IV. MOMENT-BASED APPROACHES IN CMFD

Many moment's functions have been used in the field of CMFD. Some of these moment-based approaches are given next.

##### A. Blur Moment

In 2007, moments were used for the first time in CMFD, when [12] try to use blur moment in feature extraction step. What incited them to use moment is the stability of moment invariants under additive random noise. In this method:

1. The tested image is divided into overlapping  $N \times N$  blocks.
2. Twenty four blur invariants are calculated for each block and used to create the feature vector:  $B = \langle B_1, B_2, B_3, \dots, B_{23}, B_{24} \rangle$ . Equation (6) clarifies the recursive relation that was used to calculate the blur moment.

$$B(p, q) = \mu_{pq} - \alpha \mu_{qp} - \frac{1}{\mu_{00}} \sum_{n=0}^k \sum_{i=m_1}^{m_2} \binom{p}{t-2i} \binom{q}{2i} \times B(p-t+2i, q-2i) \mu_{t=2i, 2i} \quad (6)$$

Where,

$$k = \frac{p+q-4}{2}$$

$$t = 2(k - n + 1)$$

$$m_1 = \max \left( 0, \frac{t-p+1}{2} \right), m_2 = \min \left( \frac{1}{2}, \frac{q}{2} \right)$$

$$\alpha = 1 \Leftrightarrow p \wedge q \text{ are even}, \alpha = 0 \Leftrightarrow p \vee q \text{ are odd}$$

3. Principal Component Transformation (PCT) was used to reduce the dimension of each feature vector.
4. KD-tree was used to sort the feature vectors. The selection of KD-tree depend on the fact that it requires  $O(N \log_2 N)$ .
5. The similarity between blocks is calculated using the following measures:

$$s(B_i, B_j) = \frac{1}{1 + p(B_i, B_j)}$$

where  $p$  is the distance in the Euclidean space:

$$p(B_i, B_j) = \left( \sum_{k=1}^{\dim} (B_i[k] - B_j[k])^2 \right)^{0.5}$$

For each two similar blocks -i.e.  $s$  is greater than a specific threshold (minimum required similarity), the physical distance from their upper left corners is calculated. Those blocks with distance ( $p$ ) greater than a specific threshold (minimum required physical distance) are marked as a duplicated region.

The results show the great ability of the proposed approach to detect copy-move forgery in spite of the presence of blur, noise, JPEG compression or contrast changes in the copied areas.

The disadvantage of the proposed method is the computational time. The average run time of the implemented experimental on  $640 \times 480$  RGB images with parameters  $N = 20$  (block size) and  $s = 0.97$  (similarity threshold) on a 2.1 GHz processor and 512 MB RAM is 40 minute. Note that, the runtime is not the same for all images with the same size; it depend on image characteristics.

##### B. Hu Moment

Liu, et al propose a solution to CMFD by using Hu Moment [19]. The main steps in this method were as follows:

1. Images decomposed by Gaussian pyramid. This technique involves creating a series of images, which are weighted down using a Gaussian blur and scaled down. When this technique is used multiple times, it creates a stack of successively smaller images.
2. The produced low frequency sub-image is divided into many overlapping circular blocks.
3. The Hu Moments are extracted from the circular blocks and used as a matching feature. The Hu moments are computed only in the circle region; discarding the pixels outside the inscribed circle will have a little effect on the false alarm of the detection algorithm. Here, the circular block mode and the Hu moments are able to eliminate the effect of rotation.
4. The first four moments  $F = \langle f_1, f_2, f_3, f_4 \rangle$  are chosen as a feature vector for each block.
5. These  $F$  feature's vectors are then sorted according to  $f_1$  component and stored in a new array.
6. Similarity and physical distance between two blocks are used in the matching step.
7. Morphologic operation is used to remove the small and isolated regions according to a given area threshold.

In this method, the efficiency has been improved by using Gaussian pyramid decomposition (i.e. the dimension of the search space is reduced to 25% of its original amount). Moreover, using only the first four Hu moments as a feature vector for each block decrease the feature's vector dimension.

The experiments show that the proposed method has a nice robustness to rotation, blurring, noise additive and JPEG compression. Also, this method can detect multiple duplicated regions. On the other hand, it doesn't withstand scaling and cropping. The size of images for testing was  $400 \times 400$ , and the radius of circle was chosen as 7.

Hu, 1962 [20] constructed seven invariant moments, which can hold invariant against scaling, translation and rotation. The seven moments are given in Fig. 4.

$$\begin{aligned} \phi_1 &= m_{20} + m_{02} \\ \phi_2 &= (m_{20} - m_{02})^2 + 4m_{11}^2 \\ \phi_3 &= (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2 \\ \phi_4 &= (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2 \\ \phi_5 &= (m_{30} - 3m_{12})(m_{30} + m_{12}) ((m_{30} + m_{12})^2 - \\ &\quad 3(m_{21} + m_{03})^2) + (3m_{21} - m_{03})(m_{21} + m_{03})(3(m_{30} + \\ &\quad m_{12})^2 - (m_{21} + m_{03})^2) \end{aligned}$$

$$\begin{aligned}\phi_6 &= (m_{20} - m_{02})((m_{30} + m_{12})^2 - (m_{21} + m_{03})^2) + \\ 4m_{11}(m_{30} + m_{12})(m_{21} + m_{03}) \\ \phi_7 &= (3m_{21} - m_{03})(m_{30} + m_{12})((m_{30} + m_{12})^2 \\ &\quad - 3(m_{21} + m_{03})^2) \\ &\quad - (m_{03} - 3m_{12})(m_{21} \\ &\quad + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2)\end{aligned}$$

Figure 4. Hu Moment Invariants

### C. Krawtchouk Moment

Mustafa B., et al. (2013) propose a method for CMFD using Krawtchouk moment [21]. This method consists of the following steps:

1. The image is partitioned into overlapping blocks.
2. The Krawtchouk moments are then extracted from each block. The corresponding feature vectors are stored in a matrix.
3. In order to relocate feature vectors closer to each other, this matrix is lexicographically sorted.
4. Similarity of blocks (matrix rows) is then tested by:
  - a) The mean square error (*MSE*) between the current row and *r* adjacent rows is calculated. If two vectors are similar (i.e. *MSE* is less than a specific threshold) then go to the next test.
  - b) The physical distance between those blocks is calculated. If the distance between two blocks is greater than a pre-determined threshold value, then increment the corresponding counter in the shift vector. (Shift vector between two blocks is designated by a pair of integers corresponding to distances between upper left *x* and *y*-axes coordinates of these blocks).
  - c) Blocks are marked as forged if the numbers of suspicious block pairs that have the same shift vectors exceed a certain threshold.

The  $(n+m)^{\text{th}}$  Krawtchouk moment  $Q_{nm}$  for an image  $f(x, y)$  of size  $N \times M$  can be calculated using (7).

$$Q_{nm} = \sum_{x=1}^{N-1} \sum_{y=0}^{M-1} \overline{K_n}(x; p_1, N-1) \times \overline{K_m}(y; p_2, M-1) f(x, y) \quad (7)$$

$K_n$  is  $n^{\text{th}}$  order classical Krawtchouk polynomial.  $K_n$  is given in (8).

$$k_n(x; p; N) = \sum_{k=0}^N a_{k,n,p} x^k = 2^{F1(-n, -x; -N; \frac{1}{p})} \quad (8)$$

Where  $x, n = 0, 1, 2, \dots, N$  and  $N > 0$ ,  $p \in (0, 1)$ ,  $2^{F1}$ , is the hyper-geometric function given in (9)

$$2^{F1(a,b;c;d)} = \sum_{k=0}^n \frac{(a)_k (b)_k z^k}{(c)_k k!} \quad (9)$$

The definition of the Pochammer symbol denoted by  $(a)_k$  is given in (10).

$$(a)_k = a(a+1) \dots (a+k-1) \quad (10)$$

Experiments on  $512 \times 512$  8-bit gray level forged show that the proposed method can detect the duplicated region even if the replaced region has an irregular shape. Moreover, the experiments show the robustness of the method for both Gaussian blurring and additive Gaussian noise. For experiment parameters: maximum order of  $m$  is assumed to be 5 for all experiments, the value of similarity threshold and physical distance are determined to be 4 and 16 respectively.

### D. Exponenti-Fourier and Histogram Moment

Z. Le, et al. has proposed a method for solving CMFD. Here forgery is detected using mix of two moments: Exponenti-Fourier moment and Histogram moment [22]. Exponential moments are linear invariance for translation, scale and rotation. The histogram moments also are translation, rotation and scaling invariance. The main steps are as follows:

1. Gaussian pyramid transform was used to extract the low-frequency information from the image.
2. The low frequency part of the images is divided into overlapping blocks.
3. Mixed moments are then calculated for each block. The corresponding feature vectors are stored in a matrix. Features vector can be defined as  $X = (E_{1,1}, E_{2,2}, E_{3,3}, f_1, f_2, f_3, f_4)$ , where  $E_{nm}$  are Exponenti-Fourier moments and  $f_i$  are Histogram moments.  $X$  can accurately express the information of image block.
4. The eigenvector of block composed by the Exponenti-Fourier moments and Histogram moments is lexicographically sorted.
5. Euclidean distance and physical distance were used to determine the similarity of blocks.

The experiment results show that this method can detect the forged part that affected by translation, rotation, scaling and mixed operation tamper when the image is changed by brightness variation and contrast adjustment. The method in this paper improves the representation accuracy, but it had some limitations on the smaller tamper region.

### E. Zernike Moment (ZM)

[23], [24] propose using of ZM for CMFD. ZM have been shown to be superior to the others in terms of their insensitivity to image noise, information content, and the ability to provide faithful image representation. Moreover, the magnitude of ZM is algebraically invariant against rotation. Detecting steps were as follow:

1. The image is partitioned into overlapping blocks, assuming that the pre-defined block size is smaller than the tampered region.
2. Zernike moments of particular degree  $n$  are calculated for each block and vectorized. These vectors are stored in an array  $Z$ .

3.  $Z$  is then lexicographically sorted.
4. The Euclidean distance between adjacent pairs of  $Z$  and the physical distance were used as similarity measurements.

Zernike moments of order  $n$  with repetition  $m$  for a continuous image function  $f(x, y)$  is calculated using (11):

$$A_{nm} = \frac{n+1}{\pi} \sum \sum_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(\rho, \theta) \quad (11)$$

where

- $V_{nm}^* = R_{nm}(\rho)e^{-im\theta}$ ,  $\rho \leq 1$
- $R_{nm}(\rho) = \sum_{\alpha=0}^{\frac{n-|m|}{2}} -1^\alpha \frac{(n-\alpha)!}{\alpha! (\frac{n+|m|}{2}-\alpha)! (\frac{n-|m|}{2}-\alpha)!} \rho^{n-2\alpha}$
- $n = \theta, 1, 2, \dots, m = -n, \dots, n$
- $n \in \mathbb{Z}^+, |m| \leq n$
- $n - |m|$  is even
- $V_{nm}(\rho)$ : The Zernike basis polynomial
- $V_{nm}^*(\rho)$ : The conjugate of Zernike basis polynomial
- $R_{nm}(\rho)$ : The radial part of the Zernike basis polynomial
- $e^{-im\theta}$ : The angular part of the Zernike basis polynomial
- $(\rho, \theta)$ : The corresponding polar coordinate for  $(x, y)$

The proposed method can detect a forged region even though it is rotated. This scheme is also resilient to the intentional distortions such as additive white Gaussian noise, JPEG compression, and blurring. Although it concerned several attacks, this method is still weak against scaling or the other tampering based on affine transform.

The most distinctive thing in method presented in [23] is the use of local sensitive hashing (LSH) for block matching procedure in order to remove falsely matched block pairs by inspecting phase differences of corresponding Zernike moments. Authors found that the proposed method outperforms prior art in particular when duplicated regions are smooth. Experiments indicate high robustness against JPEG compression, blurring, additive white Gaussian noise, and moderate scaling. It is clear that detectors based on Zernike moments are unable to localize duplicated areas that underwent strong affine transformations other than rotation.

Table (I) provide robustness-based comparisons between all moment-based methods that mentioned in this section. All these methods represent important results in CMFD although each method has its own strengths and drawbacks.

TABLE I. COMPARISON BETWEEN MOMENT-BASED METHODS

	Robustness Against				
	Rotation	Scaling	JPEG	Noise, Blurring	Brightness, contrast change
Blur [12]	? <sup>a</sup>	?	Yes	Yes	Yes

<b>Hu</b> [19]	Yes	No	Yes	Yes	?
<b>Krawtchouk</b> [21]	?	?	?	Yes	?
<b>Exponenti-Fourier and Histogram</b> [22]	Yes	Yes	?	?	Yes
<b>Zernike</b> [23]	Yes	moderate scaling	Yes	Yes	?
<b>Zernike</b> [24]	Yes	No	Yes	Yes	?

a. No information available

## V. MOMENT BASED METHODS COMBINED WITH OTHER TECHNIQUES

Since moment-based approaches in CMFD have some limitations, some researchers suggest combining moments with another technique to detect duplicated region in order to increase the precision and the robustness of the detection method. Following are some examples:

### A. Zernike moment and SIFT Features

Zernike moments are invariant against rotations and degradations such as additive white Gaussian noise, JPEG compression and blurring and it is efficient for flat regions but failed for scale changing. On the other hand, key-point based method (methods that compute the features only on image regions with high entropy) such as SIFT-based method failed to detect low entropy regions (i.e. flat regions) but its invariant against rotation and scale changing. For example, SIFT and Zernike moments are used to detect duplicated regions. SIFT features is applied to detect all copied region that were geometrically changed or rotated. and Zernike moments is used to detect flat copied region[25], [26].

### B. Blur and Affine Moment Invariants

A combination between blur moment and affine moment invariants is proposed in [27] in order to enhance the robustness against affine transformation (i.e. combination of rotation, translation and scaling).

Experiments on gray-scale images,  $512 \times 512$  resolution, and block size of  $8 \times 8$  show that the proposed method gets effective detection. This method is an effective way to detect the duplication regions under some simple affine transforms and blur degradations blindly.

### C. Undecimated Wavelets and Zernike Moments

In [28] authors proposes a CMFD method using undecimated wavelets transform (UWT) and Zernike moments (ZM). UWT is translation invariant, while ZM is rotation invariant. In this method, UWT is applied on the image to find its approximation (LL). Then ZMs are extracted from each block. Similar blocks (using Euclidean distance) are then labeled as duplicated blocks. Experimental results show that the proposed image forgery detection method performs better

for rotated and scaled blocks. And it is robust to noise and smoothening

## VI. CONCLUSION

CMFD is still a hot topic; there still much work to be accomplished. The most important thing in these methods is the robustness in front of all types of attacks that may be used to deceive the detection method. Most existing methods succeeded in some cases and failed in another cases (e.g. succeeded in front of blurring but failed in front of rotation). Even, those methods that use invariants features are practically weaker than what is declared by the theoretical definition. Succeed in any future work must overcome the current challenges that face CMFD and must overcome the obstacles to the development of more reliable and robust detection algorithms. Moreover, suitable computational complexity for these methods is a very important aspects..

## ACKNOWLEDGMENT

This research is funded by the deanship of scientific research at Zarqa University/Jordan.

## REFERENCES

- [1] J. Flusser, T. Suk, and B. Zitová, *Moments and Moment Invariants in Pattern Recognition*. 2009.
- [2] J. A. Redi, W. Taktak, and J. L. Dugelay, "Digital image forensics: A booklet for beginners," *Multimed. Tools Appl.*, vol. 51, no. 1, pp. 133–162, 2011.
- [3] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 3, pp. 507–518, 2015.
- [4] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 6, pp. 1841–1854, 2012.
- [5] J. Fridrich, D. Soukal, and J. Lukáš, "Detection of Copy-Move Forgery in Digital Images," *Int. J.*, vol. 3, no. 2, pp. 652–663, 2003.
- [6] N. Muhammad, M. Hussain, G. Muhamad, and G. Bebis, "A non-intrusive method for copy-move forgery detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6939 LNCS, no. PART 2, pp. 516–525, 2011.
- [7] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-based forensic method for copy-move attack detection and transformation recovery," *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3 PART 2, pp. 1099–1110, 2011.
- [8] B. L. Shivakumar and S. S. Baboo, "Detection of Region Duplication Forgery in Digital Images Using SURF.," *Int. J. Comput. Sci. Issues*, vol. 8, no. 4, pp. 199–205, 2011.
- [9] J. Hu, H. Zhang, Q. Gao, and H. Huang, "An improved lexicographical sort algorithm of copy-move forgery detection," in *Proceedings - 2nd International Conference on Networking and Distributed Computing, ICNDC 2011*, 2011, pp. 23–27.
- [10] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image regions," *Dept. Comput. Sci., Dartmouth Coll. Tech. Rep. TR2004-515*, no. 2000, pp. 1–11, 2004.
- [11] S. Bayram, H. T. Sencar, and N. Memon, "AN EFFICIENT AND ROBUST METHOD FOR DETECTING COPY-MOVE FORGERY," *Image (Rochester, N.Y.)*, pp. 1053–1056, 2009.
- [12] B. Mahdian and S. Saic, "Detection of copy-move forgery using a method based on blur moment invariants," *Forensic Sci. Int.*, vol. 171, no. 2–3, pp. 180–189, 2007.
- [13] Y. Cao, T. Gao, L. Fan, and Q. Yang, "A robust detection algorithm for copy-move forgery in digital images," *Forensic Sci. Int.*, vol. 214, no. 1–3, pp. 33–43, 2012.
- [14] A. Gupta, N. Saxena, and S. K. Vasistha, "Detecting Copy move Forgery using DCT," *Int. J. Sci. Res. Publ.*, vol. 3, no. 5, pp. 3–6, 2013.
- [15] G. L. G. Li, Q. W. Q. Wu, D. T. D. Tu, and S. S. S. Sun, "A Sorted Neighborhood Approach for Detecting Duplicated Regions in Image Forgeries Based on DWT and SVD," *Multimed. Expo, 2007 IEEE Int. Conf.*, pp. 1750–1753, 2007.
- [16] Z. Ting and W. Rang-Ding, "Copy-move forgery detection based on SVD in digital image," in *Proceedings of the 2009 2nd International Congress on Image and Signal Processing, CISIP'09*, 2009, no. 2, pp. 0–4.
- [17] S. X. Liao, "Image Analysis by Moments," The University of Manitoba, Canada, 1993.
- [18] C. W. Chong, P. Raveendran, and R. Mukundan, "Translation invariants of Zernike moments," *Pattern Recognit.*, vol. 36, no. 8, pp. 1765–1773, 2003.
- [19] G. Liu, J. Wang, S. Lian, and Z. Wang, "A passive image authentication scheme for detecting region-duplication forgery with rotation," *J. Netw. Comput. Appl.*, vol. 34, no. 5, pp. 1557–1565, 2011.
- [20] M.-K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Inf. Theory*, vol. 8, pp. 179–187, 1962.
- [21] M. Bilgehan and M. Uluta, "Detection of Copy-Move Forgery Using Krawtchouk Moment," in *8th International Conference on Electrical and Electronics Engineering (ELECO)*, 2013, pp. 311–314.
- [22] Z. Le and W. Xu, "A robust image copy-move forgery detection based on mixed moments," in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 2013, no. 208098, pp. 381–384.
- [23] S. J. Ryu, M. Kirchner, M. J. Lee, and H. K. Lee, "Rotation invariant localization of duplicated image regions based on zernike moments," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 8, pp. 1355–1370, 2013.
- [24] S. Ryu, M.-J. Lee, and H. Lee, "Detection of Copy-Rotate-Move Forgery Using Zernike Moments," in *Information Hiding*, vol. 6387, 2010, pp. 51–65.
- [25] Z. Mohamadian, "Image Duplication Forgery Detection using Two Robust Features," *Res. J. Recent Sci.*, vol. 1, no. 12, pp. 1–6, 2012.
- [26] W. Yan, R. Xu, P. Luo, X. Yang, and H. Qin, "A SIFT and Zernike Moment Based Copy-move Forgery Detection Algorithm Description," *J. Comput. Inf. Syst.*, vol. 15, no. 2013, pp. 5661–5671, 2015.
- [27] T. Wang, J. Tang, and B. Luo, "Blind detection of region duplication forgery by merging blur and affine moment invariants," in *7th International Conference on Image and Graphics, ICIG 2013*, 2013, pp. 258–264.

- [28] G. Muhammad and M. Hussain, "Passive detection of copy-move image forgery using undecimated wavelets and zernike moments," *Inf.*, vol. 16, no. 5, pp. 2957–2964, 2013.

#### AUTHORS PROFILE

Khaled Walid Mahmoud received a BSc in Computer Science from Jordan University in 1992, MSc in Computer Science (Artificial Intelligence) from Jordan University in 1998 and a PhD in Print Security and Digital Watermarking from Loughborough University (uk) in 2004. This was followed by academic appointments at ZARQA University (Assistance Professor in computer Science). His areas of interest include Information Security, Digital watermarking, Image processing, AI and Arabic Language processing.

Arwa Husien Abu Al-Rukab received the BA degree in computer science from the Jordan University of science and technology, Jordan in 2008. She received MSc degree of computer science from Zarqa University , Jordan, in 2016. currently, she worked at Colleges of Computing and Information Society. Her research interests are in Image Processing and Machine learning

# A New Approach to Predict Stock Big Data by combination of Neural Networks and Harmony Search Algorithm

Kiarash Aghakhani

Young Researchers and Elite Club  
Arak Branch, Islamic Azad University  
Arak, Iran

Abbas Karimi\*

Department of Computer Engineering  
Faculty of Engineering, Arak Branch, Islamic Azad  
University, Arak, Iran

**Abstract**— Nowadays, due to the vast volume and complicated interrelation of daily stock data, the prediction of the stock price is very crucial in order to earn the highest profit of the shareholder's investment is the main target. For these purposes, data mining techniques such as correlation analysis and prediction, and likewise data modeling and pattern recognition are utilized.

Since the stock market is a chaotic and nonlinear system, the exact prediction of the massive data exchange, requires intelligent and advanced tools such as neural networks and meta-heuristic algorithms. This purpose method is conducted on the stock data of IBM, Apple and Dell companies and gold price in the global market. Moreover, the prediction error is compared with results of ARIMA<sup>1</sup>, ANN<sup>2</sup>, HMM<sup>3</sup> ANN-ICA<sup>4</sup>, ANN-GA<sup>5</sup>, ANN-PSO<sup>6</sup>, HMM-Fuzzy<sup>7</sup>, HMM-ANN-GA<sup>8</sup> methods. The comparison indicates that the purposed method provides remarkable improvement in the prediction performance.

**Keywords-** *Data mining, Big Data, Predict Stock Price, Artificial Neural Network, Harmony Search Algorithm*

## I. INTRODUCTION

Human population growth and increasing demand for goods and services, at the same time spirit of consumerism, also the limitation of the traditional means in providing needed financial resources to develop and increase production capacity, further provides a context to use the new mechanisms in order to attract the financial resources and lead it into production [26,35].

Continues development of the economy of society has caused a rapid increase in capital markets in different countries. As a result, all investors need the robust and reliable tools to predict stock prices [34]. There are various methods to predict Stock market which divides into two groups: traditional methods and modern methods. Traditional methods consist fundamental and technical analysis. Fundamental analysis includes stating some factors such as global economy, political situation, the annual

budget of the country, supply and demand in related markets, market share and so on. The technical analysis predicts future changes in stock prices based on the previous events. Technical analysis believes that the history will repeat itself and future changes in stock price can be determined considering previous stock prices [22, 23, 37].

Just like all of the forecasting methods, these methods can't have an exact Prediction of change because stock big data is very wide spreading. Inability to predict stock price due to different reasons causes offering the "Efficient market hypothesis". According to this hypothesis, pricing securities in the market is influenced by the sellers and buyers reaction toward the latest information and the company future [15].

In general, these methods are based on the statistical data while the stock market is a nonlinear and convulsive one with a wide range of big data that depends on political, economic and emotional factors. Therefore, implementing traditional analyzing tools to make exact decisions about the stock will be so hard and inefficient. In recent years, following the developments in computer technology, artificial intelligence and recognizing chaotic relation in nonlinear time series, different countries try to do some activities to forecast the securities stock price [21]. Artificial intelligence techniques that consist neural networks, Fuzzy logic and Meta-heuristic algorithm including Genetic algorithm, Harmony search algorithm, Firefly algorithm, Hill-climbing and optimization mass particle algorithm and etc., leads to successful results in this field [2, 25]. Some of the meta-Heuristic algorithms are practically faster than others, as a result, users are keener on using this kind of algorithm rather than others. Artificial neural network as an intelligent system can recognize the nonlinear relation between input and output base on a set of data and understanding the basic relations among them. Therefore, in this study, we used artificial intelligence techniques besides harmony search Algorithm and neural network as a modern and smart technique to predict stock market big Data for the first time [1].

<sup>1</sup> Auto Regressive Integrated Moving Average

<sup>2</sup> Artificial Neural Network

<sup>3</sup> Hidden Markov Model

<sup>4</sup> Combination of ANN and Imperialist Competitive Algorithm

<sup>5</sup> Combination of ANN and Genetic Algorithms

<sup>6</sup> Combination of ANN and Particle Swarm Optimization

<sup>7</sup> Combination of HMM and Fuzzy System

<sup>8</sup> Combination of HMM and ANN and GA

## II. LITERATURE

White [38] used a neural network for the first time to Predict stock Market. He focused on this question: "are the neural networks able to identify nonlinear rules in time series and unknown rules in changes of properties and stock prices?" White purpose of presenting this paper was to show one feed forward neural network can do this task .He proved this issue by showing an example of daily prices of IBM.

The neural network entered the financial domain after White's primary study in 1988. There have been many types of research about that; as there were totally 213 scientific activities on the neural network about this case, during 1988 and 1993 [36].

Chiang et al. [8] used one Propagation network to predict the price of a company's net asset at the end of the year. They compared their findings with the obtained result of traditional econometrics techniques. They realized that neural networks act better than Regression methods in a meaningful way.

Aiken [3] used one feed forward neural network trained by Genetic Algorithm (GA) method to predict the interest rates on the treasury of the USA. They conclude that neural network can be useful for this.

Garliuskas [11] has predicted the time series in the stock market using computational algorithm related to kernel function and forecasting method of returning an error. He believes that predicting financial time series by means of the neural network can be better than classical statistics methods and others.

Chan [7] predicted the financial time series using feed forward neural network based on daily data of stock exchanges in shanghai. He used descent Gradient algorithm in order to have higher speed and convergence, also to determine the weights he used multiple linear regression. He concludes the neural network is able to predict the time series more satisfactory. Moreover, choosing weights method led to less computational costs [7].

Lendasse [31] foresaw the index using neural networks. He found out neural networks are better than linear methods.

Egeli et al. [9] forecast the daily stock market index of Istanbul (ISE). The results have shown that neural networks can predict MA (Moving Average) more precise.

Hadavandi et al. [16] predicted stock using artificial neural networks by genetic fuzzy. Here genetic fuzzy was used to decrease the future complexity of price time series.

## III. THE PROPOSED METHOD

We implement the scheme using programming in a complete software content MATLAB version R2o12a and a manual coding, as well [1].

### A. Input

To recognize the variables we studied a lot and considering the limitation. We have chosen our input variables including

open price, high price, Low price and close price in daily stocks of three companies such as Apple Computer Inc., International Business Machines Corporation (IBM), Dell Inc., and the output variable closing stock price on the day after.

It should be noted that to compare the scheme model with other models, we use the train and test dataset as used in other articles [16-19]. Train dataset is from 10 February 2003 till 10 September 2004. Test dataset is from 13 September 2004 till 21 January 2005. AS a whole, test dataset includes 91 data and train dataset also include 400 data [1].

For the date of gold price in global market 80% of data was for training and 20% for test, completely randomly. These data have been considered daily during 2003 and 2014. [2]

### B. Input Preparation

The used data as input for the model should be normalized and calibrated. In other words, their vibrations and noise must be reduced or we should change the data in a way that be used as input.

For data preparation, considering the fact that data normalization in the range [-1, +1] distance, we use Eq. (1).

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \times 2 - 1 \quad (1)$$

After normalization by this Equation, the neural network will be designed.

### C. Neural Network Design

Multi-layer Perceptron (MLP) neural network are one of the most beneficial neural network used in the most researchers. A propagation algorithm for training this feed-forward multilayer network via stimulus differentiable functions can be used to predict, recognize and classify the pattern [24].

In this paper after necessary studies and comparing various neural networks, we decided to utilize multilayer perceptron neural network. For training, we used to feed forward and harmony search Meta-heuristic algorithm [1].

### D. Activation Function

Activation Function clarifies the relation between input and output in a node or a network. This function gives the network a grade of nonlinearity which is very important for most of the neural networks [13]. The best function here for the middle layer is Sigmoid Function.

### E. Training Neural Network

In the first Phase of the scheme, our model uses the propagation algorithm. First, it's assumed that the network weights are selected randomly. In every step, the output is calculated according to its difference with the ideal output. Moreover, the weights will be corrected. At the end, this error

changes to a minimum. The activation function for every nerve, in propagation algorithm, is considered as the weights sum of inputs related to that nerve. As a result, considering this assumption that W is corresponding weights between the input layer and next one, we can introduce Eq. (2):

$$A_j(\bar{x}, \bar{w}) = \sum_{i=0}^n x_i w_{ji} \quad (2)$$

Clearly we can see that activation function output of nerve is just dependent on the corresponding input weights. Therefore, we should change the weights therewith change the output. As we mentioned before, training goal is to achieve an ideal output. So, first, we should define the Error function for every neuron.

This error will be obtained by calculating the difference between the actual output and expected the output of the network: "Eq. (3)"

$$E_j(\bar{x}, \bar{w}, d_j) = [o_j(\bar{x}, \bar{w}) - d_j]^2 \quad (3)$$

Selecting the square of the difference between actual output  $o_j$  and desired output  $d_j$  is controversial from several aspects. First, by using square, the error value will always be Positive. Second, if the difference between actual and desired output is noticeable, then the square will cause this number become larger. Conversely, if this difference is Low, the square will cause the number become smaller. Therefore, the total error of network can be calculated by the total error of every single nerve in output layers. So, we will have "Eq. (4)"

$$E(\bar{x}, \bar{w}, \bar{d}) = \sum_j E_j(\bar{x}, \bar{w}, d_j) = \sum_j [O_j(\bar{x}, \bar{w}) - d_j]^2 \quad (4)$$

Now we should analyze the relation between error and inputs, weights and outputs. There are different methods for these, which the most important ones are a Gradient method, Meta-Heuristic algorithm, Newton's method, Gross-entropy method and etc. [6, 27, 32].

So, we used meta-Heuristic algorithm here in this paper. Harmony search (HS) Algorithm is one of the most simple and the newest Meta-Heuristic methods, which is used in the optimization process that is inspired by the simultaneous playing of the orchestra music. In other words, there is a similarity between finding the optimal Problem solution and the Process of playing the music [39].

Because of its easy operation, little parameters, simple concepts, little mathematical calculations and being practical for newton and continuous optimization problems, Harmony search Algorithm has changed to one of the most useful ones during recent years [30].

Subscribe to the neural network with the meta-Heuristic algorithm, is that phase where after neural network structure design, their training process ends with an optimization problem.

The second phase of the scheme includes 5 steps:

Step 1: instead of  $f(x)$  in Eq. (5), we use the error function in Eq. (4):

$$\text{Minimize } f(x) x_i \in X_i, i = 1, 2, \dots, N \quad (5)$$

In this step, these Parameters values are calculated: the Harmony Memory Size (HMS) or the number of solution vectors in the harmony memory, Harmony Memory Consideration Rate (HMCR), Pitch Adjustment Rate (PAR) and Number of Improvisation (NI).

Step 2: creation and shaping of Harmony Memory which is set based on the matrix in Eq. (6).

$$HM = \left[ \begin{array}{cccc|c} x_1^1 & x_2^1 & \dots & x_n^1 & f(x^1) \\ x_1^2 & x_2^2 & \dots & x_n^2 & f(x^1) \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ x_1^{HMS} & x_2^{HMS} & \dots & x_n^{HMS} & f(x^{HMS}) \end{array} \right] \quad (6)$$

In this step as you can see the matrix of Harmony memory is randomly formed by real performance function  $f(x)$  and generating solve vector, which acts as the memory from now on.

Step 3: this step is the most important step of a harmony search algorithm because all of the changes in present harmonies will happen here.

Considering Eq. (7), HMCR clarifies that in forming New Harmony how much inner harmony memory should be used and (1-HMCR) shows the probability of creating the new Random harmony.

$$\dot{x}_i \leftarrow \begin{cases} \dot{x}_i \in \{x_i^1, x_i^2, \dots, x_i^{HMS}\} & \text{w.p HMCR} \\ x_i \in X_i & \text{w.p (1 - HMCR)} \end{cases}, HMCR \in [0,1] \quad (7)$$

For example, one HMCR 95% shows that the harmony search algorithm, chased 95% of saved values in harmony memory and only 5% will be random.

The value of 1 for HMCR isn't recommended. Because the total improvement of the solution will work through saved values in harmony memory to offer the best solution.

When one value is chosen from inside the memory, it can change based on PAR probability. "Eq. (8)"

$$\dot{x}_i \leftarrow \begin{cases} \text{Yes} & \text{w.p PAR} \\ \text{No} & \text{w.p (1 - PAR)} \end{cases}, PAR \in [0,1] \quad (8)$$

The value of (1-PAR) sets the rate of doing nothing. If the pitch adjustment decision for  $\dot{x}_i$  is YES,  $\dot{x}_i$  is replaced as follow: "Eq. (9)"

$$\dot{x}_i \leftarrow \dot{x}_i \pm rand \times BW \quad (9)$$

Where, BW is an arbitrary distance bandwidth and is rand a random number between [0,1].

Step 4: In this step, if the New Harmony is better than the worst member in memory, the New Harmony will be replaced with the old one, afterward, the worst harmony will be deleted.

We sort harmony memory based on the best member at the top. As a result, we can update memory in this way.

Step 5: Termination of algorithm happens in this stage. If

the termination doesn't happen, stages 3 and 4 will be repeated again. However, we can adjust the end up the condition of the algorithm to a certain optimal value so repeat the steps of algorithm till the end. (See Fig. 1)

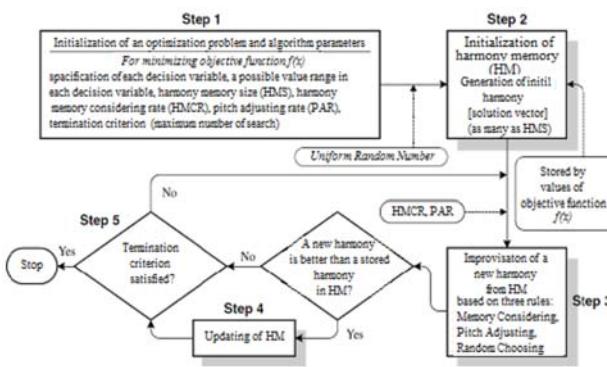


Figure 1. Flow chart of Harmony Search Algorithm

In the proposed phase certain optimal value of Harmony Search Algorithm to achieve the minimum error and the most exact answer are shown in table 1.

TABLE I. CERTAIN OPTIMAL VALUE OF HARMONY SEARCH ALGORITHM

Parameters	Value
Maximum Number of Iterations	1000
Harmony Memory Size	500
Number of New Harmonies	250
Harmony Memory Consideration Rate	0.5
Pitch Adjustment Rate	0.1
Fret Width (Bandwidth)	$0.02 \times (\text{var}_{\max} - \text{var}_{\min})$
Fret Width Damp Ratio	0.995

#### F. Network structure

After identifying the type and method of the network in neural network structure, we should clarify the number of input neuron, the number of hidden (middle) layers and hidden neurons and the number of output neurons.

Selecting the number of inputs is very important. Most of the researchers used trial and error method. To calculate the number of neurons in this paper the number of input neurons is chosen exactly as the amount of network input which means 4 neurons. Moreover, the number of output neurons is one, since dependent variable here is predicting the closing stock price in the next day.

Also, the number of hidden neuron layers plays a very important role in neural network success. The neurons in the hidden layer will help the neural network to discover the characteristics of data.

A neural network having one hidden layer can model every continuous and dependent function. Our desired model can

consist of every needed number of neurons (10, 100 ...) we introduced Eq. (10) and Eq. (11) related to this issue that is very helpful: [5,6]

$$\text{No of hidden node} = \sqrt{\text{input} \times \text{output}} \quad (10)$$

$$\text{No of hidden node} = \ln(\text{No of nodes in previous layer}) \quad (11)$$

#### IV. CRITERION FOR EVALUATING THE PERFORMANCE OF SCHEME

To predict the issues, we used some of the performance criterions to show the relation between data, which is usually related to the error of predicted output and desire real output. We used 5 criterions in this research which are shown in table 2.

TABLE II. PERFORMANCE EVALUATION CRITERIA IN OUR PAPER

Equation	Concept	Name
$\frac{\sum  e_t }{N}$	Mean Absolute Error	MAE
$MSE = \frac{\sum_{i=1}^n  y_i - \hat{y}_i ^2}{n}$	Mean Squared Error	MSE
$RMSE = \sqrt{\frac{\sum_{i=1}^n  y_i - \hat{y}_i ^2}{n}}$	Root Mean Square Error	RMSE
$MAPE = \left( \frac{1}{n} \sum_{i=1}^n \frac{ p_i - A_i }{A_i} \right) \times 100$	Mean Absolute Percentage Error	MAPE
$R^2 = 1 - \left[ \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (\hat{y}_i)^2} \right]$	Coefficient of Determination (R squared)	R2

#### V. IDENTIFYING THE NUMBER OF NEURONS IN HIDDEN LAYER

To identify the exact number of nodes in hidden Layer, we operated our scheme on the data of IBM Company in America, considering Eq. (10), Eq. (11) and error test in order to obtain the optimal answer. The results are shown in Table 3.

TABLE III. COMPARING THE RESULT OF SCHEME TO IDENTIFY THE NUMBER OF NEURONS IN HIDDEN LAYERS [1]

MAE	MSE	R2	RMSE	MAPE	Hidden Layer
0.0736	0.0091	0.9589	0.0956	0.6998	(4-17-1)
0.0721	0.0090	0.9599	0.0949	0.4009	(4-19-1)
0.0712	0.0082	0.9687	0.0906	0.3092	(4-21-1)
0.0771	0.0104	0.9423	0.1021	0.6555	(4-23-1)
0.0776	0.0099	0.9580	0.0994	0.6390	(4-25-1)

As you can see in Table 3 and Figs 2, 3 and 4 we can say that whenever the number of neurons in the middle layer is 21 there is the least error, using MAE, RMSE, and MAPE evaluation criterions.

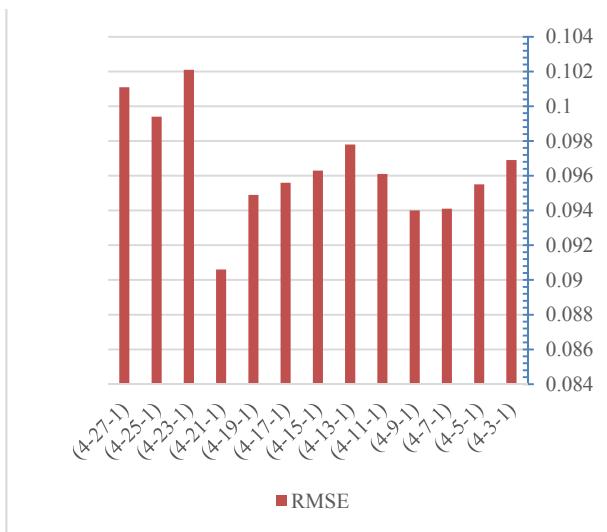


Figure 2. Comparing the number of neurons in the hidden layer based on RMSE evaluation criterion [1]

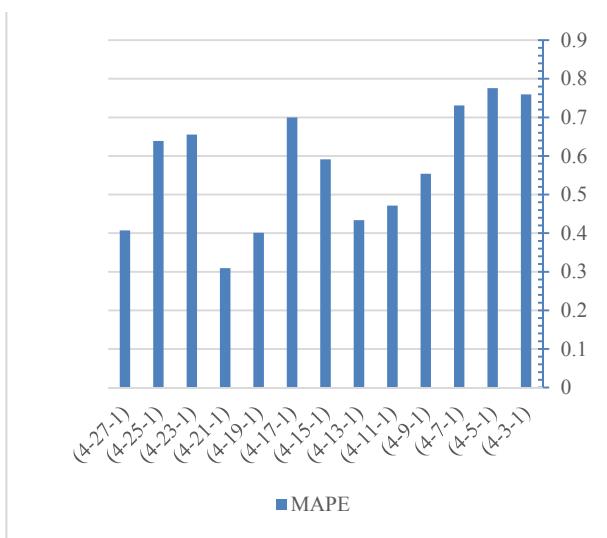


Figure 3. Comparing the number of neurons in the hidden layer based on MAPE evaluation criterion [1]

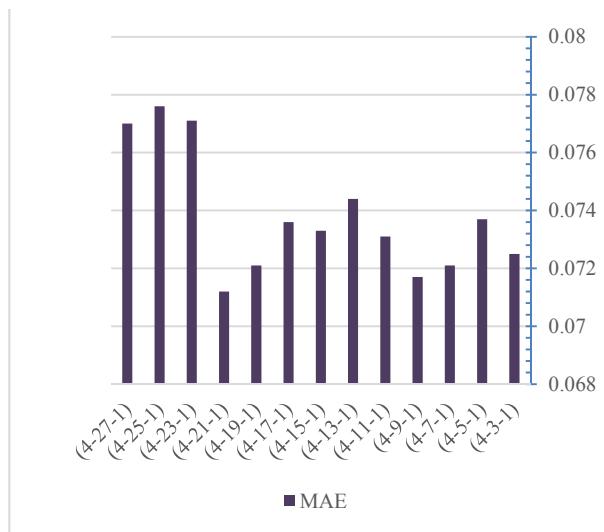


Figure 4. Comparing the number of neurons in the hidden layer based on MAE evaluation criterion [1]

It's necessary to mention that in Fig. 5 and also considering R2 evaluation criterion or the coefficient of determination in (4-21-1) status with 21 neurons in hidden layer there was the highest amount, we know that whenever a Coefficient determination is closer to 1, the model is more accurate and precise.



Figure 5. Comparing the number of neurons in the hidden layer based on R2 evaluation criterion [1]

As a result, that mentioned before, the number of ideal layers in this research is 3 (one input layer, one hidden layer, and one output layer) with (4-21-1) number of neurons.

## VI. APPLYING THE SCHEME TO STOCK DATA OF IBM, APPLE, AND DELL COMPANIES

After a precise identification of neurons, we applied our scheme to stock data of IBM company. Then we applied the same data to the Combination of ANN with ICA, GA, and PSO. The outcomes are shown in Tables 4 and Figs. 6.

TABLE IV. COMPARING THE SCHEME FOR THE COMBINATION OF ANN WITH ICA, GA AND PSO FOR IBM'S STOCK DATA [1]

<b>MAE</b>	<b>MSE</b>	<b>R2</b>	<b>RMSE</b>	<b>MAPE</b>	<b>Model</b>
0.1037	0.0175	0.9566	0.1324	0.6010	ANN-ICA
0.0979	0.0147	0.9636	0.1213	0.5936	ANN-GA
0.0966	0.0142	0.9650	0.1190	0.6214	ANN-PSO
0.0955	0.0101	0.9796	0.1006	0.3092	ANN-HS <sup>a</sup>

a. Proposed Method

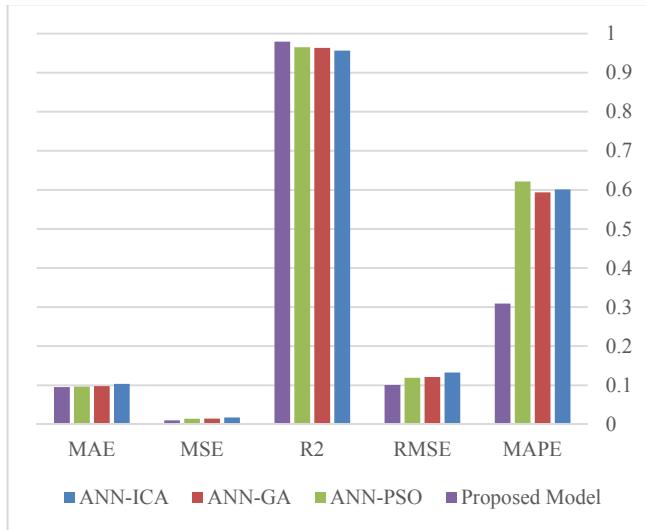


Figure 6. Comparing the scheme for the Combination of ANN with ICA, GA and PSO for IBM's stock data [1]

We applied our scheme to stock data of Dell Company. Then we applied the same data to the Combination of ANN with ICA, GA, and PSO. The outcomes are shown in Tables 5 and Figs. 7.

TABLE V. COMPARING THE SCHEME FOR THE COMBINATION OF ANN WITH ICA, GA, AND PSO FOR DELL'S STOCK DATA [1]

<b>MAE</b>	<b>MSE</b>	<b>R2</b>	<b>RMSE</b>	<b>MAPE</b>	<b>Model</b>
0.1042	0.0237	0.9504	0.1539	0.1928	ANN-ICA
0.1034	0.0226	0.9527	0.1503	0.1968	ANN-GA
0.0897	0.0185	0.9614	0.1358	0.1602	ANN-PSO
0.0833	0.0139	0.9731	0.1178	0.1447	ANN-HS <sup>a</sup>

a. Proposed Method

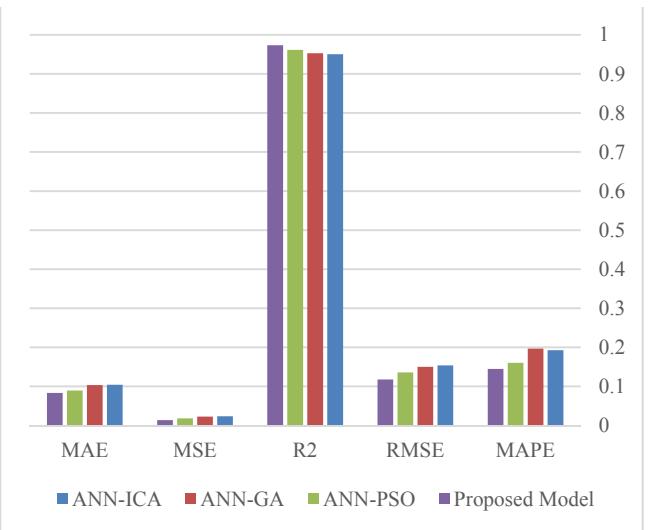


Figure 7. Comparing the scheme for the Combination of ANN with ICA, GA, and PSO for Dell's stock data [1]

We applied our scheme to stock data of Apple Company. Then we applied the same data to the Combination of ANN with ICA, GA, and PSO. The outcomes are shown in Tables 6 and Figs. 8.

TABLE VI. COMPARING THE SCHEME FOR THE COMBINATION OF ANN WITH ICA, GA, AND PSO FOR APPLE'S STOCK DATA [1]

<b>MAE</b>	<b>MSE</b>	<b>R2</b>	<b>RMSE</b>	<b>MAPE</b>	<b>Model</b>
0.0918	0.0146	0.9646	0.1210	0.3916	ANN-ICA
0.1346	0.0249	0.9399	0.1577	0.5810	ANN-GA
0.0678	0.0098	0.9762	0.0992	0.2311	ANN-PSO
0.0660	0.0069	0.9808	0.0830	0.2074	ANN-HS <sup>a</sup>

a. Proposed Method

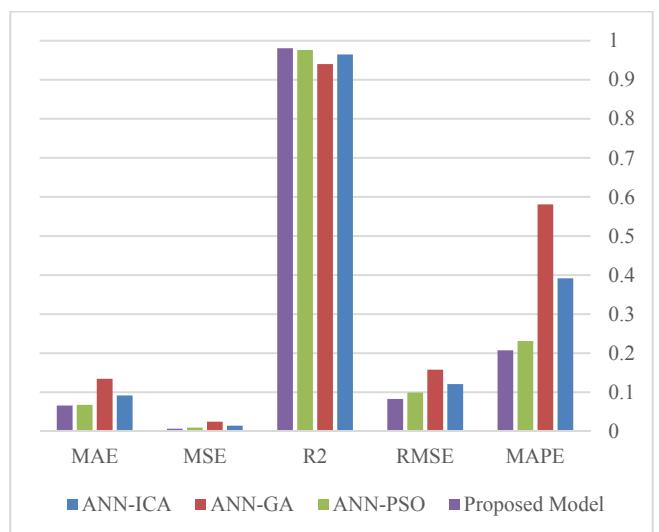


Figure 8. Comparing the scheme for the Combination of ANN with ICA, GA, and PSO for Apple's stock data [1]

## VII. APPLYING THE SCHEME TO THE PRICE OF GOLD IN GLOBAL MARKET

We applied our scheme to data of the price of gold in global market. Then we applied the same data to the neural network based on ICA, GA and PSO. The outcomes are shown in table 7 and Fig. 9.

TABLE VII. COMPARING THE SCHEME FOR THE COMBINATION OF ANN WITH ICA, GA AND PSO FOR GOLD PRICE IN GLOBAL [2]

<b>MAE</b>	<b>MSE</b>	<b>R2</b>	<b>RMSE</b>	<b>MAPE</b>	<b>Model</b>
0.0414	0.0026	0.9919	0.0557	0.7715	ANN-ICA
0.0708	0.0066	0.9798	0.0815	0.4750	ANN-GA
0.0612	0.0038	0.9811	0.0623	0.3301	ANN-PSO
0.0403	0.0025	0.9925	0.0500	0.1309	ANN-HS

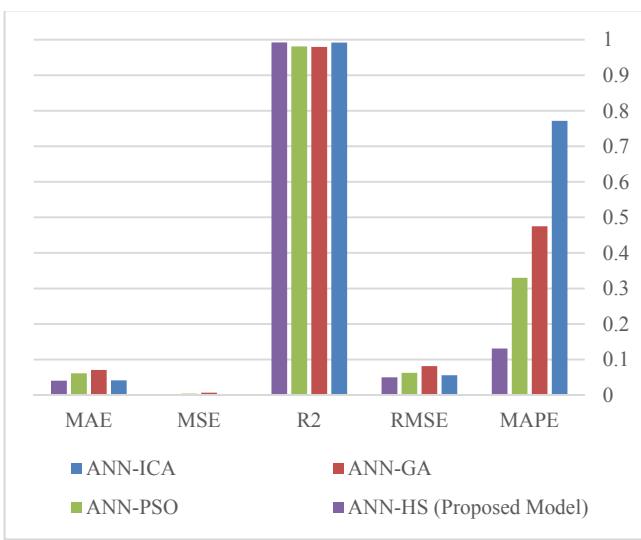


Figure 9. Comparing the scheme for the Combination of ANN with ICA, GA and PSO for gold price in global market [2]

## VIII. COMPARING THE SCHEME WITH OTHER ALGORITHMS

Considering the fact that predicting the stock price of IBM, Apple and Dell companies has been studied in several types of research till now, we will briefly review the two important papers on data of those mentioned companies to predict the issue. First, in Hassan's [17] Paper there is a new combination of Hidden Markov model and Fuzzy model to predict the stock price. In this paper, first, Markov Model is used to recognize data patterns. Second for predicting the stock Price for next day they used Fuzzy Logic to obtain fuzzy rules .Comparing this combined model with ANN and ARIMA shows the superiority of Markov model.

Second, Hadavandi et al. [16] predict the stock of those companies, introducing a new smart combined model. The prediction of the stock price in this paper is done by designing an expert fuzzy system. The authors extracted the database of the expert fuzzy system by means of Genetic Algorithm. The result was considerably improved in comparison with other papers [17-19]. Table 8 and Fig. 10 compared the scheme with other algorithm used in mentioned papers based on the performance criterion of MAPE which shows the superiority of scheme.

TABLE VIII. COMPARED THE SCHEME WITH OTHER ALGORITHM BASED ON MAPE [1]

<b>Dell</b>	<b>IBM</b>	<b>Apple</b>	<b>Model</b>
0.660	0.972	1.801	ARIMA
1.012	1.219	2.837	HMM-based
0.405	0.779	1.796	HMM-fuzzy
0.699	0.849	1.925	HMM-ANN-GA
0.660	0.972	1.801	ANN
0.1928	0.6010	0.3916	ANN-ICA
0.1968	0.5936	0.5810	ANN-GA
0.1602	0.6214	0.2311	ANN-PSO
0.1447	0.3092	0.2674	ANN-HS

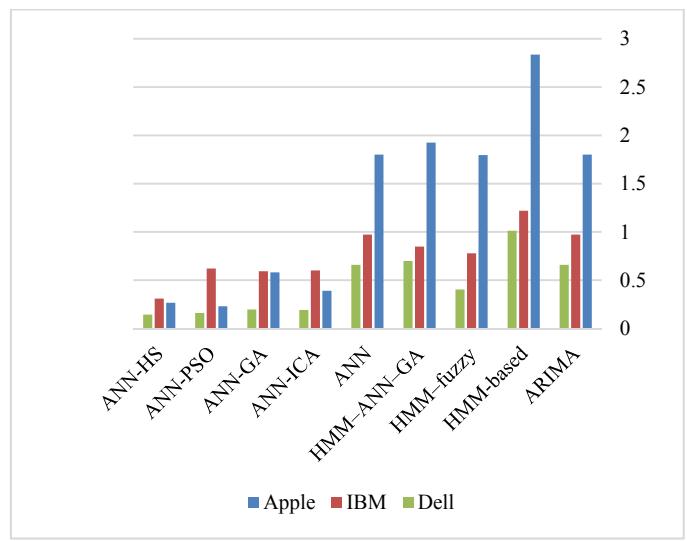


Figure 10. Compared the scheme with other algorithm based on MAPE [1]

Finally, Figs. 11 and 12 we can see that, adaptation of the predicted values to the actual values, confirms the superiority of scheme model.

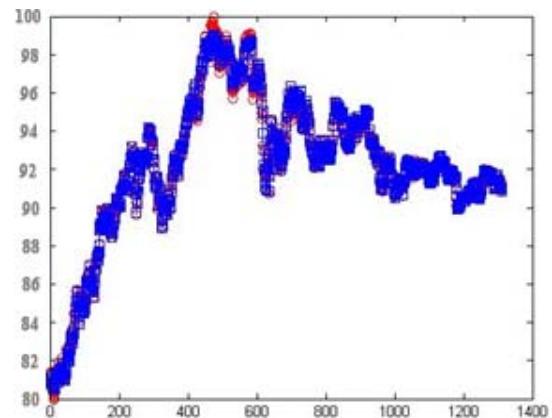


Figure 11. Comparing the prediction based on scheme and real values (far view)

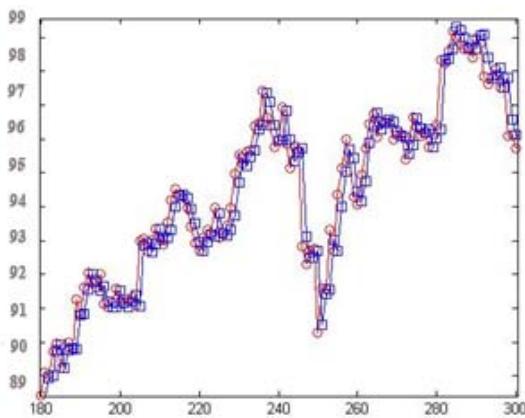


Figure 12. Comparing the prediction based on scheme and real values (close view)

## IX. CONCLUSION

- According to table 8 and Figs. 10 the amount of predicted error by scheme based on performance evaluating criteria MAPE, is more superior rather than other methods. As a result, we could obtain a precise prediction using Meta-Heuristic harmony search.
- According to tables 4, 5 and 6 and Figs. 6, 7 and 8 the amount of prediction error by means of our scheme based on 5 performance evaluation criteria is less than other methods. Therefore, the scheme of predicting the time series of stock Price for IBM, Apple and Dell companies is more accurate rather than other methods in neural network training such as ICA, GA, PSO.
- Considering table 7 and Figs. 9 there were very good result using our scheme. Therefore, predicting the price of gold in global market is considerably accurate using our scheme.
- Number, 1, 2 and 3 of results show that the scheme algorithm on 3 famous active companies in international stock and also on the price of gold in global market, was perfectly predicted. Moreover the Estimation error here was less than other methods. AS a result, our scheme can be used for other time series such as active companies in Iran or international stock.
- As the scheme model includes artificial neural network and harmony search algorithm, is considerably capable to recognize the data patterns. Moreover, it is more superior to other algorithms or methods based on those 5 performance evaluating criteria. Furthermore, results Show that our model has a unique fast convergence, high accuracy and an ability for Approximation function. Also, it is very suitable to predict the stock price index.
- All of the artificial intelligence methods have shown better results rather than traditional and classical methods. Moreover, our scheme method is superior to the classic methods.
- In the scheme method, normalization of data in internal  $[-1, +1]$  was very effective in improving data. Therefore, first the data were normalized, then after entering the

trained neural network with HS algorithm, they were outside the normal output and were returned to the primary domain.

- Considering the Figs. 11 and 12, the adaptation of predicted values to the actual values, shows the superiority of the scheme model.

## X. FUTURE RESEARCH FIELDS

- The approach of neural network based on HS algorithm is a strong method for predicting different issues. As a result, we recommend using this method for other issues including demand prediction, quality control, currency prediction, oil price prediction, medical issues prediction, etc.
- In this paper, we used MLP neural network. Moreover, other neural networks can be used. Such as Radial Basis Functions or RBF or even fuzzy neural network etc.
- In this research, we used the Meta-Heuristic Harmony Search (HS) algorithm in neural network training phase. Moreover, instead of this algorithm, every kind of Meta-Heuristic algorithms such as Gravitational Search Algorithm, Simulated Annual, Hunting Search Optimization and etc. can be used.

## REFERENCES

- [1] K. Aghakhani, A. Karimi, A new approach to predict stock big data by combination of neural network and harmony search algorithm, 5th International Conference on Computer Science, Electrical and Electronics Engineering, Malaysia (2016).
- [2] K. Aghakhani, A. Karimi, Provide a new method for time series prediction using Artificial Neural Network and harmony search meta-heuristic algorithm, Proc. 1st International Conference on Advances in Engineering and Basic Sciences (2014).
- [3] M. Aiken, Using a neural network to forecast inflation, Industrial Management & Data Systems 7 (1999) 296–301.
- [4] M. Asghari Oskoei, Time Series Prediction by Neural Nets, Iranian Economic Research Journal, V.12, (2002), 69-96.
- [5] M. Bashiri and A. Geranmayeh, Tuning the parameters of an artificial neural network using central composite design and genetic algorithm, Scientia Iranica, E, Ind. Eng. 18(6) (2011) 1600-1608.
- [6] R. Bill and T. Jackson, meeting neural networks, translated by Dr. Alborzi, Mahmoud, Scientific publication of San'ati Sharif University (2001).
- [7] M. C. Chan, C. C. Wong, and C. C. Lam, Financial Time Series Forecasting by Neural Network Using Conjugate Gradient Learning Algorithm and Multiple Linear Regression Weight Initialization, Computing in Economics and Finance (61) (2000).
- [8] W. C. Chiang, T. L. Urban and G. W. Baldridge, A neural network approach to mutual fund net asset value forecasting, Omega, Int.j.mgmt Sci. 24(2) (1996) pp. 205-215.
- [9] B. Egeli, M. Ozturk and B. Badur, Stock market prediction using artificial neural networks, Proc. 3rd International Conference on Business, (Hawaii, June 2003), pp. 1-8.
- [10] M. H. Fazel Zarandi, M. Avazbeigi and M. Alizadeh, A Neuro-Fuzzy Expert System Trained by Particle Swarm Optimization for Stock Price Prediction, Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies, IGI-Global, (2012) 633-650.

- [11] A. Garliuskas, Neural Networks Chaos and computational algorithms of forecast in finance, Proc. IEEE SMC conference, Man and Cybernetics, (1999), pp. 638-643.
- [12] Z. W. Geem, J. H. Kim, and G. V. Loganathan, A new heuristic optimization algorithm: Harmony search, Simulation 76 (2001) 60–68.
- [13] M. Ghiasi, and A. Karimi , Fault Diagnosis Method for Mobile Ad-hoc Network by Using Smart Neural Networks, Procedia Computer Science 42 (2014) 222-227.
- [14] M. H. Gholi Zadeh, Gh. Vahid Pur, Stock price prediction by using fuzzy regression, Research in Economic Sciences, No.12, pp. 106-128.
- [15] C. W. J. Granger, Forecasting Stock market prices, Lessons for casters, Working paper, University of California, San Diego, Department of Economics, (1991), p. 178-179
- [16] E. Hadavandi, H. Shavandi, and Ghanbari, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting, Knowledge-Based Systems, Vol. 23, (2010), pp. 800-808.
- [17] M. R. Hassan, A combination of hidden Markov model and fuzzy model for stock market forecasting, Neurocomputing, Vol. 72, (2009), pp. 3439-3446.
- [18] M. R. Hassan and B. Nath, Stock market forecasting using hidden Markov model: a new approach, Proc. 5th international conference on intelligent system design and application, (Poland, 2005), pp. 192-196.
- [19] M. R. Hassan, B. Nath, and M. Kirley, A fusion model of HMM, ANN and GA for stock market forecasting, Expert Systems with Applications, Vol. 33, (2007), pp. 171-180.
- [20] N. Hatami, H. Mirzazadeh and R. Ebrahimpour, Combine neural network for stock price prediction, Journal of Economics Sciences, Vol. 10, No. 2(39), (February 2011), pp. 61-80.
- [21] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall, (1999).
- [22] J. R. Jang, C. Sun and E. Mizutani, Neuro-Fuzzy, and Soft Computing, Prentice-Hall, (1997).
- [23] P. Jones, Investment; Analysis and management, Jane Wiley and sons, Inc. 7th edition, New York. 99, 300-380.
- [24] I. Kaastra, and M. Boyd, Designing a Neural Network for Forecasting Financial and Economic Time Series, Neurocomputing, Vol 10, (1996), pp. 215-236.
- [25] A. Karimi., et al., Cluster head selection using fuzzy logic and chaotic based genetic algorithm in wireless sensor network, Journal of Basic and Applied Scientific Research 3 (2013) 694-703.
- [26] A. Karimi, F. Z., S.A.R. Al-Haddad, S. Morshed, PFA: Parallel Filtration Algorithm for Query Technology in Spatial Database, J. Basic Appl. Sci. Res. 4(1), (2014) 21-26.
- [27] S. V. Kartalopoulos, Understanding neural networks & fuzzy logic – Basic concepts & applications, Prentice Hall of India Pvt. Ltd., (New Delhi, 2004).
- [28] H. Khaloozadeh, A. Khaki- Sedigh, C. Lucas, On the predictability of price fluctuations in Tehran stock Exchange: A correlation dimension estimation approach, Esteghlal Journal of Engineering, Isfahan University of Technology, Vol. 18, No. 1, (Sep 1999).
- [29] M. Khashei, S. R. Hejazi and M. Bijari, A new hybrid artificial neural networks and a fuzzy regression model for time series forecasting. Fuzzy Sets and Systems, 159, (2008), pp. 769-786.
- [30] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by simulated annealing, Science, Vol. 220, (1983), pp. 671-680.
- [31] Lendasse, A. et al. , non-linear financial time series forecasting application to Bell 20 stock market index, European Journal of Economic and social system, 14, No1, (2000), PP .81-91.
- [32] M. B. Manhaj, Foundations of Neural Networks, Amirkabir University, Tehran, Iran, (2000),(In Persian)
- [33] Q. K. Pan, P. N. Suganthan, M. Fatih Tasgetiren, J. J. Liang, A self-adaptive global best harmony search algorithm for continuous optimization problems, Applied Mathematics, and Computation. 216, (2010) 830-848.
- [34] R. Sundaresw, R. Steve, Sh. Philip, The financial analyst forecasting literature: taxonomy with suggestions for further research, International Journal of Forecasting, Vol.24, Issue 1, (January-March 2008) 34-75.
- [35] J. Shahrebi and V. S. Niaz, Data Mining, Jahad Daneshgahi, Inc., Tehran, (2009).
- [36] T. Stengos, E. Panas, Testing the efficiency of the Athens stock exchange: Some result from the Banking Sector, Empirical Economics, No.17 (2), (1992), pp .239-252.
- [37] M. H. Tayarani and M. R. Akbarzadeh, Magnetic Optimization Algorithms a new synthesis, Proc. IEEE Congress on Evolutionary Computation, (2008), pp. 2659-2664.
- [38] H. White, Economic Prediction Using Neural Networks: The case of IBM Daily stock Returns, Proc. IEEE International conference on Neural Networks II (1989).
- [39] X. S. Yang, Harmony Search as a Metaheuristic Algorithm in Music-Inspired Harmony Search Algorithm: Theory and Applications, (Editor Z. W. Geem). Studies in Computational Intelligence. Springer Berlin, Vol. 191, (2009), pp. 1-14.

# Effective Techniques for Reduction of Impulse, Gaussian and Speckle Noises

Md. Golam Moazzam, Tanzila Rahman and Mohammad Shorif Uddin

Department of Computer Science and Engineering

Jahangirnagar University

Savar, Dhaka, Bangladesh

**Abstract**—Noise is a common phenomenon and usually introduce during acquisition and transmission of images. Reduction and removal of noise from digital images is a prerequisite for subsequent analysis and recognition. Hence, nowadays it becomes an active area of research. Different types of noise can be added with digital images, such as impulse noise, Gaussian noise, speckle noise and so on. Impulse noise can be defined by replacing the intensity of an image point with random value of either higher-end or lower-end. Gaussian noise can be described by randomly adding values with zero mean maintaining Gaussian distribution to the intensities of image points. With a view to eradicate of these noises in this paper we briefly describe some important noise reduction methods. On the other hand speckle is a multiplicative noise that usually occurs in SAR and ultrasound images. For effective reduction of this noise here we have modified an existing technique and perform experimentation to confirm its superiority.

**Keywords-** *Gaussian noise, median filter, fuzzy filter, frost filter, mean square error.*

## I. INTRODUCTION

Unsystematic variation of luminance or color information in digital images is a common phenomenon that can be induced during acquisition, transmission, reception, storing and retrieval. For improving the performance of image analysis, object detection, classification and so on, noise reduction as well as removal is very important in image processing. Due to the contradiction of noise removal and detail preservation, it becomes more challenging task. The problem nature depends on the types of noise in images. There are several types of noises such as Gaussian, impulse, speckle, shot and quantization, uniform, film grain, anisotropic etc. Among these, usually impulse, Gaussian and speckle noise are very common. Hence, our main focus is on these noises. In impulse noise, the pixel value is 0 or 255 [1]. Usually transmission errors cause this type of noise. The primary goal of impulse noise reduction is to reduce the noise as well as preserving detail information [2], [3]. Some well-known median filters are SMF (Standard median filter) [4], AMF (Adaptive median filter) [5], HMF (Hybrid median filter) [6], DBMF (Decision-based median

filter) [7], [8], VMF (Vector median filter) [9][10] and so on. An existing image pixel becomes noise free by applying these filters globally on the whole image. However, these filters are suitable for removing low density impulse noise. Recently, fuzzy-based filters showed promising results for removing high density impulse noise [1], [11], [12].

Gaussian noise adds random intensities in image points maintaining zero mean Gaussian distribution. Many researchers proposed diverse methods, such Standard mean filter [4], Wiener filter [13], [14], Geometric mean filter [15], Harmonic mean filter [16] and so on. Hence, in this paper, we highlight some methods which are very useful in reducing impulse and Gaussian noise from grayscale and color images. On the other hand, for removing speckle noise here we have proposed a modified Frost filter.

The remainder of this paper is arranged as follows. Section II describes different types of noise modeling, Section III contains methodology, Section IV presents results and discussions, and, finally, conclusions are drawn in section V.

## II. NOISE MODELING

Various types of noise can affect digital images at the time of acquisition and transmission. Undesirable effects like artifacts, unrealistic edges, lines, corners and so on can be produced. To reduce these unwanted phenomena we need to know the details of these noises for the proposition of a filter. During acquisition of images using charge coupled device (CCD) several factors such as CMOS sensors, temperature, light levels, PSF (point spread function) and MTF (modulation transfer function) can affect the density of noise in the image. Noisy image can be modeled as follows:

$$c(x, y) = a(x, y) + b(x, y) \quad (1)$$

Here,  $a(x,y)$ ,  $(x,y)$  and  $c(x,y)$  are the input image, noise and resultant corrupted digital image, respectively. As mentioned earlier, here we are considering the three following very common types of noise in digital images.

#### A. Impulse noise(Pepper and salt noise):

It is also known as data drop noise as it drops original data values statistically [17]. Two possible values  $a$  and  $b$  with the probability of each is less than 0.2 exists. For an 8-bit digital image, the intensity value is 0 and 255 for pepper and salt noise, respectively.

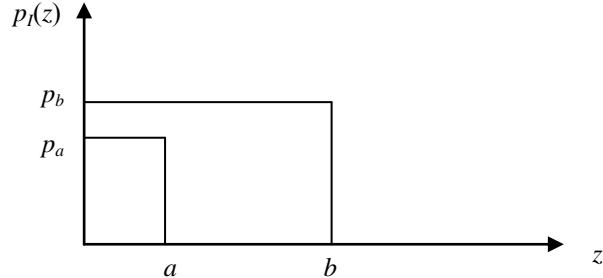


Figure 1: Probability density function for impulse noise model.

The PDF can be characterized by the following equation:

$$PDF = \begin{cases} p_a, & g = a(\text{pepper}) \\ p_b, & g = b(\text{salt}) \end{cases} \quad (2)$$

#### B. Gaussian Noise:

Gaussian noise usually arises in amplifiers or detectors, thus it is also called electronic noise. For the zero-mean Gaussian distribution characteristic we can remove noise through averaging pixel values [5][18]. Normally gray values are disturbed by Gaussian noise. Hence, Gaussian noise with PDF can be defined as follows:

$$p(g) = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\frac{(g-\mu)^2}{2\sigma^2}} \quad (3)$$

Where,  $g$ ,  $\sigma$  and  $\mu$  are gray value, standard deviation and mean, respectively.

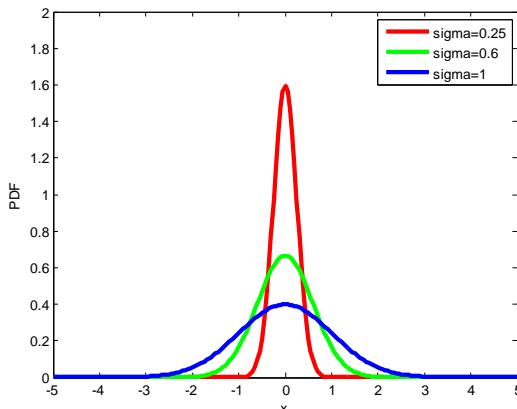


Figure 2: Probability density function for Gaussian noise.

#### C. Speckle Noise:

This is a type of granular noise [19] which degrades image quality in radar, synthetic aperture radar (SAR), ultrasound and tomographic images. It degrades fine details, contrast resolution, edges and so on. By multiplying image pixel with random values, speckle noise can be modeled as follows:

$$g(n, m) = f(n, m) * u(n, m) + \xi(n, m) \quad (4)$$

Where,  $g(n, m)$  = observed image,  $u(n, m)$  = multiplicative component of speckle noise and  $\xi(n, m)$  = additive noise.

### III. METHODOLOGY

We discussed some conventional as well as fuzzy filters for removing impulse and Gaussian noises. Moreover, for speckle noise we proposed a modified Frost filter.

#### A. Impulse noise:

##### 1) Conventional filters:

a) *Standard Mean filter [1]*: It is a linear low pass filter. With the reduction of intensity variation of neighbouring pixels it smooths the image. Basically each pixel in the image is replaced by the mean value of neighboring and the pixel itself. The mean filtering process carries out with the computation of the straightforward convolution of an image with appropriate kernel.

b) *Adaptive Mean filter [5]*: This type of filter is used to determine the corrupted pixel with the help of spatial processing. The whole process is done by comparing each pixel with its surrounding neighbor pixels in the image whereas the size of neighbor pixel is adjustable.

c) *Hybrid Median filter [6]*: In this filter median value of a 3x3 or 5x5 or 7x7 kernel is replaced with the current pixel in digital image. Outward are top and bottom edges pixels and the side edge pixels are wrapped around to complete the edge bound kernels.

d) *Decision based Median filter [7]/[8]*: The filter start processing with the detection of impulse noise. For this, first check whether the pixel is noise free or not. If the pixel is noise free, it remains unchanged otherwise the pixel value is replaced with the median value.

e) *Vector Median filter [9]*: VMF can be defined as the derivation of two multidimensional probability density functions whereas maximum-likelihood-estimate approach is

used. The PDFs are exponential and the filter has the similar property as median filter.

2) *Fuzzy based filter*: Filtering process can be divided into two steps:

a) *1st Step[1]*: In case of color image, converting RGB to gray may cause loss of information. Thus, to avoid this problem, color components R, G and B need to be processed separately. If input image is grayscale there is no need to apply this step.

b) *2nd Step[11]*: This step is divided into three parts. In case of color image each RGB component uses these stages.

i) *Estimation of noise[11]*: For estimating noise of the corrupted image a  $3 \times 3$  or  $5 \times 5$  window is taken and find whether the center lies within trimming range or not. If not, need to replace with the maximum and minimum gray values. Otherwise, it is considered that pixel lies within the trimming range.

ii) *Degree of corruption calculation[11]*: Fuzzy membership value can be found by calculating the amount of corruption. For each corrupted pixel the following (5) and (6) are used to find fuzzy membership value [6]:

$$\mu[w(i,j)] = \max \left( 1 - \frac{D}{E(i,j)} |J(i,j) - E(i,j)| \right) \quad (5)$$

$$D = \text{median} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |J(i,j) - E(i,j)| \quad (6)$$

Where,  $\mu[w(i,j)] \in [0,1]$  and  $E(i,j)$  are the fuzzy membership and estimated value of each pixel, respectively.

iii) *Restoration of original image[11]*: The following output is generated based on the above membership value,

$$Y(i,j) = E(i,j) + \mu[w(i,j)] \times [J(i,j) - E(i,j)] \quad (7)$$

For color image, finally, R, G, B components are combined to get the original color image.

## B. Gaussian noise:

### 1) Conventional filters

a) *Wiener filter[13]/[14]*: Wiener filter is a linear filter that removes additive noise by linear time-invariant property. It

produces less mean-squared error. Noise degraded frequency components are unable to reconstruct by Wiener filter. It can only suppress them. Moreover, blurring effect caused by bandlimiting can not be reduced by this filter.

b) *Geometric mean filter[15]*: In this filter each pixel value is replaced by geometric mean value of the neighboring pixels. The main drawback of this type of filter is image blurring. It works better for removing Gaussian noise with preserving detail information.

c) *Harmonic mean filter [16]*: Also known as subcontrary mean filter that works for monochrome. Like geometric mean filter it gives better output for Gaussian noise and preserving image details.

### 2) Fuzzy based Existing filter [12]

According to the fuzzy filtering the output for input image  $f_p$  is given as,

$$g(p) = \frac{\sum_p F_p f_p}{\sum_p F_p} \quad (8)$$

Where  $F_p$  is known as 8-neighbouring function.  $F_p$  can be determined by the following function.

$$F_p = \exp \left( -\frac{(f_p - f_{\max})^2}{2\sigma} \right) \quad (9)$$

Here,  $f_{\max}$  is the maximum intensity value of 8-neighbours and  $\sigma$  is the standard deviation of the filtering window.

### 3) Modified fuzzy filter[12]

Existing fuzzy filter uses a  $3 \times 3$  filtering window. Thus it is more suitable to calculate the membership function  $F_p$  based on the effects of 8 neighbors. Therefore,  $F_p$  can be determined using the following modified equation:

$$F_p = \begin{cases} 1 & \text{if } f_p = 0 \text{ or } 255 \\ \exp \left( -\frac{(f_p - f_{\max})^2}{2 \times 8 \times \sigma} \right), & \text{Otherwise} \end{cases} \quad (10)$$

## C. Speckle noise:

### 1) Conventional filters

a) *Frost filter[17]*: Frost filter is a type of adaptive filter that uses negative exponential distribution for removing speckle noise. It calculates weight of each cell using statistics of local image where the weights vary based on the distance of

the center cell. For minimizing MSE (Mean squared error) the filter calculate the average weight of the cell values in the filtering window. As a result homogeneous area becomes more smooth. The weighting function and the filtering formula of frost filter are given in equations(11) and (12), respectively.

$$W(x, y) = Ke^{-k_d C(x, y)\sqrt{x^2 + y^2}} \quad (11)$$

And  $I' = I * W \quad (12)$

Here, K is the constant, C is the standard speckle index and  $K_d$  is called damping factor.

b) *Kuan filter*[17]: This filter uses maximum likelihood probability for calculating signal value of the center cell in the filtering window. Moreover, it converts multiplicative noise model into an adaptive noise model and preserves edges.

c) *Lee filter* [18], [19]: Lee filter is used to smooth speckled data which are related to the image intensity as well as additive and multiplicative image components. Basically, it is developed based on the standard deviation. It effectively preserves image edges and details while suppressing noise. In this filter, noise affected pixel value is replaced by the calculated value of the surrounding pixels.

d) *Wavelet thresholding* [20]: Speckle noise is one of the high frequency component which appears in wavelet coefficient. In this procedure, at first it requires to calculate DWT of the corrupted image and threshold wavelet coefficients. For denoising we need to calculate IDWT. Frequently used thresholding functions are hard and soft thresholding. The hard thresholding can be defined as follows:

$$\eta_1(w) = wI(|w| > T) \quad (13)$$

And the soft thresholding can be expressed as:

$$\eta_2(w) = (w - sgn(w)T)I(|w| > T) \quad (14)$$

Here,  $w$  is the wavelet coefficient,  $T$  is the threshold value and  $sgn$  is the sign function of the image.

## 2) Modified Frost filter

Existing frost filter [17] use a  $5\times 5$  mask with square structuring element. As frost filter uses average weight mask thus it is more convenient to use a  $7\times 7$  filtering window with disk type structuring element. Again, calculated weighting function gives small value. Hence, if we multiply square of

pixels value by 8, the value becomes more suitable for further processing. Now, the modified weighting function becomes:

$$W(x, y) = Ke^{-k_d C(x, y)\sqrt{8(x^2 + y^2)}} \quad (15)$$

## IV. SIMULATION REPORTS AND RESULTS

All algorithms are tested a set of  $512\times 512$  test images of both color and grayscale which are corrupted by impulse, Gaussian and speckle noise. For speckle noise, we add noise ranging from 10% (low density) to 80% (high density). For measuring the performance and effectiveness of the above methods we calculate PSNR and MSE.

### A. Performance Measurement

PSNR can be defined as follows:

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) dB \quad (16)$$

Where MSE is the mean squared error, and is defined as,

$$MSE = \frac{\sum_x \sum_y (I(x, y) - Y(x, y))^2}{MN} \quad (17)$$

Here  $I$  and  $Y$  are the corrupted image and restored image of size  $M\times N$  pixels, respectively.

### B. Comparison Results

The proposed modified method for reducing speckle noise is tested on different images such as Lena, Baboon, Parrot, Barbara (both grayscale and color images) and so on. Also for comparing the performance of the proposed method we use different conventional filter such as mean filter, hybrid median filter, frost filter, Kuan filter and lee filter where added noise ranges from 10% (low-density) to 80% (high-density). Fig. 3 and Fig.4 show the comparison graphs of the PSNR values after applying different filters on corrupted gray and color images. From these comparison graphs we can decide that proposed method gives better result for removing noise as well as preserving detail information.

Moreover, the visual and qualitative results for speckle noise are presented in Figs. 5 and 6. Both results confirm that modified method for speckle noise reduction gives better result than that of the traditional filtering methods.

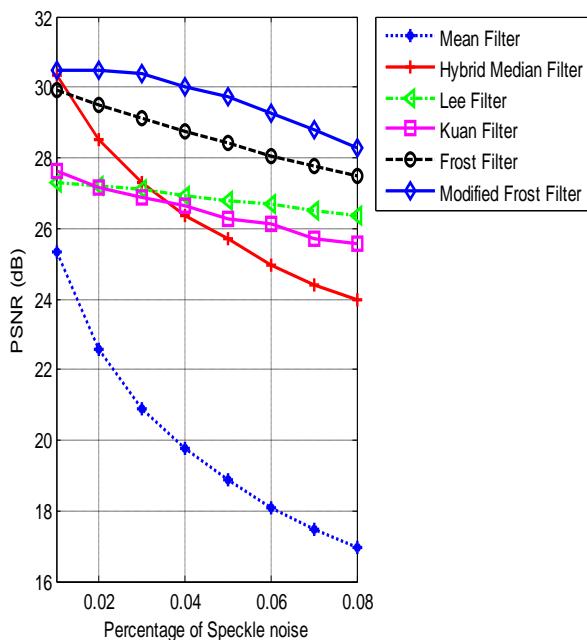


Figure 3. PSNR graph showing the comparison results of different filters for "Lena" gray image at different speckle noise densities.

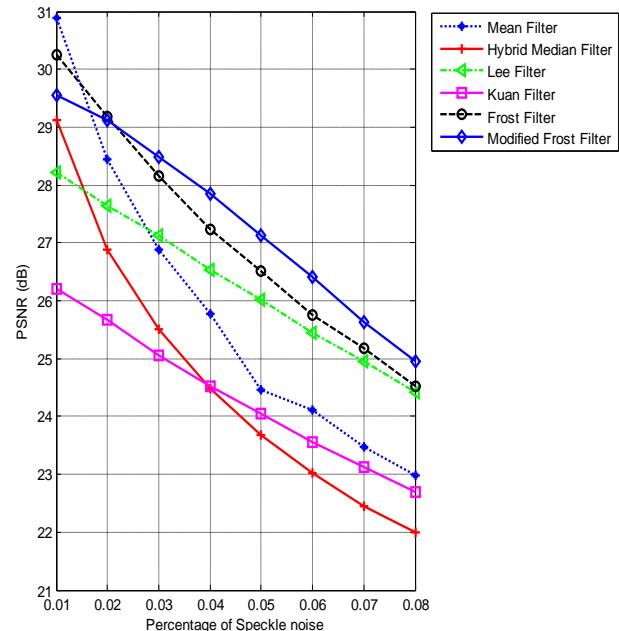


Figure 4. PSNR graph showing the comparison results of different filters for "Lena" color image at different speckle noise densities.

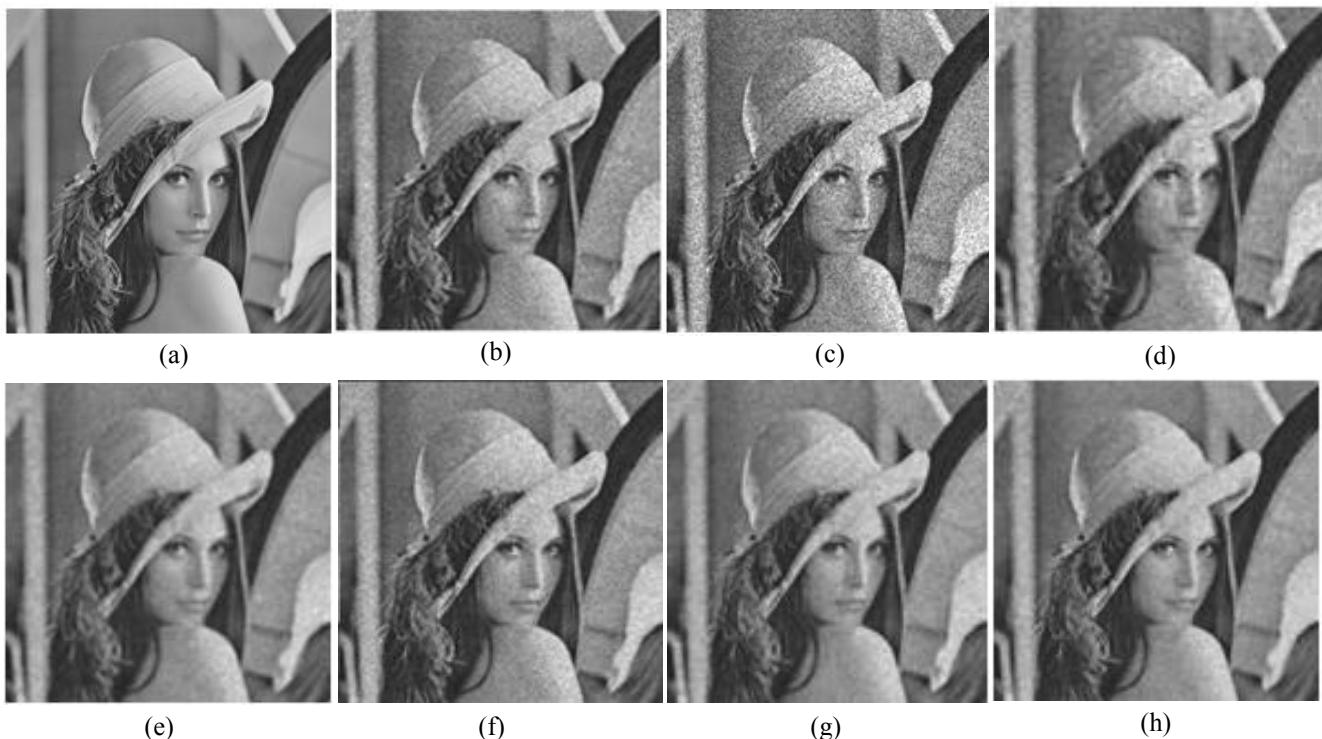


Figure 5. Experimental results of different filters where input image is corrupted by speckle noise: (a) Input grayscale image; (b) corrupted image by 80% speckle noise; (c) Mean filter (PSNR: 16.94 dB); (d) Hybrid Median filter (PSNR : 24.00 dB); (e) Lee filter (PSNR : 26.35 dB); (f) Kuan filter (PSNR : 25.59 dB); (g) Frost filter (PSNR : 27.48dB); (h) Modified Frost filter (PSNR : 28.29 dB).



Figure 6. Experimental results of different filters where image is corrupted by speckle noise: (a) Input color image; (b) corrupted image by 80% Speckle noise; (c) Mean filter (PSNR : 22.98 dB); (d) Hybrid Median filter (PSNR : 21.99 dB); (e) Lee filter (PSNR : 24.41 dB); (f) Kuan filter (PSNR : 22.70 dB); (g) Frost filter (PSNR : 24.52 dB); (h) Modified Frost filter (PSNR : 24.98 dB).

## V. CONCLUSION

In this paper, some effective methods have been described to remove impulse and Gaussian noises in digital images. Moreover, we proposed a modified frost filter for reducing speckle noise. To demonstrate the performance of this modified filter, experiments have been conducted on grayscale and color images with a wide range of noise densities (from 10% to 80%). Experimental results show that the proposed methods exhibit superiority in comparison with the state-of-the-arts filtering methods in terms of both visual and quantitative (PSNR figures) points of view. Our future work will focus on reduction of shot noise, stripping noise and so on for both grayscale and color images.

## ACKNOWLEDGMENT

Authors are very grateful for the partial support of Department of Computer Science and Engineering at Jahangirnagar University.

## REFERENCES

- [1] Tanzila Rahman and Mohammad Sharif Uddin, "Removal of High Density Impulse Noise from Color Images Using an Adaptive Fuzzy Filter," International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), 2014.
- [2] G.R. Arce, N. C. Gallagher, T. Nodes, "Median filters: Theory and applications," in Advances in Computer Vision and Image Processing, T. Huang, Ed. Greenwich, CT:JAI, 1986.
- [3] A.C. Bovik, T. Huang, D.C. Munson, "A generalization of median filtering using linear combinations of order statistics," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-31, pp. 1342-1350.
- [4] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Digital Image Processing Using MATLAB, Prentice-Hall, 2004.
- [5] V.Jayaraj , D.Ebenezer, K.Aiswarya, "High Density Salt and Pepper Noise Removal in Images using Improved Adaptive Statistics Estimation Filter", IJCSNS International Journal of Computer Science an Network Security, Vol .9, No.11, November 2009.
- [6] Hybrid median filter, Available online: <http://rsb.info.nih.gov/ij/plugins/hybrid2dmedian.html>, Access on : January 13, 2015.
- [7] S.Gopi Krishna, T. Sreenivasulu Reddy, G.K.Rajini, "Removal of High Density Salt and Pepper Noise Through Modified Decision Based Unsymmetric Trimmed Median Filter", International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 1, pp.090-094, Jan-Feb 2012.
- [8] K.S. Srinivasan and D. Ebenezer , "A New Fast and Efficient Decision-Based Algorithm for Removal of High-Density Impulse Noises", IEEE signal processing letters, Vol. 14, No. 3, 2007, pp.189 -192.
- [9] J. Astola, P. Haavista, Y. Neuvo, "Vector median filters", Proceedings of the IEEE, Vol. 78, Issue 4, 1990.

AUTHORS PROFILE

- [10] T.A. Nodes and N.C. Gallagher, "The output distribution of median type filters," IEEE Trans. Commun., vol. COM-32, pp. 532-541, 1984.
- [11] Madeena Sultana, Mohammad Shorif Uddin, Farhana Sabrina, "High Density Impulse Denoising by A Novel Adaptive Fuzzy Filter", 2nd International Conference on Informatics, Electronics and Vision (ICIEV13), 17-18 May, 2013.
- [12] Tanzila Rahman, Mohammad Reduanul Haque, Liton Jude Rozario and Mohammad Shorif Uddin, "Gaussian Noise Reduction in Digital Images Using a Modified Fuzzy Filter," 17th Int'l Conf. on Computer and Information Technology, December 2014.
- [13] The Wiener filter, Available online: [http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\\_COPIES/VELDHUIZEN/node15.html](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/VELDHUIZEN/node15.html), Access on : January 14, 2015.
- [14] Ajay Kumar Boyat and Brijendra Kumar Joshi, "A Review Paper: Noise Models in Digital Image Processing," Signal & Image Processing : An International Journal (SIPIJ), vol.6, no.2, April 2015.
- [15] Ling Guan, Rabab K. Ward, "Restoration of stochastically blurred images by the geometrical mean filter", Optical Engineering, Vol. 29, No. 4, April 1990.
- [16] V. Murugan, T. Avudaiappan and Dr. R. Balasubramanian, "A Comparative Analysis of Impulse Noise Removal Techniques on Gray Scale Images", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol.7, no.5, 2014.
- [17] T. Sun and Y. Neuvo, "Detail-Preserving Median Based Filters in Image Processing," Pattern Recognition Lett., Vol. 15, No. 4, pp. 341-347, April 1994
- [18] R. Marudhachalam and Gnanambal Ilango, "Fuzzy Hybrid Filtering Techniques for Removal of Random Noise from Medical Images," International Journal of Computer Applications, vol. 38, no. 1, January 2012.
- [19] Yonghong Huang and J.L. van Genderen, "Evaluation of Several Speckle Filtering Techniques for ERS-1 & 2 Imagery," International Archives of Photogrammetry and Remote Sensing, vol. XXXI, 1996.
- [20] S.Sudha, G.R.Suresh and R.Sukanesh, "Speckle Noise Reduction in Ultrasound Images by Wavelet Thresholding based on Weighted," International Journal of Computer Theory and Engineering, vol. 1, no.1, April 2009.

**Md. Golam Moazzam** completed his B.Sc (Hons) in Electronics and Computer Science and M.S. in Computer Science and Engineering from Jahangirnagar University, Bangladesh in 1997 and 2001 respectively. He joined as a lecturer in the Department of Computer Science and Engineering, Jahangirnagar University, in 2001. Currently, he is serving as a Professor of this department. His research interests include Digital Image Processing, Artificial Intelligence, Computer Vision and so on.

**Tanzila Rahman** has completed her B.Sc. honors from Jahangirnagar University, Savar, Dhaka and M.S. from the same university. She worked as a Software Engineer in HawarIT Software Service Limited (A Dutch Bangladeshi Company) from March 2011-January 2014. Later she joined in Secure Link Service BD Limited from February 2014. Now she is working as a Lecturer at Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka. Her research field is Image Processing, Computer Vision, Machine Learning, Biomedical imaging and Pattern recognition.

**Mohammad Shorif Uddin** received his PhD in Information Science from Kyoto Institute of Technology, Japan, Master of Education in Technology Education from Shiga University, Japan, Bachelor of Science in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET). He joined in the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka in 1992 and currently he is serving as a Professor and also Chairman of this department. He began his teaching career in 1991 as a Lecturer of the Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology (CUET). He undertook postdoctoral researches at Bioinformatics Institute, A-STAR, Singapore, Toyota Technological Institute, Japan, Kyoto Institute of Technology, Japan, Chiba University, Japan and University of Bonn, Germany. His research is motivated by applications in the fields of imaging informatics, computer vision and image velocimetry. He has published more than 80 papers in peer-reviewed international journals and conference proceedings and also delivered keynote speeches in international conferences in home and abroad. He has two patents for his scientific inventions. He received the Best Paper Award in the International Conference on Informatics, Electronics & Vision (ICIEV2013), Dhaka, Bangladesh and Best Presenter Award from the International Conference on Computer Vision and Graphics (ICCVG 2004), Warsaw, Poland. He is the co-author of three books. He is a Fellow of Bangladesh Computer Society and also a senior member of IEEE and IACSIT.

# Performance Evaluation of Femtocell Based LTE Network under the Concept of Cross-layer Optimization

Jesmin Akhter

Associate Professor, Institute of Information Technology,  
Jahangirnagar University  
Dhaka, Bangladesh

Md. Imdadul Islam

Professor, Department of Computer Science and  
Engineering, Jahangirnagar University  
Dhaka, Bangladesh

M. R. Amin, Member IEEE

Professor, Electronics and Communications Engineering,  
East West University, Dhaka, Bangladesh

**Abstract**—To achieve end-to-end maximum throughput, the wireless Internet access requires (a) sufficient SNR at evolved Node-Bs and UEs (user equipment) at physical layer and (b) congestion control algorithm in determining appropriate window size at transport layer. Considering above, the paper deals with both the layers for Femto cellular LTE network and relates the fading parameters of physical layer and congestion parameters of transport layer. One of the promising approaches of 4G mobile cellular network is to incorporate Femto cell inside macro cells to get access of a MS (Mobile Station) within few meters. This approach is adopted to combat the small scale fading of wireless link so that a MS can achieve optimum throughput, otherwise huge capacity of a mobile cellular network is lost under fading environment. This paper deals with the relation among outage probability, density of Femto cell, threshold link capacity, threshold SNR (signal to noise ratio) and mean congestion window size under fading environment. We found that Nakagami-m environment provides better result compare to Rayleigh case (because of several direct link in Nakagami-m environment) at the same time path loss exponent is a vital factor for such network. Next we analyze the performance of end-to-end TCP (Transmission Control Protocol) link under the concept of congestion window control with newly developed state transition chain. The impact of fading parameter on outage probability, mean transmission rates, mean window size and throughput are analyzed explicitly for such network.

**Keywords-component;** 4G mobile, small scale fading, optimum throughput, outage probability, Nakagami-m fading and moment generation function.

## I. INTRODUCTION

The service area is populated with Base Stations (BSs) in a continuous fashion in Mobile Communications Network (MCN). The Long Term Evolution (LTE) considered as forerunners of 4G has introduced the concept of Femtocells as the access point of home users. Femto cells are the home base stations with short-range pave the way for accessing Internet through the backbone network at low low-power transmission

discussed in [1]. Since Femtocells use the same spectrum of overlaid macro cell hence huge interferences occurs between them and the performance of the network is declined. Various frequency reuse schemes in LTE to mitigate the cross tier interference between macrocells and Femtocell networks are studied in [2].

The main objective of this paper is to derive the outage probability of a user under macro-Femto environment. An analysis of above environment is found in [3], although the authors did the analysis for cognitive radio network. Here the authors investigated aggregate interference at the primary receiver from cognitive radios distributed in a finite ring. We adopt the concept to find the aggregate interference on a Femtocell user surrounded by several Femtocells within a circular ring. Authors showed that the aggregate interference is approximately Gamma distributed where the entire analysis is done based on moment generating function. The extension of the paper is found in [4]; where two approaches: no power control and distance based power control is used to derive aggregate interface on a user. Both the paper is further extended in [5], showed that by locating the beacon transmitter at the PU (Primary User) receiver can enhance the performance in context of capacity outage probability of the PU. The performance of the PU in the presence of the SU's (Secondary User) interference is also found in [6-8] with complete statistical analysis. On the other hand since Femtocells are able to use the available sub-bands using the cognitive radio techniques, a graph colouring aided sub-channel allocation algorithm is developed in [9] using graph theoretic approach to get the optimum Femtocell throughput in dense Femtocellular network environment. Here the authors tried to suppress the interference among the Femto users by grouping them into different clusters. Dynamic Frequency Planning (DFP) is another approach used in 4G Femtocell of WiMAX to avoid interference at the expense protocol complexity mentioned in [10]. Efficient handling of handover calls is the key for successful Femtocell/macrocell integration. A large neighbour Femtocell list causes unnecessary scanning for the handover.

Therefore, an appropriate and optimal neighbour Femtocell list is essential for dense Femtocellular network deployment. For this purpose in [11], the authors propose a call admission control algorithm to create an optimum neighbour cell list for handover. Here, for dense Femtocellular network deployment, the frequency for each of the Femto access points is allocated on the basis of the neighbouring overlapping Femtocells. Thus, the overlapping of the two Femtocells does not use the same frequency to avoid interference [12].

In this paper we combined outage probability of a user under fading channel of LTE with the congestion control operation of TCP. Our aim is to analyze the traffic performance of wireless TCP operation of LTE network. The idea of flow control of TCP; based on congestion window is available for both wires and wireless communication in [13-16]. TCP has variable congestion control schemes, such as Cubic, Reno, Vegas, Westwood and WinSock. In this paper we chose congestion window based control. Among them TCP Westwood (TCPW) is a sender-side modification of the TCP congestion window algorithm that improves upon the performance of TCP Reno in wired as well as wireless networks [13]. While a proactive congestion control mechanism is proposed to improve TCP performance in paper [14] and showed that TCP Vegas performs better than TCP Reno. Hybrid TCP can eliminate drawbacks of conventional methods as long as its mode switching is carried out in a deliberate manner, the analytical framework on hybrid TCP performance are shown in [15]. The direct employment of the existing protocols cannot achieve the requirements of LTE-Advanced due to the large-bandwidth and low-latency links. An enhanced congestion control mechanism is used in [16] to improve the performance of TCP over LTE-Advanced and developed to be able to transfer high rate of packet over large bandwidth low-latency platform. By controlling path diversity of cellular networks, data transport using multipath TCP is a promising solution discussed in [17-18]. For simplicity of analysis we use ‘slow start’ algorithm of [19].

The paper is organized as follows: section II provides analytical model of outage probability of a Femto user, under aggregate interferences of surrounding cells based on moment generating function and the concept is applied in congestion window based TCP network, section III provides the results based on theoretical analysis of section II and a Markovian chain is modeled to get the probability states of congestion window under steady condition, finally, section IV concludes the paper.

## II. SYSTEM MODEL

In this section we have derived outage probability of a user under the interferences of surrounding Femtocells based on the concept of spectrum sensing of cognitive radio network. Here we consider exponential path loss model due to short distance and moment generating function under aggregate shot noise. Fig.1 shows several Femto BSs inside the coverage of a macro cell of radius  $R_m$ . A MS inside the coverage of a Femtocell experiences interference from surrounding Femto BSs within the circle of radius  $R_E$ , where the radius of each Femto cell is  $R_f$ .

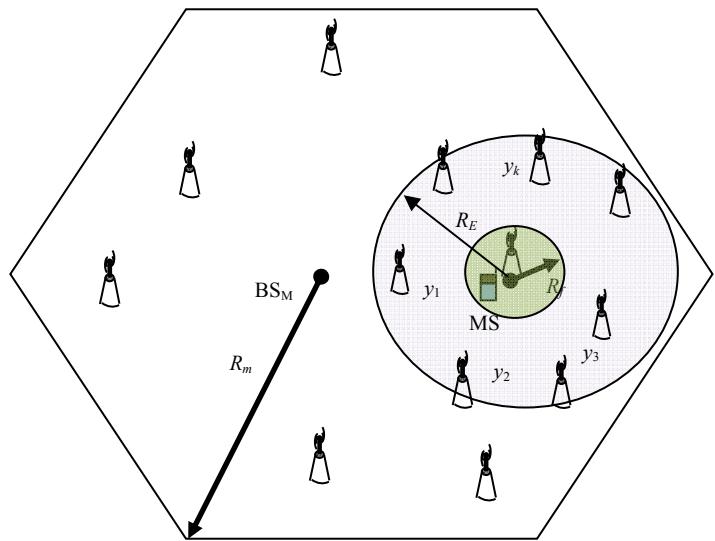


Fig.1 Test MS under a set of interfering Femtocells

The useful parameters of analytical model are:

Transmit power of Macro BS =  $P_m$

Transmit power of Femto BS =  $P_f$

Radius of Macro cell =  $R_m$

Radius of Femto cell =  $R_f$

Location of Macro cell MSSs,  $\varphi_m = \{X_i\}$

Location of Femto cell BSs,  $\varphi_f = \{Y_i\}$

Taking Femto BSs of fig.1 which make interference on MS we can write,

$$\varphi_f^p = \{Y : \text{Those make interference on MS}\}; \quad \text{Where}$$

$$\varphi_f^p \subset \varphi_f$$

Let the test MS lies inside the Femtocell located at  $Y_0$  and any the Femto BS belongs to the set  $\varphi_f^p$  will make interference to test MS.

Now, the interference made by Femto BS at  $Y_i$  will be [3]:

$$I_i = G_{Y_i} P_f d^{-\alpha} \quad (1)$$

Where  $G_{Y_i}$  is the channel gain of Femto BS at  $Y_i$  and the test MS and  $\alpha$  is the path loss exponent. The total interference:

$$I = \sum_{Y_i \in \varphi_f^p} I_i \quad (2)$$

Let the distribution of location of Femto cells within the Macrocell follow the Poisson's pdf.

For circular area of fig.1, if the number of Femto BS is  $N = n$ , then using the idea of Poisson's pdf of network traffic [20-21],

$$P(N = n) = \frac{(\lambda A_I)^n}{n!} e^{-\lambda A_I} \quad (3)$$

Where,  $\lambda$  is density of Femto BS and  $A_I$  is area of region which makes interference with test Femto MS; where  $A_I = \pi(R_E^2 - R_f^2)$  and  $R_E$  is the maximum distance to make interference on the test MS.

In equation (1), let us consider a fading random variable  $X_i$  instead of BS's gain  $G_{yi}$ .

Now from eq.(1), the moment generation function, MGF is obtained from [22-23] where authors considered the two r.v.s independent for simplicity of analysis,

$$\begin{aligned} M_I^i(s) &= E[e^{-sI_i}] \\ &= \int_0^\gamma \left\{ \int_{R_f}^{R_E} f_D(d_i) e^{-sI_i} d_{d_i} \right\} f_{X_i}(x) dx \end{aligned} \quad (4)$$

;  $\gamma$  is the instantaneous SNR wireless link. Again the pdf of distance  $d_i$ ,

$$f_D(d_i) = \begin{cases} \frac{2\pi d_i}{A_I}; R_f < d_i < R_E \\ 0; \text{Otherwise} \end{cases} \quad (5)$$

And  $f_{X_i}(x)$  is the pdf of fading random variable  $x$ .

Taking the average of eq. (4) using eq. (1) we get like [3, 24],

$$M_I(s) = \exp\{\lambda A_I(M_I^i(s) - 1)\} \quad (6)$$

Above equation can be simplified like,

$$\begin{aligned} M_I(s) &= \exp\{-\lambda A_I(1 - M_I^i(s))\} \\ &= \exp\{-\lambda A_I(1 - E[e^{-sI_i}])\} \\ &= \exp\{-\lambda A_I(1 - E[e^{-sP_l(u)}])\} \\ &= \exp\left\{-\lambda A_I\left(1 - \frac{\mu}{\mu + sl(u)}\right)\right\} \\ M_I(s) &= \exp\left\{-\lambda A_I \frac{sl(u)}{\mu + sl(u)}\right\} \end{aligned} \quad (7)$$

Where the transmit power of Femto cell is taken as a r.v.  $P$  with mean value  $E[P] = P_f = I/\mu$  and  $l(x,y) = l(|x-y|) = l(u) = u^\alpha$  is the path loss between point  $x$  and  $y$ . Therefore the interference  $I_i = P.l(u)$  and  $E[e^{-sP_l(u)}] = \frac{\mu}{\mu + sl(u)}$ .

Integrating the infinitesimal circular area of interference,  $dA_I = 2\pi u du$  of (7),

$$\begin{aligned} M_I(s) &= \exp\left[-\lambda \int_0^\infty 2\pi u \frac{sl(u)}{\mu + sl(u)} du\right] \\ &= \exp\left[-2\pi \lambda \int_0^\infty \frac{su^{-\alpha}}{\mu + su^{-\alpha}} u du\right] \\ &= \exp\left[-2\pi \lambda \int_0^\infty \frac{udu}{1 + \frac{\mu}{su^{-\alpha}}}\right] \end{aligned}$$

$$M_I(s) = \exp\left[-2\pi \lambda \int_0^\infty \frac{udu}{1 + \frac{\mu}{su^{-\alpha}}} \right] \quad (8)$$

Above equation will be used later to derive the expression of outage probability. Now the signal to noise plus interference ratio of the test Femto user at the cell boundary,

$$SINR = \frac{P_f G_f R_f^{-\alpha}}{N_0 + I} \quad (9)$$

The eq. (9) provides the minimum value of SNR since a user at the cell boundary is maximally affected by interferences. The SNR of other users can be found only changing  $R_f$  with their distance.

The probability of SNR greater than or equal to the threshold  $\Gamma_{th}$ ,

$$\begin{aligned} P(SINR_{Test\_user} \geq \Gamma_{th}) &= P\left(\frac{P_f G_f R_f^{-\alpha}}{N_0 + I} \geq \Gamma_{th}\right) \\ &= P\left\{G_f \geq \frac{\Gamma_{th}(N_0 + I)}{P_f R_f^{-\alpha}}\right\} \\ &= e^{-\Gamma_{th} N_0 / P_f R_f^{-\alpha}} \cdot M_I\left(\frac{\Gamma_{th} I}{P_f R_f^{-\alpha}}\right) \end{aligned} \quad (10)$$

The derivation of last line of (10) is given in detail in [7].

Putting,  $s = \frac{\Gamma_{th} I}{P_f R_f^{-\alpha}}$  in (8)

$$M_I\left(\frac{\Gamma_{th} I}{P_f R_f^{-\alpha}}\right) = \exp\left(-\hat{\lambda}_f P_f^{2/\alpha} s^{2/\alpha} K_\alpha\right) \quad (11)$$

; Where  $K_\alpha = 2\pi^2/\alpha \sin(2\pi/\alpha)$  and  $\hat{\lambda}_f$  is the maximum allowed density of Femtocell.

Combining (10) and (11) we get,

$$P(SNR \geq \Gamma_{th}) = e^{-\Gamma_{th} N_0 / P_f R_f^{-\alpha}} \cdot e^{-\hat{\lambda}_f R_f^2 \Gamma_{th}^{2/\alpha} K_\alpha} = 1 - \varepsilon \quad (12)$$

; where  $\varepsilon$  is the outage probability.

Solving (12) we get,

$$\hat{\lambda}_f = \frac{-\ln(1-\varepsilon) - \frac{\Gamma}{P_f R_f^{-\alpha}} N_0}{R_f^2 \Gamma^{2/\alpha} K_\alpha} \quad (13)$$

The outage probability of the test user under SNR constraint will be [25],

$$1 - P(SNR \geq \Gamma_{th}) = 1 - e^{-\Gamma_{th} N_0 / P_f R_f^{-\alpha}} \cdot e^{-\hat{\lambda}_f R_f^2 \Gamma_{th}^{2/\alpha} K_\alpha} = \varepsilon \quad (14)$$

Now we will determine outage probability of the test user threshold capacity of the channel using Gamma pdf on

aggregate interferences  $I = \sum_{i=1}^n I_i$ . In Gamma pdf we need two

parameters,  $\theta = \frac{Var[I]}{E[I]}$  and  $n = (E[I])^2/var[I]$ . The pdf of  $I$  is expressed as [5],

$$f_I(i, n, \theta) = i^{n-1} e^{-i/\theta} / \theta^n \Gamma(n) \quad (15)$$

Now

$$E[I] = nE[I_i] = n \int_{R_f}^{R_E} PGd_i^{-\alpha} f_D(d_i) dd_i; \quad \text{Where}$$

$$\begin{aligned} I_i &= PGd_i^{-\alpha} \\ &= PGn \int_{R_f}^{R_E} d_i^{-\alpha} \frac{2\pi d_i}{\pi(R_E^2 - R_f^2)} dd_i \\ &= \frac{2PGn}{R_E^2 - R_f^2} \int_{R_f}^{R_E} d_i^{1-\alpha} dd_i \\ &= \frac{2PGn}{R_E^2 - R_f^2} \cdot \frac{d_i^{1-\alpha+1}}{1-\alpha+1} \Big|_{R_f}^{R_E} \\ &= \frac{2PGn}{R_E^2 - R_f^2(2-\alpha)} \cdot [R_E^{2-\alpha} - R_f^{2-\alpha}] \end{aligned} \quad (16)$$

The 2nd moment of  $I$ ,

$$\begin{aligned} E[I^2] &= nE[I_i^2] = nP^2 G^2 \int_{R_f}^{R_E} d_i^{-2\alpha} f_D(d_i) dd_i \\ &= \frac{2nP^2 G^2}{R_E^2 - R_f^2} \int_{R_f}^{R_E} d_i^{1-2\alpha} dd_i \\ &= \frac{2nP^2 G^2}{(R_E^2 - R_f^2)(2-2\alpha)} \cdot [R_E^{2-2\alpha} - R_f^{2-2\alpha}] \\ &= \frac{n P^2 G^2 [R_E^{2-2\alpha} - R_f^{2-2\alpha}]}{(R_E^2 - R_f^2)(1-\alpha)} \end{aligned} \quad (17)$$

Using (16) and (17),

$$\theta = \frac{Var[I]}{E[I]} = \frac{n(E[I^2] - E^2[I_i])}{nE[I_i]} = \frac{E[I^2] - E^2[I_i]}{E[I_i]} \quad (18)$$

Now the capacity outage probability will be like [8],

$$\begin{aligned} P_{out} &= E[P_r[C_I \leq C_{th}|I]] \\ &= E[P_r[\log_2 \left(1 + \frac{|h_0|^2 P_0}{I + \sigma^2}\right) \leq C_{th}]] \\ &= E\left[P_r\left(\frac{|h_0|^2 P_0}{I + \sigma^2} \leq (2^{C_{th}} - 1)\right)\right] \end{aligned}$$

$$\begin{aligned} &= E\left[P_r\left(|h_0|^2 \leq \frac{(2^{C_{th}} - 1)(I + \sigma^2)}{P_0}\right)\right] \\ &= E\left[P_r\left(|h_0|^2 \leq \frac{(I + \sigma^2)}{P_r}\right)\right] \end{aligned} \quad (19)$$

; Where  $C_{th}$  is the threshold capacity of a user under interference  $I$  and  $P_r = \frac{P_0}{2^{C_{th}} - 1}$

Assuming the fading channel gain  $|h_0|^2$  has exponential pdf.

$$\begin{aligned} P_{out} &= 1 - e^{-\frac{\sigma^2}{P_r}} \cdot e^{-\frac{I}{P_r}} \\ &= 1 - e^{-\frac{\sigma^2}{P_r}} \int_0^\infty e^{-\frac{i}{P_r}} f_I(i; \theta, k) di; \\ &= 1 - e^{-\frac{\sigma^2}{P_r}} \left(1 + \frac{var[I]}{P_r E[I]}\right)^{-[E[I]/var[I]]} \\ &= 1 - e^{-\frac{\sigma^2}{P_r}} \left(1 + \frac{\theta}{P_r}\right)^{-n} \\ &= 1 - e^{-\frac{\sigma^2}{P_r}} \frac{i^{k-1} e^{-i/\theta}}{\theta^k \Gamma k} \end{aligned} \quad (20)$$

; where  $f_I(i; \theta, k) = \frac{i^{k-1} e^{-i/\theta}}{\theta^k \Gamma k}$  is the Gamma pdf of total interference  $I$ .

We applied above analysis in congestion window based TCP network since TCP is used in wireless network discussed [26-27] even in two-hop wireless network [28]. In TCP under congestion window, initially window size starts from unity then grows exponentially (congestion window of size  $W$  increased by  $1/W$  after getting an acknowledgement) till a threshold level. Beyond threshold the size of congestion window increases linearly (congestion avoidance state) till another threshold and maintain that size until any unavoidable circumstances. In case of time out the congestion window algorithm begins from an initial state (initial size of congestion window size is 1) and grows exponentially till the threshold level which is half of previous threshold level, beyond threshold the algorithm reaches congestion avoidance state and rises linearly towards the ultimate maximum threshold. In case of triple duplicate the algorithm remains in congestion avoidance state but the starting window size is just half of previous level.

Let us consider a duration  $t_d$  between two consecutive triple duplicate which consist of  $R_i$  rounds. Let  $a_{ith}$  packet is lost in  $R_{ith}$  round due to triple duplicate when the size of window is  $W_i$ . Let  $b_i - 1$  packet beyond  $a_i$  is lost due to the triple duplicate therefore including  $a_i$  that  $b_i$  packets have to be transmitted in next triple duplicate period. If the size of the window is  $W_i$  when triple duplicate take place then  $a_i$  has the average position in the middle of  $W_i$  therefore the number of successfully transmitted packet,

$$V_i = a_i + W_i - 1$$

Taking the expected value:

$$E[V] = E[a] + E[W] - 1 \quad (21)$$

If the number of round is  $R_i$  in  $i$ th triple duplicate period then we can also express the number of successfully transmitted packet like [29],

$$V_i = \sum_{k=0}^{R_i-1} \left( \frac{W_{i-1}}{2} + k \right) + b_i$$

Again taking expected value,

$$E[V] = \frac{3E^2[W]}{8} + \frac{E[W]}{4}, \quad E[R] = \frac{E[W]}{2} \quad \text{and}$$

$$E[b] = \frac{E[W]}{2} \quad (22)$$

Here, we consider packet losses occur in any round independently when the received SNR falls below the threshold level. If the packet dropping probability is  $P$ , then the expected number of lost packet,

$$E[a] = \sum_{k=1}^{\infty} k(1-P)^{k-1} P = \frac{1}{P} \quad (23)$$

From equation (21) and (23) we get the expected number of packets transmitted is:

$$E[V] = \frac{1-P}{P} + E[W] \quad (24)$$

From equation (22) and (23) we get the expected value of window size,

$$E[W] = 1 + \sqrt{\frac{8}{3P} - \frac{5}{3}} \quad (25)$$

Since Packet  $a$  is lost in the penultimate (second before last) round, therefore there are a total of  $X+1$  rounds in the triple duplicate period. Let the duration of a round is equal to the  $RTT$  called Round Trip Time explained in [30-31].

So expected duration of the triple duplicate period  $t_d$  is:

$$E[t_d] = RTT \left( \frac{3}{2} + \sqrt{\frac{2}{3P} - \frac{5}{12}} \right) \quad (26)$$

Average sending rate:

$$\bar{x}(p) = \frac{E[V]}{E[t_d]} = \frac{\frac{1}{P} + \sqrt{\frac{8}{3P} - \frac{5}{3}}}{RTT \left( \frac{3}{2} + \sqrt{\frac{2}{3P} - \frac{5}{12}} \right)} \quad (27)$$

According to Amherst model, Triple Duplicate Period ends with timeout. Considering Triple Duplicate (TD) and Time Out (TO) period [32-33], the sending rate can be expressed using the possibility  $Q$  of Amherst model [29]:

$$\bar{x}(p) = \frac{E[V] + Q \cdot E[V^{TO}]}{E[t_d] + Q \cdot E[Z^{TO}]} \quad (28)$$

Where,  $E[V^{TO}]$  is the expected value of the number of packets sent,  $E[Z^{TO}]$  is the expected value of the duration of

a sequence of timeouts and  $Q = 1/n$ , the probability that packet loss results in a time out period. Now,  $E[V^{TO}] = \frac{1}{1-P}$

From equation (24), (25) we get,

$$\bar{x}(p) = \frac{\frac{1-P}{P} + E[W] + \frac{1}{n} \cdot \frac{1}{1-P}}{E[t_d] + \frac{1}{n} \cdot \frac{T_o f(P)}{1-P}} \quad (29)$$

Where, the expected number of timeout is:  $f(P) = 1 + P + 2P^2 + 4P^3 + 8P^4 + \dots$

The eq. (29) now can be expressed as,

$$\bar{x}(p) = \frac{\frac{1-P}{P} + 1 + \sqrt{\frac{8}{3P} - \frac{5}{3}} + \frac{1}{n} \cdot \frac{1}{1-P}}{RTT \left( \frac{3}{2} + \sqrt{\frac{2}{3P} - \frac{5}{12}} \right) + \frac{T_o}{n} \cdot \frac{1+P+2P^2+4P^3+8P^4+\dots}{1-P}} \quad (30)$$

The equations (1)-(20) are related to fading environment of physical layer where as equations (21)-(30) are related to congestion control technique of transport layer. The relation among two set of equation lies in the packet dropping probability which is taken as the probability of SNR falls below the threshold from the first set of equations. The combination of small scale fading of physical layer and end-to-end congestion window control of transport layer is called cross layer model of wireless network. Such concept is available for 4G network in [34-37] and for wireless video transmission in [38]. In next section we will provide the profile of both wireless link and traffic parameters and their relation in different combinations.

### III. RESULTS

For Fig. 2(a) we consider free space path loss exponent  $\alpha = 2.2$ , and other parameters as:  $R_f = 15\text{m}$  and  $N_0 = 10^{-12} \text{ W/Hz}$ . The outage probability increase with increase in density of Femto BS and also increases with increase in threshold SNR at receiving end. Similar analysis is done against threshold link capacity for Nakagami-m and Rayleigh fading channel shown in Fig. 2(b). Here we take the typical values of path loss model as:  $n = 12$ ,  $\sigma = 10^{-5}$ ,  $\beta = 0.001$ ,  $G = 3$ ,  $R_E = 100 \text{ m}$ ,  $P_f = 0.01$  and the path loss exponent is 2. Outage probability is found maximum for Rayleigh fading channel since there is no strong link between BS and MS. For Nakagami-m under  $m = 7$  shows promising result which is found very closed to free space path loss mode because  $m$  strong link between BS and MS.

Next we will depict the profile of different traffic parameters like [39]. Fig. 3(a) and 3(b) depicts the variation of mean window size of TCP congestion window algorithm under congestion avoidance state for free space path loss and fading environment respectively. Under fading environment we are concern about threshold link capacity instead of density of Femto cell since small variation of density of BS under fading environment deteriorates the performance drastically. The mean window size decreases with increase in density of Femto BS and that of threshold link capacity. The mean window size is found larger under Nakagami-m fading case than that of

Rayleigh, shown in fig.3 (b) i.e. the result is consistent with the convention of fading environment of wireless communication like Fig.2 as well.

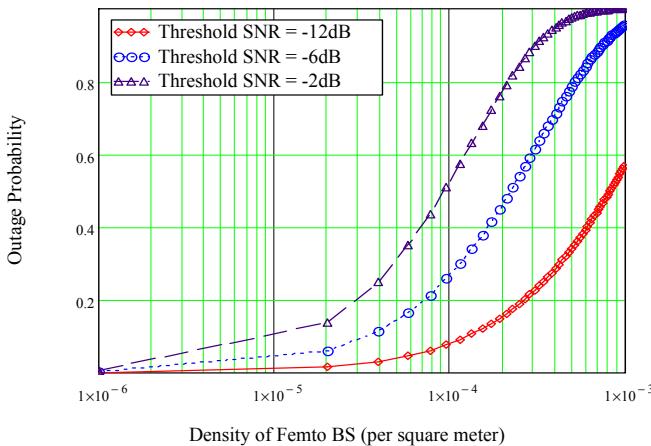


Fig. 2(a) Variation of outage probability against density of Femto BS taking threshold SNR as a parameter

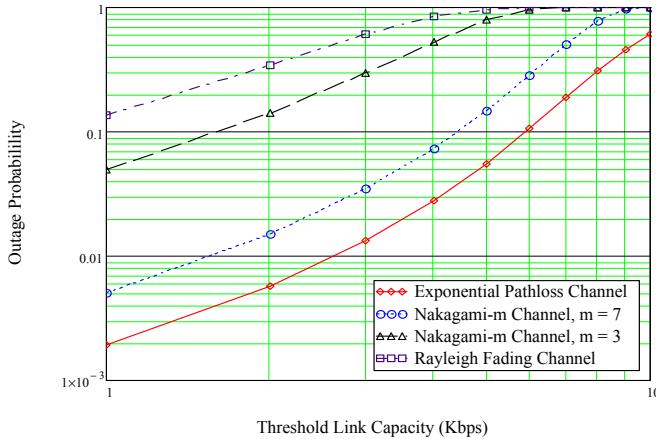


Fig. 2(b) Variation of outage probability against threshold link capacity under fading channel

Similar result is found in Fig. 4(a) and 4(b) where the transmission rate is plotted against the same parameters and reveals the similar performance. Here we consider the parameter of TCP congestion window as: round trip time (*RTT*) 5ms, *TO* = 50ms,  $\bar{n} = 3$ .

Finally throughput of the network under both congestion avoidance (only triple duplicate occurs frequently) and time out (*TO*) is plotted in Fig. 5(a) and 5(b) where the performance is found worse than the Fig.4 of congestion avoidance since *TO* has deteriorate the performance. If the probability of successful transmission of exponentially rising state, probability triple duplicate and probability time out are  $P_s$ ,  $P_d$  and  $P_{to}$  respectively then we can represent the TCP communication system with the state transition chain of Fig.6 using the concept of state transition diagram of [40].

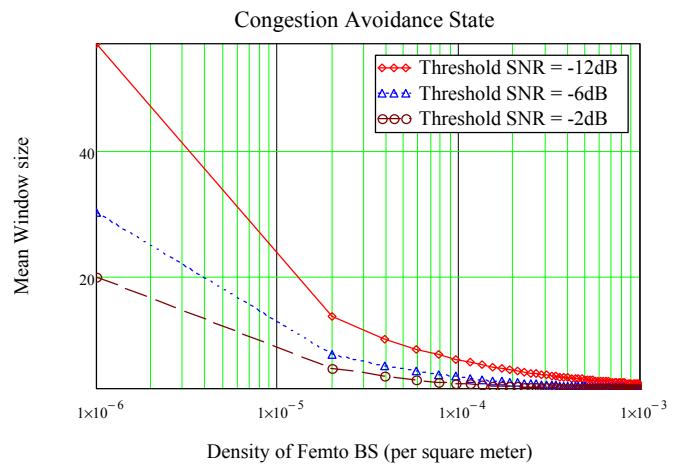


Fig. 3 (a) Variation of mean window size against density of Femto cell under free space path loss environment

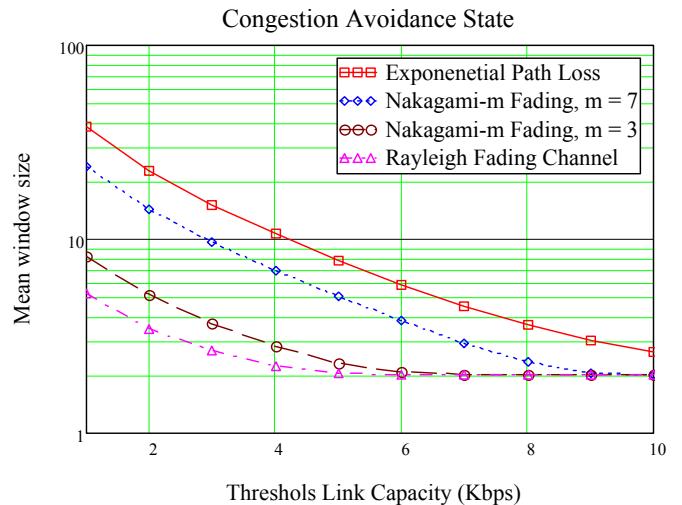


Fig. 3 (b) Variation of mean window size against threshold link capacity under fading channel

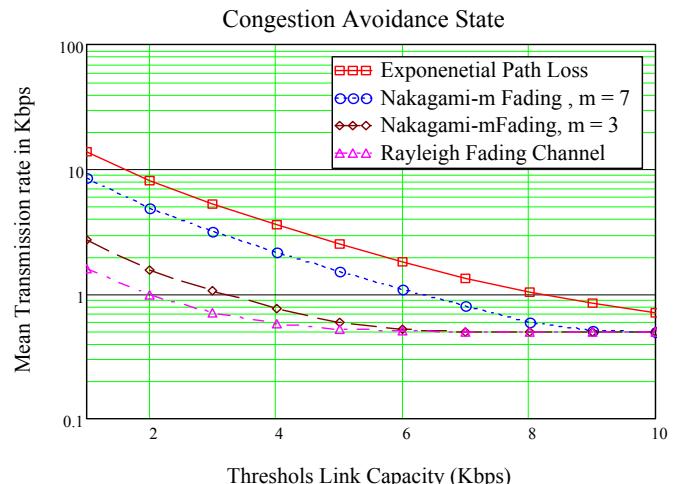


Fig. 4 (a) Variation of transmission rate against threshold link capacity under fading channel

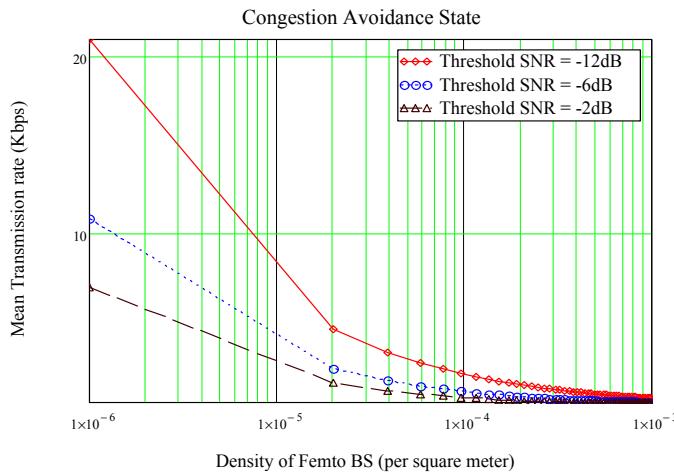


Fig. 4 (b) Variation of transmission rate against density of Femto cell under fading channel

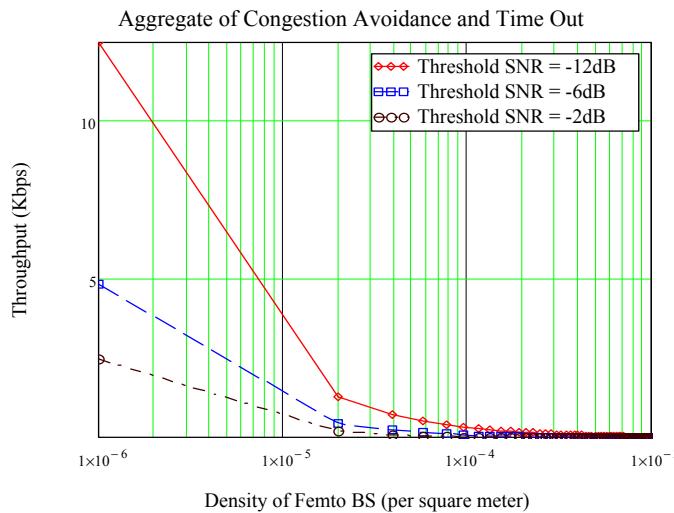


Fig. 5 (a) Variation of throughput against density of Femto cell under free space path loss environment

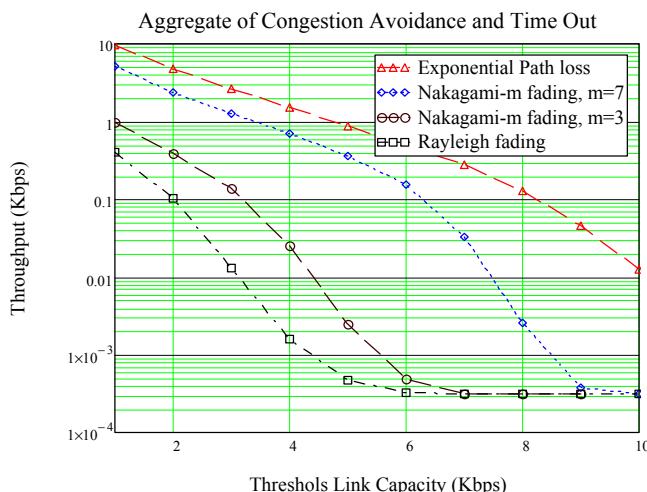


Fig. 5 (b) Variation of throughput against threshold link capacity under fading channel

Applying node equation and normalization of states we get the relation:

$$\begin{bmatrix} \Pi_i \\ \Pi_e \\ \Pi_l \\ \Pi_s \end{bmatrix} = \begin{bmatrix} P_s & -(P_s + P_{td} + P_{to}) & 0 & 0 \\ 0 & P_s & -(P_s + P_{to}) & P_{td} \\ 0 & 0 & P_s & -(P_s + P_{to}) \\ 1 & 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \quad (31)$$

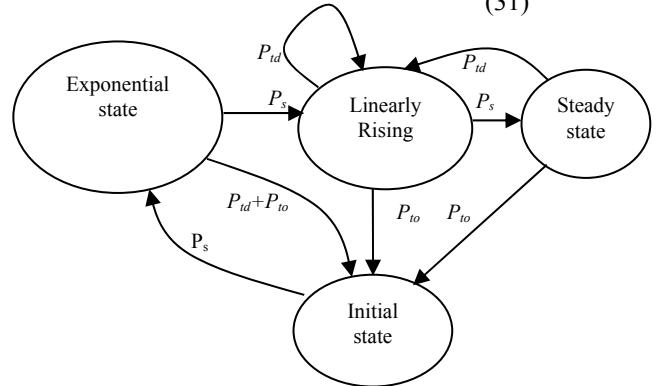


Fig. 6 State transition of congestion window based TCP communication system

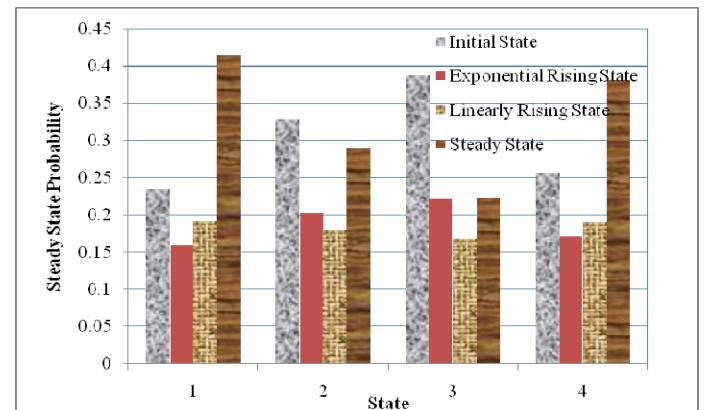


Fig. 7 Steady probability state of congestion window based TCP for 4 different values of  $p$

Taking  $TO = 5\text{ms}$ ,  $P = 0.02, 0.05, 0.08, 0.2$ ; we get the value of  $P_s$ ,  $P_{td}$  and  $P_{to}$  like [41]. Using (8) we evaluate the steady probability of initial state  $\Pi_i$ , exponential rising state  $\Pi_e$ , linearly rising state  $\Pi_l$  and steady state  $\Pi_s$  shown in Fig. 7. Except the linearly rising state the other three states follow the profile of Poisson's pdf but the linearly rising state shows almost steady.

#### IV. CONCLUSIONS

In this paper we integrated the concept of fading environment of Femto cell with the traffic performance of TCP link under congestion window. The impact of fading parameter on TCP performance is analyzed and the result section reveals the relation among outage probability, mean congestion window, mean distribution rate, threshold SNR, threshold link capacity and fading environments. Such analysis is helpful for a network planner to choose the condition of maximum throughput. We also design a Markovian chain to get the steady

probability states of congestion window. Such work can be extended using two dimensional Markov chain. Here we did not consider the impact of equalization and diversity scheme because of short wireless link but above analysis can be included to observe the amount of enhancement of performance of the network.

## REFERENCES

- [1] V. Chandrasekhar and J. G. Andrews, "Femtocell networks: A survey," IEEE Commu. Mag., vol. 46, no. 9, pp. 59–67, Sept. 2008.
- [2] Bouras, C., Kavourgias, G., Kokkinos, V., Papazois, A., "Interference Management in LTE Femtocell Systems Using an Adaptive Frequency Reuse Scheme," Wireless Telecommunications Symposium (WTS), pp. 1 – 7, London, 18-20 April, 2012
- [3] Sachitha Kusaladharma, Chintha Tellambura, "Aggregate Interference Analysis for Underlay Cognitive Radio Networks," Wireless IEEE Communications Letters, vol.1, Issue 6, pp.641 – 644, 2012
- [4] Sachitha Kusaladharma and Chintha Tellambura, "On Approximating the Cognitive Radio Aggregate Interference," IEEE Wireless Communications Letters, vol. 2, no. 1, pp. 58-6, February 2013
- [5] Mahsa Derakhshani, Tho Le-Negoc, "Aggregate interference and Capacity-Outage Analysis in a Cognitive Radio Network," IEEE Transactions On Vehicular Technology, vol. 61, no. 1, pp.196 – 207, Jan 2012
- [6] Risala T. Khan, Tanzilah Noor Shabnam, Md. Imdadul Islam, and M. R. Amin, "Enhancement of Performance of Cognitive Radio Network with Incorporation of MRC Scheme at Secondary Receiver," IACSIT International Journal of Engineering and Technology, vol. 4, no. 4, pp.495-499, August 2012
- [7] Shin-Ming Cheng, Weng Chon Ao, and Kwang-Cheng Chen, "Downlink Capacity of Two-tier Cognitive Femto Networks," 21<sup>st</sup> Annual IEEE international Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1301-1306, 2010
- [8] Mahsa Derakhshani, Tho Le-Negoc and Mai Vu, "Interference and Capacity-Outage Analysis in a Cognitive Radio Network with beacon," 25th Biennial Symposium on Communications (QBSC), pp. 261 – 264, 2010
- [9] Gang Ning, Qinghai Yang, Kyung Sup Kwak, Hanzo, L., "Macro- and Femtocell Interference Mitigation in OFDMA Wireless Systems," Global Communications Conference (GLOBECOM), pp. 5068 – 5073, 3-7 Dec. 2012
- [10] D. Lopez-Perez et al., "Interference Avoidance and Dynamic Frequency Avoidance and Dynamic Frequency Planning for WiMAX Femtocells Networks," International Conference on Communication Systems, NOV. 2008.
- [11] Mostafa Zaman Chowdhury and Yeong Min Jang, "Handover management in high-dense Femtocellular networks," EURASIP Journal on Wireless Communications and Networking, A Springer Open Journal, vol. 2013, no. 6,7, pp. 1-21, January 2013
- [12] MZ Chowdhury, YM Jang, ZJ Haas, "Cost-effective frequency planning for capacity enhancement of Femtocellular networks," Wirel. Personal. Commun.60(1), pp.83–104, 2011
- [13] Claudio Casetti, Mario Gerla, Saverio Mascolo, M.Y. Sanadidi and Ren Wang, "TCP Westwood: End-to-End Congestion Control for Wired/Wireless Networks," Kluwer Academic Publishers, Manufactured in the Netherlands. Wireless Networks 8, pp.467– 479, 2002
- [14] Ghassan A. Abed, Mahamod Ismail and Kasmiran Jumari, "Characterization and observation of (transmission control protocol) TCP-Vegas performance with differentparameters over (Long term evolution) LTE networks," Scientific Research and Essays, vol. 6(9), pp. 2003-2010, 4 May 2011
- [15] JiroKatto, Kazumine Ogura, Yuki Akae, Tomoki Fujikawa, Kazumi Kaneko and Su Zhou, "Simple Model Analysis and Performance Tuning of Hybrid TCP Congestion Control," pp.1-6, New Orleans, LO, Nov. 30 2008-Dec. 4 2008
- [16] Ghassan A. Abed, Mahamod Ismail, KasmiranJumari, "Improvement of TCP Congestion Window over LTE Advanced Networks," International Journal of Advanced Research in Computer and Communication Engineering, vol. 1, Issue 4, , pp. 185-192, June 2012
- [17] Breeson Francis1, Venkat Narasimhan2, Amiya Nayak1, Ivan Stojmenovic, 'Techniques for Enhancing TCP Performance in Wireless Networks,' 32nd International Conference on Distributed Computing Systems Workshops, IEEE Computer Society, pp. 222-230, 2012
- [18] Yung-Chih Chen, Don Towsley, Erich M. Nahum, Richard J. Gibbens, Yeon-sup Lim, "Characterizing 4G and 3G Networks: Supporting Mobility with Multi-Path TCP," UMass Amherst Technical Report: UM-CS-2012-022, [https://people.cs.umass.edu/~yungchih/publication/12\\_mtcp\\_4g\\_technical\\_report.pdf](https://people.cs.umass.edu/~yungchih/publication/12_mtcp_4g_technical_report.pdf)
- [19] Van Jacobson, 'Congestion Avoidance and Control,' Computer Communication Review, ACM SIGCOMM, vo1.18, no. 4, pp.158-173, August 1988
- [20] Md. Shahid Hossain and M. I. Islam, 'A proposed 2-D queuing model of PCT-I traffic,' 6<sup>th</sup> International Conference on Computer and Information Technology (ICCIT)J.U., pp.114-118, Dhaka, Dec'2003
- [21] D. Bear, "Principle of Telecommunication Traffic Engineering," Second Edition, Peter Peregrinus Ltd, 1988
- [22] T. Veerarajan, 'Probability, Statistics and Random Process,' ISBN-0-07-049482-7, chap-9, pp.468-516, Tata McGraw-Hill Publishing Company Limited, 2003
- [23] Sheldon M. Ross, 'Introduction to Probability Models,' 7<sup>th</sup> Edition, ISBN-81-7867-055-0, pp.216-225, Academic Press, A Harcourt Science and Technology Company, San Diego, USA, 2001
- [24] Ali Jemmalai, Mohammad Torabi and Jean Conan, 'Performance Analysis of MIMO Schemes in 3GPP Long Term Evolution System,' Wireless Pers Commun, 82:pp. 1107–1125, April 2015
- [25] Joydev Ghosh, Varsha Panjwani and Sanjay Dhar Roy, 'Outage Impact in Cognitive-Femtocell Deployed Macrocell Network,' Journal of Algorithms, Computer Network, and Security, Vol.1 No.1, pp.25-32, January 2016
- [26] KiWon Sung, Harald Haas and Stephen McLaughlin (EURASIPMember) "A Semi-analytical PDF of Downlink SINR for Femtocell Networks", EURASIP Journal on Wireless Communications and Networking, Volume 2010, Article ID 256370, pp. 1-9
- [27] Ian F. Akyildiz , David M. Gutierrez-Estevez, Elias Chavarria Reyes, "The evolution to 4G cellular systems: LTE-Advanced," Physical Communication 3 (2010), pp.217–244
- [28] Martin Haenggi and Radha Krishna Ganti, "Interference in Large Wireless Networks," Now- Foundations and Trends in Networking, vol. 3, no. 2, 2008, pp. 127–248
- [29] Jae-Hyun Hwang and Yoo, C., 'Formula-based TCP throughput prediction with available bandwidth,' IEEE Communications Letters, vol.14 , no. 4, pp.363 – 365, April 2010
- [30] Jitendra Padhye, Victor Firoiu, Don Towsley, Jim Kurose, 'Modeling TCP throughput: A simple model and its empirical validation,' ACM SIGCOMM Computer Communication Review, Volume:28, Issue:4 , pp. 303-314, 1998
- [31] Qi He, Constantinos Dovrolis, Mostafa Ammar, "On the Predictability of Large Transfer TCP Throughput," Computer Networks, Volume 51, Issue 14, Pages 3959-3977, October 2007
- [32] Prasanthi Sreekumari and Meejeong Lee, 'TCP NRT: a new TCP algorithm for differentiating non-congestion retransmission timeouts over multihop wireless networks,' EURASIP Journal on Wireless Communications and Networking, pp.1-20, 2013, <http://jwcn.eurasipjournals.com/content/2013/1/172>
- [33] P Mi-Young, C Sang-Hwa, Detecting TCP retransmission timeouts nonrelated to congestion in multi-hop wireless networks,' IEICE Transactions on Information and Systems, vol. E93-D, no.12 pp.3331-3343, Dec' 2010
- [34] D. Vinayagam, R. Kurinjimalar, D. Srinivasan, ' Performance Evaluation of Cross Layer QoS scheduling for Long Term

- [35] Evolution Network,' International Journal of Advanced Computer Research, vol.2,no.3, pp.75-83, September-2012  
Ghassan A. Abed and Samir I. Badrawi, 'Augmentation Opportunity of Transmission Control Protocol Performance in Wireless Networks and Cellular Systems,' International Journal of Computer, Electrical, Automation, Control and Information Engineering vol.8, no.5, pp 915-919, 2014
- [36] Bharti Jaglan and Neha Pawar, 'A Review of congestion control variants of TCP over IEEE 802.16 standard networks,' SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE)-EFES, pp.11-16, April 2015, <http://www.internationaljournalsrg.org/IJCSE/2015/Special-Issues/EFES/IJCSE-EFES-P103.pdf>
- [37] Onur Ozturk and Nail Akar, 'A novel queue-aware wireless link adaptation mechanism and its fixed-point analytical model,' EURASIP Journal on Wireless Communications and Networking (2015), 2015:248, pp.1-20, 2015
- [38] Scott Pudlewski, Nan Cen, Zhangyu Guan, and Tommaso Melodia, 'Video Transmission Over Lossy Wireless Networks: A Cross-Layer Perspective,' IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 1, pp.6-22, February 2015
- [39] Raja Murali Prasad and Pentamsetty Satishekhar, 'Joint Routing, Scheduling and Admission Control Protocol for WiMAX Networks,' The International Arab Journal of Information Technology, vol. 10, no. 1, pp.85-94, January 2013
- [40] Peng Li, Yunjian Jia1, Mingjun Feng, Changrong Ye1, Fei Chen, and Huifang Fan, 'A Real-Time Software Defined Radio Platform for LTE-Advanced Heterogeneous Networks,' *Journal of Communications* Vol. 11, No. 3, pp.263-275, March 2016
- [41] John A. Gubner and Kei Hao, "A Computable Formula for the Average Bit Error Probability as a Function of Window Size for the IEEE 802.15.3a UWB Channel Model," IEEE Transactions on Microwave Theory and Techniques, vol. 54, no. 4, pp. 1762-1768, April 2006



**Md. Imdadul Islam** has completed his B.Sc. and M.Sc Engineering in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 1993 and 1998 respectively and has completed his Ph.D degree from the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh in the field of network traffic in 2010. He is now working as a Professor at the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. Previously, he worked as an Assistant Engineer in Sheba Telecom (Pvt.) LTD (A joint venture company between Bangladesh and Malaysia, for Mobile cellular and WLL), from Sept.1994 to July 1996. Dr Islam has a very good field experience in installation and design of mobile cellular network, Radio Base Stations and Switching Centers for both mobile and WLL. His research field is network traffic, wireless communications, wavelet transform, OFDMA, WCDMA, adaptive filter theory, ANFIS and array antenna systems. He has more than hundred and fifty research papers in national and international journals and conference proceedings.



**M. R. Amin** He has received his B.S. and M.S. degrees in Physics from Jahangirnagar University, Dhaka, Bangladesh in 1984 and 1986 respectively and his Ph.D. degree in Plasma Physics from the University of St. Andrews, UK in 1990. He is a Professor of Electronics and Communications Engineering at East West University, Dhaka, Bangladesh. He served as a Post-Doctoral Research Associate in Electrical Engineering at the University of Alberta, Canada, during 1991-1993. He was an Alexander von Humboldt Research Fellow at the Max-Planck Institute for Extraterrestrial Physics at Garching/Munich, Germany, during 1997-1999. Dr. Amin awarded the Commonwealth Postdoctoral Fellowship in 1997. Besides these, he has also received several awards for his research, including the Bangladesh Academy of Science Young Scientist Award for the year 1996 and the University Grants Commission Young Scientist Award for 1996. He is a member of the IEEE. His current research is in the broad field of wireless communications and in the nonlinear plasma physics.



**Jesmin Akhter** received her B.Sc. Engineering degree in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka, Bangladesh in 2004 and M.Sc Engineering degree in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka, Bangladesh in 2012. Since 2008, she is a faculty member having current Designation "Associate Professor" at the Institute of Information Technology in Jahangirnagar University, Savar, Dhaka, Bangladesh. Her research areas are on network traffic, complexity and algorithms and software engineering. Now she is pursuing PhD at the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh in the field of 4G wireless networks.

# Data Mining Approach to Extract the Interdependency among Different Attributes of Cardiac Patients

Rao Muzamal Liaqat, Bilal Mehboob, Nazar Abbas Saqib, Muazzam A Khan

*College of E&ME, National University of Sciences and Technology, Islamabad, Pakistan*

**Abstract--**Nowadays we are surrounding with large data related to patient history, test results and reports. Usually, doctors diagnose the disease on the basis of recommended tests. A final recommendation about patient health may involve a lot of factors including patients test results and doctor experience. In this paper, we will use the data mining approach to extract the dependency among different tests recommended by practitioners, as well as relations of important parameters in cardiac patient's dataset. In this paper, we have used ID3, CHAID, Random Tree, Random Forest, Decision Tree and Decision Stump to extract the interdependency among different attributes in cardiac patients. We have performed the comparative analysis of these algorithms; according to analysis, ID3 give the best result. In this paper we have used the dataset provided by AFIC (Armed Force Institute of Cardiology), our dataset consist of 1500 records along with 36 attributes.

**Keywords:** Data Mining; Cardiac Patients; Supervised Learning; DT (Decision Tree), ID3

## I. Introduction

Data mining approach is used to extract the hidden pattern, relationship, and knowledge from the data which is not possible by using traditional statistical methods of information extraction [1]. Heart failure is leading reason of death from last decades, according to "WHO" more than 500,000 die every year due to heart diseases [2].

It is the common practice that a large of data remains unexplored in hospital due to patient ignorance and unavailability of respective doctors that raises significant problems in healthcare domain. Then certain question arises e.g. "How we can get the useful information from the data, is there any hidden

relation between the data that reveals the some specific pattern to practitioner so that they can take some wise decision". All these questions can be answered by using data mining and machine learning algorithms to indicate the unseen or hidden pattern [3]. Nowadays we are surrounding with a large dataset related to patient history [4]. However the current database of patients is not so informative to extract any useful information or to track the patient diseases [5]. Researchers have used the statistical approach to analyze the medical data. We can extract the useful data by using the different statistical tools, software to analyze the data and extract the useful information [6].

In our work we will use the data mining algorithms which are more reliable as compared to statistical model; we will also compute the performance of different algorithms. Basically, there are two types of algorithms that are commonly used in data mining. One is known as supervised learning algorithms (in supervised learning we have trainee dataset e.g. SVM, Naïve Bayes). Second is known as unsupervised learning (in which we have no trainee dataset or label attribute e.g. K-Mean, DBSCAN). Data Mining plays an important role in heart disease prediction as well as extraction of hidden pattern [5]. In this paper we have used different data mining algorithms such as DT (Decision Tree), Random Forest, DBSCAN, ID3, CHAID and Decision Stump to extract the hidden pattern from the data.

The main focus of this paper is hidden pattern extraction and to find out the dependency level among different attributes that will assist the practitioners to write a wise and better prescription

for heart patients. In this paper we have used different algorithms and compute their performances for comparative analysis and find out the dependency level among different attributes for heart patients.

The remaining paper is divided into 5 sections. Section 2 describes the literature review. Section 3 describes the methodology and algorithm selection is detailed in section 4; performance of results is carried out in section 5. Conclusion and future work is detailed in section 6.

## II. Related Work

In literature, a lot of work has been carried out for medical data analysis to discover the hidden pattern and extraction of useful information from large data by applying data mining techniques [7]. In conventional methods for information extraction from data, Professional's manual method was used, which has no worth when dataset increases in volume as well as in dimension. To deal such data we need some computing technologies [8].

In medical domain most of the work is carried out on cardiac image segmentation, feature extraction, pattern recognition as well as correlation [9, 10]. A Decision tree is a widely used algorithm that is used to mine the hidden information and backtrack the root cause in medical data. In decision tree we have a root node and leaf nodes, leaf nodes represent concrete knowledge according to label attribute. Commonly used decision tree algorithms are ID3, CHAID, Random Forest and Decision Stump which are mostly used for mining the useful information [11]. Many intelligent systems have been developed to assist the practitioners in cardiac diseases [12]. Researchers have used the Naïve Bayes, ANN and decision tree to extract the hidden pattern and correlation among attributes [13].

Our main objectives are to process the data to get the useful information and explored the hidden patterns and interdependency among different factors in cardiac patients. In this paper, we have used the dataset provided by AFIC (Armed force Institute of Cardiology). Preprocessing steps and performance of different unsupervised learning classifiers are described in the methodology section.

### III. Proposed Methodology

Our methodology to extract the interdependency among different attributes of cardiac patient and their impact on heart failure is based on data mining approach that is depicted in fig 1. The model is divided into 6 phases; each phase may involve the certain input, output, and operations. We will explain the each phase in detail.

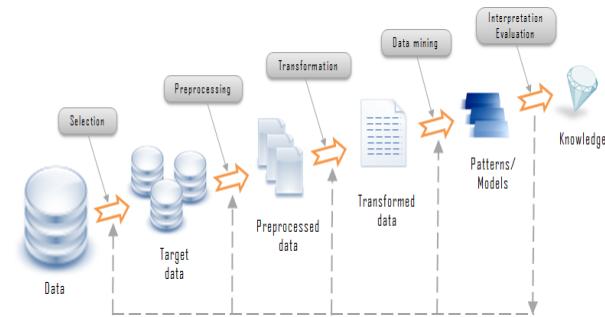


Figure 1: Data Mining Process Model

#### *a. Data Acquisition*

Mostly we have the medical data in the form of medical reports, lab reports and doctor review from all kind of data can be categorized as an unstructured form of data [13]. We get the data in report form from Armed Force Institute of Cardiology (AFIC). Raw data consist of 1500 records with 50 attributes. Then we get the target data from raw data by applying feature selection on the basis of attribute weight and expert opinion.

*b. Target Data (Attribute Selection)*

Target data is actually the data of our interest that is mined from raw data. We have selected the target attributes from raw data by assigning weights to attributes using correlation matrix and the experts' consensus. Correlation operator applied on cardiac patient's data is shown in figure 2.

Attribut...	Patient...	Age	Gender	Protocol	BMI	Knowm_...	Knownw...	Knownw...	FatMet_T...	Angiop...	Zocho...	Zocho...	Zocho...	Zocho...	Zocho...
Patient_ID	1	0.026	-0.132	0.270	-0.151	-0.327	-0.057	-0.118	-0.200	0.121	-0.043	0.027	0.037	-0.131	-0.119
Age	0.025	1	-0.078	-0.250	-0.160	0.065	0.158	0.008	0.189	0.031	0.041	-0.100	0.069	0.117	0.000
Gender	-0.132	-0.078	1	-0.055	0.091	0.043	0.043	-0.102	-0.052	-0.324	-0.012	-0.085	-0.123	-0.081	-0.175
Protocol	0.270	-0.250	-0.055	1	-0.119	-0.166	-0.166	-0.178	-0.237	-0.097	-0.114	-0.210	-0.085	-0.319	0
BMI	-0.151	-0.160	0.091	-0.118	1	0.167	0.080	0.199	-0.080	-0.043	0.122	-0.147	-0.006	-0.069	-0.086
Knownm...	-0.027	0.065	0.043	-0.196	0.107	1	0.242	0.157	-0.037	-0.329	-0.070	0.104	0.045	-0.079	-0.036
Knownw...	-0.057	0.158	0.043	-0.166	0.080	0.242	1	0.432	-0.021	-0.200	0.186	-0.037	0.051	-0.048	-0.031
Knownw...	-0.118	0.008	-0.102	-0.178	0.119	0.167	0.432	1	-0.088	-0.154	-0.100	0.022	0.174	-0.031	0.138
FatMet_T...	0.220	0.189	-0.052	-0.237	-0.180	-0.037	-0.021	-0.088	1	0.076	-0.019	0.251	0.388	0.294	0.244
Angiop...	0.121	0.031	-0.324	0.007	-0.043	-0.329	-0.200	-0.154	0.076	1	-0.003	-0.029	-0.039	0.050	0.155
Zocho...	-0.043	0.041	-0.012	0.186	-0.122	-0.070	0.108	-0.100	0.019	-0.003	1	-0.141	-0.082	0.290	0.063
Zocho...	-0.027	-0.010	-0.095	-0.014	-0.147	0.104	-0.037	0.002	0.021	-0.029	-0.141	1	0.278	0.005	0.459
Zocho...	0.037	0.089	-0.123	-0.210	-0.006	0.045	0.051	0.174	0.388	-0.039	-0.082	0.276	1	0.067	0.293
Zocho...	-0.131	0.117	-0.081	-0.085	-0.069	-0.079	-0.048	-0.031	0.280	0.055	0.290	0.005	0.067	1	0.069
Zocho...	-0.119	0.080	-0.175	-0.319	-0.086	-0.036	-0.031	0.138	0.294	0.156	0.083	0.459	0.263	0.069	1

Figure 2: Correlation Matrix

Now we can see the different values of weight assigned to attributes by using the correlation matrix. Weight values against each attribute are depicted by table 1.

Table 1: Weight Assigned by Correlation Matrix

Attribute	Weight	Attribute	Weight
Patient ID	0.504	Heart_Rate_MA_BP_M	0.443
Gender	0.571	LV_Mayocardium	0.278
Age	0.437	I_LVEF	0.728
BMI	0.685	Report Category	0.725
Known_Disease_1	0.520	Defect Size	0.024
Known_Disease_2	0.516	Defected Area Size	0
Known_Disease_3	0.553	BP BI upper Limb	0.203
Angiography Result	0.494	BP MA upper Limb	0.338
2D_Echo_Result	0.404	BP BI lower Limb	0.359
ECG_Result	0.243	BP MA lower Limb	0.209
Heart_Rate_BI_B	0.356	Affected Area I	0.279
PM			
Defect Segment	0.065	Affected Area 2	0.334
2D Echo Result 1	0.293	Affected Area 3	0.476
FstMI Type	0.302	2D Echo Result 2	0.375

With the help of weight assigned by correlation matrix and expert opinion, we have selected 16 attributes. Now we will extract the hidden pattern among these attributes by using the different data mining algorithms.

#### c. Preprocessed Data

In this step, we make our data compatible with machine learning algorithms by applying some preprocessing steps. Usually, we have missing values in our data to remove these values we apply filtering so that more reliable result can be extracted from the data. We have also converted the numeric data form into polynomial form because the ID3 algorithm doesn't work on such types of data. In the "Report Category", we have Normal, Moderate, Risk and Critical labels which are replaced by numeric values 0, 1, 2 and 3 respectively.

#### d. Transformed Data

Data transformation is carried out by using certain scripts on data, basically, data transformation is related to data preprocessing steps such as data cleansing (in which we make the data smooth by applying some filtering to mitigate the abrupt changes in data). Data reduction is also an important step in data transformation which is used to remove or exclude the certain column

that has redundant behavior or zero effect on the overall result as shown in figure 3.

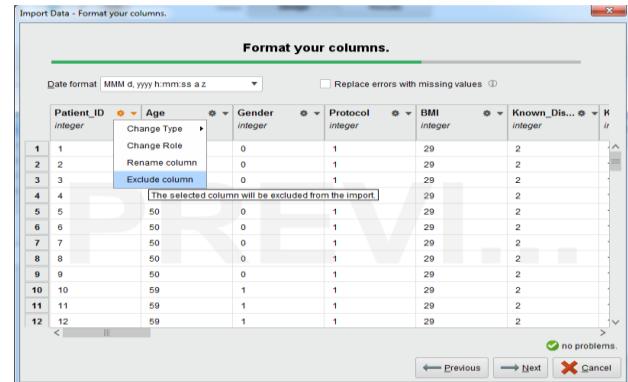


Figure 3: Transform Data to Exclude Column

#### e. Patterns/Models

This phase describes the hidden pattern extracted from data. We will briefly explain the hidden pattern in result and discussion section before that we have to make some assumptions for better understanding and visualization of results. These assumptions are made according to universal standards and expert recommendations.

### IV. Algorithm Selection

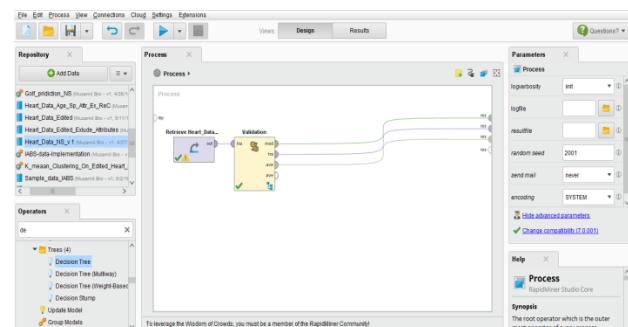


Figure 4: Algorithm Selection

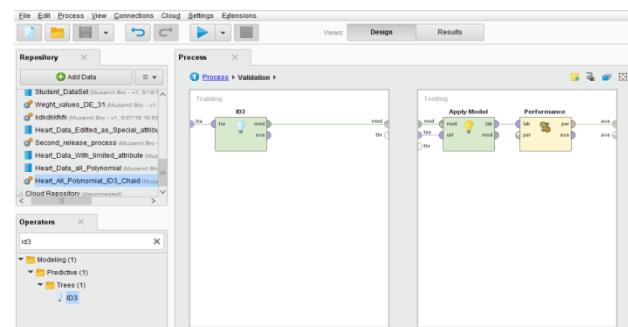


Figure 5: Sub process Of Selected Algorithm

## V. Comparative Analysis

In this paper we have used the six algorithms, we have selected the ID3 algorithm on the basis of Accuracy,

precision and Recall. The names of all algorithms and their obtained results are depicted by table 2.

Table 2: Comparative Analysis of Different Algorithms

Algorithm	Accuracy	Precision Critical	Recall Critical	Precision Risk	Recall Risk	Precision Normal	Recall Normal
Decision Tree	95.80	96.05	94.81	97.56	85.11	97.42	97.93
Decision Stump	61.60	0.00	0.00	0.00	0.00	93.57	67.88
Random Tree	51.20	88.89	10.39	0.00	0.00	49.07	95.34
Random Forest	75.60	98.00	63.64	100.00	10.64	74.61	98.96
ID3	97.60	96.05	94.81	97.83	95.74	97.44	98.45
CHAID	74.40	91.84	58.44	67.86	40.43	69.20	89.64

Dependency Description by using ID3 Algorithm

```

IsDefected = 0
| 2DEchoResult_Part1 = 0
| | Defected_AreaSize = 0
| | | LV_Myocardium = 0
| | | | FstMI_Type = 0
| | | | | 2DEchoResult_Part = 0
| | | | | | I_LVEF = 0
| | | | | | | AffectedArea1 = 0
| | | | | | | | BMI = 22: Normal {Critical=0, Moderate=0, Normal=3, Risk=0}
| | | | | | | | BMI = 25: Normal {Critical=0, Moderate=0, Normal=3, Risk=0}
| | | | | | | | BMI = 26: Moderate {Critical=0, Moderate=8, Normal=0, Risk=0}
| | | | | | | | BMI = 27: Moderate {Critical=0, Moderate=1, Normal=0, Risk=0}
| | | | | | | | BMI = 29: Normal {Critical=0, Moderate=0, Normal=4, Risk=0}
| | | | | | | | BMI = 30: Moderate {Critical=0, Moderate=3, Normal=0, Risk=0}
| | | | | | | | BMI = 31: Normal {Critical=0, Moderate=0, Normal=13, Risk=0}
| | | | | | | | AffectedArea1 = 29: Moderate {Critical=0, Moderate=4, Normal=0, Risk=0}
| | | | | | | | I_LVEF = 27: Moderate {Critical=0, Moderate=2, Normal=0, Risk=0}
| | | | | | | | I_LVEF = 35: Moderate {Critical=0, Moderate=2, Normal=0, Risk=0}
| | | | | | | | I_LVEF = 43: Moderate {Critical=0, Moderate=3, Normal=0, Risk=0}
| | | | | | | | I_LVEF = 49: Moderate {Critical=0, Moderate=8, Normal=0, Risk=0}
| | | | | | | | I_LVEF = 51: Moderate {Critical=0, Moderate=4, Normal=0, Risk=0}
| | | | | | | | I_LVEF = 53
| | | | | | | | | Age = 57: Normal {Critical=0, Moderate=0, Normal=5, Risk=0}
| | | | | | | | | Age = 68: Moderate {Critical=0, Moderate=4, Normal=0, Risk=0}
| | | | | | | | I_LVEF = 54
| | | | | | | | | Age = 46: Moderate {Critical=0, Moderate=3, Normal=0, Risk=0}
| | | | | | | | | Age = 48: Normal {Critical=0, Moderate=0, Normal=1, Risk=0}
| | | | | | | | I_LVEF = 55: Normal {Critical=0, Moderate=0, Normal=17, Risk=0}
| | | | | | | | I_LVEF = 56: Normal {Critical=0, Moderate=0, Normal=5, Risk=0}

```

```
| | | | |     I_LVEF = 57: Critical {Critical=1, Moderate=0, Normal=0, Risk=0}
| | | | |     I_LVEF = 60: Normal {Critical=0, Moderate=0, Normal=2, Risk=0}
| | | | |     I_LVEF = 61: Normal {Critical=0, Moderate=0, Normal=9, Risk=0}
| | | | |     I_LVEF = 63: Normal {Critical=0, Moderate=0, Normal=2, Risk=0}
| | | | |     I_LVEF = 64: Normal {Critical=0, Moderate=0, Normal=5, Risk=0}
| | | | |     I_LVEF = 66: Normal {Critical=0, Moderate=0, Normal=6, Risk=0}
| | | | |     I_LVEF = 67: Normal {Critical=0, Moderate=0, Normal=3, Risk=0}
| | | | |     I_LVEF = 71: Normal {Critical=0, Moderate=0, Normal=39, Risk=0}
| | | | |     I_LVEF = 86: Normal {Critical=0, Moderate=0, Normal=4, Risk=0}
| | | | |     2DEchoResult_Part = 1
| | | | |         Age = 67: Critical {Critical=2, Moderate=0, Normal=0, Risk=0}
| | | | |         Age = 78: Normal {Critical=0, Moderate=0, Normal=2, Risk=0}
| | | | |     2DEchoResult_Part = 11: Moderate {Critical=0, Moderate=3, Normal=0, Risk=0}
| | | | |     2DEchoResult_Part = 45: Normal {Critical=0, Moderate=0, Normal=2, Risk=0}
| | | | |     FstMI_Type = 1: Moderate {Critical=0, Moderate=18, Normal=0, Risk=0}
| | | | |     FstMI_Type = 3: Moderate {Critical=0, Moderate=6, Normal=0, Risk=0}
| | | | |     FstMI_Type = 8: Normal {Critical=0, Moderate=0, Normal=6, Risk=0}
| | | | LV_Myocardium = 1
| | | |     Known_Disease3 = 0
| | | |     RestingECGResult1 = 2
| | | |         Gender = 0: Moderate {Critical=0, Moderate=4, Normal=0, Risk=0}
| | | |         Gender = 1: Normal {Critical=0, Moderate=0, Normal=7, Risk=0}
| | | |     RestingECGResult1 = 3: Normal {Critical=0, Moderate=0, Normal=27, Risk=0}
| | | |     RestingECGResult1 = 4: Moderate {Critical=0, Moderate=3, Normal=0, Risk=0}
| | | |     Known_Disease3 = 3: Critical {Critical=1, Moderate=0, Normal=0, Risk=0}
| | | | LV_Myocardium = 2: Risk {Critical=0, Moderate=0, Normal=0, Risk=5}
| | | | LV_Myocardium = 3: Moderate {Critical=0, Moderate=7, Normal=0, Risk=0}
| | | | LV_Myocardium = 4

| | | | |     Gender = 0
| | | | |         BP-MA_mmHg-uppLim = 100: Normal {Critical=0, Moderate=0, Normal=5, Risk=0}
| | | | |         BP-MA_mmHg-uppLim = 130: Moderate {Critical=0, Moderate=6, Normal=0, Risk=0}
| | | | |     Gender = 1: Risk {Critical=0, Moderate=0, Normal=0, Risk=4}
| | | | |     LV_Myocardium = 6: Critical {Critical=5, Moderate=0, Normal=0, Risk=0}
| | | | |     LV_Myocardium = 7
| | | | |         BP-BI_mmHg-uppLim = 110: Risk {Critical=0, Moderate=0, Normal=0, Risk=4}
| | | | |         BP-BI_mmHg-uppLim = 120: Normal {Critical=0, Moderate=0, Normal=8, Risk=0}
| | | | |         BP-BI_mmHg-uppLim = 131: Risk {Critical=0, Moderate=0, Normal=0, Risk=1}
| | | | |         BP-BI_mmHg-uppLim = 160: Risk {Critical=0, Moderate=0, Normal=0, Risk=3}
| | | | |     LV_Myocardium = 8
| | | | |         Age = 54: Moderate {Critical=0, Moderate=1, Normal=0, Risk=0}
| | | | |         Age = 68: Critical {Critical=4, Moderate=0, Normal=0, Risk=0}
| | | | |     Defected_AreaSize = 1: Normal {Critical=0, Moderate=0, Normal=4, Risk=0}
| | | | |     Defected_AreaSize = 2
| | | | |         2DEchoResult_Diseasel = 0: Critical {Critical=3, Moderate=0, Normal=0, Risk=0}
| | | | |         2DEchoResult_Diseasel = 1: Moderate {Critical=0, Moderate=7, Normal=0, Risk=0}
| | | | |         2DEchoResult_Diseasel = 2: Risk {Critical=0, Moderate=0, Normal=0, Risk=5}
| | | | |     Defected_AreaSize = 3
| | | | |         Age = 61: Moderate {Critical=0, Moderate=2, Normal=0, Risk=0}
| | | | |         Age = 67: Risk {Critical=0, Moderate=0, Normal=0, Risk=3}
| | | | |     Defected_AreaSize = 4: Critical {Critical=6, Moderate=0, Normal=0, Risk=0}
| | | | |     Defected_AreaSize = 5: Critical {Critical=2, Moderate=0, Normal=0, Risk=0}
| | | | |     2DEchoResult_Part1 = 14: Critical {Critical=5, Moderate=0, Normal=0, Risk=0}
| | | | |     2DEchoResult_Part1 = 2
| | | | |         Protocol = 1: Moderate {Critical=0, Moderate=3, Normal=0, Risk=0}
| | | | |         Protocol = 2: Normal {Critical=0, Moderate=0, Normal=10, Risk=0}
| | | | |     2DEchoResult_Part1 = 23: Risk {Critical=0, Moderate=0, Normal=0, Risk=1}
| | | | |     2DEchoResult_Part1 = 3: Critical {Critical=9, Moderate=0, Normal=0, Risk=0}
```

```
| 2DEchoResult_Part1 = 6: Critical {Critical=1, Moderate=0, Normal=0, Risk=0}
| 2DEchoResult_Part1 = 7: Critical {Critical=4, Moderate=0, Normal=0, Risk=0}
IsDefected = 1
| 2DEchoResult = 0
| | Known_Disease3 = 0
| | | Known_Disease2 = 0
| | | | LV_Myocardium = 0
| | | | | 2DEchoResult_Part1 = 0
| | | | | | BP-BI_mmHg-lowLim = 70: Moderate {Critical=0, Moderate=16, Normal=0, Risk=0}
| | | | | | BP-BI_mmHg-lowLim = 80
| | | | | | | Age = 54: Critical {Critical=2, Moderate=0, Normal=0, Risk=0}
| | | | | | | Age = 69: Moderate {Critical=0, Moderate=2, Normal=0, Risk=0}
| | | | | | 2DEchoResult_Part1 = 7: Critical {Critical=2, Moderate=0, Normal=0, Risk=0}
| | | | | LV_Myocardium = 3
| | | | | | Angiography_Result1 = 0
| | | | | | | Age = 48: Moderate {Critical=0, Moderate=2, Normal=0, Risk=0}
| | | | | | | Age = 49: Moderate {Critical=0, Moderate=4, Normal=0, Risk=0}
| | | | | | | Age = 53: Moderate {Critical=0, Moderate=10, Normal=0, Risk=0}
| | | | | | | Age = 56: Moderate {Critical=0, Moderate=6, Normal=0, Risk=0}
| | | | | | | Age = 59: Critical {Critical=2, Moderate=0, Normal=0, Risk=0}
| | | | | | | Angiography_Result1 = 11: Moderate {Critical=0, Moderate=5, Normal=0, Risk=0}
| | | | | | | Angiography_Result1 = 12: Moderate {Critical=0, Moderate=9, Normal=0, Risk=0}
| | | | | | | Angiography_Result1 = 2: Risk {Critical=0, Moderate=0, Normal=0, Risk=4}
| | | | | | | Angiography_Result1 = 3: Moderate {Critical=0, Moderate=2, Normal=0, Risk=0}
| | | | | | | Angiography_Result1 = 4
| | | | | | | | Age = 61: Risk {Critical=0, Moderate=0, Normal=0, Risk=3}
| | | | | | | | Age = 72: Critical {Critical=2, Moderate=0, Normal=0, Risk=0}
| | | | | | | Angiography_Result1 = 5: Moderate {Critical=0, Moderate=4, Normal=0, Risk=0}
| | | | | | | LV_Myocardium = 6: Moderate {Critical=0, Moderate=13, Normal=0, Risk=0}
```

### Graphical View for Interdependency

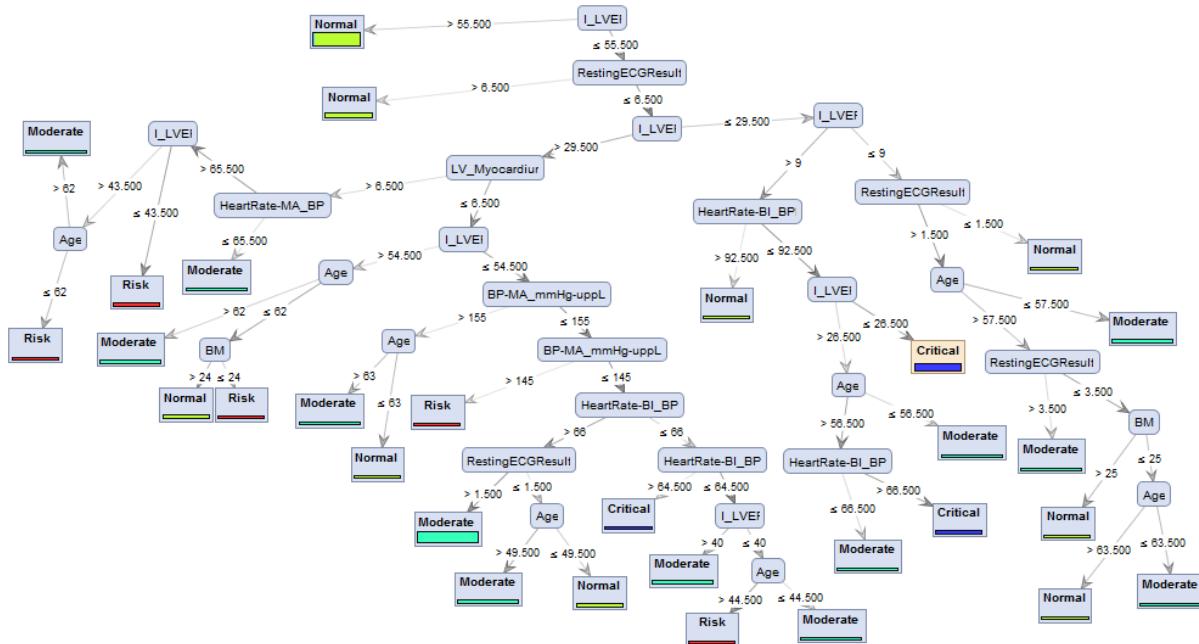


Figure 6: Graphical View for Interdependency

## VI. Conclusion & Future Work

In this we have used six algorithms DT, ID3, Random Forest, Random Tree, Decision Stump and CHAID to extract the dependency level among different test values in cardiac patient's data set. We have selected the ID3 algorithm on the basis of Accuracy, Precision and Recall. ID3 gives the best results (Accuracy 97.60) as shown in comparative analysis table. According to the nature of data we have divided the cardiac patient's data set into four categories Critical, Risk Moderate and Normal patients. Initially we have 36 attributes in our data Set we reduce these attributes into 16 attributes by using correlation Matrix and expert opinion. To understand the dependency level among these reduced attributes we have used the ID3 algorithms to understand the interdependency level among different attributes. In this paper we also indicate the on certain value of different test patient lie in which category (Critical, Risk, Moderate or Normal).In dependency description by using ID3 we also indicate no of objects that lie in Risk, Moderate, Critical and Normal category based on certain numeric value of-

different test. Dependency Description by using ID3 helps us to understand the following points.

- Patients who have  $I\_LVEF > 55$  mostly were categorized Normal.
  - No patient were in Critical or Risk category who ECG Value  $> 6.5$
  - Patients age  $> 56$  and  $Heart\_Rate\_BI\_BPM > 66.5$  were in critical zone.
  - Patients who have  $I\_LVEF < 20$  were categorize in critical zone.
  - Patients Age  $> 42$  And  $I\_LVEF < 40$  categorized as Risk.

This work will help the practitioners to extract the dependency level among different attributes in cardiac patient's data set. Graphical view will also help the doctors to backtrack the collective dependent behavior of different attributes in cardiac patient's data set. In medical domain for identification of certain disease doctor have to see all the parameter that can affect the disease. In future this work will be

foundation to automatically guide and indicate the possible dependency parameters related to cardiac disease. In this way practitioner can cure the patients precisely and wisely.

#### Acknowledgement

I am grateful to AFIC, Pakistan for providing me dataset for research study. I am thankful to my HoD Dr Shoab A Khan for providing me dataset for this work. I am also thankful to Dr Aqib Malik RMO, EME College for assisting me in this research.

#### References

- [1].Han, j. and M. Kamber, Data Mining Concepts and Techniques. 2006: Morgan Kaufmann Publishers.
- [2].Rajkumar, A. and Reena, G.S.: Diagnosis of Heart Disease Using Datamining Algorithm. In: Global Journal of Computer Science and Technology, Vol. 10 (2010).
- [3].Palaniappan, S., Awang, R.: Intelligent Heart Disease Prediction System Using Data Mining Techniques. 978-1-4244-1968-5/08/ ©IEEE (2008)
- [4]Abu Khousa, E.; Campbell, P., "Predictive data mining to support clinical decisions: An overview of heart disease prediction systems," Innovations in Information Technology (IIT), 2012 International Conference on , vol., no., pp.267,272, 2012.
- [5]. Rao, R. B., Krishnan, S., & Niculescu, R. S. (2006), Data mining for improved cardiac care. ACM SIGKDD Explorations Newsletter, 8(1), 3-10.
- [6].Kajabadi, A., Saraee, M. H., & Asgari, S. (2009, October). Data mining cardiovascular risk factors. In Application of Information and Communication Technologies, 2009.AICT 2009. International Conference on (pp. 1-5). IEEE.
- [7]. Giudici, P.: "Applied DataMining: Statistical Methods for Business and Industry", New York: John Wiley, 2003.
- [8]. Wamiq M. Ahmed, (2008) Knowledge representation and data mining for biological imaging, Purdue University Cytometry Laboratories, Bindley Bioscience Center, 1203 W. State Street, West Lafayette, IN 47907, USA
- [9]. J.J. Sychra, D.G. Pavle, E. Olea,(1988) , Classification Images Of Cardiac Wall Motion Abnormalities
- [10]. R. Bharat Rao, Glenn Fung, BalajiKrishnapuram, (2010), Mining Medical Images
- [11] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, USA, 2011.[http://docs.rapidminer.com/files/rapidminer/RapidMiner\\_OperatorReference\\_en.pdf](http://docs.rapidminer.com/files/rapidminer/RapidMiner_OperatorReference_en.pdf)
- [12]. Palaniappan, S. &, Awang, R., "Intelligent heart disease predication system using data mining technique".IJCSNS International Journal of Computer Science and Network Security.Vol. 8, No. 8,2008.
- [13]. Ms. Ishtake S.H , Prof. Sanap S.A., Intelligent Heart Disease Prediction System Using Data Mining Techniques, International J. of Healthcare & Biomedical Research, Volume: 1, pp. 94-101,2013.

# Predicting Student Performance and Risk Analysis by Using Data Mining Approach

Bilal Mehboob<sup>1,2</sup>, Rao Muzamal Liaqat<sup>1</sup>, Nazar Abbas Saqib<sup>1</sup>

<sup>1</sup>Department Of Computer Engineering, College of EME, National University of Sciences and Technology

(NUST), H-12 Islamabad, Pakistan; <sup>2</sup>se@ts global, 54 Orchard, Paragon city Lahore, Pakistan

**Abstract** – Today we are surrounding with large data related to student performance (class participation, attendance, pre student history, quiz result, subject dependency, student CGPA till to final semester). In this paper we will evaluate the reason of student failure basis on the previous data, predict the risk of failure for next course so that students may be mentally prepare for offered course as well dependency level of the course. In engineering it is common practice if a student doesn't know about the basic course he/she can't perform well in advance courses of same scopes. In this paper we will back trace the failure cause with the help of six algorithms. This work will also help out to estimate the risk in early phase, which can help the teachers to design an effective planning for the students who are at risk. We have used the six algorithms for prediction and risk analysis and ID3 algorithm gives the best results as compared to other five algorithms. In this paper we have used the data set of CEME, NUST. Our dataset consists of 450 records extracted from five degrees (DE-29, DE-30, DE-31, DE-32, and DE-33).

**Keywords:** Data Mining, ID3, Risk, Performance Prediction

## I. Introduction

Education plays an important role for the development of a country, especially for underdeveloped countries like Pakistan. However it is important to find out the failure reason of students to improve educational growth as well as gap holes in this domain. Traditionally teachers predict the student performance on the basis of their experience; they had well understood the student nature and temperament. Educational Data Mining is an

emerging field it helps us to explore the knowledge from the stored educational databases. EDM provides us a set of techniques that can be used as a learning experience to improve the quality of education. In this paper we will use data mining approach to predict the student performance as well as identification of failure risk. Data mining techniques can be applied in various fields such as marketing, trades, sales, business, web engineering and real estate etc. [1]. Use of data mining in education domain is rapidly increasing. It helps us in determining student's performance level, time and subject related complexities and constraints. Data mining is also related with knowledge discovery, in which we extract useful information from the data [2]. We have different tools that are used in data mining such as rapid miner, weka, spss etc. We have used the rapid miner studio 7 [3], because it provides us a large no of clustering algorithm and preprocessing operators for clustering. Moreover it is an integrated environment which is widely used in data mining, machine learning, text mining, and business predictive analytics. In data mining two important concepts, called unsupervised learning and supervised learning are used. In unsupervised learning we use clustering mechanism to extract the useful information from the data. Famous algorithm for unsupervised learning are K-NN, K-Mean, DBSCAN, SVM (Support Vector Machine) and Expectation Maximization Clustering (EMC). In supervised learning we have trainee data. The training data consist of set of training examples. Our main focus in this paper is on supervised learning.

Here we will use the decision tree, Random Forest and ID3 algorithms for student performance prediction and risk analysis. Data mining concept in education is known as Educational Data Mining (EDM), it was defined by International society of educational data mining [4], and this educational society also deals with different types of data that is gathered from educational domain. Krina Parmar et.al. has used decision tree and random forest in “Performance Prediction of students using distributing data”[5].

We have used the Decision Tree, Random Tree, Random Forest, CHAID, ID3 and Decision Stump algorithm for prediction of the student performance and to evaluate the failure reason, in addition to the methods used by Krina Parmar. In this paper we have select the ID3 algorithm on the basis of Accuracy, Precession and Recall. We have performed the comparative analysis of these algorithms to evaluate the results.

Rest of the paper is divided into V sections. Related work in this domain is describes in Section II. Problem definition is given in Section III and proposed methodology is described in Section IV. Introduction of used algorithms is given in Section V, Experimental result and evaluation is described in Section VI Comparative analysis of different algorithm is also carried out in Section VI. Conclusion and future work is given in Section VII.

## II. Related Work

Data mining has emerged as an active research in educational domain because we can expose lot of things by using data mining approach. In this way we can analyze the students' performance, faculty performance, difficulty level among different subjects as well as the failure reason among students. According to survey thousands of students are dropped out due to their poor performances in academics [6]. We can extract the strength and weakness of students by using data mining approach [7]. Data mining has been used to extract the hidden pattern and useful knowledge from the data [8]. Remero and Ventura give us an exhaustive overview of using data mining approach by different researcher from 1995 to 2005 for educational data mining [9]. Rayan Baker has described the state of data mining in

educational data [10]. Traditional statistical method have been used to calculate the student performance, these methods doesn't give satisfactory results in student performance prediction based on previous data [11]. Researcher has used the tree based structure to obtain the useful information from data, in tree like structure we have root nodes and leaf nodes information between these nodes is depicted in the form of layers [10]. Clustering algorithm has been used for performance prediction [12]. Dr. S. Hari and J James Manoharan has used the K-Mean clustering to predict the student performance, they have divided the data into different sets of clusters [12]. Dr. V.Shirividya and Keerthana have used both classification and clustering techniques to predict the student performance [13].

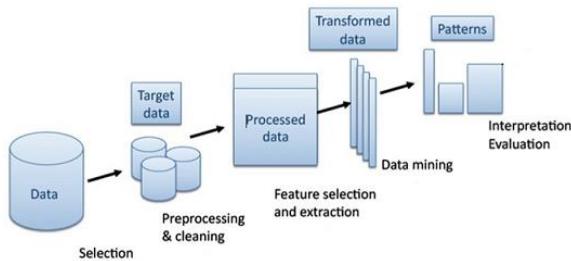
D. A. Carnegie and C. Watterson has used the school record for prediction of student's GPA in first year of engineering [14]. Dekker and M Pechenizkiy have used the data mining approach to find out the dropout information and student performance [15]. H. Bydzovska and Bayer give the concept to use the student data along with social data to predict the student performance more precisely [16]. Dr. Sangeeta and T Mishra has used the Random Tree and J48 algorithms to predict the student performance [17]. Lopez and C Romero introduce the concept of meta classifier for clustering and used the EM (Expectation Maximization) algorithm to measure the academic performance of students [18]. Pallamreddy et al. have used (DT) Decision Tree algorithm on dataset, this algorithms give the tree like model that helps to understand the decisions and consequences based on the nature of data [19].

## III. Problem Definition

In educational domain students drop outs due to their poor academic performances. Failure of student may be due to difficulty level of subject or its dependency on other subject or any additional factors (such as lack of teaching skills and competence, absence of good tools and lab support) that influence the academic performance of the student. In this work we will predict the student performance based on previous grades and will investigate the root causes and relations, if any, among the root causes leading to the poor performance of the students. This prediction

will help the students as well university management to evaluate their success or failure chances on the basis of previous historical performance and adjust their circumstances in the future in order to obtain improved education and research performances.

#### IV. Methodology



**Figure 1: Methodology**

Figure 1 above provides schematic representation of the methodology adopted for the research investigations here. The components of Figure 1 are explained below.

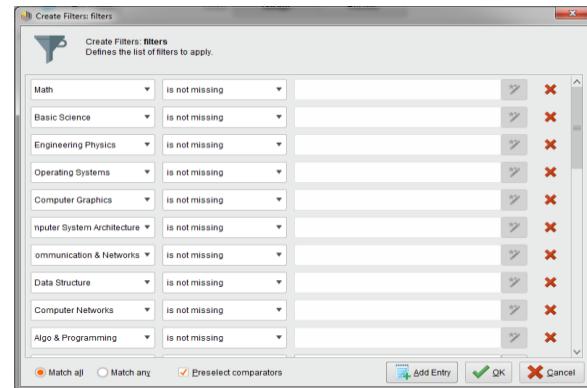
##### A. Data Selection

In this paper we have used the data set provided by College of Electrical and Mechanical Engineering CEME, NUST. Data set consist of 450 records extracted from degree DE\_29, DE\_30,..., DE\_33. We have selected these degrees due to complete access of profile data as well as individual academic records.

##### B. Preprocessing & Cleaning

In this step we convert raw data into machine understandable data by applying some preprocess steps. We have to convert the data according to nature of algorithms e.g. to run ID3 and CHAID we convert the data into polynomial form. To run DT we have to assign the label attribute. In preprocessing of the data we exclude the missing value to make the data compatible with algorithms by taking the average mechanism to reduce the effect of missing values on overall data. We have applied the filtering mechanism to remove the missing values in the data, filters also give us facility to extract the data of our interest by using the built in operators in filters. If a column reveals low information according to label attributes then we can exclude it from data. Filtering

mechanism and data formatting are represented in Figure 2 and Figure 3, respectively.



**Figure 2: Filtering Mechanism**

Student record	Category	Mathematic...	Mathematic...	Mathematic...	Mathematic...	Mathematic...	n
89	polynomial	real	real	real	real	real	2,000
90	2010-NUST-BE	Change Role	3,000	3,500	3,500	2,000	2,500
91	2010-NUST-BE	Rename column	3,000	3,000	2,000	2,500	2,500
92	2010-NUST-BE	Exclude column	2,000	4,000	1,000	2,000	2,000
93	2010-NUST-BE	NS	3,000	3,000	1,000	2,500	2,500
94	2010-NUST-BE	NS	2,500	0,000	1,250	1,250	1,250
95	2010-NUST-BE	NS	4,000	4,000	4,000	3,500	3,500
96	2010-NUST-BE	NS	3,000	3,000	2,000	2,500	2,500
97	2010-NUST-BE	NS	3,500	3,500	3,000	3,600	3,600
98	2010-NUST-BE	NS	3,500	3,500	3,500	3,500	3,500
99	2010-NUST-BE	NS	2,000	3,500	3,500	3,000	3,000
100	2010-NUST-BE	NS	2,500	3,500	2,500	3,000	3,000

**Figure 3: Data Formatting**

##### C. Feature Extraction

Feature extraction is an important step in which we select the most important attributes that have the direct effect on label attribute. There are different methods that are used for feature extraction. We have used the entropy, IG (information Gain), reducts and core to extract the useful features. In this paper we will use the correlation matrix to find out the most important attribute. In Figure 4, which represents correlation matrix, each attribute is assigned a numeric value known as weight. Minimum value of assigned weight is "0" and maximum value is "1". High value of weight reveals the high importance of attribute and vice versa. On the basis of assigned weights we can apply the threshold value given by related expert to reduce the attributes. Weight assigned to different attributes by using correlation attributes are shown by Table 1.

Attribute	Student	Category	Mathem_1	Mathem_2	Mathem_3	Mathem_4	Mathem_5	Mathem_6	Mathem_7	Mathem_8	Math	Applied	Engines.	Engines.	Basic_S.	Pakista_1	Interne_1	S
Student..	1	-0.049	0.022	-0.247	0.093	-0.052	0.049	-0.071	-0.043	0.050	-0.099	0.136	0.628	0.079	0.057	0.154	0.001	0.000
Category	-0.040	1	-0.064	0.003	-0.149	-0.023	-0.087	-0.009	-0.094	0.053	0.032	-0.179	-0.037	0.102	-0.022	0.000	0.000	0.000
Mathem_1	-0.022	-0.094	1	0.499	0.530	0.569	0.519	0.520	0.750	0.415	0.536	0.415	0.560	0.251	0.362	0.000	0.000	0.000
Mathem_2	-0.247	-0.083	0.499	1	0.544	0.645	0.493	0.589	0.796	0.415	0.551	0.459	0.641	0.169	0.342	0.000	0.000	0.000
Mathem_3	-0.093	-0.148	0.530	0.544	1	0.600	0.479	0.518	0.767	0.332	0.450	0.517	0.544	0.175	0.280	0.000	0.000	0.000
Mathem_4	-0.052	-0.023	0.569	0.645	0.600	1	0.613	0.857	0.607	0.412	0.519	0.473	0.623	0.200	0.483	0.000	0.000	0.000
Mathem_5	0.048	-0.057	0.519	0.493	0.476	0.613	1	0.823	0.771	0.434	0.530	0.407	0.607	0.279	0.389	0.000	0.000	0.000
Mathem_6	-0.071	-0.008	0.520	0.589	0.518	0.857	0.623	1	0.800	0.393	0.453	0.433	0.510	0.122	0.319	0.000	0.000	0.000
Math	-0.143	-0.084	0.759	0.776	0.788	0.858	0.771	0.809	1	0.489	0.985	0.583	0.750	0.281	0.445	0.000	0.000	0.000
Applied_c.	0.098	0.063	0.415	0.415	0.332	0.412	0.434	0.351	0.409	1	0.532	0.416	0.770	0.388	0.332	0.000	0.000	0.000
Engineer_c.	-0.089	0.032	0.508	0.851	0.450	0.554	0.530	0.453	0.685	0.532	1	0.452	0.840	0.295	0.330	0.000	0.000	0.000
Engineer_s.	0.138	-0.179	0.415	0.409	0.517	0.471	0.467	0.403	0.583	0.410	0.452	1	0.775	0.101	0.323	0.000	0.000	0.000
Pakista_1	0.008	-0.037	0.569	0.540	0.546	0.923	0.907	0.519	0.750	0.779	0.948	0.775	1	0.279	0.948	0.000	0.000	0.000
Pakistan Studies	0.079	0.102	0.251	0.186	0.175	0.280	0.279	0.212	0.281	0.368	0.200	0.101	0.270	1	0.029	0.000	0.000	0.000
Score	0.057	-0.022	0.382	0.342	0.280	0.463	0.399	0.319	0.445	0.332	0.338	0.325	0.408	0.429	1	0.000	0.000	0.000

Figure 4: Correlation Matrix

Table I: Weights by Using Correlation

Attribute	Weight	Attribute	Weight
Category	1	Database Engineering	0.158
Mathematics_1	0.256	OOP	0.172
Mathematics_2	0.344	Algorithm and computing	0.329
Mathematics_3	0.318	Data Structure	0.298
Mathematics_4	0.228	PL&E	0.247
Mathematics_5	0.242	Computer Networks	0.156
Engineering Mechanics	0.215	Mobile networks	0.223
Pakistan Studies	0.467	Network Analysis	0.226
Digital system Design	0.121	Digital communication	0.064
Electronic Circuit	0.242	Computer Aided drawing	0.546
Control System	0.139	Design project	0.390
AI	0.125	Metric/O level	0.867
Engineering Economics	0.288	FSC/A level	0.886

Software Engineering	0.171	Games/Activities	0.878
----------------------	-------	------------------	-------

#### D. Data Mining

Now we will apply the DT (decision tree), Random forest, Random tree, ID3, CHAID and decision Stump to extract the useful knowledge from the data. To get the better understanding from data we have mapped the CGPA of 8<sup>th</sup> semester in “Student performance prediction Label”. We have calculated the performance of each algorithm based on Accuracy, Precision and Recall. The definitions of the Accuracy, Precision and Recall are given below and comparative analysis of each algorithm is shown in Table II.

#### E. Accuracy

Accuracy is predictive measure of a model. It defines how many times model is correct when we apply data. It can be calculated by using following formula. Where, TP represents the “True Positive”, FP represents “False Positive”, FN represents “False Negative” and TN represents “True Negative” respectively.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \quad (1)$$

#### F. Precision

It is also called Positive Predictive Value (PPV) which can be calculated by using following formula.

$$\text{PPV} = \text{TP} / (\text{TP} + \text{FP}) \quad (2)$$

#### G. Recall

Recall is also known as sensitivity or True Positive Rate (TPR). It can be calculated by using following formula.

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

## V. Introduction of used Algorithms

### A. Decision Tree

A decision tree is a widely used algorithm, which is based on root node and leaf nodes. In decision tree root node is decided based on information gain. We

calculate the information gain of each attribute that is based on entropy. Attribute that has the maximum value of information gain is selected as a root node. Decision tree give the easy interpretation of results in graphical tree like structure [9].

#### B. Random Forest

As the name indicate “Random Forest” algorithm creates the random sets of trees. Value of random set of trees depends upon the random tree operator. These random set of trees make a model. Each model in random forest has a specified no of trees decided by tree operator. Random Forest may consist of many models; a model is selected on the basis of voting mechanism [14].

#### C. Random Tree

Leo Breiman et al introduced the concept of random tree. Working of random tree is same as Decision tree; in random tree we have a random subset of attributes to deals with limitation of decision tree. Value of random subset is based on operator, in this way we can solve the classification as well as regression problem [5].

#### D. ID3

ID3 algorithm was invented by Ross Quinlan, basically, it is extended form of C4.5 algorithms with some modifications. ID3 construct a decision tree by using exemplary data. Then we categorized the remaining samples on the basis of this tree. Data set may contain the large no attributes and instances. ID3 build the tree by using the trainee dataset, remaining instances or samples are categorized in yes or no category based on trainee dataset. In ID3 algorithm class name is represented by leaf node whereas a non-leaf node represents the decision node. ID3 is widely used algorithm in data mining for useful information extraction [3].

#### E. CHAID

Working of “CHAID” is similar as Decision Tree, in decision tree we use the information gain or gain ratio to select the root attribute. Whereas in CHAID we use the chi squared based criterion to select the root attribute. CHAID algorithm can't work on numeric data type for that we have to convert it into different form [3].

#### F. Decision Stump

Decision Stump algorithm is used for build a decision tree, generation of decision tree in by using this operator based on single split mechanism. Unseen examples are classified by using this tree. Efficiency of this this algorithm is depends on AdaBoost operator. Dataset may contain multiples attributes and instances. Each instances/example set is classified in yes or no category based on generated tree. In Decision Stump class name is represented by using leaf nodes whereas a non-leaf node represents the decision node in example set. [3].

### VI. Results & Interpretation

We have used Studio 7 to analyze our data in reference to the six algorithms explained above and the results are explained below. Each result is extracted from the ID3 algorithm description. We have divided the student result into four categories on the basis of CGPA of 8<sup>th</sup> semester. If the CGPA was  $\geq 3.0$  we have categorized them as “Above Average”, CGPA (2.3 to 2.5) named as “Below Average”, CGPA (2.6 to 2.9) labeled as “Average” and students who have CGPA  $\leq 2.2$  classified as “Risk”. A descriptive output of ID3 helps us to extract following results from the data which is shown in figure 5 & 6 respectively.

- Students who were at Risk in Mathematics\_1 were also at Risk in Mathematic\_2.
- Students who were at Average in Digital Image processing were also at Average in Digital Signal Processing.
- Students who were above average in DBMS, they performed well in Design Projects.
- Students who have the status of “Hafiz e Quran” performed well as compared to others.
- Students who score low grades in DBMS, they also got low grades in Database Engineering.
- Mostly students were above average in Microprocessor Based Design, those were above average in computer Architecture.
- Students who were below average or at risk in Math\_1, they got low grades in Math\_5.

**Table II: Comparative Analysis of used Algorithms**

Algorithm	Accuracy	Average		Risk		Below Average		Above Average	
		P	R	P	R	P	R	P	R
Decision Tree	55.52	42.86	5.17	58.82	50.00	25.00	15.62	59.20	59.20
Random Tree	54.11	39.13	15.52	75.00	15.00	30.00	18.75	58.48	58.48
Random Forest	61.97	40.00	12.07	72.22	60.00	36.36	28.12	65.85	65.85
ID3	79.23	78.23	87.00	88.00	82.00	91.21	85.00	93.50	93.50
CHAID	49.50	0.00	0.00	0.00	0.00	0.00	0.00	49.54	49.54
Decision Stump	50.95	0.00	0.00	50.00	6.00	0.00	0.00	50.94	50.94

**Keyword: Precession P, Recall R**

```

Digital Image Processing = 1
| Category = NS: Risk {Average=0, Above Average=0, Risk=2, Below Average=0}
| Category = PC: Below Average {Average=0, Above Average=0, Risk=0, Below Average=1}
Digital Image Processing = 1.3: Below Average {Average=0, Above Average=0, Risk=0, Below Average=1}
Digital Image Processing = 1.5
| Pakistan Studies = 2.5: Below Average {Average=0, Above Average=0, Risk=0, Below Average=2}
| Pakistan Studies = 3: Risk {Average=0, Above Average=0, Risk=2, Below Average=0}
| Pakistan Studies = 3.5: Average {Average=1, Above Average=0, Risk=0, Below Average=0}
Digital Image Processing = 1.6: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
Digital Image Processing = 2
| Gender = female: Above Average {Average=0, Above Average=3, Risk=0, Below Average=0}
| Gender = male
| | Design Projects = 3: Risk {Average=0, Above Average=0, Risk=2, Below Average=0}
| | Design Projects = 3.5
| | | Category = EC
| | | | Mathematics-2 = 1: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
| | | | Mathematics-2 = 2.5: Below Average {Average=0, Above Average=0, Risk=0, Below Average=1}
| | | | Category = NS: Below Average {Average=0, Above Average=0, Risk=0, Below Average=5}
| | | | Category = PC: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
| | | Design Projects = 4: Average {Average=1, Above Average=0, Risk=0, Below Average=0}
Digital Image Processing = 2.1
| Mathematics-2 = 2: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
| Mathematics-2 = 3: Below Average {Average=0, Above Average=0, Risk=0, Below Average=1}
Digital Image Processing = 2.25: Below Average {Average=0, Above Average=0, Risk=0, Below Average=1}
Digital Image Processing = 2.3: Below Average {Average=0, Above Average=0, Risk=0, Below Average=3}
Digital Image Processing = 2.5
| planning Engineering/Project Management = 0: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
| planning Engineering/Project Management = 1: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}

```

**Figure 5: Image of results of ID3 Algorithm (1)**

```

| planning Engineering/Project Management = 2
| | Algorithms and Computing = 2.5
| | | Category = EC: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
| | | Category = NS: Below Average {Average=0, Above Average=0, Risk=0, Below Average=1}
| | Algorithms and Computing = 3: Average {Average=3, Above Average=0, Risk=0, Below Average=0}
| planning Engineering/Project Management = 2.3: Risk {Average=0, Above Average=0, Risk=2, Below Average=0}
| planning Engineering/Project Management = 2.5
| | Gender = female: Average {Average=1, Above Average=0, Risk=0, Below Average=0}
| | Gender = male
| | | Professional Ethics = 2: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
| | | Professional Ethics = 2.5: Below Average {Average=0, Above Average=0, Risk=0, Below Average=5}
| | | Professional Ethics = 3: Below Average {Average=0, Above Average=0, Risk=0, Below Average=3}
| planning Engineering/Project Management = 2.6: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
| planning Engineering/Project Management = 3
| | Gender = female: Risk {Average=0, Above Average=0, Risk=1, Below Average=0}
| | Gender = male
| | | Matric/o-level = O level: Below Average {Average=0, Above Average=0, Risk=0, Below Average=1}
| | | Matric/o-level = matric
| | | | Logic and Sequential Circuit design = 2.5: Average {Average=2, Above Average=0, Risk=0, Below Average=0}
| | | | Logic and Sequential Circuit design = 3: Average {Average=6, Above Average=0, Risk=0, Below Average=0}
| | | | Logic and Sequential Circuit design = 3.5: Average {Average=1, Above Average=0, Risk=0, Below Average=0}
| | | | Logic and Sequential Circuit design = 4: Above Average {Average=0, Above Average=1, Risk=0, Below Average=0}
| planning Engineering/Project Management = 3.5: Average {Average=1, Above Average=0, Risk=0, Below Average=0}
| planning Engineering/Project Management = 4: Above Average {Average=0, Above Average=1, Risk=0, Below Average=0}
Digital Image Processing = 2.6: Average {Average=1, Above Average=0, Risk=0, Below Average=0}
Digital Image Processing = 2.8
| Mathematics-3 = 2.5: Below Average {Average=0, Above Average=0, Risk=0, Below Average=2}
| Mathematics-3 = 3.5: Average {Average=1, Above Average=0, Risk=0, Below Average=0}

```

**Figure 6: Image of results of ID3 Algorithm (2)**

#### A Probable Cause of Risk

In this work, through data mining and data analysis, we have identified subject complexity, lack of student attention and subject's dependency as the causes of student risk and failure. For example student who performs well in Math\_1 and Math\_2 also got good grades in Math\_3 and Math\_4 and vice versa. Students who pay more attention in basic courses got good grades in advance courses of that domain e.g. students who perform well in Digital Image Processing (basic course) were outstanding in Digital Signal Processing (advance course). As a result student should pay more attention in basic of subjects in this way they can reduce the complexity of subjects and ensure the success in advance courses of same domain. It is to mention that we have not found deficiencies in the skills and teaching competencies of the teachers and in the provision of basic infrastructure (such as lab, tools, state-of-the-art presentation and experiment techniques) that can be regarded as contributing cause for the low student performance.

#### VII. Conclusion & Future Work

In this paper, data mining approaches for predicting the performance of a student has been discussed. We have divided the students into four categories on the basis of their CGPA (Average, Risk, below Average and Above Average). Students who were at a certain level in previous semester they were at same level in advance course of same domain. In result we can predict students who will be at risk in coming semesters. This prediction is useful in introducing and then focusing specific strategies, which could possibly reduce the future risks associated to the low student performance or failure. In this work we also find out the subject dependency in term of student performance. We have used the six algorithms; results obtained by using each algorithm are mentioned in comparative analysis table. ID3 gives the best results as compare to other on the basis of accuracy and precision for performance prediction and risk analysis. This work will help the students to reduce the risk and failure chances by emphasizing student attention level in basic subjects and reduce the complexity level of students for next advance

courses being offered. In this way we can predict the departmental growth based on student performance.

In future we will design a tool that will predict the next semester subject's grades and Risk Analysis on the base of current result.

### Acknowledgement

Authors greatly acknowledge the reviews of Dr. Qamar Mahboob of Siemens AG and Prof. Dr. Ashiq Anjum of Derby University.

Funding from se@ts global, Paragon city Lahore, Pakistan is acknowledged.

### References

- [1] Mehmed Kantardzic "Data Mining: Concepts, Models, Methods, and Algorithms" John Wiley & Sons, 05-Jan-2011.
- [2] Ryan S.J.d and Baker, "The state of educational data mining "in 2009: A review and future visions.
- [3] <https://my.rapidminer.com/nexus/account/index.html#download>
- [4] Educational Data Mining Society, Available: <http://www.educationaldatamining.org/>
- [5] Krina Parmar, "Performance prediction of students using distributed Data mining" Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015
- [6]. L. D. Hammond, M. B. Zielezinski, and S. Goldman, "Using technology to support at-risk students' learning," in Alliance for Excellent Education, Stanford Center for Opportunity Policy in Education, 2014. [Online]. Available:<https://edpolicy.stanford.edu/sites/default/files/scope-pub-using-technologyreport.pdf>
- [7]. "Personal learning plan," The Glossary of Educational Reform, Great Schools Partnership, Portland, ME, USA. [Online]. Available: <http://edglossary.org/personal-learning-plan/>
- [8] Fayyad U., Piatetsky-Shapiro G., and Smyth P., From Data Mining to Knowledge Discovery in Databases, AI Magazine, Fall 1996.
- [9] C. Romero and S. Ventura, "Educational data mining: a survey from 1995 to 2005," Expert Systems with Applications, no. 33, pp. 135–146, 2007
- [10] R.S.J.D Baker and K.Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions" , Journal of Educational Data Mining, 1, Vol 1, No 1, 2009.
- [11] Ashkan Sharabiani, Fazle Karim, Anooshiravan Sharabiani, Mariya Atanasov, Houshang Darabi," An Enhanced Bayesian Network Model for Prediction of Students' Academic Performance in Engineering Programs", 2014 IEEE Global Engineering Education Conference,pp 832-837.
- [12] Jamesmanoharan, J.Ganesh, S.H, Felciah, M.L.P.Shafreenbanu, A.K." Discovering Students' Academic Performance Based on GPA Using K-Means Clustering Algorithm" Computing and Communication Technologies (WCCCT), 2014 World Congress on Feb. 27 2014-March 1 2014,pp. 200 – 202.
- [13] G.Keerthana , Dr. V.Sridhya , "Performance Enhancement of Classifiers using Integration of Clustering and Classification Techniques "in International Journal of Computer Science Engineering (IJCSE) , 03 May 2014,pp.200-203.
- [14]. D. A. Carnegie, C. Watterson, P. Andreae, and W. N. Browne, "Prediction of success in engineering study," in 2012 IEEE Global Engineering Education Conference (EDUCON), 2012, pp. 1–9.
- [15]. G. W. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers. (2009). "Predicting Students Drop Out: A Case Study," in Proceedings of the 2nd International Conference on Educational Data Mining, Cordoba, Spain, vol. 9, pp. 41–50.
- [16]. J. Bayer, H. Bydzovská, J. Géryk, T. Obsivac, and L. Popelinský. (2012). "Predicting drop-out from social behaviour of students," in Proceedings of the 5th International Conference on Educational Data Mining-EDM 2012, Chania, Greece, pp. 103–109.
- [17] Piatetsky-Shapiro, Gregory (1991), Discovery, analysis, and presentation of strong rules, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA.
- [18] M.I. Lopez, C. Romero, S. Ventura, and J.M. Luna, "Classification via clustering for predicting final marks starting from the student participation in Forums", ;in Proc. EDM, 2012, pp.148-151.
- [19] Pallamreddy.venkatasubbareddy, Vuda Sreenivasarao," The Result Oriented Process for Students Based On Distributed Data Mining", International Journal of Advanced Computer Science and Applications, Vol. 1, No. 5, November 2010,pp.22-25.

# Android-Based Health Care Management System

Fazal Masud Kundi

Institute of Computing and Information Technology  
Gomal University, D.I.Khan, Pakistan

Ammara Habib

Institute of Computing and Information Technology  
Gomal University, D.I. Khan, Pakistan

Anam Habib

Institute of Computing and Information Technology  
Gomal University, D.I. Khan, Pakistan

Muhammad Zubair Asghar

Institute of Computing and Information Technology  
Gomal University, D.I. Khan, Pakistan

**Abstract—Objective:** The primary goal of this study is to develop an android-based healthcare application, which can assist the users to monitor their health-related conditions for improving their health. **Methods:** The application is developed using android operating system environment. A Visual block programming language, namely MIT App Inventor is used to develop the system. The modification is presented as: (1) integration of different modules and their offline usage, (2) history facility, (3) user friendly. The qualitative method is used to study the objective. **Findings:** The research paper depicts a brief study of existing systems and the new development that has made in the application and also it is better in the manner that it works as a guide to control risk factors. The descriptive analysis point outs that the application is effective to deal with health related issue. **Applications/Improvement:** Integration of modules is performed on the android platform of different applications that are located on different websites, the storage facility is added by using Tiny DB, guidance in the form of charts and text is provided to the users. Such features are not provided in the previous work.

**Keywords—**Health Care; App Inventor; Android; Diabetes; Target Heart Rate.

## I. INTRODUCTION

The Expert System (ES), namely Computer Assisted diagnosis for red eye (CARDE) is proposed by [1], which assists the patients in the treatment of Red eye disease. It works like an ophthalmologist and it is not limited to only red eye diseases, but can be extended to diagnose other diseases.

A Web based expert system is proposed by [2]to diagnose red eye disease and to provide prescription with it. This system typically diagnoses disease, of the eye in which red eye is a common symptom. It has an attractive and easy to use graphical user interface.

[3] Proposed an ES to diagnose skin diseases. This system can diagnose almost 13 types of skin disease. This web based expert system can be enhanced to diagnose all types of skin diseases.

An automated alarm ringing system is developed and its center of interest is the interaction between doctor and patients. The description of medicines, date and time can be set by patients through an alarm. They received the notification through an email or messages [4].

There is a persistent disease known as diabetes mellitus, increasing globally that is caused due to the relative deficiency of insulin. Therefore, android based diabetes management health care application is developed, which helps in diagnosis and treatment of diabetes as well hypertension [5].

Chronic health patients suffer from multiple ailments, however, different patients have different such ailments. The objective of this project was to design and prototype a health monitoring system that has a capability to monitors multiple diseases [6].

An application, namely “smart carb”, is developed based on an Android OS for the management of Type2 diabetes. If the patients do not manage their diabetic level then it will lead to many complications; and if it is not treated properly, that it may even lead to death. In order to manage diabetes to avoid these complication, this application was developed [7].

Due to the refinement of wireless mobile technologies in the erstwhile years, the need for mobile data services has been aggravated dramatically. The location of the user can also be obtained to provide better facilities to the users by the service provider. However, it also has some issues like needing an approval of user privacy, standardization, and accessibility of smart services [8].

There are many systems on the health-related content analysis in the context of opinion mining and sentiment analysis [9, 10, 11,12], however, most of such studies are web-based and address the user generated contents. In addition to aforementioned studies, there are recent works [13, 14, 15, 16, 17, 18, 19, 20] performed for developing healthcare applications, which assist the users in taking care of their health. In this work, we present the development of an android-based health care application using the MIT App inventor software [21]. Nowadays, health related issues are getting common due to hectic daily routine and unbalanced diet. Therefore, it is an important task and a need to develop an android application that could assist the users to keep themselves aware of their daily activities including diet, exercise, and glucose level, B.P reading etc.

We have integrated different modules into one android application that were located on different websites such as calorie level, Target heart rate, blood volume, diabetes [22].

We also provide the data storage facility using tiny DB, which assists the user to retrieve the previous records easily, such facility is not provided in the previous work [23]. In the web based calorie application, there is no facility for the basic calorie needs so we included the Caloric chart in the calorie level module which guides the users about their caloric need, [24]. In the target heart rate module, we have given the information in the form of text so that user can easily understand the application as this feature is not present in the foregoing heart rate application, [25]. In the previous diabetes application, a diet chart facility is not provided so we address this issue in our application so that the user can maintain their diet [26].

The rest of the paper is organized as follows. Section 2 gives a detail of Material and Method. In section 3, we present Result and Discussion of proposed approach, which evaluates the effectiveness of the proposed system. The final section concludes the work with a discussion on a future extension.

## I. MATERIAL AND METHOD

The materials used to develop the software are as follows: (1) Window 8.1 Haier laptop, (2) MIT App Inventor 2 Software, (3) Samsung Tablet and (4) Infinix X551Android Cell Phone.

The experimental setup section presents detail about the implementation and evaluation of the proposed system. As described earlier, we developed the software using MIT App inventor and tested the apps in Bluestack emulator. To evaluate the effectiveness of proposed system, a web-based survey is conducted. The proposed system is given below.

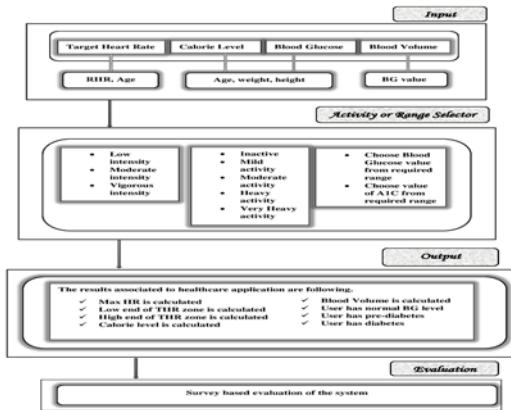


Figure 1. The proposed system

**Target Heart rate:** In the following code blocks Fig. 2, we used two text boxes to enter age, and resting heart rate, where age and resting heart rate are the variables. Also, a procedure is used to display the output: upper and lower limit of target heart rate; and a button is used to call a procedure. The clock component is used to display the current date and time.

Figure 2. Code block for input and output of target heart rate

**Calorie Level:** The Spinner component provides the choice of male and female to the user. There are three variables age, height, weight that are initialized. Button 7 is used to call the procedure. The value of age, height and weight are entered in the textbox. A label is used to display the output that is used to determine the calorie level in the human body. The clock component is used to present the current date and time, below is a partial set of coding Fig. 3.

Figure 3. Code block for input and output of Calorie level

**Blood Volume:** In Fig. 4 the partial code blocks for blood volume module is presented, where a button is used to call a procedure to initialize three variables: cm, height and weight. The procedure textbox invites users to enter height and weight spinners for selection of gender, and labels to display that how much blood is in the human body.

Figure 4. Code block for input and output of Blood volume

**Diabetes:** The diabetes module code blocks has three list pickers to display a list of items for assisting the user to make a selection from a list. A button is an event handler in which variables, list picker, and labels are used to exhibit the output in terms of blood glucose level, which is either normal, pre diabetes or diabetes. Similarly, for Random blood glucose level, three list picker are provided and a button event handler, which executes a sequence of commands; and also for HbA1c test type, a button contains spinner component that gives a list of choices to the user for making a selection. The partial code of blocks is shown in Fig. 5.

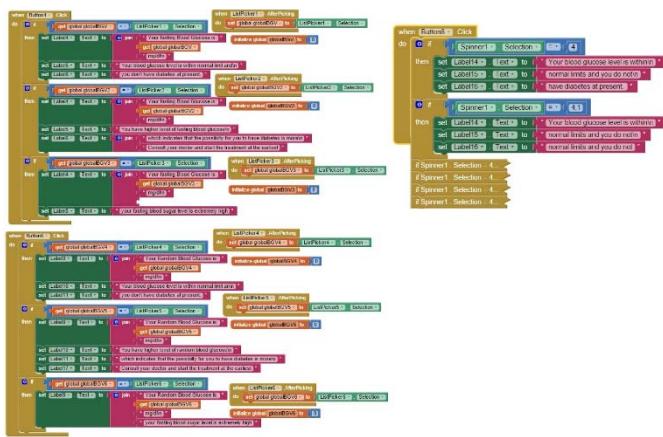


Figure 5. Code block for fasting blood glucose, random blood glucose and hba1c test type

**Data Viewer Screen:** Below is the partial code blocks Fig. 6, which are used to store data by using tiny DB. It involves Data viewer screen and a button event handler. Labels are used to display the results of all the modules that are stored in tiny DB. It helps the user to organize the history of their records.

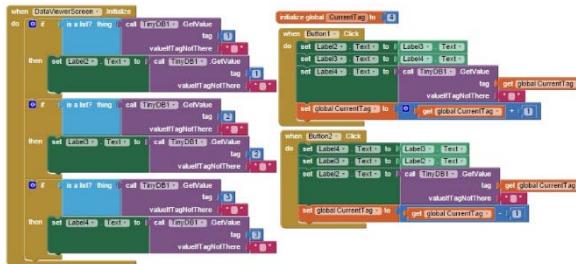


Figure 6. Code block for data viewer screen

#### A. Methodology

The proposed system is comprised of four modules, namely (1) target heart rate, (2) calorie level, (3) blood volume, and (4) diabetes. The flow chart of proposed system is described in Fig. 7.

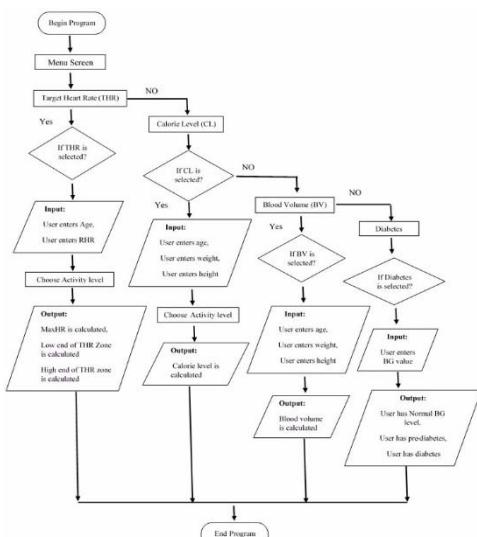


Figure 7. Flow chart of proposed system

#### 1) Target Heart Rate

The target heart rate module allows the user to get information about the different readings related to heart-beat levels, which assists in keeping the heartbeat level at desired level. Firstly, the user has to give certain inputs like resting heart rate, age and activity level. These inputs are then calculated to get the desired output, such as it computes the extreme pulse rate and the higher and lower pulse rate limits. The pseudo code of this module is given below.

#### Algorithm 1.Computation of Target Heart rate

**Objective:** The goal of this pseudo code is to calculate the Target Heart Rate.

**Input:** RHR, Age

**Output:** Display MaxHR, Display Low End of THR Zone, Display High End of THR Zone

#### Begin:

```

1.If workout intensity=Low Intensity (50-60%) then
{
2.   MaxHR ← 206.9 – (0.67 * age)
3.   HRR ← MaxHR – RHR
4.   TR1 ← HRR * 0.5
5.   Low End of THR Zone ← TR1 + RHR
6.   TR2 ← HRR * 0.6
7.   High end of THR Zone ← TR2 + RHR
}
8.If workout intensity=Moderate Intensity (60-70%) then
{
9.   MaxHR ← 206.9 – (0.67 * age)
10.  HRR ← MaxHR – RHR
11.  TR1 ← HRR * 0.6
12.  Low End of THR Zone ← TR1 + RHR
13.  TR2 ← HRR * 0.7
14.  High end of THR Zone ← TR2 + RHR
}
15.If workout intensity=Vigorous Intensity (75-85%) then
{
16.   MaxHR ← 206.9 – (0.67 * age)
17.   HRR ← MaxHR – RHR
18.   TR1 ← HRR * 0.75
19.   Low End of THR Zone ← TR1 + RHR
20.   TR2 ← HRR * 0.85
21.   High end of THR Zone ← TR2 + RHR
}
End

```

#### 2) Calorie Level

The second module determines the caloric demands of the user based on his/her age, weight, height and activity level, and gives recommendations accordingly. Age, weight, height and activity level are the inputs, required from the user. These inputs are used in the calculation of final result, reflecting how much calorie is in the human body. The pseudo code of calorie level is given as follows,

**Algorithm 2.** Computation of Calorie level

**Objective:** The aim of this pseudo code is to calculate the Calorie level.

**Input:** age, weight, height

**Output:** Display the Calorie level

**Begin**

1. If gender=male then
2.     if workout intensity=inactive then
3.         Calorie level=  $(9.99 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} + 5) * 1.2$
4.     else if workout intensity=mild active then
5.         Calorie level =  $(9.99 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} + 5) * 1.375$
6.     else if workout intensity=moderate active then
7.         Calorie level =  $(9.99 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} + 5) * 1.55$
8.     else if workout intensity=heavy active then
9.         Calorie level =  $(9.99 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} + 5) * 1.7$
10.    else workout intensity=very heavy active then
- Calorie level =  $(9.99 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} - 161) * 1.9$
11. If gender=female then
12.    if workout intensity= inactive then
13.         Calorie level =  $(9.99 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} - 161) * 1.2$
14.    else if workout intensity=mild active then
15.         Calorie level =  $(9.99 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} - 161) * 1.375$
16.    else if workout intensity=moderate active then
17.         Calorie level =  $(9.99 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} - 161) * 1.55$
18.    else if workout intensity=heavy active then
19.         Calorie level =  $(10 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} - 161) * 1.7$
20.    else workout intensity=very heavy active

```

then
{
21.   Calorie level =  $(10 * \text{weight} + 6.25 * \text{height} - 5 * \text{age} - 161) * 1.9$ 
}
End

```

*3) Blood Volume*

The blood volume component aims at determining the quantity of blood in a human body subject to height and weight. It requires inputs including age, height and weight from the users required for calculating the blood volume. The pseudo code is given below.

**Algorithm 3.** Computation of Blood volume

**Objective:** The goal of this pseudo code is to calculate the Blood Volume.

**Input:** age, weight, height

**Output:** Display the Blood volume

**Begin**

- 1.f gender=male then
2.     BD =  $0.3669 * \text{height} + 0.03219 * \text{weight} + 0.6041$
- 3.If gender=female then
4.     BD =  $0.3561 * \text{height} + 0.0338 * \text{weight} + 0.1833$

**End**

*4) Diabetes*

This module records the reading of blood sugar to assist the users for tracking their diet. The user first has to choose from one of the three test types, namely (1) Fasting blood glucose level, (2) Random blood glucose level, and (3) hemoglobin A1C. When the user selects the test type of Fasting Blood Glucose, then he chooses the blood glucose value from required ranges given to the user. The user then gets informed about his blood glucose value that either it is in the normal range, pre-diabetes or diabetes. Similarly, when the user selects the test type of Random Blood Glucose or hemoglobin A1C, then it intimates the user about their blood glucose value, i.e. whether is in the normal range, pre diabetes or diabetes. The pseudo code is given below.

**Algorithm 4.**Determination of Blood Glucose (BG)

**Objective:** The goal of this pseudo code is to calculate the Blood Glucose.

**Input:** BG value

**Output:** Your blood glucose level is within normal limit and you don't have diabetes at present, You have higher level of fasting blood glucose which indicates that the

possibility for you to have diabetes is more consult your doctor and start the treatment at the earliest, Your fasting blood sugar level is extremely high,

## Begin:

```

1. If test type=fasting blood glucose then
{
2.     if BGV=70 to 99 then
    {
3.         Display "Your blood your blood glucose level is
within normal limit and you don't have diabetes at
present"
    }
4.     else if BGV=100 to 125 then
    {
5.         Display "You have higher level of fasting blood
glucose which indicates that the possibility for you to
have
diabetes is more consult your doctor and start the
treatment at the earliest"
    }
6.     else BGV > 126 then
    {
7.         Display" your fasting blood sugar level is
extremely high"
    }
8.     End if

9. Else If test type=Random blood glucose then
{
10.     If BGV=70 to 139 then
    {
11.         Display "your blood your blood glucose level is
within normal limit
and you don't have diabetes at present"
    }
12.     else If BGV=140 to 199 then
    {
13.         Display" You have higher level of random blood
glucose which indicates that the possibility for you
to have diabetes is more consult your doctor
and start the treatment at the earliest"
    }
14.     else BGV >=200 then
    {
15.         Display" your fasting blood sugar level is
extremely high"
    }
16.     End if
    }

17. Else If test type=HemoglobinA1C then
{
18.     If BGV 4 to 5.6 then {
19.         Display "Your blood glucose level is within
normal limits
and you do not have diabetes at present"
}

```

```
20.      } else If BGV 5.7 to 6.4 then
21.      {
22.          Display "Your blood glucose level is above
normal
and this is seen in pre-diabetes"
23.          }
24.      else BGV 5.7 to 6.4 then
25.          {
26.              Display "You have higher level of fasting
blood glucose which indicates that the possibility
for you to have diabetes is more"
27.          }
28.      End if
29.  }
30. End if
End
```

## II. RESULTS AND DISCUSSION

We executed our healthcare application using android based platform, which encourages users to nourish their health and improves their healthy habits. Visual block programming language is used for the development of the application. Fig. 8 shows menu screen of our application, Fig. 9 to Fig. 13 show input and output of our application and Fig. 14 shows the data storage screen of our application.



Figure 8. Main menu screen



(a) (b) (c)  
Figure 9. Target heart rate module screen (a) input1 (b) input2 (c) output



Figure 10. Calorie level module (a) input (b) output (c) Calorie chart



Figure 14. Storage of data

#### A. Quantitative Evaluation

The qualitative evaluation consists of basic statistical analysis of the survey.

TABLE I. SHOWING THE GENDER-WISE BASIC STATISTICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	4.00	2.00	1.83	0.37

The minimum and maximum mean the smallest and largest number answer choice that collects not less than one response. It is useful to find the range of answer by subtracting the minimum and maximum. In Table I, minimum (1) and maximum (2) presents that there were 5 responses in the uppermost answer (i.e. Male) and 25 responses in the lowermost answer (i.e. female). The answer choice that is in the center of all responses shows a median, means there is 50% response before median are smaller and 50% response after median are larger. The median of 2.00 (higher than the 1.83 mean) shows that there were more respondents who were Female than respondents who were Male. The mean gives the average of entire responses by adding all number answer choices and then divide them by total amount of number. In this case, a mean of 1.83 represents the overall respondents came in somewhere between Male, and the Female. Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.37.

#### What is your gender?

Answered: 30 Skipped: 0

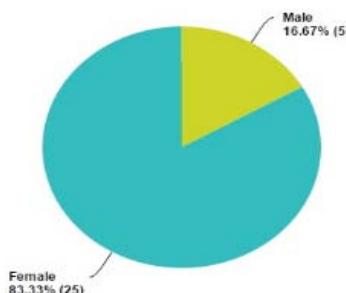


Figure 15. Pie Chart of gender

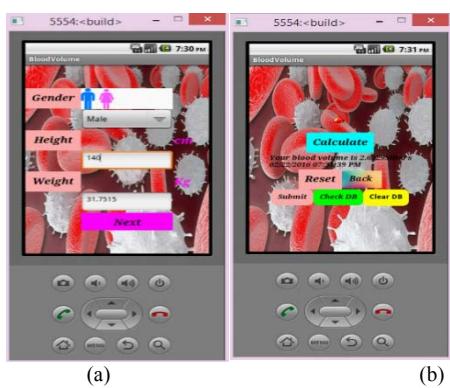


Figure 11. Blood volume (a) input (b) output

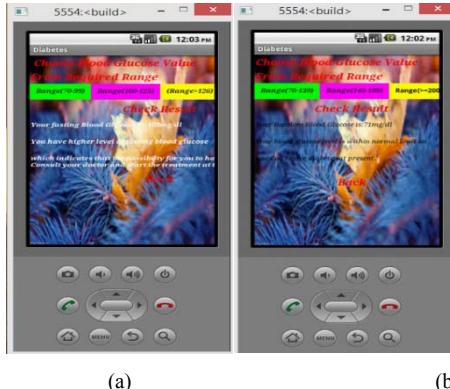


Figure 12. (a) Fasting blood glucose (input & output) (b) Random blood glucose (input & output)



Figure 13. (c) HbA1C test (input & output) (d) Unit converter (e) Diet chart

The Fig. 15 shows that there were total 5 (16.67%) male respondents and 25 (83.33%) female respondents in the survey and the total respondents were 30.

TABLE II.

SHOWING THE AGE-WISE BASIC STATISTICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	3.00	2.00	1.67	0.60

In Table.II, minimum (1) and maximum (3) presents that there were 12 responses in the uppermost answer (i.e. age 18 to 24) and 3 responses in the lowermost answer (i.e. age 45 to 54).The median of 2.00 (higher than the 1.67 mean) shows that there were more respondents who were in age (25 to 34) than respondents who were in age (18 to 24).A mean of 1.67 shows that overall respondents came in somewhere between age (18 to 24), and the age (25 to 34).Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.60.

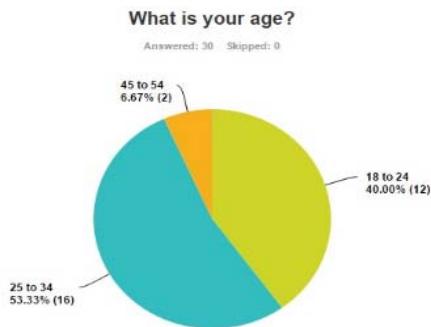


Figure 16. Pie Chart of age

Fig. 16 shows that there were total 12(40.00%) respondents whose age is between 18 to 24, and 16(53.33%) respondents whose age is between 25 to 34 while the respondents whose age is between 45 to 54 were 2(6.67%).

TABLE III.

SHOWING THE IMPORTANCE OF EXERCISE-WISE BASIC STATISTICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	4.00	2.00	2.50	0.99

In Table.III, minimum (1) and maximum (4) presents that there were 5 responses in the uppermost answer (i.e. extremely important) and 6 responses in the lowermost answer (i.e. slightly important). The median of 2.00 (lower than the 2.50 mean) shows that there were more respondents who said exercise is very important for them. In this case, a mean of 2.50 shows that overall respondents came in somewhere between very important, and the moderately important. The mean gives the average of entire responses. Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.99.

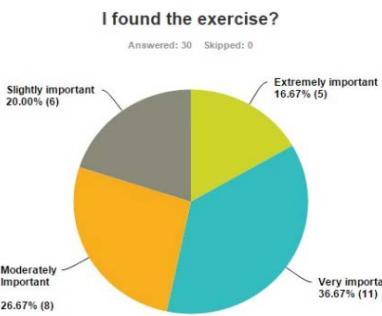


Figure 17. Pie Chart about importance of exercise

The Fig. 17shows that there were total 5(16.67%) respondents who consider that exercise is extremely important for them, and the exercise that is very important for the respondents were 11(36.67%) while the respondents who said that exercise is moderately important for them were 8(26.67%), the exercise that is slightly important for the respondents were 6(20.00%).

TABLE IV.

SHOWING THE LEVEL OF BASIC STATISTICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	4.00	2.00	2.43	1.05

In the Table. IV, minimum (1) and maximum (4) presents that there were 6 responses in the uppermost answer (i.e. lift weights) and 7 responses in the lowermost answer (i.e. Aerobics).The median of 2.00 (lower than the 2.43 mean) shows that there were more respondents who mostly do walk for exercise.).In this case, a mean of 2.43 shows that overall respondents came in somewhere between exercise (walk), and the exercise (run).Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 1.05.

I most often do for exercise?

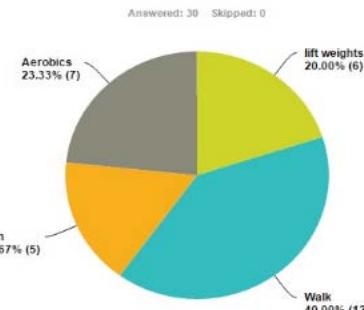


Figure 18. Pie Chart for level of exercise

The Fig. 18shows that there were total 20.00% respondents and for them the level of exercise is just lifting weights, 12{40.00%} respondents do walk for exercise, 5(16.67%) do running, and 7(23.3%) perform Aerobics exercise.

TABLE V.

SHOWING THE SIGNIFICANCE OF BG APP BASIC STATISTICS ICS

	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	4.00	2.00	2.13	0.72

In the Table.V minimum (1) and maximum (4) presents that there were 5 responses in the uppermost answer (i.e. strongly agreed) and 1 response in the lowermost answer (i.e. Disagree). The median of 2.00 (lower than the 2.13 mean) show that there were more respondents who were agreed. In this case, a mean of 2.13 shows that overall respondents came in somewhere between agreed, and the satisfactory. Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.72

I found application helpful to control blood glucose level.

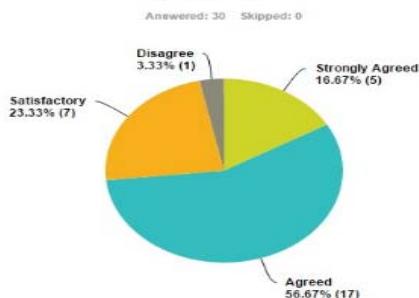


Figure 19. Pie Chart for Significance of BG app

The Fig. 19 shows that there were total 20.00% respondents who were strongly agreed with the statement, 17(56.67%) respondents were agreed while 7(23.33%) have satisfactory views about BG app and only 1(3.33%) respondent disagree.

TABLE VI.

SHOWING THE MAINTAIN DIET BASIC STATISTICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	3.00	2.00	1.80	0.65

In the Table.VI, minimum (1) and maximum (3) presents that there were 10 responses in the uppermost answer (i.e. strongly agreed) and 4 responses in the lowermost answer (i.e. Satisfactory).The median of 2.00 (higher than the 1.80 mean) shows that there were more respondents who were agreed than respondents who were strongly agreed. In this case, a mean of 1.80 shows that overall respondents came in somewhere between strongly agrees, and the agreed. Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.65.

I found this application helpful to maintain the diet?

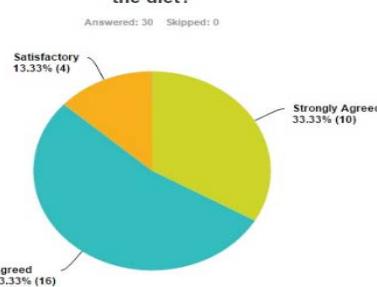


Figure 20. Pie Chart of maintain diet

Fig. 20 shows that there were total 10(33.33%) respondents who were strongly agreed with the statement, 16(53.33%) respondents said that application is helpful to maintain their diet and 4(13.33%) found application satisfactory to maintain their diet.

TABLE VII.

SHOWING THE GUIDANCE ABOUT BASIC CALORIC NEEDS BASIC STATISTICS ICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	3.00	2.00	1.60	0.61

In Table.VII, minimum (1) and maximum (3) presents that there were 14 responses in the uppermost answer (i.e. strongly agreed) and 2 responses in the lowermost answer (i.e. Satisfactory).The median of 2.00 (higher than the 1.60 mean) shows that there were more respondents who were agreed than respondents who were strongly agreed. In this case, a mean of 1.60 shows that overall respondents came in somewhere between strongly agreed, and the agreed. Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.61.

I found application provide guidance about basic caloric needs?

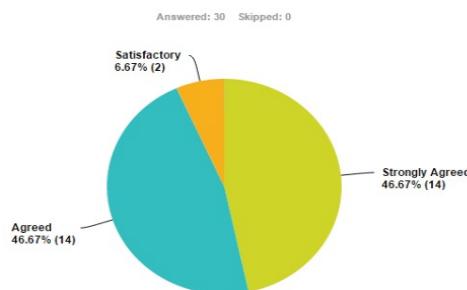


Figure 21. Pie Chart of Guidance for Basic Caloric needs

The Fig. 21 shows that there were total 14(46.67%) respondents who were strongly agreed with the statement, 14(46.67%) respondents said that application is helpful to provide basic information about caloric needs and 2(6.67%) respondents found application satisfactory.

TABLE VIII. SHOWING THE INTEGRATION OF MODULE BASIC STATISTICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	3.00	1.00	1.53	0.76

In the Table VIII, minimum (1) and maximum (3) presents that there were 19 responses in the uppermost answer (i.e. strongly agreed) and 2 responses in the lowermost answer (i.e. Satisfactory). The median of 1.00 (less than the 1.53 mean) show that there were more respondents who strongly agreed with the statement. In this case, a mean of 1.53 shows that overall respondents came in somewhere between strongly agreed, and the agreed. Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.76.

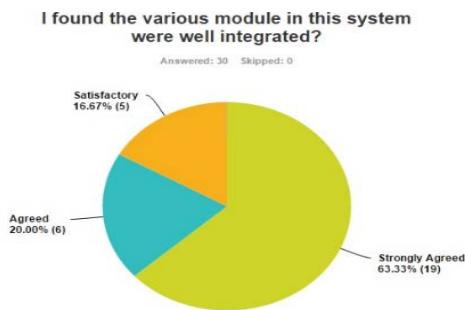


Figure 22. Pie Chart for Integration of module

The Fig. 22 shows that there were total 19(63.33%) respondents who were strongly agreed with the statement, 6(20.00%) respondents agreed that the modules well integrated and 5(16.67%) respondents found the integration of modules in an application is satisfactory.

TABLE IX. SHOWING THE USER-INTERFACE BASIC STATISTICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	3.00	2.00	1.63	0.60

In the Table.IX, minimum (1) and maximum (3) presents that there were 13 responses in the uppermost answer (i.e. strongly agreed) and 2 responses in the lowermost answer (i.e. Satisfactory).The median of 2.00 (higher than the 1.63 mean) shows that there were more respondents who were agreed than respondents who were strongly agreed. In this case, a mean of 1.63 shows that overall respondents came in somewhere between strongly agreed, and the agreed. Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.60.

I found the system user-friendly?

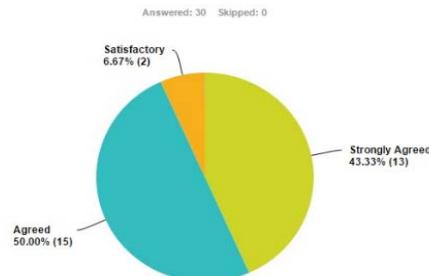


Figure 23.Pie Chart of system user interface

Fig. 23 shows that there were total 13(43.33%) respondents who were strongly agreed with the statement, 15(50.00%) respondents agreed that the system is user friendly and 2(6.67%) respondents have satisfactory views about the user interface of an application.

TABLE X.

SHOWING THE RATE APPLICATION BASIC STATISTICS

Sr.no	Basic Statistics				
	Minimum	Maximum	Median	Mean	Standard deviation
1.	1.00	3.00	2.00	1.87	0.85

In Table.X, minimum (1) and maximum (3) presents that there were 13 responses in the uppermost answer (i.e. reliable) and 9 responses in the lowermost answer (i.e. Useful). The median of 2.00 (higher than the 1.87 mean) show that there were more respondents who said that the application is of high quality. In this case, a mean of 1.87 shows that overall respondents came in somewhere between reliable, and the high quality. Finally, the standard deviation shows the growth or alteration of your responses, so here the standard deviation is 0.85.

I would use following words to describe the application.

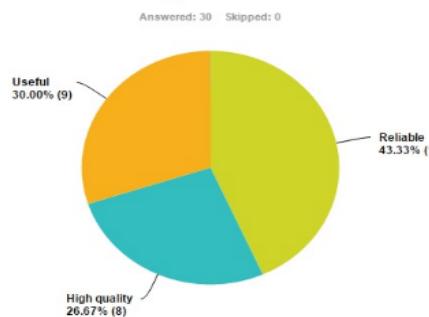


Figure 24. Pie Chart of rate application

The Fig. 24 shows that there were total 13(43.33%) respondents found the application reliable and 8(26.67%) respondents said that application is of high quality while 8(26.67%) respondents said that the application is useful.

Sr. no	Excel Data Analysis												
	Mean	Standard Error	Median	Mode	Standard Deviation	Sample Variance	Kurtosis	Skew- ness	Range	Min- imum	Maxi- mum	Sum	Count
1.	1.896	0.109506	1.815	1.6	0.346288	0.119916	-0.49301	0.9092	0.97	1.53	2.5	18.6	10

TABLE XI. SHOWING THE EXCEL DATA ANALYSIS

Table XI calculations are obtained from the **Excel data analysis** by using the mean values of all the questions.

### B. Questionnaire

Fig. 25 shows the questionnaire of an individual respondent in order to obtain the feedback and also to analyze the result. The respondent chooses one option from multiple choice.

Figure 25. Questionnaire of individual person

Some of the major findings during analysis are listed below

- A question was asked from the respondent to gather information about the performance of application that whether application helps them to keep track of their diet. So, 33.3% respondents were strongly agreed that the application is helpful to maintain their diet, while 53.3% respondent were agreed with the statement, 33.3% respondents were satisfied and there are no respondents that disagree with the statement.
- The objective of the Second question was to get respondents views about the integration of module. The separated modules are combined or coordinated into main application so respondents look at the integration that it is well organized or not. 62.23% respondents were strongly agreed with the statement, 20.00% respondent were agreed with the statement that module were well integrated while at the same time 16.67% respondents were satisfied with the statement moreover no respondent were found who disagree with the statement.

respondents find the application helpful to control blood glucose level or not.

16.67% respondents strongly agreed that application is helpful to control blood glucose level. There are 56.67% respondents who consider that application is helpful to control blood glucose level, 23.33% respondents found application satisfactory to control blood glucose level, 3.33% respondents disagree with the statement.

From the above questions we conclude that the users find that the application is beneficial to maintain their health

### III. CONCLUSION AND FUTURE WORK

The main purpose and focus of developing the healthcare application is to help people to maintain their health. This healthcare application includes the four modules, namely (1) Target heart rate, (2) calorie level, (3) blood volume, and (4) diabetes app.

The first module describes the pulse rate (in beats per minute) that allows the user to exercise safely while getting the maximum benefits from your workout. It includes THR zones which range from low to vigorous i. e (50 to 85) % of MaxHR.

The second module is the calorie level, all essential process of our body, uses this measurement unit of energy. In order to

encounter the energy needs of our body the speed at which the calorie is used alters continually. Throughout different phases of life, it changes from individual to individual. It is used to determine the caloric needs based on the age, weight, and height and activity level.

The third module is the blood volume, which reflects the amount of the blood in human body. This app assists in answering about how much blood is in the human body, more precisely in your own body depending on the height and weight.

The fourth module is the diabetes app tells about that when the body does not properly use or store glucose. Its records, the

• T  
hird  
questio  
n aims  
to  
know  
whethe  
r

blood sugar readings, and assists users to track their diet properly.

**Future Work:** In the future, we will integrate more apps to our main application to make it a more sophisticated auto-help tool and to provide a wide range of facilities to the end user. These apps will include: (1) Measuring blood pressure and Measuring Weight of the body, (2) Provide reminders to users about their medications which help them to take medicine on time. Therefore, through these reminders, the user can take care of their health, and (3) Graphs of the output obtained will help the user to keep track of the changes in diabetes-related readings and to manage their diet and health in a more effective way.

#### ACKNOWLEDGMENT

In the name of Allah, the Most Gracious and the Most Merciful, Alhamdulillah, all praises to Allah for the strengths and His blessing in completing this research paper.

We would like to express our deepest gratitude to our Supervisor, Dr. Muhammad Zubair Asghar, for his excellent guidance, caring, patience, and providing us with an excellent atmosphere for doing our research work.

#### REFERENCES

- [1] Asghar MZ, Khan A R, Asghar M J. Computer assisted diagnoses for red eye (CADRE). International Journal on Computer Science and Engineering. 2009, 1(3), pp. 163-70.Date accessed: 23/01/2015
- [2] Asghar D, Zubair M, Asghar MJ. Expert System for Online Diagnosis of Red-Eye Diseases. International Journal of Computer. Date accessed: 23/01/2015
- [3] Science & Emerging Technologies (IJCSET). 2010, 1(2), pp. 35-39. Date accessed: 24/01/2015
- [4] Top 15 Android medical apps for health care professionals. <http://www.imedicalapps.com/2011/01/top-free-android-medical-apps-healthcare-professionals/>. Date accessed: 24/01/2015.
- [5] Get help applying for health insurance. <https://www.healthcare.gov/apply-and-enroll/get-help-applying/>. Date accessed: 13/01/2016
- [6] Multi health-care: <http://sdc.csce.uark.edu/projects/modhealth/>. Date accessed: 12/12/2015
- [7] A mobile nutrition self-management application for people with diabetes.<http://munin.uit.no/bitstream/handle/10037/4233/thesis.pdf>Date accessed: 22/11/2015
- [8] Asghar D, Zubair M, Ahmad D. A Review of Location Technologies for Wireless Mobile Location-Based Services. Journal of American Science. 2014, 10(7), pp. 110-18. Date accessed: 25/11/2015
- [9] Asghar MZ, Qasim M, Ahmad B, Ahmad S, Khan A, Khan IA. Health Miner: Opinion Extraction From User Generated Health Reviews. International Journal of Academic Research. 2013 Nov, 5(6), pp. 279-84.Date accessed: 30/11/2015
- [10] Asghar MZ, Ahmad S, Marwat A, Kundi FM. Sentiment Analysis on YouTube: A Brief Survey.MAGNT Research Report. 2015 Nov, 3(1), pp. 1250-57.Date accessed: 01/12/2015
- [11] Asghar MZ, Khan A, Kundi FM, Qasim M, Khan F, Ullah R, Nawaz IU. Medical opinion lexicon: an incremental model for mining health reviews. International Journal of Academic Research. 2014 Jan, 6(1), pp. 295-302.Date accessed: 12/12/2015
- [12] KeunYoo Lee, HakJin Moon, Ye Seul Han, Soon Ryun Lim. The Factors affecting Health Behaviors of a Mother with Infants and Toddlers, Indian Journal of Science and Technology, 2015, 8(35), pp. 1-9.Date accessed: 26/12/2015
- [13] Jaeyeon Kang, SunjuSohn.Limited Access to Health Care and the Impact there of on Married Women's Mental Health, Indian Journal of Science and Technology, 2015, 8(20), pp. 1-7.Date accessed: 09/01/2016
- [14] Sung-Soo Kim. A Study on the Acceptance Factor for Telehealth Service According to Health Status by Group, Indian Journal of Science and Technology, 2015, 8(1), pp. 542-50Date accessed: 14/01/2016
- [15] Hwansoo Kang, Jinhyung Cho, Heechern Kim. Application Study on Android Application Prototyping Method using App Inventor, Indian Journal of Science and Technology, 2015, 8(19), pp. 1-5.Date accessed: 15/01/2016
- [16] Hwansoo Kang, Jinhyung Cho. Case Study on Efficient Android Programming Education using Multi Android Development Tools, Indian Journal of Science and Technology, 2015, 8(19), pp. 1-5.Date accessed: 11/02/2016
- [17] Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, and Bashir Ahmad. SUBJECTIVITY LEXICON CONSTRUCTION FOR MINING DRUG REVIEWS. Science International 26(1) 2014, pp 145-149.Date accessed: 18/02/2016
- [18] Asghar, Muhammad Zubair, et al. "AndorEstimator: Android based Software Cost Estimation Application." arXiv preprint arXiv:1605.02304 (2016).Date accessed: 12/05/2016
- [19] Dr. Muhammad Zubair Asghar, Ulfat Batool, Farheen Bibi, Sadia Ismail, et al. "Financial Studio: Android Based Application for Computing Tax, Pension, Zakat and Loan"International Journal of Academic Research Vol. 4 Iss. 2 (2016) p. 96 - 117 ISSN: 2075-4124 Available at: <http://works.bepress.com/drzubair/21/> Date accessed: 12/05/ 2016
- [20] Dr. Muhammad Zubair Asghar, Iqra Sana, Hina Iqbal and Khushboo Nasir. "Quizzy: Quiz Application Development Using Android-Based Platform" (2016) Available at: <http://works.bepress.com/drzubair/28>
- [21] Mit App Inventor. <http://appinventor.mit.edu/explore/>. Date accessed: 22/07/2016.
- [22] Health Calculator Free. <https://play.google.com/store/apps/details?id=com.hashinclude.healthcalculator>. Date accessed: 06/01/2014.
- [23] The Calculator. <http://www.thecalculator.co/health/Blood-Calculator-67.html>. Date accessed: 23/12/2014.
- [24] Calorie Calculator. <http://www.calculator.net/calorie-calculator.html?ctype=metric&cage=22&csex=m&cheightfeet=5&cheightinch=10&cpond=160&cheightmeter=160&ckg=27.2155&cactivity=1.2&printit=0&x=84&y=16>.Date accessed: 09/10/2015.
- [25] Find Your Target Heart Rate.<http://exercise.about.com/od/cardiotraining/ss/findtargetheart.htm#showall>. Date accessed: 05/04/2016.
- [26] Blood-Sugar Chart.[http://www.medindia.net/patients/calculators/bloodsugar\\_chart.asp](http://www.medindia.net/patients/calculators/bloodsugar_chart.asp).Date accessed: 18/04/2016.

# Genetic Algorithm based Novel approach for Load Balancing problem in Cloud environment

Dr. Surjeet Dalal

Department of Computer Science & Engineering,  
SRM University  
Haryana, India  
profsurjeetdalal@gmail.com

Shilpa Kukreja

Department of Computer Science & Engineering,  
SRM University  
Haryana, India  
kukreja.shilpa@gmail.com

**Abstract**—Cloud computing has come up as one of the most promising & reliable technologies in the IT sector. However presently there exists a major issue of load balancing in the cloud computing environment. This paper consists of a solution for optimizing the load using genetic algorithm. Genetic algorithm which follows the evolutionary mechanism is able to develop a solution close to optimal solution. The proposed algorithm is developed by merging two existing algorithms by considering cost value as the fitness function. The workload is balanced by the considering the combination of both the load percentage and cost value of the resources. Allocation of resources is performed by taking the best fit value and reducing the response time and overall cost. Simulation results are shown using the cloud analyst simulator.

**Keywords-** Cloud computing; genetic algorithm; load balancing; fitness value; load percentage;

## I. INTRODUCTION

Cloud computing is the new word for distributed computing on a large scale platform. This describes the large association of computing resources with virtualization as its key technology for providing a reliable computing platform for users seeking various IT resources for fulfilling different requirements for hardware, software and platform related needs. In the coming years it will be seen as the most effective and reliable means to outsource various IT requirements. Due to various outstanding features it provides, many businesses resort to its services. It provides services according to the user's demand to cater their needs. The user does not need to be in a specific place to gain access to the services. Cloud applications can be accessed from any part of the world provided there exists internet equipped device. The businesses can provision resources according to their requirement without the extra burden of infrastructure maintenance costs.

There are various cloud providers providing three types of services namely:

- Infrastructure as a Service
- Platform as a Service
- Application as a Service

Today cloud computing is the evolutionary factor behind all existing technologies and has given all the businesses new dynamic changes in the computing environment. It has the potential to impact the companies, businesses, organisations in the IT sector by providing reliable services such as

availability, scalability, on-demand access to computing resources. It is because of the features it provides that this technology is gaining recognition and will remain the effective means for carrying out various business related operations and fulfilling several IT resources requirements in the coming several years[1].

There are basically four cloud deployment models:

- Public Cloud
- Private Cloud
- Hybrid Cloud
- Community Cloud [1]-[3].

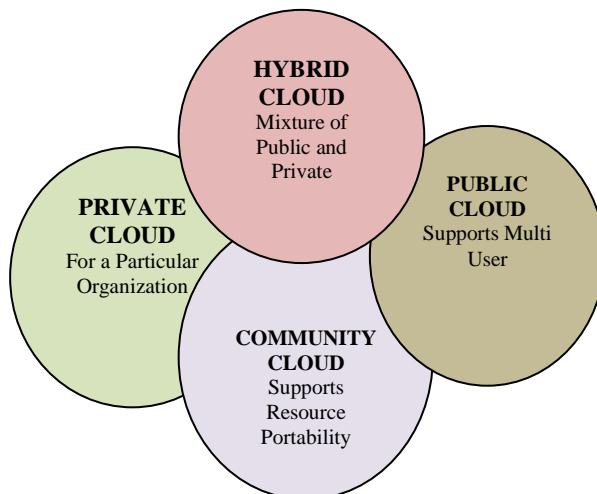


Figure 1 Cloud models

Even though this technology has attained extensive adoption throughout the IT industry, there still exist a no. of challenges and issues which need to be addressed and resolved for enhancing the performance of the cloud system. The cloud computing environment is expanding at a rapid rate with the no. of users growing each day. Therefore with the increasing users, the question of maintaining the numerous requests with the delivery of high performance remains a problem waiting to be resolved. Due to the existing issues and challenges, there is an opportunity to carry out various researches in the cloud computing. Several studies have been done so far in the various facets of this technology.

Load balancing is an activity which aims to efficiently balance the workload amongst various processing units of a

cloud environment to facilitate high end performance to its users. The workload can be characterised in terms of CPU utilisation, memory utilised and the network load for a particular virtual machine. The job of load balancing involves creation of an effective load balancing algorithm that targets to avoid any over utilization and under utilization of resources. This is an important issue that needs to be addressed to avoid any extra revenue costs and improve the overall performance factor. Although many contributions have been made to resolve the issue of load balancing, but the scope for improved results is still open.

There exist two approaches to balance the load among the processing units

- a) Centralized
- b) Decentralised

In the centralised approach one node which is central to the overall system is responsible for balancing the load among various processors. Although this approach sounds appealing but it fails in the situation where the central node crashes or fails due to any system problem.

In the decentralised approach, each node is responsible for balancing the load among them. This is the most preferred form of load balancing since a single node failure will not result in the whole system failure for balancing load.

Several researches have been done to create and implement new algorithms for load balancing. Algorithms play an important role in ensuring the effective balancing of load in the cloud system. Various parameters are targeted through the use of these algorithms to enhance the performance of the cloud system.

#### **a) Resource Utilisation**

This improves the utilisation factor of all the resources

#### **b) Response Time**

This refers to the time taken in execution of client requests

#### **c) Cost**

Total cost incurred in the cloud system

#### **d) Migration time**

Time taken for migrating client requests between different processing units

Genetic algorithm is a form of approach based on the idea of genetics and evolution mechanism. The proposed algorithm uses the concept of two existing genetic algorithms.

Further the paper consists of section 2 which presents the previous work done by researches using genetic algorithm, section 3 consists of detailed information about the genetic algorithm, section 4 consists of the proposed algorithm and further simulation results are shown using cloud analyst simulator.

## **II. RELATED WORK**

Chandrase et al. (2013) proposed a genetic algorithm to balance the load effectively among all the virtual machines. The load of machines was analyzed under stable and variant load conditions. The load of physical machines is determined by the CPU load and the memory load. Simulation was performed using open nebula software. Load percentage was

calculated by comparing with round robin and greedy algorithm. Results were illustrated under stable conditions and then under variant conditions. Further results were illustrated with values for CPU and memory both for memory oriented application and CPU oriented application. Simulation showed that genetic algorithm did better allocation of VM & distributed load amongst all machines effectively than other two algorithms [4].

Joshi et al. (2015) presented an improved version of genetic algorithm. In this algorithm the cost value was taken as the fitness function. According to this fitness function the value was determined. After performing the series of steps for genetic algorithm, balancing of load was performed amongst machines. Simulation was performed using the MATLAB toolkit. Results were compared with round robin algorithm and original existing genetic algorithm taking response time as the parameter. Improved genetic algorithm showed tremendous reduction in response time when compared to the other two algorithms Round Robin and existing genetic algorithm [5].

Portalwi et al. (2014) considered the challenge of reducing the power consumption whilst providing high performance. The approach was to reduce the cloud system power consumption. The genetic algorithm was based on a energy efficient strategy using genetic algorithm approach. The task was to find optimal solution and allocate resources efficiently. This was a multi-objective genetic algorithm which aimed to generate non dominated solutions out of various experiments performed. Taking different parameters, it was observed that as the no. of servers increased power consumption and time was reduced [6].

Dasgupta et al. (2013) proposed a strategy for load balancing using genetic algorithm. The approach was to minimize the make span of various job requests. The cost value was taken as the fitness function. Depending on the values of fitness function, decision regarding balancing of load between various physical machines was done. Results regarding the response time parameter were shown using the cloud analyst simulator. The proposed algorithm showed better results when compared to previous algorithms [7].

Shahjahan et al. (2015) showed the entire procedure of genetic algorithm. Various operators and their functionality was discussed. Proposed algorithm was developed using the operations of genetic algorithm. The population was converted into binary strings and considering the time complexity, results were evaluated. Simulation was shown differently using one, two, three up to five data centers and results were compared with three algorithms namely round robin, stochastic hill climbing and First cum first serve. The proposed algorithm showed minimum response time [8].

Chun-Cheng et al. (2014) proposed an algorithm for multimedia system in cloud. Balancing of multimedia load(images, audio, video) among all the servers efficiently with minimum cost is a task in itself. Dynamic scenario for multimedia service was taken modeling as integer linear

programming problem. Immigrant scheme for solving dynamic problems was followed. The author considered a centralized hierarchical cloud multimedia service model. Simulation was performed using genetic algorithm and it was demonstrated that best cost values were obtained [9].

Jain et al. (2012) explained an approach for balancing the workload among processing units using brute force approach and dynamic programming. Both classical and evolutionary method was used to ensure efficient and effective balancing of load. Experiment was performed using JAVA language. When demonstrated using genetic algorithm, optimization of workload was done effectively [10].

Kaur et al. (2012) presented an approach for scheduling in cloud computing environment. The algorithm was developed with the combination of two existing algorithms. Simulation was performed using JGAP(Java based Genetic algorithm Package). Results were shown with two parameters: average make span and execution cost and were better than existing approaches [11].

Suraj et al. (2013) proposed a resource allocation mechanism based on capacity of a node taking node weight of each processing node. Adaptive Genetic algorithm used both future prediction and node weight to allocate resources and aimed to solve the issue of VM migration which had no criteria for migration. CPU utilization and memory factor were used for indicating performance of a machine. Cloud Booster algorithm was used to find a node's capacity. Results were shown in terms of communication cost, idle time& waiting time which turned out to be better and close to optimal solution [12].

Zhao et al. (2014) proposed a load balancing strategy using bayes theorem with deployment of a heuristic and clustering based strategy. The heuristic approach was followed to select the appropriate physical hosts for assignment of various job requests to them. In this algorithm a constraint value was set on the basis of which a set for physical hosts which had remaining resource amount value greater than the constraint value was created. Two resources were taken to determine the remaining resource amount of a physical host CPU resource amount and memory resource amount. The bayes theorem was used to determine the posterior probability of the physical hosts. Simulation was carried out to illustrate efficient balancing of workload amongst various physical hosts. The parameters on which simulation was carried were make span, standard deviation, throughput, failure number of tasks& incremental percentage of standard deviation values [15].

Nahir et al. (2013) presented an approach for load balancing in distributed server systems. The focus was laid on the communication overhead that was caused when data was collected for taking decision on scheduling and balancing of workload. The overhead was eliminated from the critical path of any service request. The approach revolved around the creation of several replicas of every job request and further every replica was sent to a different server. The removal of replicas was done after intimation from the head of the queue

to the servers. Simulation was performed to illustrate reduced overhead when different servers and different job requests were taken. Simulation illustrated the efficient results of the proposed algorithm [16].

Cao et al. (2016) developed an optimal power allocation and load distribution strategy for multiple heterogeneous multi core server processors in cloud environments. The optimization strategy was for optimizing power and performance. The research proposal took two cases of core speeds. One case was where the core worked at zero speed and other case was where the core speed worked at a constant model. A queuing model was also described for different heterogeneous multi core servers [17].

### III. GENETIC ALGORITHM

Genetic algorithm use the concept of evolution and their process is derived from the Darwinian's theory of survival of fittest & evolution concept. It is basically a technique to derive solutions for complex problems through the search and optimization mechanism. The process of optimization includes alteration of input values to obtain optimum output values by defining fitness function or a cost function. Genetic algorithms consist of the basic unit chromosomes or genes on which the selection of population is made. Further applying various operators of genetic algorithm, a solution close to the optimal solution is derived [12].

Following are the basic operators of genetic algorithm:

- Selection
- Crossover
- Mutation

#### Selection

After generation of a population set, a set of chromosomes are selected for crossover to generate a new chromosome. Several methods are available for selection namely

- a) Roulette wheel selection
- b) Boltzmann selection
- c) Rank selection
- d) Steady state selection

Roulette wheel selection assigns the part of wheel to individuals as per their fitness value

Rank selection method ranks the population by some mechanism. After ranking fitness value is assigned as per the ranking. This method doesn't not present fair results when fitness value has a high varying value.

#### Crossover

After selection of parent chromosomes, combination of these two is performed to obtain a new child chromosome. The technique of crossover is performed to produce a new offspring with better traits originating from both the parent chromosomes.

#### Mutation

Another operator in genetic algorithm is the mutation operator which is responsible for modifying the values of genes in chromosomes. After modification of gene values,

new better solution may be obtained. Following are the types of mutation operator

- Flip
- Bit Boundary
- Uniform
- Non uniform
- Gaussian[2]

#### IV. PROPOSED ALGORITHM

The proposed algorithm is developed by combining two existing algorithm. The solution is optimized by merging these two algorithms.

#### Mathematical Model

The mathematical model for the proposed algorithm can be explained with the following notations. Let F be a set of following symbols:

$$F_t = (D, VM, L, C, \beta_t, U, G)$$

in which :

D represents the set of all data centers in the cloud environment,

VM represents the set of all virtual machines assigned to different data centers,

L represents the amount of combined load in terms of CPU, memory utilization on a virtual machine,

C represents the cost value associated with the resource,

$\beta_t$ : LUC represents the function that combines the load and cost value of a resource at a particular time t,

U represents the set of users requesting a resource,

G represents the link between the data center and the user  $m \in U, n \in D$ .

At a particular time t, the  $\beta_t$  should be minimized for a particular user so that the minimum value is assigned to the user for its job request.

$$\text{Minimize } \sum_{n \in D} \beta_t / \sum_{m \in U} \beta_t$$

#### Existing Algorithm

1. This algorithm aims to optimize virtual machine migration and balance the load amongst all virtual machines by determining the load percentage of the virtual resources. The proposed approach calculated the load percentage of all resources and balanced the workload effectively

a. If there exist a set of physical hosts  $H = \{h_1, h_2, h_3, \dots, h_n\}$

b. N being the total no. of nodes in the cloud system,

VM is the set of virtual machines assigned to the physical hosts  $VM = \{VM_1, VM_2, \dots, VM_i\}$

c.  $H_i$  is the load on a physical host, L is the load on

$$\text{cloud } L = \frac{\sum_{i=1}^n H_i}{n}$$

d.  $F(1)$  fitness function =  $100 / \sum_{i=1}^n (L - P_i)$  [2]

This function is optimized to calculate the best solution and optimizing the solution

2. This algorithm takes the cost value as the fitness function and optimizes the solution by taking MIPS parameter
  - a.  $F(2) = Wt1 * (EC)(N/MIPS) + Wt2 * (DC)/i$  [5]
  - b. EC=Execution cost
  - c. DC=Delay cost
  - d. MIPS=million instructions per second
  - e. N=total no. of instructions
  - f. I = job no.

$$\text{Fitness Function} = \text{Optimize} [\text{Best fit (Cost, Load \%)}] \\ \text{Best Fit } (f1, f2)$$

Steps for genetic algorithm

- Creating a new random population
- Computing the fitness of individuals using the fitness function
- Selecting the chromosome with the least value of fitness twice & removing the chromosome with fitness value (**selection operation**)
- (**crossover**) Applying single point crossover to obtain a new offspring
- (**mutation**) Mutation is applied with a probability of (0.01)
- including the new offspring into the population generated
- Determining if best solution is provided
- Loop [Start from computing fitness function] [6] [14].

#### V. EXECUTION AND SIMULATION

The proposed algorithm for load balancing is simulated using the Cloud Analyst tool. It is the part of Cloud Sim tool package and is created upon its toolkit for simulation purpose. It has a list of attractive features over cloud sim. The tool is based on java language only as cloud sim but the job of programming is not much and the researcher can focus on the simulation section

Following are some of the features of cloud analyst

- User friendly
- Use of graphics for simulation
- Reiterating experiments
- Provision of saving configuration files and results of simulation

The operation of this simulator is very user friendly. Simulation results include graphical outputs such as graphs and tables. While performing simulation there is a need to reiterate the experiments which is the feature provided by cloud analyst. There also exist a provision for saving the configuration files and using them later on for repeating the experiments. It also provides the feature of saving the simulation results [13].

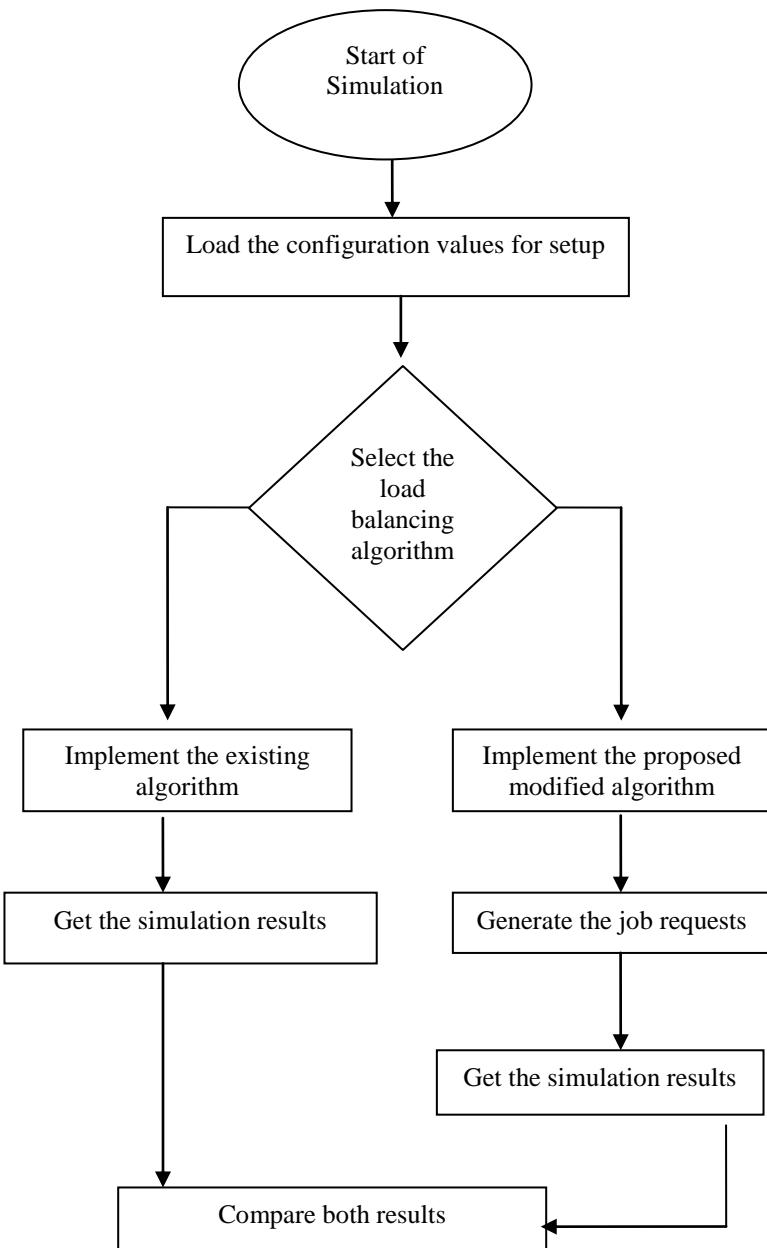


Fig 2 Flowchart of Simulation Steps

There are three built in load balancing policies in Cloud Analyst round robin, throttled, equally spread current execution load. A new load balancing policy, genetic algorithm is added and the algorithm is implemented in the tool by adding it to the required class file in the tool.

Several experiments were carried out in performing the simulation. The user requests from different regions of the world were processed by taking 5 data centers which were allotted with 5, 50, 80 110 and 140 virtual machines. Different user bases were taken for all the six regions of the world given in the tool. The simulation was performed for the genetic algorithm and following results were obtained for different user bases.

Table 1 Userbase Response Time

Userbase	Avg(ms)	Min (ms)	Max(ms)
UB1	50.44	40.38	62.13
UB 2	52.63	42.26	64.76
UB 3	201.92	150.01	251.01
UB 4	50.17	39.16	59.16
UB 5	51.77	39.02	62.77
UB 6	50.33	39.38	62.63

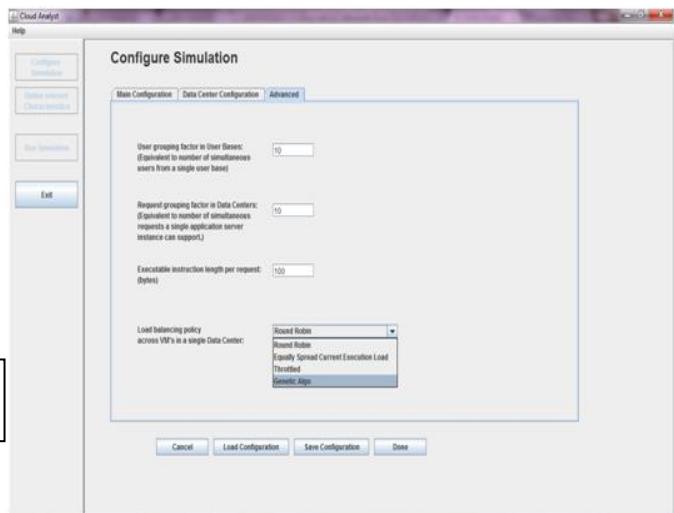


Fig 3 Load Balancing Algorithm



Fig 4 Overall Response Time

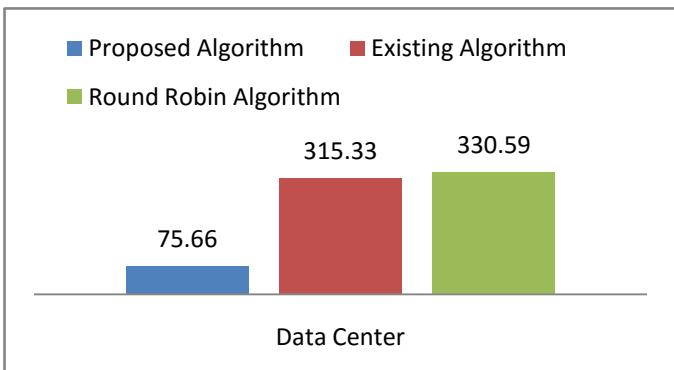


Fig 5 Comparison graph

## VI. CONCLUSION

In this paper a genetic based strategy was developed to present a solution to the problem of load balancing in cloud computing. The proposed approach was able to balance and direct the user request appropriately by combining two factors load percentage and the cost value. The simulation results revealed that the strategy was able to reduce the overall response time and cost value. The workload was balanced in an efficient and effective manner which resulted in reduction of the two parameters.

## REFERENCE

- [1] Fahrukh shahzad. (2014). "State of the art Survey on Cloud Computing Security Challenges Approaches and Solutions", in Proc The 6th International Symposium on Adhoc and sensor networks Procedia Science Direct publication.
- [2] Qi Zhang. Lu Cheng. Raouf Boutaba. (2010) Cloud Computing: a Perspective Study. *Journal of Internet Service Application* 1: pp 7–18.
- [3] Nick Antonopoulos, Lee Gillam.( 2010) "Cloud Computing Principles, Systems and Applications" Springer International Edition ISBN 978-81-322-0443-5
- [4] Chandrasekaran K. Usha Divakarla. (2013) "Load Balancing of Virtual Machine Resources in Cloud Using Genetic Algorithm" in Proc. *ICCN 2013*, pp. 156–168. Elsevier Publications.
- [5] Garima Joshi, SK Verma. (2015) Load balancing approach in cloud computing using improvised Genetic Algorithm: A soft Computing Approach. *International Journal of Computer applications*. 122, pp 24–28.
- [6] Giuseppe Portalwi. (2014) A power efficient Genetic Algorithm for resource allocation on cloud computing data centres, in Proc. IEEE 3rd International Conference on Cloud Networking(Cloudnet) pp 58-63.
- [7] Kousik Dasgupta. Brototi Mandala (2013). "A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing" A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing" in Proc. International Conference on Computational Intelligence: Modeling Techniques and Applications Procedia Technology pp 341-347.
- [8] Md. Shahjahan Kabir. Kh. Mohaimenul Kabir. Dr. Rabiul Islam. (2015). Process of Load Balancing In Cloud Computing using Genetic Algorithm. *Electrical & Computer Engineering: An International Journal (ECIJ)*. 4(2), pp 57-65.
- [9] Chun Cheng Lin. Hui Hsin Chin. (2014). Dynamic Multiservice load balancing in cloud based multimedia system. *IEEE systems journal*. 8, pp. 1-10.
- [10] Ashish Jain. Narendra S.Chaudari. (2012). Genetic Algorithm based concept design to optimize network balance. *ICTAT Journal on soft computing*. 2(4), pp. 357-360
- [11] Shaminder Kaur. Amandeep Verma. (2012). An efficient approach to genetic algorithm for task scheduling in cloud computing environment. *International Journal of Informational Technology & computer science*. pp 74-79.
- [12] SR Suraj. R Natchadalingam. (2014). Adaptive Genetic Algorithm for efficient Resource management in cloud computing. *International Journal of Emerging Technology & Advanced engineering* 4(2). pp. 21-25.
- [13] CloudAnalyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale Cloud Computing Environment pp. 433-659, Distributed Computing project esse department University of Melbourne. Bhathiya Wickremasinghe
- [14] Rahul Malhotra. Narinder Singh. Yaduvir Singh. (2011). Genetic Algorithms: Concepts, Design for Optimization of Process Controllers. *Computer and Information Science*. 4(2). pp 39-54.
- [15] Jia Zhao. Kun Yang. (2016). A Heuristic Clustering-Based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment. *IEEE Transactions on Parallel And Distributed Systems*. 27(2), pp 305-316.
- [16] Amir Nahir. (2016). Replication-Based Load Balancing. *IEEE Transactions on Parallel And distributed Systems* 27(2), pp. 494-507.
- [17] Jia Zhao. Kun Yang. (2016). A Heuristic Clustering-Based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment. *IEEE Transactions on Parallel And Distributed Systems*, 27(2), pp 305-316.
- [18] Chadi Assi. Sara Ayoubi. Samir Sebbah. (2014). Towards Scalable Traffic Management in Cloud Data Centers. *IEEE Transactions on Communication Systems*. 62(3), pp 1033- 1045.
- [19] Fei Xu. Fangming Liu. (2014). iAware: Making Live Migration of Virtual Machines Interference-Aware in the Cloud. *IEEE Transactions on Computers* , 63(12), pp. 3012-3025.
- [20] Jianying Luo. (2015). Spatio-Temporal Load Balancing for Energy Cost Optimization in Distributed Internet Data Centers. *IEEE Transactions on Cloud Computing*, 3(3), pp 987-997.

## AUTHORS PROFILE



Dr. Surjeet Dalal has been working as an Associate Professor in SRM University. His current research area is Artificial Intelligence, Multi-agent system, Case-based reasoning and Cloud Computing. He has presented various papers in the national/international conferences.

He has published more than thirty papers in the national and international journals.

# ***Real Time Algorithm for the Smart Home Automation Based on the Internet of Things***

Salman Ali Khan<sup>1</sup>, Arhsad Farhad<sup>2</sup>, Muhammad Ibrar<sup>2</sup>, Muhammad Arif<sup>1</sup>

<sup>1</sup>Department of computer science, City University of Science & Information Technology, Peshawar, Pakistan

<sup>2</sup>Department of computer science, COMSATS Institute of Information Technology, Sahiwal, Pakistan

**Abstract:**—Internet of Things (IoT) is enabled by the advancements in the latest technologies including sensors, Radio Frequency Identification (RFIDS), internet protocols and communication technologies. The most important premises of IoT is to connect devices and sensors without human intervention. The proposed smart home automation system differs from other systems by allowing the user to access and operate the system from anywhere around the world through internet along with decision controls according to the needs. In this paper, we propose an algorithm for smart home automation system based on IoT using sensor nodes which are directly connected to Arduino Nano. The algorithm perform some basic local functions such as; Turning ON/OFF the lights based on the motion sensor and generating the alarm based on the gas sensor. In the proposed algorithm the Arduino Mega is connected to the internet using Wi-Fi module to monitor the power consumption of different home appliances and can be controlled from anywhere on the internet. The objective of the proposed system is to provide a low cost and efficient solution for home automation system by using IoT. Results show that the proposed system is able handle all controlling and monitoring function of home.

**Keywords**— *Smart home system, Internet of Things, Motion sensor, Gas sensor, Alarm system*

## I. INTRODUCTION

The Internet of things [1] is a network of physical objects, devices, buildings and other items embedded with electronics devices, sensors and network connectivity that enable these objects to collect and exchange data. By means of IoT, we can control a door, lights, fans and other embedded electronics appliances which are connected to internet even if we are not present in the building.

IoT has revolutionize the business strategies, through this enterprise companies built a proprietary system to collect and organize data that is secure and complaint, and uses connected devices to transmit more data to the system. Parcel service companies use sensors on their vehicle to monitor the speed of a car, mileage, and number of stops. IoT has wide range of application, these applications are categorized as follows:

**Building and Home Automation:-** Buildings and homes are automated by means of temperature. Lights, AC and Fans are controlled based on room temperature observed. Energy Optimization is one of the prime concern of IoT. If we accidentally left our home's lights we can switch them off with IoT technology. Connected Appliances is another application

based on IoT, which includes; smart refrigerators, smart vacuum cleaner etc.

**Smart Cities:-** Residential E – meters is used to improve the efficiency of service and meet residents meet by telling us the amount of electricity consumed and its demand. Smart Grids is used to control the energy generating and/or consuming entities within the electricity network. Smart Street Lights are used to monitor switching, voltage, power, alarm, energy consumption and also set parameters and power line, lamp failure alarm etc. Surveillance cameras used for security purpose, operate in hidden mode, capture live events and save as video file, switching on and off, broadcast video live on internet.

**Manufacturing:-** IoT plays a vital role in manufacturing, these applications include: Real time inventory, the sale and inventory reports, more efficient manufacturing decisions, increase manufacturing agility, save money, boost sales. Asset Tracking help to control asset location, optimizing asset availability, increase efficiency by minimizing stock out cases.

**Wearable's:-** Entertainment google glass is one example which uses exiting google apps like Gmail and google. Healthcare some medical instruments when connected to patient's body helps doctor to monitor him/her if he/she is not physically present there; e.g. smart stethoscope, B.P. apparatus etc. Smart Watch allows to check the metabolism and notifies when metabolism is disturbed by using internet. Location and Tracking/Pet Tracking uses microchips which are placed on person/pet body to track its location easily.

**Health care:-** Remote Monitoring smart medical apparatus help doctor to monitor the patient details and history directly when apparatus are attached to patient's body even when doctor is not available at particular place. Hospital Asset tracking smart cabins helps us to track any particular patient data/record and tells the location of different asset placed in smart cabins. Access Control, a smart medical instrument when applied to patient body tells him/her about present condition and recommended him/her about the type of specialist doctor. Ambulance Telemetry, having connection with internet and treat/diagnose patient in a community as well as in the hospital. Drug Tracking contain sensors in the pill which enables doctors to have more clear and better insights of human body. Predictive Maintenance device connected to patient's body and predict that when and how much medication is needed.

**Automotive:-** Automotive applications are related to smart vehicles Infotainment provide us information about

vehicle and travelling and also entertainment contents. Smart wire corrector helps us to control its wiring and its replacement for better functioning of vehicle.

**Retail:-** Shopping Applications smart screens provide buyers deeper info about the items they are looking at, including up sells. A smart cabin tells us about over stock, out of stock. Supply Chain Management we can connect the supplying machine with the internet and control it with our smart phones from anywhere.

**Environment:-** Water Pollution sensors placed in the water channels, they detect the harmful entities above the safe limit so we can apply remedies. Air Pollution detectors placed at different places to detect the air pollution and alarm at specific air pollution and configuration control. Weather Monitoring Ethernet analogue sensor is used to monitor weather.

The IoT is an emerging technology that is providing a privilege to communicate all around the world. Its objective is Anything, Anyone, Anytime, Anyplace, Any service and any network.

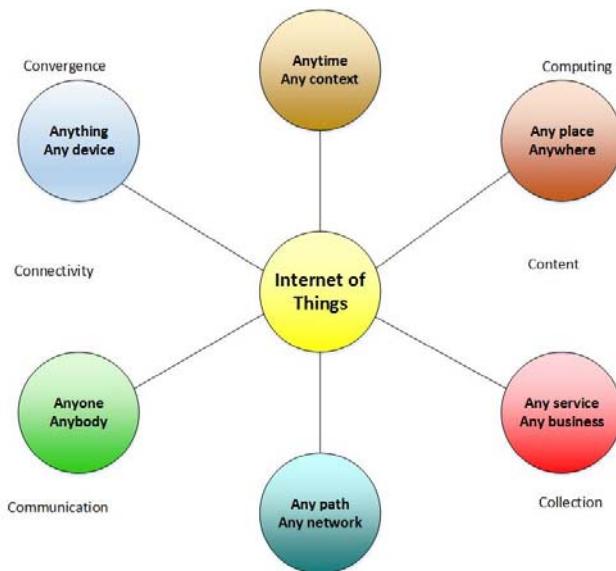


Figure 1. C'S and A's of IoT

Figure 1 describes the combination of C's and A's that discloses, peoples and things can be connected anytime, anyplace with anything and anyone by using any available network and using any service. This elaborates addressing of elements such as convergence, content, repositories, computing, communication and connectivity in the framework where there is continuous interconnection between people and things or between things and things so the A and C elements are tightly coupled.

The main challenges faced by smart home systems are high cost of ownership, poor security, poor manageability, and inflexibility. However, the existing solutions are costly in terms of upgradation and installation. Therefore, an efficient

and cost effective algorithm for the smart home system is needed to provide a feasible solution.

The rest of this paper is organized as follows; related works is discussed in section II, section III outlines the internet of things. Prototype of proposed smart home system is presented in section IV. Section V describes the performance evaluation of the proposed algorithm, whereas the last section concludes this paper.

## II. RELATED WORK

This section presents the existing techniques of smart home system based on IoT, ZigBee, Bluetooth and smart phones.

In [2] the author aims to control the heat, light and temperature of house/office via smart bulbs, smarts heat detectors with the help of IoT. Furthermore, it can also clean house through smart vacuum cleaner and can turn OFF/ON the lights even when users are away at a distance from home. The work presented in [3] targets to find the lost things in house that we forgot where we placed them, thus with the help of trackers these objects are tracked easily and alarm is generated when these objects are moved from its appropriate place. An SMS notification is sent to the owner in case of misplacement. In [4], a car is replaced with a robot to automate the car by monitoring the path way, tracking the location by GPS, as a result, reduces the accidental risks. By using NFC tags [5], visual markers and numeric identifier are placed on posters and panels to get the needed information about the cost, number of seats availability, to automate the ticket system. The work presented in [6] aims to automate the availability of restaurants, their menu and rooms availability. People can reserve hotel rooms directly being staying at home. Patient monitoring system is presented in [7], the goal of this research to automate the medical instruments to receive clinical data for diagnosis purposes. In [8], smart homes are introduced for the comfort, security, convenience and to provide energy efficiency to its occupants. In [9], author describes that with advancement in sensor networks there comes a rapid increase in automation. Everything going to be automated rather than manual. IoT is an emergent network of daily objects which is being monitored and controlled using mobile or internet. In [10], the author states that by making home environment intelligent enough, it can make life easy for disabled and elderly personals. In last few years there is much increase in home automation because of rapid increase in smart phone usage. In [11], the author suggests an idea that in order to make smart home low cost and flexible micro web server based on Arduino Ethernet, hardware interface modules and android based application is used. Using this system authorized users can remotely control and monitor home devices connected through 3G/4G or Wi-Fi. In [12], author proposes smart appliances based on Bluetooth. The smart appliances are controlled by Bluetooth technology with the use of Wi-Fi to efficiently monitor maximum energy consumption appliances. Therefore, to overcome the maximum use of energy consumption by turning OFF/ON those appliances. An information framework for creating a smart city through IoT is presented in [13], aims to enhance the life style. In [14] the author suggests an innovative, detached and flexible ZigBee based smart home system. The system is flexible and

scalable that allows extra home appliances designed by multiple vendors to securely add to the home network with the minimum amount of extra work. The system allows its owners to monitor and control the connected devices locally, through multiple controls like any Wi-Fi enabled device which supports Java or using ZigBee based remote control. Moreover, in this system a common home gateway is used to integrate ZigBee based home automation system and Wi-Fi network. The network is interoperable, simple and flexible due to common home gateway that provides user interface, and remote access to the system.

### III. INTERNET OF THINGS

Machine to Machine (M2M) communication and mobile technologies are the current revolution of the internet. It is the leading phase towards the IoT. IoT [15] will empower the physical objects in a way that they can hear, think and see. These objects can accomplish jobs by talking with each other and by knowing the status of other connected physical devices. Devices are connected so that they can share data and information to co-ordinate decisions. IoT has transforms the objects from traditional to smart by using existing technologies such as pervasive computing, communication technologies, sensor networks, ubiquitous computing and internet protocols.

To get better insight into the real meaning and functionality of the IoT there is a need to understand its building blocks which are as follows and as shown in figure 2:

- i. Identification
- ii. Sensing
- iii. Communication
- iv. Computation (hardware and software)
- v. Services
- vi. Semantics



Figure 2. IoT basic building blocks

According to a new data from the juniper research [16] the number of IoT connected devices in the future will have an increase of more than 285% from 13.4 billion at the end of 2015 and 38.5 billion at the end of 2020.

Home automation is needed to be enhanced by the integration of above mentioned technologies in the real world environment. This enhancement is the part of emerging concept called smart home system. Smart home system will change people's life radically with the new ubiquitous computing and communication technologies. It will provide the devices and systems supported with the smart technologies. It will have the rapid response towards the change in the circumstances without human intervention. This new system will be able to learn from these circumstances.

Smart environment [17] is just like a small world where sensor enabled and network devices are integrated to work and

collaborate continuously to make the lives of its inhabitants more comfortable as it was before.

Every Smart community has important requirements which make that community smart [18]. These requirements are:

**Sensible:** Sensor can sense the environment.

**Connectable:** All the networking devices should be connected for information sharing.

**Ubiquitous:** It should be accessible for the user through the web

**Shareable:** Object should be addressable and accessible as well as data.

**Visible/Augmented:** Information should be visible.

RFID is getting prominence in the identification technology due to its small size, low power, low price, light weight and inexpensive maintenance rates. It is going to be used in many advance fields like pharmaceuticals, manufacturing and retail. It is now in consideration to use RFIDs with the emerging technologies including ubiquitous computing.

In general almost every home is connected to the internet and each home has nearly same objects which can be converted into the smart objects with the meaning of IoT such as doors, windows, fans, meters, security system and automation. All these objects can be converted into the smart objects by using sensors, Quick Response (QR), RFID, Near Field Communication (NFC) and by giving them significant level of intelligence [17]. The purpose is to allow operation of actuators and even power of decision making. All these characteristics of objects can transform a classical home into the smart home system.

### IV. SMART HOME AUTOMATION ALGORITHM

In this section network assumptions, network components and operation of the proposed algorithm is explained.

#### A. Network Components

Automation part is built by using following components: Arduino mega is used for controlling the whole automation part of the smart home automation. It is connected with the other modules used in automation as shown in figure 3. It actuates different devices on the basis of sensors data. Sensors attached to Arduino Mega consists of: Temperature and humidity sensor, Motion Sensor, Electromagnetic door sensor, Gas Sensor, Electromagnetic Relays are used to control and automate the electrical appliances on the basis of sensed data.

Energy monitoring part is used to monitor and control the energy consumption of home appliances especially heavy appliances by using web page and on an LCD screen as shown in figure 4. This part consists of following modules: CC3000 Wi-Fi module is attached to Arduino mega; which aims to provide data on the web page for further processing, actions and controlling.

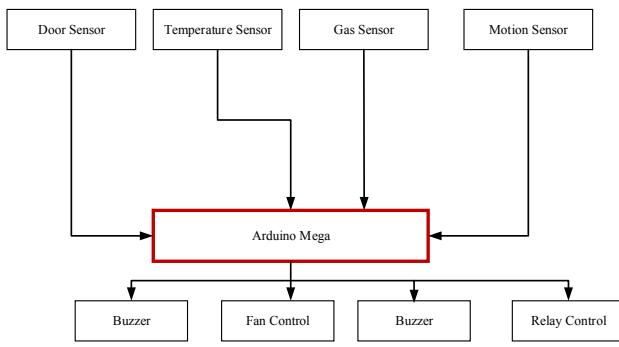


Figure 3. Automation flow

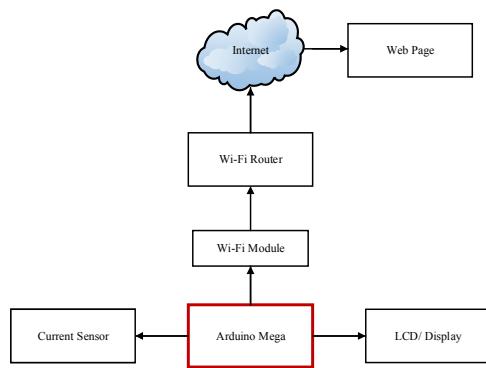


Figure 4. Energy consumption flow

#### B. Network Model

This section briefly discusses the network model of the proposed system. The proposed model is mainly based on two parts; Smart automation and Smart energy monitoring. These models are discussed in the rest of this section. The detailed prototype of proposed model is presented in Figure 5.

**Automation:-** In this portion sensors are connected to the controller (Arduino Mega) and provides automation features of objects such as; Light is turned ON/OFF based on motion sensed by the sensor or any activity observed by the sensor. Furthermore, an alarm is generated when gas leakage is

observed. Also, notifies when the main door is left open for at least 30 seconds.

**Energy monitoring:-** In this portion temperature and current sensors are connected to the controller (Arduino Mega) temperature sensor is used to automate the fan in the room as the fan will automatically turned ON/OFF when the temperature rises to certain value and the fan speed will gradually increase with the increase in temperature. Current sensor is used to monitor the energy consumption of the appliances at home and a Wi-Fi module is used to send the data to the internet and is accessed on a web page. The values of energy consumption and temperature are shown on web page and the control of the appliance is also connected with the web page which can be accessed globally and cab be controlled.

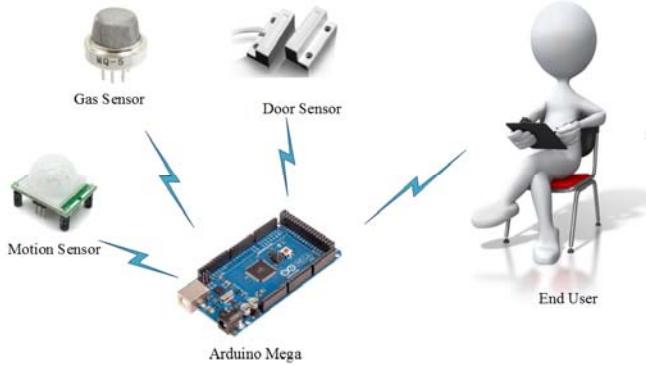


Figure 5. Prototype of smart home automation

**C. Algorithm Working**

The algorithm is written on Arduino Mega chip [19] for the smart home system. In smart home system, the Passive Infrared Sensor (PIR) motion sensor is installed on the top of door to monitor the motion when a person enter into the room. When a person enters into the room the sensor detects motion as a result the lights turned ON. Otherwise, sensor continuously

sense the motion (lines 1-5). A gas sensor is installed in a kitchen for the safety purpose to handle the critical situation. The MQ5 sensor sense the data, if its values is greater than a predefined threshold (1050) as a result alarm is turned ON to notify the user that there is leakage or high amount of gas is detected in the kitchen (lines 6-10). If an intrusion detection is observed, which means the door is observed open for more than 30 seconds then alarm is turned ON to notify the user about the

door (lines 11-16). Lines 16-20 describes the working of Fan which is based on room temperature. If the temperature detected less or equal to 24°C then the Fan is kept OFF, on the other hand, if the temperature exceeds from 24 °C the Fan is turned ON and the speed of Fan is directly proportional with the temperature. Therefore, with the increase of temperature the speed of Fan is increased. The detailed working of algorithm is shown in Algorithm 1.

---

**Algorithm 1: Smart Home System Algorithm**

---

```

1.      if motion sensed by the PIR sensor
        then
2.          Turned ON Light
3.      Else
4.          Keep sensing
5.      end if
6.          if MQ5 gas value greater than or
        equals to 1050 then
7.              Start Alarm
8.          else
9.              Keep sensing
10.         end if
11.         if electromagnetic door sensor
        lost the line of sight connection for
        30 sec then
12.             Start Alarm
13.         else
14.             Keep checking
15.         end if
16.         if temperature less than or
        equals to 24°C then
17.             Turned OFF Fan
18.         else
19.             if temperature greater
        than 24°C then
                Turned ON Fan
                (Speed of Fan increased with the
                increase in temperature)
            end if
20.
21.     end if

```

---

## V. PERFORMANCE EVALUATION

This section briefly describes the performance assessment of the proposed algorithm on a webpage for the temperature and power consumption, while the motion sensor, gas sensor and door sensor works locally.

Figure 6 shows the energy consumption of a 100W bulb from 9am to 11am. Power is calculated by using ampere and voltage as under:

$$Power = Voltage * Current (amperes)$$

Both these values are graphically represented on IoT webpage working on the embedded static IP of the Wi-Fi module. If some appliances is consuming more power and exceeding the threshold (a maximum limit), the user can control the appliance through IoT webpage.

The real time temperature observed on the webpage against time is shown in figure 7. Thus, on the basis of temperature data the fan speed is automatically controlled. On the other, hand the

user can control the socket from the IoT webpage to turn the air conditioner ON or OFF after getting temperature values.

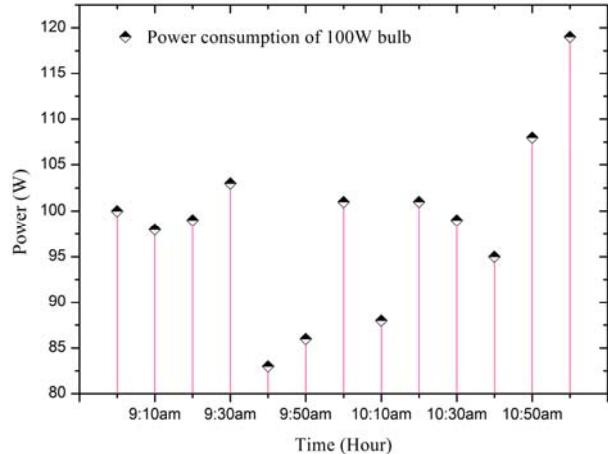


Figure 6. Power consumption variation of 100W bulb

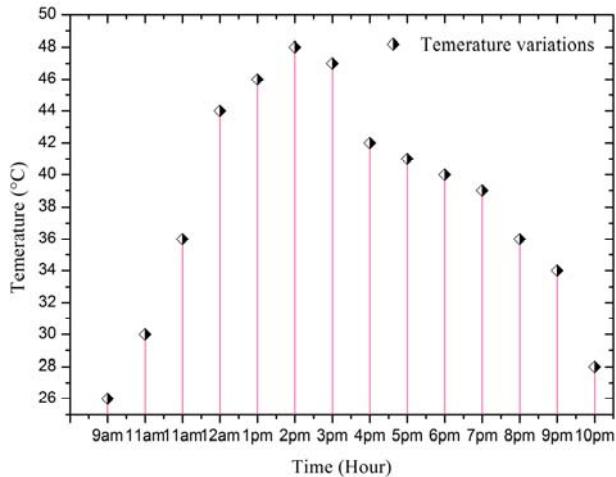


Figure 7. Temperature variation against time

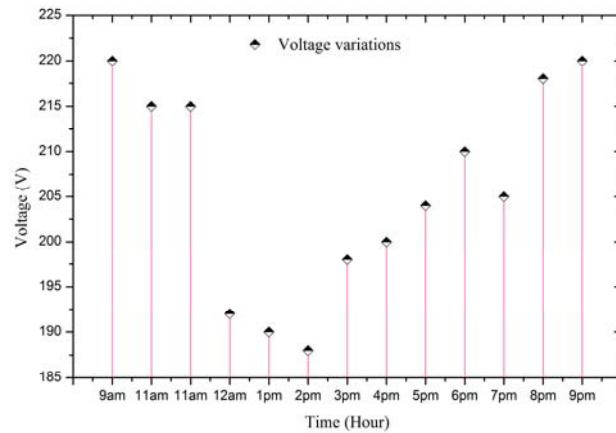


Figure 8. Voltage variations against time

Figure 8 shows the input voltage fluctuation for a 100 watt electric bulb from 9am to 9pm. It is observed that in busy hours

the voltage drops to the minimum limit. As voltage vary, the power consumption of the appliances also vary.

## VI. CONCLUSION

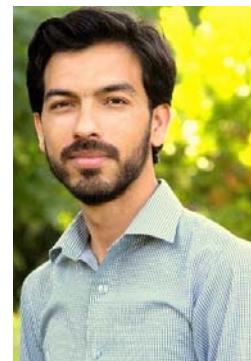
In this paper, an IoT based algorithm is proposed for the smart home system to automate the Fan, monitor the gas leakage and notify by means of an alarm, intrusion detection and energy monitoring. The proposed algorithm was practically implemented on Arduino mega for the testing purpose. The result shows that, the algorithm is capable to observe the motion of a human being, to observe the intrusion by monitoring the line of sight communication between door and sensor. The temperature and power consumption are monitored on a web page globally and can be monitored and controlled being away from home. Simulation results show that, the system is efficient and cost effective in terms of providing reliable information and automation. In future, this work can be to implement in a real world home to automate it as smart home.

## REFERENCES

- [1] Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15), 2787-2805.
- [2] C. Buckl, S. Sommer, A. Scholz, A. Knoll, A. Kemper, J. Heuer, A. Schmitt, Services to the field: an approach for resource constrained sensor/actor networks, in: Proceedings of WAINA'09, Bradford, United Kingdom, May 2009
- [3] R. Yuan, L. Shumin, Y. Baogang, Value Chain Oriented RFID System Framework and Enterprise Application, Science Press, Beijing, 2007.
- [4] G. Broll, E. Rukzio, M. Paolucci, M. Wagner, A. Schmidt, H. Hussmann, PERCI: pervasive service interaction with the internet of things, IEEE Internet Computing 13 (6) (2009) 74–81.
- [5] D. Reilly, M. Welsman-Dinelle, C. Bate, K. Inkpen, Just point and click? Using handhelds to interact with paper maps, in: Proceedings of ACM MobileHCI'05, University of Salzburg, Austria, and September 2005
- [6] D. Niyato, E. Hossain, S. Camorlinga, Remote patient monitoring service using heterogeneous wireless access networks: architecture and optimization, IEEE Journal on Selected Areas in Communications 27 (4) (2009) 412–423.
- [7] Piyare, R. and Lee, S.R., 2013. Smart home-control and monitoring system using smart phone. ICCA, ASTL, 24, pp.83-86.
- [8] K. S. M. Vinay sagar K N, "Home Automation Using Internet of Things," International Research Journal of Engineering and Technology (IRJET), vol. 02, no. 03, pp. 1965-1970, 2015.
- [9] Joshi, M., & Kaur, B. (2015). Web Integrated Smart Home Infrastructure Using Internet of Things. *International Journal of Engineering Research and General Science*, 3(6).
- [10] Piyare, R., 2013. Internet of things: Ubiquitous home control and monitoring system using Android based smart phone. International Journal of Internet of Things, 2(1), pp.5-11
- [11] P. McDermott-Wells, "What is Bluetooth?" *IEEE Potentials*, vol. 23, no. 5, pp. 33–35, Jan. 2005.
- [12] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 112–121, Apr. 2014.
- [13] D. Uckelmann, M. Isenberg, M. Teucke, H. Halfar, and B. Scholz-Reiter, "Autonomous control and the Internet of Things: Increasing robustness, scalability and agility in logistic networks," *Unique Radio Innovation for the 21st Century*, pp. 163–181, 2010.
- [14] Gill, K., Yang, S.H., Yao, F. and Lu, X., 2009. A zigbee-based home automation system. Consumer Electronics, IEEE Transactions on, 55(2), pp.422-430.J. Hurtado-López and E. Casilar, "An adaptive algorithm to optimize the dynamics of IEEE 802.15.4 networks," *Mobile Networks and Management*. 2013, pp. 136–148.
- [15] Al-Fuqaha, A., et al., "Internet of things: A survey on enabling technologies, protocols, and applications", in *Communications Surveys & Tutorials*, IEEE, 2015. 17(4): p. 2347-2376.
- [16] Steffen sorell, "The internet of things" in consumer, industrial and public services 2015-2020. 2015
- [17] Cata, M., *Smart University ,a new concept in the Internet of Things*. IEEE computer society.
- [18] Adamkó, Attila, Tamás Kádek, Lajos Kollár, Márk Kósa, and János Pánovics. "New Challenges in Smart Campus Applications."
- [19] Arduino Mega, <https://www.arduino.cc/en/Main/ArduinoBoardMega>



**Salman Ali Khan** is currently Pursuing his MS degree in Computer Science from City University of Science and Information technology Peshawar, Pakistan. He received his BS degree in Computer Science from City University of Science and Information technology Peshawar, Pakistan in 2014. His research interests include IOT (Internet Of things), wireless sensor network.



**Arshad Farhad** currently working as Lecturer in department of computer, COMSATS Institute of Information Technology, Sahiwal, Pakistan. He has completed his MS degree in Telecommunication and Networking from Bahria University, Islamabad, Pakistan in 2015 with distinction as a Silver Medalist. He received his BS degree in Information Technology from University of Peshawar, Peshawar, Pakistan in 2012. His research interests include design and performance evaluation of communication protocols for wireless ad hoc, wireless body area networks and sensor networks.



**Muhammad Ibrar** currently working as Lecturer in department of computer science, COMSATS Institute of Information Technology, Sahiwal, Pakistan. He has completed his MS degree in Telecommunication and Networking from Bahria University, Islamabad, Pakistan in 2014. He received his BS degree in Telecommunication and Networking from COMSATS Institute of information technology, Abbottabad, Pakistan in 2011. His research interests include wireless ad hoc, wireless body area networks and sensor networks.



**Muhammad Arif** is a PHD scholar at COMSATS Institute of Information Technology Islamabad, Pakistan. Currently he is an Assistant Professor in the Department of Computer Science at City University of Science and Information Technology. His current interests include data warehousing, multimedia image retrieval, medical imaging, pattern recognition, image processing, and computer vision.

# Security of Dynamic and Multipoint Virtual Private Network

Ayoub BAHNASSE, *Najib EL KAMOUN*

**Abstract**—Nowadays, the solutions of virtualizing network infrastructure have become one of the most preoccupations of small, medium and large enterprises. These solutions make the extension of companies' sites possible and easier with a transparent and flexible manner. these solutions allow also the remote access to personal data, stored on several distributed sites, securely.

Dynamic and Multipoint Virtual Private Network, stands for DMVPN, is considered as a main component of these solutions, this technology involves a suite of protocols for a smooth functioning, such as : IPsec, mGRE and NHRP.

Nonetheless, even the considerable security and modularity level of DMVPN solution, this latter suffers from several security issues linked to each components' protocol, which might threaten availability, confidentiality, authentication and integrity of communications.

In this article, we will discuss the key vulnerabilities related to DMVPN technology and the possible countermeasures.

**Index Terms**— DMVPN, Security, Vulnerability, IPsec, NHRP, mGRE.

## I. INTRODUCTION

DMVPN technology [1] is considered as one of the better solutions that an enterprise can deploy to fully connect their several sites taking into account the need to ensure confidentiality, authentication and the integrity of exchanges. For these reason, companies tends to use IPsec standard [2], which, among others, ensures the confidentiality, integrity and authentication of data between two sites in depending on the used protocol : ESP [3] or AHp [4]. Via IPsec tunnels, only IP unicast packets will allowed and secured. To compensate for this shortcoming, the GRE protocol [5] can be deployed as the first encapsulating protocol, this latter supports unicast, multicast and broadcast messages in addition to IP, IPX and AppleTalk protocols, but doesn't ensure any security fundament, for this reason companies can protect GRE packet with IPsec protocol. Unfortunately GRE allows the deployment of site to site tunnels, which implies a daunting task, if not impossible, to interconnect a high numbers of sites, having dynamic assigned IP addresses.

This paper was submitted to review on 2 July 2016.

Ayoub BAHNASSE is with the Faculty of Sciences, University Chouaib DOUKALI, El Jadida, Morocco (e-mail: bahnasse.a@ucd.ac.ma).

Najib EL KAMOUN is with the Faculty of Sciences, University Chouaib DOUKALI, El Jadida, Morocco (e-mail: elkamoun@ucd.ac.ma).

DMVPN technology deals with this issues by using: mGRE protocol, considered as an extension of GRE protocol, which allows the creation of multiple tunnels between routers using a single tunnel interface. NHRP protocol [6], this protocol allows to translate a logical address (IP address of tunnel interface) to an associated public IP address. Dynamic routing protocols such as RIP [7], EIGRP [8], OSPF [9] or others allows the establishment of tunnels and the routing of users data.

To rely on various protocols is an important point, because the independence of these latters ensures a very high of modularity level, for example the three phases of NHRP protocol [10] through which the design of DMVPN architecture can take several forms; HUB to SPOKE, basic SPOKE to SPOKE or extended SPOKE to SPOKE connections, other protocols remains relatively intact.

Although multiple protocols exploitation ensures a high level of modularity and flexibility, on the other hand, as a balance, increases vulnerabilities of the whole network, obviously, vulnerabilities of each DMVPN protocol, these vulnerability can allow to perform several attacks:

- Aiming to prevent communication between DMVPN network equipment;
- Aiming to re-route all communications to hacker computer;
- Aiming to harm servers of network by a falsified or burst records;
- Allowing the hacker to infiltrate on corporate network as a legitimate equipment and access to confidential resources;
- Allowing the decrypt of a part of messages supposed to be protected by encryption mechanisms.

In this article, we will discuss the vulnerabilities of the DMVPN network allowing to perform above attacks by exploiting the weakness of each protocol component (IPsec, mGRE and NHRP). In the second section we will study some IPsec vulnerability and possible security measures to prevent against them, in third and fourth sections, we will do the same, respectively for mGRE and NHRP protocols.

## II. LIMITATIONS OF IPSEC PROTOCOL

IPsec, defined on RFC 2401, is protocol of network layer of OSI model, a complement not a successor of IP protocol, it was originally developed as part of the future IPv4 protocol (IPv6).

IPsec is a tunneling protocol, allows companies to protect against several attacks by offering different level of security such as: confidentiality, integrity, authentication and anti-replay attacks.

IPsec constitutes a main protocol of DMVPN technology, despite its high security level, it remains vulnerable to certain types of attacks aiming to compromise the availability of DMVPN tunnels and to provide additional latencies that can make some transported flows unusable.

#### A. Monotonous data

Despite its robustness in terms of encryption, IPsec can be easily exploited to decrypt certain messages supposed to be protected by the ESP, this is a challenging task but it's also doable by repetitive messages, in fact some protocols and applications generate the same data, an example is the announcements « Router Advertisement » [11] of IPv6 protocol, the Hamming distance, in other words, the number of different bits between packets is small or null, especially for applications based on UDP or IP protocols.

This similarity between packets may threaten the confidentiality of secured data by the ESP, moreover the cryptanalysis relies on this method to break the encryption key especially for conventional encryption protocols on Z/pZ these are protocols in which the discrete logarithm problem is easy to solve. If we consider two packages of which the Hamming distance is insignificant or null and given that the values of the initialization vector "IV" are explicit and the padding's construction rules are known, the only unknown information for the attacker are key to the session and the plaintext. The Hamming distance between messages is low, several attack methods can be performed. We mention as an example, the attack by forced sequences or attack by differential analysis [12], on Cypher Bloc Chaining "CBC" [13] mode, a low distance between IV or a predictable IV also makes communications vulnerable to be decrypted [14]–[16].

Some measures can be taken to complicate the task for attackers, other than decreasing the lifetime of security policy, highlights among them the data compression, this method can be used to hide the message redundancy, it serves to increase the Hamming distance between two clear messages [17]. This method introduced also for IP protocol [18, Sec. 3.1] [19], has shown its efficiency in terms of performances in addition to complicate the task for the pirate, which, in addition to breaking the encryption key will have to find the decompression methods.

#### B. Exploitation of ICMP protocol

Internet Control Message Protocol (ICMP) is a signaling protocol for IP network, knowing that IP is an unreliable and non-oriented connection protocol, i.e. it doesn't ensure any acknowledgment for transmitted data, it's important to ensure a smooth functioning of the network and to notify equipment in case of delivery failures. ICMP performs these tasks, there are currently two versions of ICMP, for IPv4 [20], [21] and IPv6 [22], ICMP messages are classified on two categories,

notification errors messages and management messages.

The ICMP protocol can be used as an attack vector against IPsec threatening:

- The performance of the entire network by reducing or increasing the path MTU;
- The response delay of some applications.

Before proceeding to the delivery of the frames, equipment should detect the PMTU (MTU size of the smallest route to take) this detection can be done by two different mechanisms PMTUD [23] and PLPMTUD [24].

In order to detect PMTU using PMTUD mechanism, source machine create a message and set the DF bit at 1, if the message is too big, intermediate router will reject it and send back an ICMP message Type 4 code 3 to the source machine, the transmitter can then correct the size of the packet and resend it again till no error message is received, further frames will respect this corrected size.

Among the major limitations of this mechanism is that some routers filters by default ICMP which sometimes makes useless the mechanism.

PLPMTUD allows to detect PMTU by using probes principle, source machine create a personalized probe and send it forward to the destination, the correct PMTU is determined once the acknowledgment is received, it is important to know that these probes and their acknowledgments vary according to the application or protocol used for probing, by default TCP protocol is used given its ACK bit and its accuracy compared to other protocols.

By exploiting these mechanisms an attacker can conduct a denial of service attack or reduce throughput between DMVPN equipment.

#### C. Denial of Service of IPsec tunnels

Let's take an example of a network as illustrated on Fig.1, an IPsec tunnel is established between both gateways UCD-GW1 and UCD-GW2, used protocol and mode are respectively ESP and tunnel.

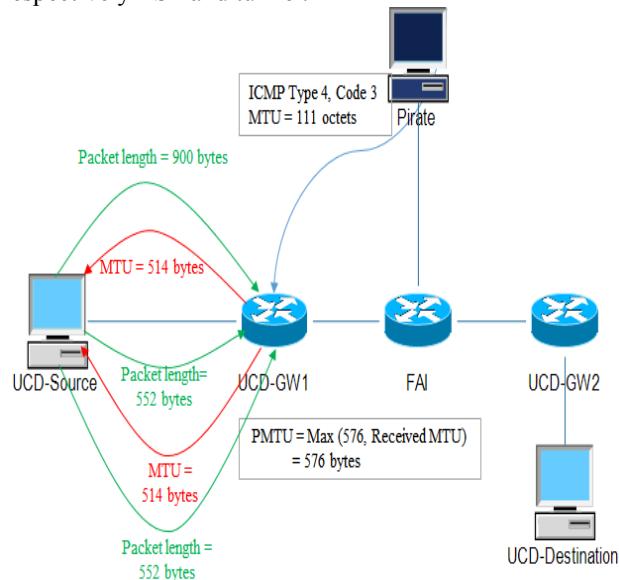


Fig. 1. Denial of service of IPsec exploiting PMTUD

To conduct a denial of service against IPsec using PMTUd as a vector of attack, hacker can follow these steps:

The hacker forge a legitimate ICMP packet Type 3 code 4 “Destination unreachable (need fragmentation)” pretending to be a legitimate IPsec gateway. The detection of exchanged messages between two gateways, can be performed by several methods among them we cite: the alteration of the routing table of the intermediate routers or both IPsec gateways if they run dynamic routing protocols or by the passive listening if the intermediate network is wireless. This forged ICMP packet will be destined to the IPsec gateway UCD\_GW1 which will record the new value of PMTU on the tunnel interface. The recorded value will be the maximum between the value of the received ICMP packet and the default value is 576 bytes (552 on IP level), for each sent packet that exceed 575 bytes, UCD\_GW1 will reject it and will inform back the user.

If both machines UCD-Source and UCD-Destination will exchange files using FTP protocol, a connection must be established, this connection will succeed because three way handshake messages size is less than 576 bytes, as a reminder, the TCP protocol set the DF bit to 1. However, once the communication is established, UCD-Source packets will be rejected because their sizes exceed 576 bytes, therefor UCD\_GW1 will send a notification message indicating that the MTU must be 514 bytes ( $576 - \text{IP} + \text{IPsec+ESP headers length}$ ).

In return, the UCD-Source change its MTU value to 552 byte, since the minimum value of the MTU is 576 bytes including the IP headers.

Such manipulation result in a denial of service and prevent the two machines to communicate.

#### D. Reduction of throughput

In the previous section we demonstrated that PMTUD mechanism may contribute to an IPsec denial of service, in this section we will detail how a hacker can reduce the throughput by exploiting the second mechanism PLPMTUD. Please refer to the following figure.

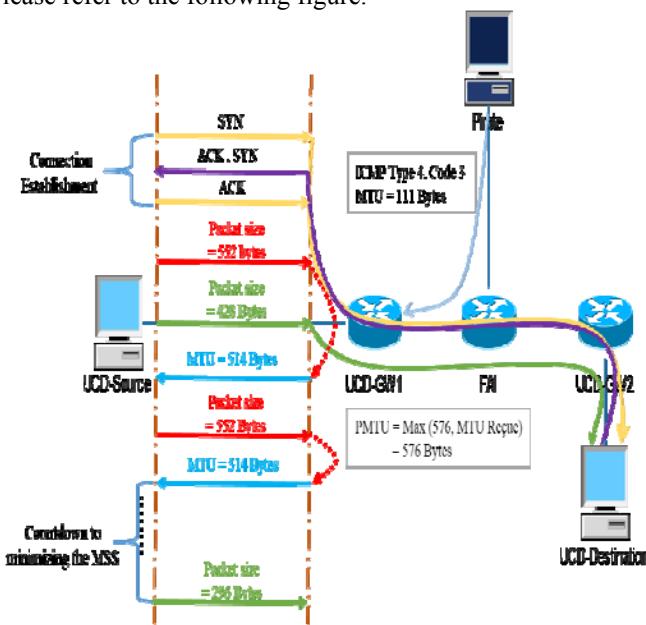


Fig. 2. Denial of service of IPsec exploiting PLPMTUD

As in the previous scenario, the attacker sends an ICMP packet type 3 code 4, to reduce the PMTU of the router. In order to send 900 bytes of FTP data, client UCD-Source must create two packets of 552 and 428 bytes each one, because the MSS size is 512 bytes and TCP (without options) plus IP headers consumes 40 bytes.

FTP is an oriented-connection protocol, it uses three way handshake process to establish the connection, the first three packets will be forwarded correctly because their size is often less than the PMTU of the UCD-GW1 gateway, Instead, the first fragment of 552 bytes exceeds the PMTU stored in security association database, so the UCD-GW1 rejects it and sends an ICMP type 3 code 4 message to the UCD-Source to decrease its PMTU.

UCD-Source will always ignore the ICMP error message because it is based on the acknowledgment of the sent probes to determine the optimum packet size, on its part UCD-GW1 will always reject the packets and return the ICMP error message. After a certain period of time UCD-Source will be forced to reduce its MSS to 256 bytes, and from that moment UCD-GW1 will deliver his packets.

As it can be seen clearly, the PLPMTUD mechanism offers some protection against denial of service but obliges end devices to reduce the rate of data transmission.

#### E. Countermeasures

As preventive measures we recommend:

- To integrate PLPMTUD on UCD-GW1 and UCD-GW2 gateways [25, Sec. 5].
- To force the router to set the DF bit to zero before IPsec encapsulation, in order to allow TCP fragmentation on both gateways [Fig.3].

```
hostname UCD-GW1
!
access-list 100 permit tcp 192.168.1.0 0.0.0.255 192.168.2.0 0.0.0.255
access-list 100 remark classification des communications LAN à LAN
!
class-map match-all NO-DF
match any
match access-group 100
!
policy-map RESET-DF
class NO-DF
set ip df 0
!
interface Serial 0/0/0
service-policy input RESET-DF
service-policy output RESET-DF
```

Fig. 3. Countermeasure against ICMP Attack– Initialization of DF bit

### III. MGRE LIMITATIONS

GRE is a tunneling protocol that allow the encapsulation of several messages as unicast, multicast and broadcast message of different protocols IP, IPX and AppleTalk into IP packet.

Although it is widely deployed today thanks to its efficiency and simplicity in terms of deployment, it suffers from the

same limitations as the IP protocol in terms of security, such as confidentiality, integrity and authentication. GRE deploys an authentication mechanism, but this mechanism remains too weak against dictionary or brute force attacks.

The GRE tunnel is a key component in the terminology of DMVPN network, it can be either point to point or point to multi point, in the first case the addresses of gateways must be statically assigned, in the second case, multi point to point, IP addresses can be dynamically, their detection is performed by the NHRP protocol.

The lack of the confidentiality and the integrity calculation mechanisms represents an opportunity for hackers to divert established tunnels, forge packets and inject them as legitimate packets or prevent communications between stations belonging to different sites or even within the same site.

In this section we will present some of the GRE protocol limitations by exploiting IP options such as "Strict Source List" or "Loose Source List" to prevent or to reset communications between two legitimate machines located behind an mGRE tunnel.

#### A. Strict and loose source routing

strict source routing is used to define all the gateways to use in order to reach a specific destination, this technique assumes that the user masters the topology because if a gateway listed is not reachable the packet will be rejected, by the cons loose routing is also based on a predefined list of routes to use, but it may use other intermediate paths to reach the next hop of the list, in both cases hacker can exploit them to access a private network from the internet. This technique was developed primarily for the following reasons:

1. Network troubleshooting: In case of failure to access to a particular destination, the administrator can record the used routes leading to failures and specify the correct path;

2. Optimizing network performance: It is advisable to forward Best Effort packets by average quality paths and reserve other routes for critical applications.

Most engineers are unaware of the severity of the source list options, and believe they are safe by deploying VPN tunnels between their remote sites, thinking that with the deployment of VPN tunnels private machines can be accessed only by legitimate and preconfigured ones of other side of the tunnel, but this assumption is false, in this section we will prove it in a DMVPN network, by:

1. Leading to a denial or degradation of service of private machines.

2. Preventing current connections between communicating machines.

#### B. Attack Scenario

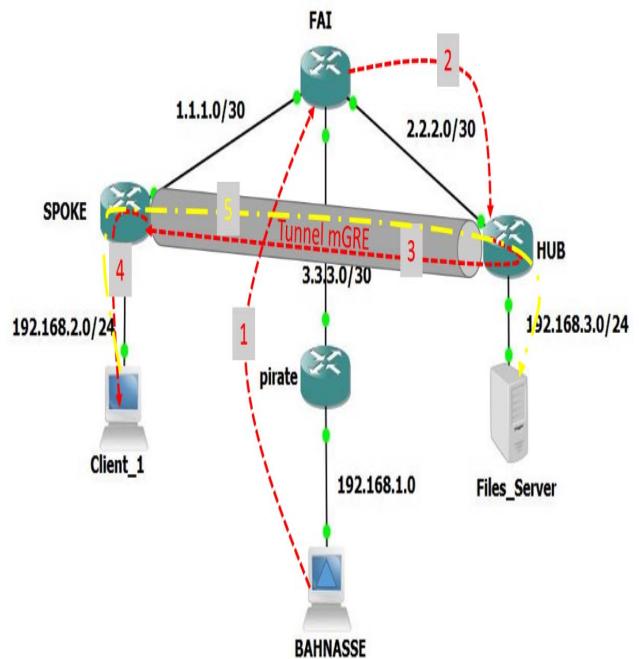


Fig. 4. Exploitation scenario of mGRE tunnels

A DMVPN network is deployed between the HUB and SPOKE routers [fig.4], the subnet address of DMVPN cloud is 172.16.1.0/24 the SPOKE router has the first address, the second address was assigned to the HUB router. Machines are also configured by the second assignable address of each private subnet.

The pirate named BAHNASSE tries to simultaneously flood the remote machines, bypassing authentication, and spoofing its identity.

Step 1: To achieve this objective BAHNASSE created a forged IP packet with the following characteristics:

- Source IP: 192.168.2.2 (Files Server address)
- Final destination: 192.168.2.2 (Client\_1 address)
- Source List: 192.168.3.1, 3.3.3.2, 2.2.2.1, 172.16.1.1, 192.168.2.2.

Step 2: Normally the packet will be delivered to the ISP, which in its turn consults the next hop (2.2.2.1) and delivers it to the HUB. Once the packet received at the HUB, it consults the destination of the packet which will require the creation of the tunnel because the destination has the address 172.16.1.1.

Step 3: A tunnel will then be created automatically, without the obligation that the hacker knowing any information about security parameters of the tunnel, the NHRP IDs and passwords.

Step 4.

```

Internet Protocol Version 4, Src: 2.2.2.1 (2.2.2.1), Dst: 1.1.1.1 (1.1.1.1)
Generic Routing Encapsulation (IP)
Internet Protocol Version 4, Src: 192.168.3.2 (192.168.3.2), Dst: 192.168.2.2 (192.168.2.2), via: 172.16.1.1 (172.16.1.1)
Version: 4
Header length: 40 bytes
Differentiated Services Field: 0x00 (DSCP 0x00: Default; ECN: 0x00: Not-ECT (Not ECN-Capable Transport))
Total Length: 80
Identification: 0x4d2f (19759)
Flags: 0x00
Fragment offset: 0
Time to live: 125
Protocol: ICMP (1)
Header checksum: 0x1c35 [correct]
Source: 192.168.3.2 (192.168.3.2)
Current Route: 172.16.1.1 (172.16.1.1)
[Source GeoIP: Unknown]
[Destination GeoIP: Unknown]
Options: (20 bytes), Strict Source Route, End of Options List (EOL)
Strict Source Route (19 bytes)
Type: 137
Length: 19
Pointer: 16
Recorded Route: 3.3.3.1 (3.3.3.1)
Recorded Route: 2.2.2.2 (2.2.2.2)
Recorded Route: 172.16.1.2 (172.16.1.2)
Destination: 192.168.2.2 (192.168.2.2)
End of Options List (EOL)
Internet Control Message Protocol

```

Fig. 5. Exploitation of DMVPN network- Forged IP packet arrived to the SPOKE router

Fig.5 illustrates the packet received at the SPOKE router, this latter will forward it to the machine 192.168.2.2, even with the presence of access control lists the packet will be accepted because this source IP address is allowed.

Step 5: The Client\_1 machine will then respond with an echo-reply to HUB Site Files Server.

Once the hacker have access to private machines, several attacks may be executed to deny or to prevent communications, in this scenario, BAHNASSE performed a set of attack: Ping of death, UDP flooding and reset TCP connections.

### C. Countermeasures

To fix related vulnerabilities of mGRE protocol, we can rely on two techniques: Unicast Reverse Path Forwarding (uRPF) to prevent the usurpation of IP address and the access control lists to verify the presence of the strict or loose source route options.

#### 1. Unicast Reverse Path Forwarding :

In order to forward a packet, routers checks routing table to ensure that the destination is reachable, uRPF technology allows to check also the source IP address, ensuring that packets come through the interface representing the best route back to the source. For this uRPF performs a search in the CEF table [26], the packet will be destroyed if the packet comes with a different interface from that indicated in the FIB.

In our scenario, we will deploy strict mode of uRPF technique, just because our architecture does not use multiple WAN or tunnel interfaces on which packets can arrive divided

due to a load balancing effect, however, if the deployed DMVPN network is a dual cloud or dual HUB architectures, the loose uRPF mode will be recommended because packets may arrive from several interfaces.

### 2. Filtering of IP option field

Access control lists technology allows to filter, i.e. permit or deny a packet based on its attributes, those latters vary from one layer to another.

In our scenario, we must reject any packet setting the option field to 131 and 137 respectively representing the loose source routing and strict source routing.

Cisco command lines, addressing this need are:

```

ip access-list extended UCD_TRANSIT_IN
    deny ip any any option ssr
    deny ip any any option lsr
    permit ip any any
!
interface Serial 0/0
    ip access-group UCD-TRANSIT-IN in

```

### IV. NHRP LIMITATIONS

The NHRP protocol defined by the ROLC workgroup acronym of Routing Over Large Clouds, is a protocol designed primarily for resolving ATM IP addresses to make connections beyond the IP, it was introduced by Cisco for DMVPN to translate a tunnel IP address into a public IP address, this resolution occurs mainly between NHC client and NHS server.

As the ARP protocol, NHRP protocol is based on a cache to store the mappings of tunnel addresses and their associated public addresses of each equipment belonging to the same DMVPN cloud, which means that this protocol is vulnerable primarily for two types of attacks: denial service and the poisoning of the NHRP cache.

#### A. NHRP Denial of service

According to [27, Sec. 5.3.4.4] all NHS are vulnerable to denial of service attacks, leading to an overload of the cache, preventing legitimate requests to be handled properly, this attack may prove difficult to perform because the attacker must know the identifier of the cloud and the authentication keys of NHRP and mGRE protocols, however hacker can proceed differently by performing a buffer overflow attack through which the process NHRP overwrites memory adjacent to a buffer that should not have been modified intentionally or unintentionally, according to Cisco Bug ID CSCin95836 [28] after this attack router's behavior becomes unpredictable, some versions of the operating systems of Cisco routers can restart or allow the attacker to execute arbitrary malicious code on the router.

#### B. NHRP cache poisoning

NHRP cache poisoning consists of changing dynamic records or adding new ones in the resolver caches, either on the NHC or NHS, this attack is commonly used to divert

traffic from one site to another intermediary, often the hacker site, in order to succeed this attack hacker must know the ID of NHRP cloud and used authentication strings of the tunnel. Unfortunately the latter are vulnerable to a brute force and dictionary attacks, in fact the length of the NHRP authentication key can't exceed 8 characters, this make task easier for the hacker who can lead a distributed attack to break the authentication strings quickly.

There are two main categories of attacks that a hacker can lead:

1. Poisoning NHRP cache of both NHC and NHS by false NBMA addresses, this manipulation can stop all current communications or reroute them to a tierce site who might not belong to the DMVPN network.
2. Man in the middle between two sites, by sending to NHC source and NHC destination false NHRP updates having as NBMA address the pirate address. These victims will always pass through the hacker to communicate each other, this latter can read, write and delete data if no security protocols are deployed.

### C. Countermeasures

In the user side, three countermeasures are possible:

1. Although it is quite difficult to block traffic transiting the network, the use of extended access control list can be useful to permit only NHRP replies coming from the legitimate DMVPN network, using the following command lines:

```
ip access-list extended NHRP_PROTECT
permit 54 host [Trusted Net] [Wildcard mask] any
permit 47 host [Trusted Net] [Wildcard mask] any
deny 54 any any
deny 47 any any
interface tunnel 1
ip access-group NHRP_PROTECT in
```

2. The above solution remains weak against spoofing attacks, It is for this reason that it is important to reinforce it by filtering options field and the use of URPF technology detailed on sections **Error! Reference source not found.** and 2

3. Upgrade the IOS with the last patched version.

## V. CONCLUSION

Through this paper, we presented the vulnerabilities related to DMVPN technology, mainly those related to DMVPN protocols such as IPsec, mGRE and NHRP. We showed that a hacker can exploit them to compromise the availability, performance, confidentiality and the integrity of the whole DMVPN network.

Concerning IPsec vulnerabilities, we presented those that can threaten the availability, reliability and confidentiality of the data encapsulated in an IPsec tunnel by exploiting ICMP variants and monotonous data, to workarounds these attacks several measures can be taken, such as the use of PLPMTUD mechanism instead of PMTUD and use data compression before encryption.

The mGRE protocol vulnerabilities have been detailed in the third section, through which we have illustrated the severity of strict and loose source route options, through these two options the hacker can penetrate the legitimate tunnels without having to authenticate and can therefore perform several attacks on private sites, as security measures we recommended the use of ACLs to filter the IP option fields, and the use of URPF technology to protect against IP spoofing.

In the last section, we discussed the vulnerabilities related to NHRP protocol, including cache poisoning and buffer overflow attack. These limits can be remedied by using the anti-spoofing mechanisms namely extended access control lists, strengthened by URPF technology.

## REFERENCES

- [1] M. L. (CCIE.), "Scaling and optimizing IPsec VPNs," in *Comparing, Designing, and Deploying VPNs*, Adobe Press, 2006, p. 523.
- [2] S. Kent and R. Atkinson, "Security Architecture for the Internet Protocol," RFC Editor, RFC2401, Nov. 1998.
- [3] S. Kent, "IP Encapsulating Security Payload (ESP)," RFC Editor, RFC4303, Dec. 2005.
- [4] S. Kent, "IP Authentication Header," RFC Editor, RFC4302, Dec. 2005.
- [5] D. Farinacci, T. Li, S. Hanks, D. Meyer, and P. Traina, "Generic Routing Encapsulation (GRE)," RFC Editor, RFC2784, Mar. 2000.
- [6] J. Luciani, D. Katz, D. Piscitello, B. Cole, and N. Doraswamy, "NBMA Next Hop Resolution Protocol (NHRP)," RFC Editor, RFC2332, Apr. 1998.
- [7] C. L. Hedrick, "Routing Information Protocol," RFC Editor, RFC1058, Jun. 1988.
- [8] R. White, J. Ng, D. Slice, S. Moore, and others, "Enhanced Interior Gateway Routing Protocol," 2014.
- [9] J. Moy, "OSPF Version 2," RFC Editor, RFC2328, Apr. 1998.
- [10] Cisco Systems, "Developmental Phases of DMVPN and NHRP," in *NHRP*, Cisco Press, 2007, pp. 6–8.
- [11] T. Narten, E. Nordmark, W. Simpson, and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)," RFC Editor, RFC4861, Sep. 2007.
- [12] S. Aumont, R. Dirlewanger, and O. Porte, "L'accès sécurisé aux données," *3ème journée des réseaux - IRES 99*, Montpellier, p. 16, Nov-1999.
- [13] R. Pereira and R. Adams, "The ESP CBC-Mode Cipher Algorithms," RFC Editor, RFC2451, Nov. 1998.
- [14] P. Karn, P. Metzger, and W. Simpson, "The ESP DES-CBC Transform," RFC Editor, RFC1829, Aug. 1995.
- [15] C. Madson and N. Doraswamy, "The ESP DES-CBC Cipher Algorithm With Explicit IV," RFC Editor, RFC2405, Nov. 1998.
- [16] S. Vaarala, A. Nuopponen, and T. Virtanen, "Attacking predictable IPsec ESP initialization vectors," in *Information and Communications Security*, Springer, 2002, pp. 160–172.
- [17] D. A. Wagner and S. M. Bellovin, "A programmable plaintext recognizer," 1994.
- [18] "hjp: doc: RFC 4301: Security Architecture for the Internet Protocol." [Online]. Available: [http://www.hjp.at/doc/rfc/rfc4301.html#sec\\_3.1](http://www.hjp.at/doc/rfc/rfc4301.html#sec_3.1). [Accessed: 15-Mar-2016].
- [19] "hjp: doc: RFC 3173: IP Payload Compression Protocol (IPComp)." [Online]. Available: <http://www.hjp.at/doc/rfc/rfc3173.html>. [Accessed: 15-Mar-2016].
- [20] F. Gont and C. Pignataro, "Formally Deprecating Some ICMPv4 Message Types," RFC Editor, RFC6918, Apr. 2013.
- [21] J. Postel, "Internet Control Message Protocol," RFC Editor, RFC0777, Apr. 1981.
- [22] A. Conta, S. Deering, and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification," RFC Editor, RFC4443, Mar. 2006.
- [23] J. C. Mogul and S. E. Deering, "Path MTU discovery." [Online]. Available: <https://tools.ietf.org/html/rfc1191>. [Accessed: 25-Feb-2016].
- [24] J. W. Heffner and M. Mathis, "Packetization Layer Path MTU Discovery." [Online]. Available: <https://tools.ietf.org/html/rfc4821>. [Accessed: 25-Feb-2016].
- [25] L. Jacquin, V. Roca, and J.-L. Roch, "ICMP: an Attack Vector against IPsec Gateways," HAL, 2013.

- [26] N. Stringfield, R. White, and S. McKee, *Cisco Express Forwarding*. Pearson Education, 2007.
- [27] J. Luciani, D. Katz, D. Piscitello, B. Cole, and N. Doraswamy, “NBMA Next Hop Resolution Protocol (NHRP),” RFC Editor, RFC2332, Apr. 1998.
- [28] Cisco Systems, “Cisco IOS Next Hop Resolution Protocol Vulnerability,” *NHRP Vulnerability*. [Online]. Available: <http://cisco.com/c/en/us/support/docs/csa/cisco-sa-20070808-nhrp.html>.

# Predominant Factors Influencing Software Effort Estimation

Sumeet Kaur Sehra <sup>1</sup>, Yadwinder Singh Brar <sup>2</sup>, Navdeep Kaur <sup>3</sup>

<sup>1</sup> Research Scholar, I.K.G. Punjab Technical University,

Jalandhar, Punjab, India

Assistant Professor, Guru Nanak Dev Engineering College,

Ludhiana, Punjab, India.

<sup>1</sup> sumeetksehra@gmail.com

<sup>2</sup> Professor, Guru Nanak Dev Engineering College,

Ludhiana, Punjab, India

<sup>2</sup> braryadwinder@yahoo.co.in

<sup>3</sup> Associate Professor, Shri Guru Granth Sahib World University

Fatehgarh Sahib, Punjab, India

<sup>3</sup> drnavdeep.sggswu@gmail.com

**Abstract**—Software effort estimation is a crucial task affecting the success of a software project. Inaccurate estimates can lead to incomplete, over-budgeted and failed projects. Accurate estimate of software development effort, which has always been a challenge for both the software industry and academia. Many models have been developed and validated by researchers to estimate the effort. But none of the models are successful for all types of projects and every type of environment. The reason is the prevalence of some fundamental issues which have a negative influence on the effort estimation process. In this paper, some of the issues affecting software effort estimation have been discussed.

**Index Terms**—Software Effort Estimation, Estimator, Factors, Environment, Dataset

## I. INTRODUCTION

Software effort estimation has become a challenging research area and has gained tremendous importance in the last two decades due to its imperative necessity in software analysis. Software being an invisible and most expensive component of IT industry has encouraged the pre-estimation process before actual development process start. Estimation is a subject to assumptions and approximation. Estimation is a process of getting a rough idea about the how much effort and cost could be required for developing a software. So it has become a part and parcel of software development life cycle, which cannot be avoided at any cost. Good estimations are essential for both developer as well as customer. So different models, processes and tools are used for the same. Although many technological advances have taken place, but effort estimation task is still an art.

the effort required for developing a software depends upon many factors, including human, technical, environmental, political etc. Sheta *et al* [1] have suggested that process of estimating the effort can be accomplished by a series of systematic steps which can provide estimates with manageable risk. Accurate and reliable estimate act as the base for

effective project planning and control as deviation from actual estimation is acceptable up to some extent but huge deviations lead to project cancellation. Jorgensen [2] has reported that Failures in accurately estimating effort lead to cancellation of nearly 40 percent of industry software projects. Jorgensen and Sheppard [3] have also reported that the inability of industry not providing accurate estimates lead to an average of 89% cost overruns. So it becomes necessary to have an accurate and excellent estimations as a single wrong estimate ultimately takes the project to a dead end. Previous research analyzed differences in accuracy of effort estimation depending on the estimation technique applied

Attempts to develop models for estimating software development effort and assessing the impact of productivity factors have been the focus of much research. In order to solve the problems of making some accurate software project predictions and supporting managers in decision making, many estimation models have been developed over the last three decades [3]. Attempts to develop such models for estimating software-development effort and assessing the impact of productivity factors have been the focus of much research.

The paper has been divided into different sections. The next section discusses about the problems encountered in software effort estimation . The related work is also discussed along with. Last section elaborates the conclusion of the review.

## II. FACTORS AFFECTING SOFTWARE EFFORT ESTIMATION

Software Estimation is really a hard task to do as a developer is dealing with an intangible component. The estimations are in the form of blueprints only, what would be the exact result is not possible to know at the start. Sometimes a very good result could be there, but on the other hand a worst result could also be there. A number of issues including, practical, measurement, and modeling contribute towards the problems

encountered in the software effort estimation process. Following is the discussion of some of the important affecting software effort estimation.

#### A. The effect of uncertainty

The estimators are supposed to have the detailed knowledge of the system to be developed before estimating the effort required to develop it. But this is inherently more difficult to understand and estimate a product or process that cannot be seen and touched [4], [5]. Estimation is a futuristic prediction of a metric to be used in the software to be developed, thus resulting in uncertainty [6]. Every software is unique in itself, its development process, requirements and demands from the customer side, everything vary from one software to another. Traditionally it was easy to estimate as software development process was easy because of few requirements and every other software's requirements get matched with one another because of low demands and functionality. But now advance functionality with advance databases, security systems, full fledge networking, high performance, high reliability, high support with internet and many more are the core requirements that lead the developers to advanced level of development at which every single requirement completely differs from another that create problems while making estimates at the initials stages. Apart from this programming method, design decisions, type of design methodology has an impact on estimations as these are not known at the beginning. Trendowicz and Jeffery [7] found that when the variance in time and budget was large, the estimations were below the actual values. Holta and Otto [8] have discussed that the accuracy of the estimate increases as a project is developed as more information becomes available as shown in figure 1.

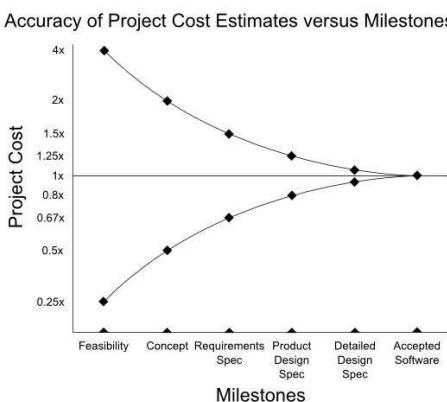


Fig. 1. Cone of uncertainty

#### B. Software Sizing

All estimation models require software size metric to estimate software development effort, cost and schedule. Size is generally measured in terms of Lines of Code (LOC) or Function Points (FP). According to Galorath and Evans [9], estimation is affected due to inaccurate size metric of a software project. As in the beginning, it is not possible to

know about how many lines it is going to take to complete the project or how many functions will be included in the software. Size estimation is complex and its results should be constantly updated with the actual tally throughout the life cycle [10]. Apart from this the style of coding also varies from programmer to programmer. Some programmer takes a single line to code while another can take more than one line. This leads to variation in LOC. There are many other factors that affect the Line of Code. In case of Function Point, number of functions also vary from one to another; also it is not known in initial level that how complex the function is going to be. So these size metrics which are the main input parameter are not initially known, create problem while estimating because estimates are based on previous experiences only. One's past experience, judgment and analytical observation play an important role in this.

#### C. Estimator Experience

Estimator's experience is considered one of the most important factors affecting estimation accuracy [11]. Generally estimations are made by senior management, but not by the persons who are going to develop it. Although they have experience of years, but they sometimes neglect the fact of feasibility [12]. Biasing in software development affects the true estimate, which ultimately leads to delay as well as over budgeting. Human biases influence estimates to great extent generally negative. Peeters and Dewey have suggested that poor and inaccurate estimates are the result of biases, which can affect the success or failure of a project [13]. Both conscious and unconscious decisions have an impact on degree of bias and degree of bias also changes for individual expert [14]. Estimators find it easier to depend on their personal memory rather than documented facts, standards, or arithmetic rules [15]. McConnell [16] has reviewed that inaccurate estimates are also the consequence of the gap between the estimators's perception of the required skill and the actual skill used in the project. Many studies have been reported in which the relation between estimator's characteristics and estimated effort have been discussed. Lederer and Prasad [17] have proposed a model depicting the relationship between estimation errors and managerial factors. A relation between team size & skills and effort has been proposed by Sehra and Kaur [18].

#### D. Incomplete and Inconsistent Data

Data is needed to construct effort estimation models and to validate them so the data produced during effort estimation is recorded by the estimators. By using recorded and documented data, estimators can apply better and model based estimation technique. Jorgensen has reported that documented data usage helps in less human and situational biases [19]. But the data collected and documented is not complete in all aspects. Strike *et al* [20] have concluded that documented estimation data sets normally have considerable missing values. Experts generally face the problem of incomplete, inaccurate and inconsistent data collected from previous projects. According to Sheppard

and Cartwright [21], a major obstacle in the effective estimation is the absence of reliable and systematic historic data. The reasons for lack of reliable data can be that it is chronophagous and tedious task. Moreover, there is no technique to ensure the accuracy, consistency and completeness of data collected. Also data is collected and recorded in different formats by many individuals, which may result in inconsistency of data. Since technology advances very frequently, another problem is that worth of collected data reduces over time. Moreover, data available during initial phases are incomplete and inaccurate, making metrics measurement quite hard [22].

#### E. Dependency on Environment

Research in software estimation has identified that validation and calibration of estimation with local data produces more accuracy than generic model [4], [23]. The Model has to be calibrated to a new environment as it is not possible for a single model to fit all situations and environments [21]. Lopez *et al* [24] have suggested that because no single software development estimation technique is best for all situations, better estimates can be produced by comparison of the results of various approaches. Since software estimation models are commonly derived from empirical data that are usually collected from diversified sources, they cannot be generalized well for all types of environments [25]. Walkerden and Jeffery [26] have concluded that software effort estimation models perform better and quickly in the environment for which they have been calibrated.

#### F. Frequent changes in a software requirements

Most of the projects have changes in requirements throughout the development process, but estimations so made are not changed for the whole project. It is going to affect the project in one or another way for sure, it can be good but the worst also. Basri *et al* [27] have suggested that acceptance of too many changes in the requirements delay project completion and result in over-budgeting. Bhatti *et al* [28] have studied the effect of requirement change to the subsequent phases of software development life cycle (SDLC). They concluded that effect of change in the requirement during implementation of existing requirements propagates to other subsequent phases of SDLC. This propagation effect directly affects the cost and effort of the software to be developed as proposed by [29]. Damian *et al* [30] have proposed a positive relationship between improved requirement engineering process and software productivity.

### III. CONCLUSION

Accurate software effort estimation has always been a challenge for researchers and academicians. Various models and techniques have been developed over the years. All the models perform differently in different environments. There is still no single method, which has established itself to the fullest to consistently deliver an accurate estimate. Many factors play role in software effort estimation process. Some of the issues inherently faced in the software effort estimation

have been discussed along with the related research. Some of the problems can be resolved by taking appropriate counter measures, but still there is some chance of inaccuracy in the estimates. The possible solution to the problems is the development of hybrid models combining different estimation techniques.

### REFERENCES

- [1] A. Sheta, D. Rine, and A. Ayesh, "Development of software effort and schedule estimation models using Soft Computing Techniques," in *IEEE Congress on Evolutionary Computation*, 2008, pp. 1283–1289.
- [2] M. Jrgensen, "Forecasting of software development work effort: Evidence on expert judgement and formal models," *International Journal of Forecasting*, vol. 23, no. 3, pp. 449–462, Jul. 2007. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S016920700700074X>
- [3] M. Jorgensen and M. Shepperd, "A Systematic Review of Software Development Cost Estimation Studies," *IEEE Transactions on Software Engineering*, vol. 33, no. 1, pp. 33–53, 2007.
- [4] B. Boehm, A. Brown, R. Madachy, and Y. Yang, "A software product line life cycle cost estimation model," in *International Symposium on Empirical Software Engineering*, Redondo Beach, California, Aug. 2004, pp. 156–164.
- [5] K. Kumari, "Software cost estimation techniques," *International Journal of Emerging Research in Management & Technology*, vol. 3, no. 4, pp. 104–108, apr 2014.
- [6] B. Kitchenham and S. Linkman, "Estimates, uncertainty, and risk," *IEEE Software*, vol. 14, no. 3, p. 69, 1997.
- [7] R. J. Adam Trendowicz, *Common Factors Influencing Software Project Effort*. S, 2014, ch. Software Project Effort Estimation, pp. 47–80.
- [8] K. Holta-Otto and C. L. Magee, "Estimating factors affecting project task size in product development—an empirical study," *IEEE Transactions on engineering management*, vol. 53, no. 1, pp. 86–94, 2006.
- [9] D. D. Galorath and M. W. Evans, *Software sizing, estimation, and risk management: when performance is measured performance improves*. CRC Press, 2006, ch. The Problem, pp. 1–24.
- [10] S. Malathi and S. Sridhar, "Analysis of size metrics and effort performance criterion in software cost estimation," *International Journal of Computer Applications*, vol. 40, no. 3, pp. 32–37, 2012. [Online]. Available: <http://research.ijcaonline.org/volume40/number3/pxc3877172.pdf>
- [11] O. Morgenshtern, T. Raza, and D. Dvir, "Factors affecting duration and effort estimation errors in software development projects," *Information and Software Technology*, vol. 49, no. 8, pp. 827–837, 2007.
- [12] R. R. Nelson and M. G. Morris, "It project estimation: Contemporary practices and management guidelines," *MIS Quarterly Executive*, vol. 13, no. 1, 2014.
- [13] D. Peeters and G. Dewey, "Reducing bias in software project estimates," *Journal of Defense Software Engineering*, vol. 13, no. 4, pp. 20–24, 2000.
- [14] K. Moløkken and M. Jørgensen, "Software effort estimation: Unstructured group discussion as a method to reduce individual biases," in *The 15th Annual Workshop of the Psychology of Programming Interest Group*, 2003, pp. 285–296.
- [15] L. C. Briand and I. Wieczorek, "Encyclopedia of software engineering," *Encyclopedia of software engineering*, 2002.
- [16] S. McConnell, "The best influences on software engineering," *IEEE Software*, vol. 17, no. 1, pp. 10–17, 2000.
- [17] A. L. Lederer and J. Prasad, "A causal model for software cost estimating error," *IEEE Transactions on Software Engineering*, vol. 24, no. 2, pp. 137–148, 1998.
- [18] N. Kaur and S. K. Sehra, "Investigating relationship between software effort estimation and team parameters," *International Journal of Advance Research in Education, Technology & Management*, vol. 2, no. 1, pp. 139–142, 2014.
- [19] M. Jrgensen, "A review of studies on expert estimation of software development effort," *Journal of Systems and Software*, vol. 70, no. 1, pp. 37–60, 2004.
- [20] K. Strike, K. El Emam, and N. Madhavji, "Software cost estimation with incomplete data," *IEEE Transactions on Software Engineering*, vol. 27, no. 10, pp. 890–908, 2001.

- [21] M. Shepperd and M. Cartwright, "Predicting with Sparse Data," *IEEE Transactions on Software Engineering*, vol. 27, no. 11, pp. 987–998, 2001.
- [22] D. Kashyap, A. Tripathi, and A. Misra, "Software development effort and cost estimation: Neuro-fuzzy model," *Journal of Computer Engineering*, vol. 2, no. 4, pp. 12–14, 2012.
- [23] S. Basha and D. Ponnurangam, "Analysis of empirical software effort estimation models," *International Journal of Computer Science and Information Security*, vol. 7, no. 3, pp. 68–77, 2010. [Online]. Available: <https://arxiv.org/ftp/arxiv/papers/1004/1004.1239.pdf>
- [24] C. Lopez-Martin, C. Yáñez-Márquez, and A. Gutierrez-Torres, "A fuzzy logic model for software development effort estimation at personal level," in *Mexican International Conference on Artificial Intelligence*. Springer, 2006, pp. 122–133.
- [25] V. Nguyen, B. Steele, and B. Boehm, "A constrained regression technique for cocomo calibration," in *Second ACM-IEEE international symposium on Empirical software engineering and measurement*. ACM, 2008, pp. 213–222.
- [26] F. Walkerden and R. Jeffery, "Software cost estimation: A review of models, process, and practice," *Advances in Computers*, vol. 44, pp. 59 – 125, 1997. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S006524580860337X>
- [27] S. Basri, N. Kama, and R. Ibrahim, "A novel effort estimation approach for requirement changes during software development phase," *International Journal of Software Engineering and Its Applications*, vol. 9, no. 1, pp. 237–252, 2015.
- [28] M. W. Bhatti, F. Hayat, N. Ehsan, S. Ahmed, A. Ishaque, and S. Z. Sarwar, "An investigation of changing requirements with respect to development phases of a software project," in *International Conference on Computer Information Systems and Industrial Management Applications*. IEEE, 2010, pp. 323–327.
- [29] D. Zowghi and N. Nurmuliani, "A study of the impact of requirements volatility on software project performance," in *Ninth Asia-Pacific Software Engineering Conference*. IEEE, 2002, pp. 3–11.
- [30] D. Damian, J. Chisan, L. Vaidyanathasamy, and Y. Pal, "Requirements engineering and downstream software development: Findings from a case study," *Empirical Software Engineering*, vol. 10, no. 3, pp. 255–283, 2005.

# Development of an Autopsy Forensics Module for Cortana Artifacts Analysis

Bernard Allen Sabernick III

*Department of Computing Security, Rochester Institute of Technology  
Rochester, NY USA  
Allen.Sabernick@rit.edu*

**Abstract**—Forensic tools are a critical component of a forensic investigator's job. As new features are added in operating systems, these tools need to adapt and be updated to analyze these new features. Microsoft recently released its Windows 10 operating system with a new voice activated personal digital assistant called Cortana. Cortana is capable of storing information about a user which could be used as evidence in criminal cases. Using the open source forensic tool Autopsy, this information is currently not being gathered in an effective manner. In order to address this problem, this paper proposes enhancements to the Autopsy tool to allow forensic investigators to collect the needed information about Cortana and analyze it more quickly.

**Keywords:** Digital Forensics, Windows 10, Cortana, Autopsy, Development

## I. INTRODUCTION

With the release of Microsoft Windows 10, security experts are faced with many new challenges in investigating criminal activities that occur on these operating systems. Forensics investigators are at a disadvantage when it comes to having the tools necessary to perform an in-depth investigation of Windows 10 because forensics tools tend to lag behind in the capabilities of new operating systems. Using the open source forensic tool Autopsy, an investigation will be conducted to determine the potential pitfalls and missing information not being collected from Windows 10 and what improvements are needed so that forensics experts are able to provide unwavering evidence in a court of law in cases involving criminal activities on Windows 10.

A study of forensics and existing research is conducted in section II. In section III a proposal for the new research techniques is made followed by section IV which briefly describes how the research will be evaluated. Lastly, section VI documents the conclusions of the research.

## II. BACKGROUND & RELATED WORK

### A. Digital Forensics

A basic understanding of computer forensics is needed before an assessment of the tools in the industry can be reviewed. The first digital forensics research workshop in 2001 defined digital forensics as the scientifically derived and proven method toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations [1]. The authors of Digital Forensics with Open Source Tools greatly reduce this to meaning that digital forensics are examinations of computer generated data [2].

As some have argued, the golden age of digital forensics is coming to an end [3]. A coming of age has happened whereby data that was once easy to obtain on a computer is now a challenge. Research suggests a few reasons for this new trend. First, data security is much more common and encrypted data files, even full disk encryption, is making it difficult for forensics investigators to do their job. Second, the media on which the data is being stored has evolved. Spinning hard drives are being replaced by solid state drives which act and behave differently than their more mechanical predecessors. Even more difficult for forensics experts is the fact that much data is now stored in the cloud. Still, both of these reasons would be mute without the third reason. Operating systems have always been the door way into how and where data is stored. The complexity of the operating system has greatly increased the difficulty in finding the information needed. With the release of Windows 10, the tools needed to examine and find relevant artifacts need improvement as well.

### B. Windows 10

Microsoft Windows 10 is the newest operating system from Microsoft. Continuing the trend of Windows 8, Windows 10 can be used as a desktop, tablet or phone operating system. Encouraging users to adopt Windows 10 on mobile devices, Microsoft added a voice activated personal digital assistant called Cortana [4]. Microsoft is even advertising Cortana as

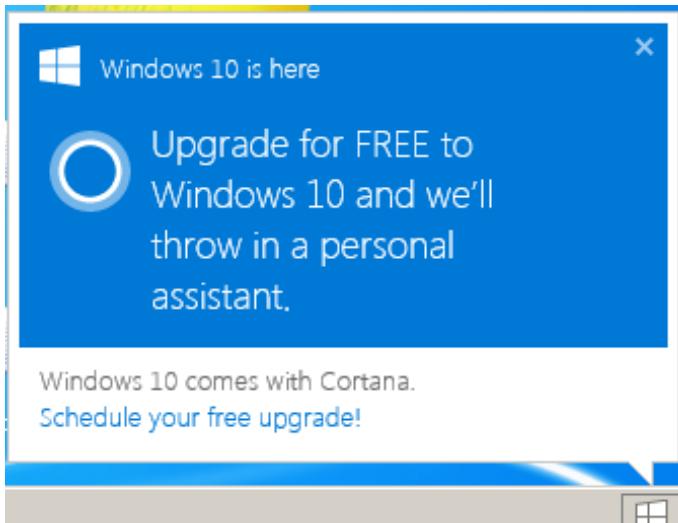


Fig. 1: Advertisement for Windows 7 users to upgrade to Windows 10 because of Cortana.

a reason for Windows 7 users to upgrade to Windows 10. As seen in figure 1, with such promotion, more and more activities we do on a computer may be possible through Cortana. Figure 2 shows a screenshot of requests being made using Cortana. Cortana is not the first voice activated personal assistant to be developed, that honor belonging to Apple's Siri, but it is the first operating system integrated assistant to be available on a desktop platform. Cortana learns a lot about a user as the user makes requests, recording everything that is asked of it and storing those artifacts on the computer [5]. From practical tasks like tracking a package, to the useless tasks like telling a joke, Microsoft is trying to make Cortana a personal companion that you feel comfortable using every day. With all the information that Cortana saves, it stands to reason that forensics investigators could learn a lot about a suspect who had been using it and what types of things they were requesting. With this new outlet for consumers to use, it is imperative that forensic experts have a tool that can properly analyze Cortana artifacts.

#### C. Autopsy Open Source Tool

Thankfully, the Autopsy open source forensics tool allows forensics investigators the ability to analyze disk images and report many types of information. Autopsy is built upon The Sleuth Kit set of command line tools. Using Autopsy, forensic examiners can conduct keyword searches, view file artifacts and fragments of files in common locations and also write custom Java and Python modules that can be easily added and shared with others. Figure 3 shows the Autopsy user interface loading up. The ability to develop custom modules and then share them with the forensic community is what makes Autopsy such a valuable tool. While Autopsy currently scans all files in an image, analysis of these files is limited to certain file types. In the case of audio files, Autopsy simply displays the file and allows the user to play the recording,

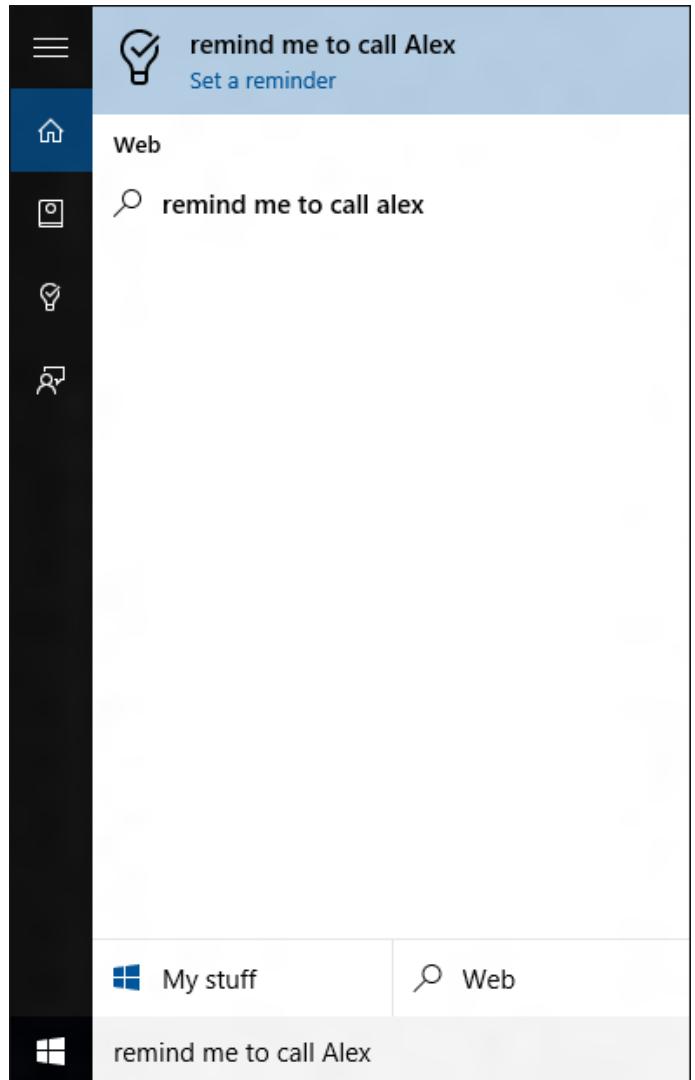


Fig. 2: Screenshot of request being made to Cortana in Windows 10.

but there is no way to currently examine these files without listening to each file. The fact that Autopsy is open source makes it easy to learn and provides the best support for implementing the solutions in this proposal.

#### D. Improving Forensic Tools

When looking for improvements in design and operation of forensics tools, much study has already been done. [6] did an analysis and explained that any improvements must be forward looking, and not rely on past techniques for success. One of the major points presented is that data filtering of erroneous data needs to begin at the data acquisition process. Attempting to conduct forensics investigations with hard drive sizes now being produced is greatly slowing the process, and will only continue to get worse. According to the 2009 RCFL report, there were 2,334TB worth of data processed in digital forensics examinations [7]. One solution to this problem is using large scale hashing tables that contain hashes of known



Fig. 3: Screenshot of Autopsy loading modules, such as the proposed Cortana search module.

files, and excluding those matching files from acquisition. The authors of [7] argue that in the future, bit by bit captures will become impractical given the size of data hard drives. With this research, it becomes clear that information needs to be presented to the forensics examiner as effectively as possible and the results need to be easily searchable.

The problem with designing and improving forensic products like Autopsy is that the information that needs to be analyzed has a relatively short life cycle. With the accuracy needed in forensics tools, development is almost one to one with the tool and what the tool is analyzing, so that if what the tool needs to analyze changes, so too does the tool itself. In [8], the author finds compelling reasons why it is difficult to make improvements in forensics tools. Digital forensics tools need to be able to analyze a wide range of data and they need to be reliable. The developers of forensics tools need to be able to support the host operating system they are trying to analyze, as well as the host operating system that they are running on. As many forensics experts have discovered, computer crime does not wait until there are appropriate tools to analyze a system. These requirements in turn mean that developing digital forensics tools take longer, and so they are slower to release to the forensics experts.

Another kind of improvement to digital forensic tools is based more on the discovered data. In [9], the authors suggest that improving digital forensics tools begins by deciding what you need to investigate further, and what does not need to be investigated further when looking at acquired data. For example, the review of chat logs or search terms in file sharing services might be found, but it is dependent on certain software being installed on the host. Such analysis is often called situational analysis and requires very careful development.

The need for improvements in very specific spaces of digital forensics, even down to the operating system and architecture type is not new. Specific tools have been developed solely for specific file systems on specific operating systems. Furthermore, a survey of forensic examiners showed that 58% said that they do not develop their own tools at all, which means that the need for writing and improving digital forensics

tools must be done by highly trained individuals and then given to the forensic community [10]. To help alleviate this problem, the commercial supporters of the Autopsy tool, Basis Technology, promote development by offering monetary rewards, and are one of the major sponsors of the Open Source Digital Forensics Conference held each year [11]. The conference has attracted forensic enthusiasts who have developed Autopsy modules for Google Analytics, improved memory analysis, and advanced registry decoders. The success of this conference and those like them show the urgent need to have the right tools for specific tasks.

### III. PROPOSAL

Previous work done on the study of improving forensic tools has provided information on types of improvements, their effectiveness and their purpose. The aim of this research proposal is to make recommendations for improvement in the Autopsy open source forensics tool for investigating Windows 10, specifically, when investigating the voice activated personal digital assistant Cortana.

When users make requests using Cortana, these requests are processed by the operating system and relevant data returned to the user. While all requests made are recorded, it is difficult to pull together all the pieces of data and analyze exactly what a user was searching for and what they got back from their search. Since forensic investigators need to be able to quickly identify relevant data, the task of finding out what information is in these recordings by searching for keywords is needed. The recommendation to improve the Autopsy open source forensic tool is therefore to develop an Autopsy plugin module that is capable of clearly identifying and presenting to the forensics investigator the requests made to Cortana inside the Autopsy user interface. Such an enhancement will greatly increase the speed and efficiency at which forensic investigators can perform examinations when voice activated personal digital assistants may have been used on computers where criminal activity occurred.

### IV. EVALUATION AND METHODS

Finding a proper method to generate, develop and then test the proposed research was paramount. The evaluation was measured upon the ease of finding the information with the developed module verse finding the information without the developed module. The measurement included the amount of time spent gathering relevant data about the requests made to Cortana, and the ease of clearly understanding the data presented. In order to determine the success of the research, a set of control requests were developed to guarantee that a response from Cortana was possible. Under normal circumstances, when Cortana does not know the answer to a request, it opens up a web browser and does a search for the request. When this occurs, the result is then the same as if a user had typed in the request without using Cortana, so it was necessary to ensure that Cortana was able to answer the requests independent of a web browser search. Without the control set, requests might be asked that Cortana could not answer, thereby reducing the means by which to perform a comparison of Cortana verse a user manually typing in the requests.

#### A. Module Development

The Autopsy forensics module was written in Java 1.8 using the Netbeans IDE. The module was composed of five Java classes. Table I lists the five classes and their descriptions. Autopsy uses The Sleuth Kit Blackboard class to post data to the user interface. All items in the user interface are a type of artifact, and each artifact has attributes. While custom artifacts can be developed, see section VI-A, such development was out of scope for this project. Instead, the TSK\_INTERESTING\_FILE\_HIT artifact was used to call out files in the Autopsy interface that were specifically related to Cortana. Figure 4 shows the developed module presenting hits in the tree view for Cortana artifacts.

To install the developed module into Autopsy, the user needs only to go to the Tools, Plugin menu and browse for the .nbm module to install it. Figure 5 shows the Autopsy installation dialog box for the developed Cortana search module. After installing the module into the Autopsy interface, the forensics examiner can then load in the drive image and select the CortanaSearch ingest data source module to run against the drive image. The developed Autopsy Cortana forensics Java module is a data source level ingest module which requires that the module itself search for the files that it needs [12]. The results are placed inside the Autopsy user interface as seen in figure 6. Figure 7 shows the contributions from each Autopsy module installed.

The Cortana artifacts were separated into two categories of files. The first category represented the raw audio files that Cortana creates when a user makes a voice request. These files are created in the user's application data folder, for example, C:\Users\Allen\AppData\Local\Packages\Microsoft.Wind ows.Cortana\_cw5n1h2txyewy\LocalState\LocalRecorder\S peech. The developed module scans the directory for .wav files. Of interest is the fact that these files overwrite each

other after 8 recordings thus reducing its usefulness as a means of investigation for any real life scenario. The files are named SpeechAudioFile\_X, where X is a number 0 through 7. The second category of file was the speech\_render[X] file, where X was the number of recording, which stores Cortana's interpretation of the user's voice request, and the response it provided. Of the two file types, the speech\_render files are the easier to search with as they are already interpreted into text. However, the results are still subject to what Cortana interpreted the user as saying, and therefore, may present problems in the legal arena.

In addition to the Cortana interpreted data placed in the speech\_render file, another speech to text library was used to confirm what was actually spoken by the user. The CMU Sphinx speech recognition algorithm is a multi-platform solution for speech recognition [13]. Located in the CortanaDataSourceIngestModule Java class, as each raw audio recording was found, it was processed through the CMU Sphinx library which returned the probable message in the voice recording. The returned text was sent to the TSK\_EXTRACTED\_TEXT artifact type, and displayed in the Autopsy tree view under Extracted Content as Extracted Text. Refer to figure 4. While Cortana will attempt interpretation, and record this in its speech\_render file, forensic tools should not rely on the operating system and must be able to independently confirm what was said. The attraction of needing to be able to analyze other binary formats and present them in a more useable form can be seen in other modules that have been developed as well, such as the video triage module which takes a video and then presents frames of the video as pictures so that the forensic investigator can quickly understand the content without having to watch the entire video [14].

When a user selects the Cortana Speech Render Files from the tree view, the files that were found will appear in the directory listing, as seen in figure 6. With the goal of clearly presenting the data to the forensics examiner, the developed module looks for the string "</script><head><!--pc--><title>" as the starting position of the request asked to Cortana and presents this cleanly to the forensics examiner. In addition, the code attempts to find the answer given back by Cortana by looking for the string "<h2 class='b\_annotation\_b\_anim'>" and appending this to the request. If no answer was given, then this is presented as well. All artifacts found in this manner are also then indexed for keyword searching. Consequently, when selecting the Extracted Text in the tree view, the CMU Sphinx library translations are shown as in figure 8.

With the ability to search and clearly present these Cortana artifacts, the developed Autopsy module provides forensics examiners with a quick diagnosis of the activity done on the system using Cortana. The data presented in section V will show the comparison of data when searching with and without the developed Autopsy module.

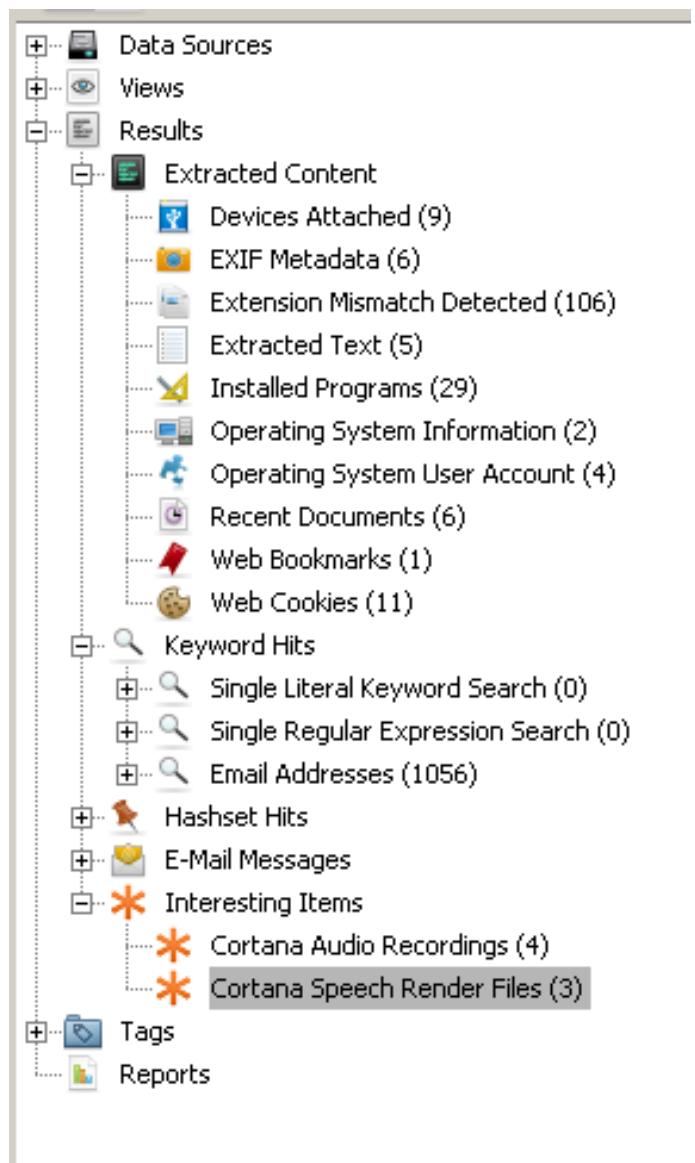


Fig. 4: Autopsy user interface showing Cortana artifacts.

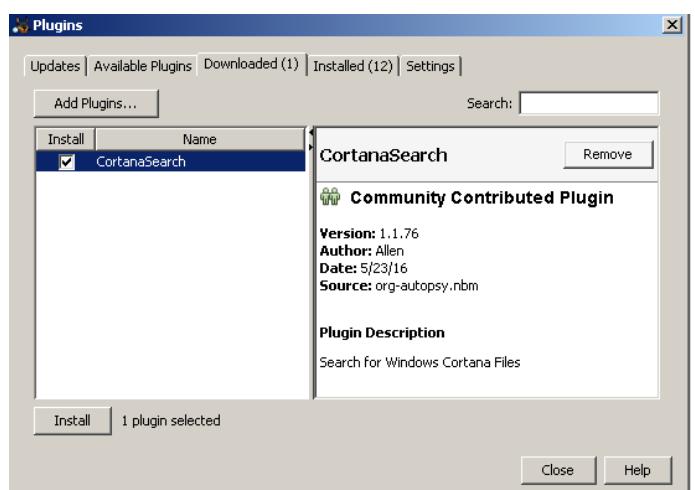


Fig. 5: Autopsy plugin installation dialog box.

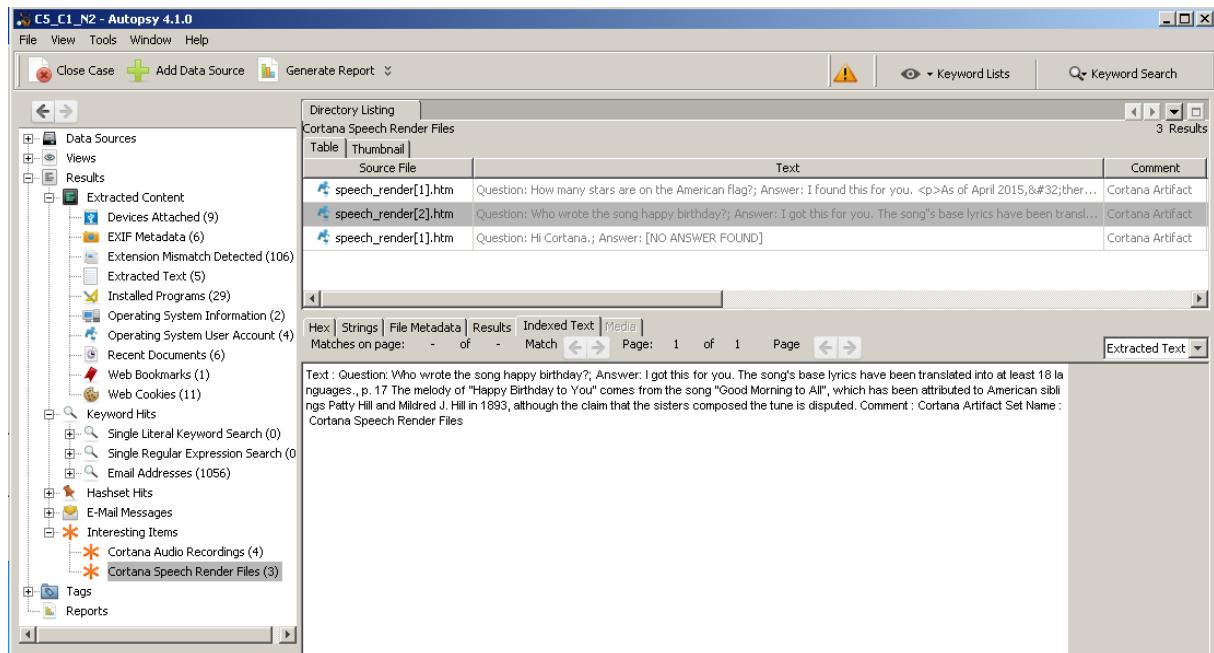


Fig. 6: Extracted contents of Cortana Speech Render Files.

CortanaDataSourceIngestModule	Code for searching for Cortana artifacts
CortanaIngestModuleFactory	The class that gets called when Autopsy loads to check for modules
CortanaFileIngestModule	Dependency Class
CortanaIngestModuleIngestJobSettingsPanel	Dependency Class
CortanaModuleIngestJobSettings	Dependency Class

TABLE I: Description of Java classes which compose the Autopsy Cortana module

Operating System	MS Windows 10 64 bit Education
Hard Drive	18GB
Memory	4GB
CPU	2

TABLE II: Windows 10 configuration

Module	Num	New?	Subject	Timestamp
Recent Activity	1		Started c5_c13.dd	18:34:11
Recent Activity	1		Finished c5_c13.dd - No errors reported	18:35:37
Recent Activity	1		c5_c13.dd - Browser Results	18:35:37
Cortana	1	•	Found 9 Cortana Audio Recordings	18:39:13
Cortana	1	•	Found 2 Cortana Speech Render Files	18:39:13

Fig. 7: Autopsy message box showing discovery of Cortana artifacts.

### B. Development of Control Requests

As mentioned in section IV, it was necessary to develop a set of control requests which Cortana could answer without needing to drop the user into a web browser to conduct the search on their own. When analyzing computers for forensic evidence, it is important to remember that it may

not necessarily be illegal activity that is being searched for, but rather gathering more information about a suspect, such as their contacts, places of interest, and activities that are of importance. As an example, Cortana would be unable to assist in modifying an image so as to insert some stegographic content, but it could certainly send an email to a user's contacts informing them that he or she were looking for a high quality printer that could print US currency. Likewise, Cortana could inform a user when flight 1234 was departing, possibly signaling travel plans the user had to leave the country. These quick and common activity requests are the basis of the control requests used to formulate the test data for Cortana.

To ensure that the requests made to Cortana were not limited to a specific type of activity, requests were classified into categories, and multiple requests made for each category. The categories were derived directly from requests that could be made to Cortana. Each request was then vetted on a third system to ensure that Cortana would directly answer the request rather than opening a web browser for external searching to be done. The requests contained keywords to further make analyzing the data easier. The list of categories and their requests is listed in Appendix VII-A.

### C. Data Sample Creation

To carry out the research, two Windows 10 client systems were used. The client systems were run using VMWare Work-

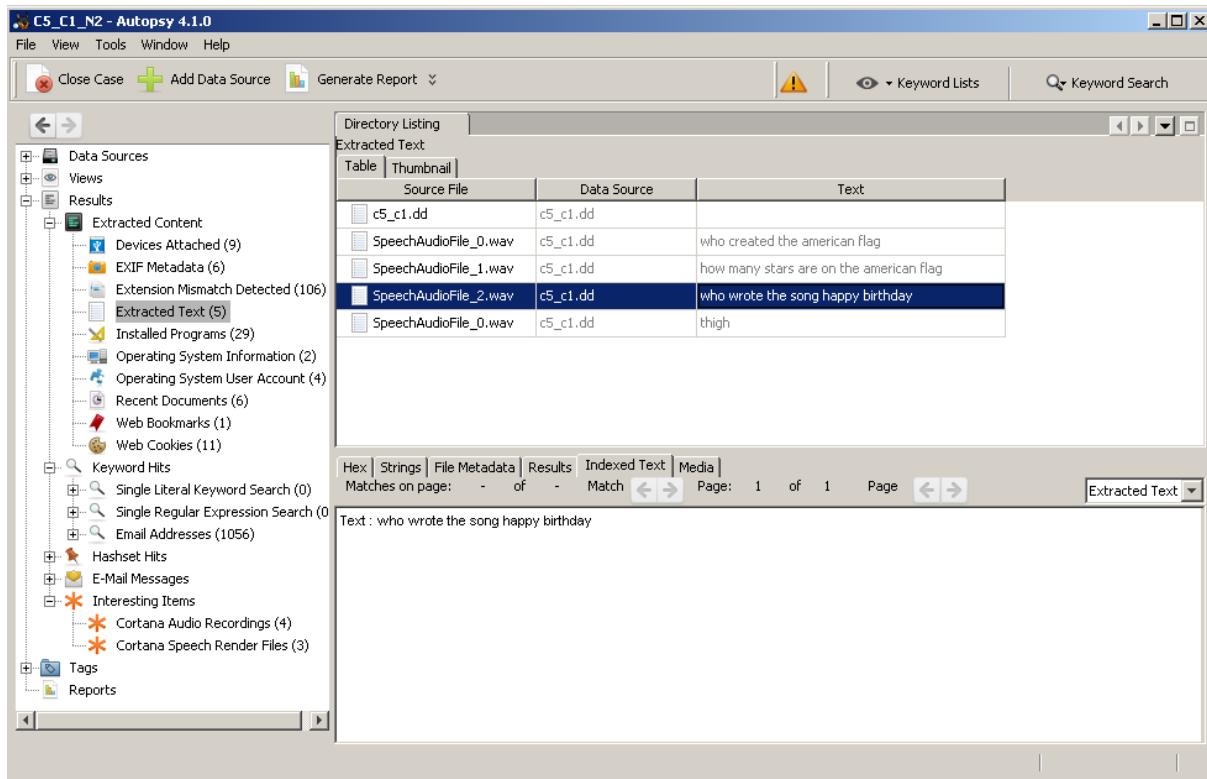


Fig. 8: Extracted contents of Cortana Audio Recordings using CMU Sphinx.

station so that snapshots could be taken and the systems rolled back for each category of requests so as not to contaminate the data from one category of requests to the next. Table II shows the configuration of both VM systems. System two was cloned directly from system one before the tests began to ensure that both systems were uniform in their configuration and build. The hard drive was set to 18GB to cut down on the time required to analyze the images. The memory was set to 4GB and the systems were given 2 CPU's, which exceeded Microsoft's system requirements for Windows 10 [15].

System one was used to make the requests to Cortana while system two was used to type the requests out on the Internet. Conducting each category of requests in parallel on system one and system two, the times were recorded in a spreadsheet to help in identifying them during analysis later. For categories where the request pertained to a user owned information asset, such as category 10 & 11, the test was run on system two by opening up the web mail in a browser and performing the request.

#### D. Data Acquisition

After each category of requests was completed on both systems, images of the drives were taken. The software used to capture the images was Back Track 5.3. Booting directly into the back track ISO, the following commands were run to capture the drive image:

```
mkdir /mnt/data
mount /dev/sdb1 /mnt/data
```

#### **dd if=/dev/sda1 of=/mnt/data/imagename.dd**

Each dd image was written to a secondary disk attached to the VM system. The process to capture each image was approximately 45 minutes.

#### E. Data Analysis

With all images captured and the developed Autopsy module installed, analysis began. In all, three rounds of image analysis occurred out of a pool of two sets, with each set containing eleven images. Set one contained the images from system one in which requests were made to Cortana. Set two contained the images from system two in which requests were typed directly into a web browser. Round one and two of image analysis analyzed sets one and two without the developed Autopsy module. Round three of image analysis analyzed set one only with the developed Autopsy module in place in order to measure the results and effectiveness of the developed Autopsy module. Set two was not analyzed twice since requests were typed in manually, and therefore, the developed Autopsy module would make no impact on the results. The set two images were used to more easily highlight the artifacts found in the set one images and show the need for the Cortana artifact module, while the set three images were used to show the benefits of the Cortana artifact module. The results are detailed in section V.

## V. RESULTS

Analyzing all the image files and conducting artifact analysis was a very time consuming process. While it would be easy to simply start searching the analyzed images for the known keywords in the control requests, the assumption was made that these control requests would not be known to the forensics investigator, and therefore, a more tedious approach must be used. Using standard forensic techniques, for each image, the goal being to find evidence of the control requests, a counter was set to record the amount of time needed to find the data. Due to time constraints, a predefined maximum of 15 minutes was established in order to find the needed data. The purpose of the developed Autopsy module is to present the Cortana artifacts quickly and correctly to the forensics examiner. In analyzing image sets one and two, the inability to find the needed evidence within 15 minutes will support the conclusion that a developed module for scanning Cortana artifacts is needed. In image set three, if the developed Autopsy module allows the forensics investigator to find the evidence in under 15 minutes, this will prove the success of the module. Again, it is assumed that a real investigator may have little to no background information about a case before viewing it, and therefore, no reason to suspect that Cortana has been used on the computer.

The following sub sections detail the results of key points that help to understand the argument of why a Autopsy Cortana module is needed.

### A. Cortana vs. User Web Browsing

Table III shows the comparison of searching for the control requests when the requests were done by Cortana verse typed in the web browser by the user. The results show that when typed in by hand, artifacts are left behind in various places in the user profile and finding files that show these control requests takes about 6 to 8 minutes on average. When Cortana is used to make the control request, very little is left behind, and using the 15 minute maximum search time cut off, the majority of control requests were not found. In order to keep the results consistent for all control requests, the same locations and order of searching was used, as seen in table IV. The list of locations searched begins with the built in artifact locations in Autopsy, followed by common best practice locations as recommended by SANS Institute [16]. Again, no keyword searches were performed since it was assumed that the forensic investigator had no background information and had to first acquire information by viewing key files. The data in table III shows the time it took to find the control request reference as well as the location (highlighted in blue) where the reference was found, and the text that was found.

The majority of artifacts from the control requests, when typed in by hand, were found in the Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active directory. Control requests for categories 3 and 4 were found in the Autopsy Recent Documents artifact section, but this data could be the result of those files being opened or touched

after creation. As category 12 is an action performed on the computer, and not a search, this could explain why information regarding this control request was found in a different location. In the end, the results of this comparison clearly show that finding artifacts left by Cortana is much more difficult and takes much more time than finding artifacts created when a user performs the request on their own.

### B. Analysis with vs without Developed Autopsy Module

With the confirmation of harder to find artifacts when Cortana is involved, the use of the developed Autopsy Cortana module to help forensics investigators resolve this problem must be validated. With the developed Autopsy module clearly presenting the Cortana artifacts to the user, as in figure 4, using the same order of searching as done for the comparison between Cortana and typed in requests, as shown in table IV, the comparison was almost a foregone conclusion. For each control request, finding the Cortana artifacts when the developed Autopsy module was installed took less than 30 seconds. In comparison with the times listed in column one of table III, which for the majority of control requests, took more than 15 minutes to find, the result that the developed Autopsy module greatly reduces the time to find Cortana artifacts is clear. Without the module, investigators are left searching for Cortana artifacts needlessly and, waste valuable time. By using the developed module, investigators can focus more on the content search, rather than trying to find the data.

### C. Cortana and CMU Sphinx Library

While presenting Cortana artifacts to forensics investigators is vital in reducing examination time, the ability to analyze those artifacts so the investigator can quickly determine if evidence is pertinent is also needed. By presenting the control request in a column next to the file in the Autopsy user interface, as seen in figure 6 and 8, the examiner will easily see if additional research into that area is required. As Cortana will place its own interpretation of the user's request in the speech\_render file, the question may become, could the interpretation be wrong, or subject to alteration. While the probability of these questions was out of scope for this project, the ability to quickly see the content of the raw audio file produced by Cortana and confirm Cortana's interpretation is needed.

Using the developed Autopsy module, the CMU Sphinx speech recognition library translated the raw audio Cortana artifacts into searchable text. Comparing the CMU Sphinx interpretation with the Cortana interpretation was surprising. The results of this comparison are in figure 9. Rows highlighted in green represent an exact match between Cortana and CMU Sphinx and were given a rating score of 1. Control requests where there was only one word difference were given a rating score of .5, while two words different were given a rating score of .25. If the interpretation varied by more than 2 words, then a 0 was given. For two of the categories, the CMU Sphinx failed to translate the raw audio file. As this may have been

Cat.	System One (By Cortana)	System Two (By Web Search)
1	Time: >15m	Time: 5m:33s. Location: Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active. Text: Who wrote the song Happy Birthday
2	Time: >15m	Time: 10m:43s. Location: Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active. Text: Flight 644
3	Time: 0m:42s. Location: Autopsy Recent Documents. Text: Mall of America	Time: 6m:48s. Location: Users\Allen\AppData\Local\Microsoft\Windows\INetCache. Text: Mall of America
4	Time: 0m:30s. Location: Autopsy Recent Documents. Text: Who won the Super Bowl	Time: 9m:42s. Location: Users\Allen\AppData\Local\Microsoft\Windows\WebCache. Text: Who won the Super Bowl
5	Time: >15m	Time: 8m:07s. Location: Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active. Text: What is in the Rochester New York
6	Time: >15m	Time: 6m:30s. Location: Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active. Text: What is the rate of inflation
7	Time: >15m	Time: 4m:15s. Location: Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active. Text: Hours for Wegmans Holt Road
8	Time: >15m	Time: 4m:06s. Location: Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active. Text: What is the meaning of yellow
9	Time: >15m	Time: 6m:55s. Location: Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active. Text: Who was the youngest president
10	Time: >15m	Time: 8m:37s. Location: Users\Allen\AppData\Local\Microsoft\Internet Explorer\Recovery\Active. Text: Camping
11	Time: 2m:41s. Location: Users\Allen\AppData\Local\Microsoft\Windows\INetCache. Text: Jenny Smith	Time: 7m:20s. Location: Users\Allen\AppData\Local\ntuser.dat.LOG1. Text: Jenny Smith

TABLE III: Results of searching for artifacts when created by Cortana vs. created by user web searching

Autopsy Built In Artifacts {Devices Attached, EXIF Metadata, Extension Mismatch Detected, Installed Programs, Operating System Information, Operating System User Account, Recent Documents, Web Bookmarks, Web Cookies, E-Mail Messages, Interesting Items}
Users\Allen\{NTUSER.DAT, ntuser.dat.LOG1, ntuser.dat.LOG2}
Users\Allen\AppData\Local\Microsoft\Internet Explorer
Users\Allen\AppData\Local\Microsoft\Windows\{History, IE*, INet*, Temporary Internet Files, WebCache}
Users\Allen\AppData\Local\Temp
Users\Allen\AppData\Local\Temporary Internet Files

TABLE IV: Locations that were searched for evidence of the control requests

a problem with the developed module, these were highlighted in blue and removed from the accuracy calculations.

When comparing Cortana verse CMU Sphinx, CMU Sphinx successfully translated the audio to text 65.38% of the time. While some fine tuning of the CMU Sphinx library may be possible, the 65% range is far too low for it to be used as a sole source speech recognition library, but could still be of value in general for searching audio files where no corresponding text string is present.

## VI. CONCLUSION

The need for the Autopsy Cortana module to bring information to the forefront when investigating is very real. The results from this research conclusively show that having such a module reduces the time to find Cortana artifacts and that without such a module, without foreknowledge of a case, forensic investigators will waste significant time trying to find evidence if the user used Cortana to commit the criminal activity. In short, the forensics tools must allow the forensics investigators to quickly and clearly find the relevant artifacts in a case, and the developed module has accomplished this.

### A. Lessons Learned

The process of developing an Autopsy forensics module has given the author of this paper a great deal of both practical and theoretical insight. As Windows 10 is still a relatively new operating system, documentation on how features like Cortana function are scarce and much of the information provided in this paper came from first hand experience and observation. As possible with any research, some of the issues encountered required significant time to address and correct.

1) *Scoping Client Requirements:* As hard drive capacities on most desktop computers today meet or exceed half a terabyte, a capacity of 60GB seemed to be a very conservative size for the hard drive of the two Windows 10 clients. After conducting the first round of requests, and then proceeding to analyze these 60GB drive images in Autopsy, it became apparent that with an analysis time of over a day, the Windows 10 clients had to be scrubbed and recreated with a significantly smaller hard disk size. As mentioned in section IV-C, this smaller size was 18GB. Even the 18GB limit took approximately six hours to analyze in Autopsy. The attention and effort invested in having to recreate the Windows 10 clients highlights the very real problem that

		CMU Sphinx	Given	Possible
1	Cortana			
2	Who created the American Flag	who created the american flag	1	1
3	How many stars are on the American Flag	how many stars are on the american flag	1	1
4	Who wrote the song happy birthday	who wrote the song happy birthday	1	1
5	When will my package arrive	No text found	0	0
6	Southwest Airlines flight 644	No text found	0	0
7	When does Delta flight 1111 arrive in Rochester	No text found	0	0
8	Give Me Directions to the mall of America	give me directions to the mall of america	1	1
9	Where was president George Washington Born	where was president george washington born	1	1
10	What is the hottest place on Earth	what is the hottest place on earth	1	1
11	Who won the super bowl	who won the super bowl	1	1
12	How many points are in a touch down	how many points are in that hot found	0	1
13	Who is the head coach for the Buffalo Bills	pool is the head coach for the bottle	0	1
14	What is the temperature in Rochester New York	what is the temperature in rochester New York	1	1
15	Do I need to wear warm clothes	do I need to wear warm clothes	1	1
16	Will it rain this weekend	william rayner this weekend	0.25	1
17	How much is Exxon Mobil stock worth	how much is exxon mobil start work	0.25	1
18	What is the exchange rate from Canada to American money?	what is the exchange rate from canada to american money	1	1
19	What is the rate of inflation	what is the rate of inflation	1	1
20	What are the show times for batman vs superman	No text found	0	0
21	When does daylight savings end	No text found	0	0
22	Hours for Wegmans Holt Road	No text found	0	0
23	Define Chocolate	defying chocolate	0.5	1
24	Translate grass in spanish	translate grants in spanish	0.5	1
25	What is the meaning of yellow	what is the meaning of yellow	1	1
26	What is the capital of Virginia	what is the capital of virginia	1	1
27	Who was the youngest president	who was the youngest president	1	1
28	What is the distance to Neptune	what is the distance the neptune	0.5	1
29	Show me the events for April 12th	a media bands for april twelfth	0	1
30	How many events do I have in my calendar	how many events do I have a mike taylor	0	1
31	Find camping in my calendar	find a campaign that in my calendar	0	1
32	Send an email to Jenny Smith	send him he melted Jenny's man	0	1
33	Open my files	open my files	1	1
34				
35		TOTAL ACCURATE	17	26 65.38%
~				

Fig. 9: Comparison of Cortana to CMU Sphinx speech recognition library

forensics investigators face when analyzing disk images from even every day ordinary computers that have massive disk sizes.

2) *Compiling Autopsy 4.1:* The initial scope of the research project was to design an Autopsy module that would clearly present the details of the Cortana artifacts to the forensics examiner and allow the examiner to be able to search for those results using the keyword search module. After completing coding for the Autopsy module, it was discovered that the current stable release of Autopsy, version 4.0.0, did not support indexing for artifacts posted to the Blackboard. Without indexing support, the effectiveness of the proposed module would be minimal, and measuring the success of the module would be extremely difficult. As a result, it became necessary to download Autopsy 4.1, which added support for indexing artifacts, and compile the code in order to proceed with the research. Several platform dependencies issues had to be overcome, which caused significant delay in conducting the research.

#### B. Future Tasks

While the results presented in this research paper are established as accurate and complete, additional work could be done to further improve on the developed module or address questions that were out of the scope of this project. As is, the developed module provides forensic investigators with the Cortana speech render and speech audio files. While these are perhaps the most important artifacts when investigating

a user's interaction with Cortana, other areas might also be beneficial. For example, a number of other artifacts exist for Cortana, such as contacts, which are stored in the contacts.json file, and suggested links, which are stored in the suggestions.json files [17] [18]. Photographic image files retrieved from a request made to Cortana could also be displayed in Autopsy giving forensics investigators visual clues into the types of requests being made. Another improvement in the developed module would be to create a unique Autopsy artifact type that specifically stored all artifacts for Cortana in its own type. This would require development within the Autopsy platform itself, but would better present the discovered artifacts to forensic investigators than is currently possible.

#### C. Recommendations

In closing, the author has learned a lot and seen the need for more people who have the skills to develop modules that meet a specific need. The constant increase in use of computers for our every day tasks is having a profound impact on forensics. As more data is stored, more data needs to be analyzed, and hence forensic examiners need better tools, and more precise tools for analysis. Microsoft Windows 10 with Cortana is just the latest in a series of technological leaps that will continue to make the forensic investigators job more difficult. In short, the solution in this proposal will not prevent the work that needs to be done tomorrow, but will ease the work that needs to be done today.

## VII. APPENDICES

### A. Control Requests

- 1) General Web Browsing
  - a) Who created the American Flag?
  - b) How many stars are on the American Flag?
  - c) Who wrote the song Happy Birthday?
- 2) Tracking
  - a) When will my package arrive?
  - b) Southwest Airlines flight 644
  - c) When does Delta flight 1111 arrive in Rochester?
- 3) Places
  - a) Give me directions to the Mall of America
  - b) Where was president George Washington born?
  - c) What is the hottest place on Earth?
- 4) Sports
  - a) Who won the Super Bowl?
  - b) How many points are in a touch down?
  - c) Who is the head coach for the Buffalo Bills?
- 5) Weather
  - a) What is the temperature in Rochester New York?
  - b) Do I need to wear warm clothes?
  - c) Will it rain this weekend?
- 6) Finance
  - a) How much is Exxon Mobil stock worth?
  - b) What is the exchange rate from Canada to American money?
  - c) What is the rate of inflation?
- 7) Show Times
  - a) Show me movie times for Batman vs Superman?
  - b) When does daylight savings end?
  - c) Hours for Wegmans Holt Road
- 8) Dictionary
  - a) Define Chocolate
  - b) Translate grass in Spanish
  - c) What is the meaning of yellow?
- 9) Facts
  - a) What is the capital of Virginia?
  - b) Who was the youngest president?
  - c) What is the distance to Neptune?
- 10) Calendar
  - a) Show me the events for April 12th
  - b) How many events do I have in my calendar
  - c) Find camping in my calendar
- 11) Email
  - a) Send an email to Jenny Smith
  - b) Open my files

### ACKNOWLEDGMENT

I wish to thank and acknowledge professor Yin Pan and Bill Stackpole for their guidance during the research. A special thank you to managers Mike and Steve who provided resources to complete this research. Most importantly, I thank my Lord

and Savior Jesus Christ who gave me strength to complete this research and for my parents and brothers and sisters in Christ who prayed for me.

### REFERENCES

- [1] D. Watson and A. Jones, *Digital Forensics Processing and Procedures: Meeting the Requirements of ISO 17020, ISO 17025, ISO 27001 and Best Practice Requirements*. Syngress Publishing, 2013.
- [2] C. Altheide and H. Carvey, *Digital Forensics with Open Source Tools*. Syngress Publishing.
- [3] S. L. Garfinkel, "Digital forensics research: The next 10 years," *Digital Investigation*, vol. 7, Supplement, pp. S64 – S73, 2010, the Proceedings of the Tenth Annual {DFRWS} Conference. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1742287610000368>
- [4] Microsoft, "What is cortana?" Online, <http://windows.microsoft.com/en-us/windows-10/getstarted-what-is-cortana>.
- [5] E. Betterts, "Cortana vs google now vs siri: Battle of the personal assistants," April 2014, <http://www.pocket-lint.com/news/128303-cortana-vs-google-now-vs-siri-battle-of-the-personal-assistants>.
- [6] G. G. Richard, III and V. Roussev, "Next-generation digital forensics," *Commun. ACM*, vol. 49, no. 2, pp. 76–80, Feb. 2006. [Online]. Available: <http://doi.acm.org.ezproxy.rit.edu/10.1145/1113034.1113074>
- [7] U. D. of Justice, "Regional computer forensics laboratory," in *Annual Report for fiscal year 2009*, 2009.
- [8] S. Garfinkel, "Lessons learned writing digital forensics tools and managing a 30tb digital evidence corpus," *Digital Investigation*, vol. 9, pp. S80–S89, 2012.
- [9] J. I. James and P. Gladyshev, "A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview," *Digital Investigation*, vol. 10, no. 2, pp. 148 – 157, 2013, triage in Digital Forensics. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1742287613000340>
- [10] H. Hibshi, T. Vidas, and L. Cranor, "Usability of forensics tools: A user study," in *IT Security Incident Management and IT Forensics (IMF), 2011 Sixth International Conference on*, May 2011, pp. 81–91.
- [11] Open Source Digital Forensics, May 2016, <http://www.osdfcon.org/about>.
- [12] *Autopsy Forensic Browser Developer's Guide and API Reference*, Basis Technology, October 2015, <http://www.sleuthkit.org/autopsy/docs/api-docs/3.1/index.html>.
- [13] CMUSphinx, May 2016, <http://cmusphinx.sourceforge.net/wiki/about>.
- [14] *Module for Autopsy*, Basis Technology, May 2016, <http://www.basitech.com/digital-forensics/autopsy/video-triage/>.
- [15] Microsoft, "Windows 10 specifications," Online, <https://www.microsoft.com/en-us/windows/windows-10-specifications#sysreqs>.
- [16] R. Lee, *Windows Artifact Analysis: Evidence of...*, SANS, May 2016, [https://uk.sans.org/posters/windows\\_artifact\\_analysis.pdf](https://uk.sans.org/posters/windows_artifact_analysis.pdf).
- [17] WIN10 - EDGE BROWSER, Syntricate, May 2016, <https://www.syntricate.com/files/computer-forensics/WIN10>
- [18] Windows 10 Artifact Locations, Syntricate, May 2016, <https://www.syntricate.com/files/computer-forensics/WINDOWS>

# An Effort Estimation Approach for Agile Software Development using Fireworks Algorithm Optimized Neural Network

Thanh Tung Khuat, My Hanh Le

**Abstract**—Software effort estimation is one of the most critical steps in the software development process. The success or failure of projects relies greatly on the accuracy of effort estimation and schedule results. Agile software development process has become prevalent in the industry and replacing the conventional approaches of software development. Nevertheless, the question of accurate estimation of effort for this novel method has still been a challenging problem with regard to researchers and practitioners. This study aims to propose a novel method to ameliorate the accuracy of agile software effort prediction process using Artificial Neural Network (ANN) optimized by Fireworks Algorithm (FWA). The performance of the proposed approach is compared to the various types of neural networks and the regression model. In addition, the role of Fireworks Algorithm in optimizing the weights and biases of the ANN is also compared with other optimization algorithms.

**Index Terms**— Software Effort Estimation, Agile Software Development, User Story, Artificial Neural Network, Fireworks Algorithm, Levenberq-Marquardt.

## I. INTRODUCTION

Software effort estimation has a critical role to play in the process of software development. Both underestimation and overestimation of effort have a negative influence on the success of projects. Therefore, it is expected to find out a method to estimate the software effort precisely. A variety of estimation techniques using data collected from past projects combined to mathematical formula to predict the project cost introduced such as COCOMO II [1], SLIM [2], PRICE-S [3]. Estimation techniques are distributed into regression-based models, expert-based methods, learning-oriented approaches, and Bayesian methods. The diversity of novel software development methodologies has resulted in the limitation of traditional approaches.

In recent years, the appearance of agile software development process has met the progress of the new software engineering methodology. The use of agile methods enables organizations to respond the volatility in the software

Thanh Tung Khuat is now with the Information Technology Faculty, University of Science and Technology - The University of Danang, Vietnam (e-mail: thanhtung09t2@gmail.com).

My Hanh Le is now with the Information Technology Faculty, University of Science and Technology - The University of Danang, Vietnam (e-mail: ltmhanh@dtu.udn.vn).

development life cycle. The application of an estimation method in the agile process is a difficult task in which we need to anticipate the size and complexity of the products to be constructed in order to specify what to do next [4]. To meet this requirement, user stories of the product need to be collected and analyzed the complexity to determine story points for components of the project. Another factor which may impact the agile software development is team velocity. That is the total number of story points that a team is able to implement in a sprint. This study uses story points and velocity to train the ANN, then the model is used to give the estimated time to complete a novel project.

The efficiency of ANN largely depends on their architecture, their training algorithm, and the selection of features utilized in training. The process of network learning optimization is to find out the weights configuration associated with the minimum output error. Many algorithms applied to training process including Ant Colony Optimization [5], Simulated Annealing and Genetic Algorithms [6], Tabu search [7]. As for Multi-layer Perceptron (MLP) neural networks, the Back-propagation (BP) algorithm and Levenberq-Marquardt (LM) are widely used for the training process [8]. The researchers prefer to use LM among the conventional approaches because of its speed of convergence and performance. However, this algorithm is easy to be stuck at the local optimum. To cope with this issue, some global optimization algorithms were employed to tune weights of MLP aiming to avoid the local optimum such as evolutionary algorithm [9], artificial bee colony (ABC) algorithm [10], ant colony optimization [5], and hybrid particle swarm optimization and gravitational search algorithm [11]. In this paper, Fireworks algorithm and LM are utilized as new training methods for the Feedforward Neural Network (FNN) in order to figure out the effectiveness of these algorithms in reducing the problems of trapping in the local minimum and the slow convergence rate of LM algorithm.

FWA proposed by Tan and Zhu [12] is a population-based algorithm inspired by the explosion process of fireworks. The FWA has a strong capability of seeking global optimal result and LM algorithm has a strong ability to find the local optimal result. Therefore, this study proposes a combination method of FWA and LM for training the FNN aiming to minimize the output error of the FNN in order to give the accurate estimation result of the effort for the agile software

development process.

The rest of this paper is organized as follows. Section II represents the previous methods for estimating the effort of agile software development process. Section III shows a model used to predict effort and introduces proposed steps to apply neural network models for the agile software effort estimation. The combination of FWA with LM applied to neural network training is presented in section IV. Section V is the experiment and the analysis of the performance of proposed model. Threats to validity of the study are discussed in section VI. Finally, section VII gives the conclusion of this paper.

## II. RELATED WORK

Keaveney and Conboy [13] figured out the applicability of conventional estimation techniques to agile development approaches by focusing on four case studies of using agile of different organizations. The authors used the main estimation techniques being expert knowledge and analogy to past projects. The obtained results revealed that the estimation inaccuracy using the proposed method was a less frequent occurrence compared to the use of traditional approaches. Coelho and Basu [14] gave an overview of the various size estimation techniques based on story points for the agile software development process. The authors showed the steps followed in the story point based approach and highlighted the area which needs to be studied further. Abrahamsson and Koskela [15] described the way to collect metrics to gauge the productivity, quality and schedule estimation, cost and effort estimation for an agile software development project using extreme programming. The authors provided evidence that agile approaches are efficient and appropriate for a variety of situations and environments. Hussain *et al.* [16] presented an approach to approximate COSMIC functional size from informally written textual requirements demonstrating its applicability in popular agile processes. Popli and Chauhan [17] introduced a model for effort and cost estimation in the agile software development by using the regression analysis. Hamouda [18] proposed a process and methodology that assure relativity in software sizing while using agile story points on the level of the CMMI organizations. Oliveira [19] provided a comparative research on Support Vector Regression (SVR), Radial Basis Function Neural Networks (RBFNs) and the linear regression for the estimation of software development effort. The experiment was conducted on NASA project data sets and the experimental results showed that SVR performed better than the RBFN and the linear regression analysis. Satapathy *et al.* [16] estimated the effort of the agile software using story point approach in which the total number of story points and project velocity were employed to predict the effort involved in developing an agile software product. The obtained results were optimized by utilizing four various support vector regression kernel methods. The authors concluded that the radial basis function kernel-based support vector regression technique outperformed other three kernel methods.

Zia *et al.* [21] developed an effort estimation model for

agile software projects. Authors introduced a method to compute the team velocity and story points from the user stories and features of the product, and then the regression analysis method was employed to give the completion time of the project. The proposed model was validated using the empirical data collected from 21 software projects. The obtained results showed that the model had good estimation accuracy in terms of the Mean Magnitude of Relative Error. Panda *et al.* [22] attempted to enhance the prediction accuracy of the agile software effort estimation process proposed by Zia. To solve this problem, various types of neural networks including General Regression Neural Network (GRNN), Probabilistic Neural Network (PNN), Group Method of Data Handling (GMDH) Polynomial Neural Network and Cascade-Correlation Neural Network were used and compared. Our study also aims at ameliorating the accuracy of estimation model of Zia by introducing an ANN optimized by the combination of Fireworks and LM algorithms.

## III. EFFORT ESTIMATION APPROACH FOR THE AGILE SOFTWARE DEVELOPMENT PROCESS

### A. An effort estimation model for agile projects

In [21], Zia *et al.* proposed a model to estimate the effort of agile software projects. This model uses story points and team velocity to predict the effort for a project.

#### 1) Computing the story point of an agile project

Story points are a number of user stories associated with their complexity completed in a unit time. In order to compute story points of an agile project, we first determine the story sizes and the complexity of each story size. Story size is an estimate of the relative scale of the work in terms of actual development effort. Each story size is assigned a value from 1 to 5 based on their scales. Value 1 indicates that the story is very small which needs tiny effort level with only a few hours of work. Value 2 shows that it is expected to finish the user story in a day or two of work meanwhile value 3 presents that we need from two to five days of work to complete the user story. Value 4 is given to the story having a very large size and requiring more than a week of work to accomplish as well as we need to take into account breaking it down into a set of smaller stories. Value 5 represents an extremely large story and it is really hard to estimate time accurately. After specifying the scale of the story, we have to consider its complexity. The complexity is also measured by five values assigned to the user story according to its nature. Value 1 states that the story requires basic programming skills to complete, and their technical and business requirements are very clear with no ambiguity. Value 5 shows that the story is extremely complex with many dependencies on other stories, systems or subsystems, and it needs a skill set or experience that is important, but absent in the team along with the extensive research and significant refactoring. The details of the user story complexity are clearly described in [21].

The total story points for  $N$  user stories of a project is computed as Eq. (1).

$$SP = \sum_{i=1}^N C_i \cdot S_i \quad (1)$$

## 2) Determining agile velocity

The initial agile velocity of a team is simply how many units of effort that this team is able to complete in a typical sprint. It can be also defined as how many story points that a team can handle in one sprint, and it is determined as follows:

$$V_i = \text{Units of Effort Completed} / \text{Sprint Time}$$

In reality, the velocity of a project is not only being simply measured by units of effort and sprint time but it also is influenced by two other factors including friction and variable or dynamic forces.

The friction forces are constants that drag on productivity and reduce the project velocity. They consist of team composition, process, environmental factors, and team dynamics. Their effects are long term, but they are easy to tackle. Table I shows four friction factors with a range of values and these values have been tuned following their risk severity [21].

TABLE I. FRICTION FACTORS

Fiction Factor	Stable	Volatile	Highly Volatile	Very Highly Volatile
Team composition	1	0.98	0.95	0.91
Process	1	0.98	0.94	0.89
Environmental Factors	1	0.99	0.98	0.96
Team dynamics	1	0.98	0.91	0.85

The value of friction (FR) can be computed as the product of all four fraction factors (FF) shown in Eq. (2).

$$FR = \prod_{i=1}^4 FF_i \quad (2)$$

The variable or dynamic forces decelerate the project or the performance of team members and bring about the project velocity to be irregular. These forces are usually unpredictable and unexpected. They include team changes, new tools requiring learning, vendor defects, responsibilities outside of the project of team members, personal issues, stakeholders, unclear requirements, changing requirements, and relocation. Table II describes variable or dynamic force factors and the values associated with them on the basis of same analogy as for size.

Dynamic Force (DF) then is computed as the product of all nine variable factors (VF) as shown in Eq. (3).

$$DF = \prod_{i=1}^9 VF_i \quad (3)$$

Deceleration of an agile project is the product of friction and dynamic forces impacting the velocity as Eq. (4).

$$D = FR \cdot DF \quad (4)$$

The final velocity of a project under the influence of friction and dynamic forces is computed as Eq. (5).

$$V = (V_i)^D \quad (5)$$

From the team velocity and story points of a project, Zia *et*

*al.* [21] used the regression method to predict the duration needed to complete the project.

TABLE II. DYNAMIC FORCE FACTORS

Variable Factor	Normal	High	Very High	Extra High
Expected team changes	1	0.98	0.95	0.91
Introduction of new tools	1	0.99	0.97	0.96
Vendor's defects	1	0.98	0.94	0.90
Team member's responsibility outside the project	1	0.99	0.98	0.98
Personal issues	1	0.99	0.99	0.98
Expected delay in Stakeholder response	1	0.99	0.98	0.96
Expected ambiguity in details	1	0.98	0.97	0.95
Expected changes in environment	1	0.99	0.98	0.97
Expected relocation	1	0.99	0.99	0.98

## B. Proposed steps for determining the predicted effort using Artificial Neural Network

This study proposes to employ the ANN optimized by the combination of Fireworks algorithm and LM for the training process of ANN. The inputs of the neural network models are a total number of story points and project final velocity and an output is the effort such as the completion time of the project. In our work, in order to enhance the accuracy of the estimation effort, FWA and LM algorithms are used to optimize the weights and biases of the ANN. The steps to estimate the effort of an agile software project are shown as follows:

**Step 1:** Collecting the total number of story points, the project final velocity, and the actual effort. In this paper, these data are taken from Zia's work [21].

**Step 2:** Normalizing the data of story points, project velocity, and actual effort values within the range of [0, 1]. Let  $X$  is the data set,  $x$  is an item of the data set then the normalized value  $x'$  of  $x$  can be computed as Eq. (6).

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (6)$$

where  $\max(X)$  and  $\min(X)$  are the minimum and maximum values of the data set  $X$  respectively.

**Step 3:** Splitting the data set into training and testing sets. In this study, the first fifteen projects are used for training and the others for testing.

**Step 4:** Training ANN: the FWA and LM are used together to optimize the weights and biases of the ANN.

**Step 5:** Testing and evaluating the performance of the proposed model using criteria shown in section V.

After the neural network implementation is completely done, the obtained results are compared to the other types of ANN as well as assessing the role of FWA in optimizing the weights and biases of the ANN.

## IV. TRAINING FEED-FORWARD ARTIFICIAL NEURAL NETWORK

### A. Feed-forward Artificial Neural Network

A feed-forward neural network (FFNN) is an artificial neural network in which connections between the units do not

form a cycle. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes, and to the output nodes without cycles or loops. Each node in the FFNN receives a signal from the nodes in the previous layer, and each of those signals is multiplied by a specific weight value. The weighted inputs are then summed and passed through an activation function which scales the output to a fixed range of values. The output of the node is broadcast to all of the nodes in the next layer. The output value of a node is computed by using Eq. (7).

$$o_i = f(\sum_{j=1}^n w_{ij} \cdot x_j + b_i) \quad (7)$$

where  $o_i$  is the output of the  $i^{th}$  node,  $x_j$  is the input of the  $j^{th}$  node,  $w_{ij}$  is the connection weight between the current node and input  $x_j$ ,  $b_i$  is the bias of the  $i^{th}$  node, and  $f$  is an activation function. The activation function is often a nonlinear function such as a Bipolar Sigmoid, a logarithmic sigmoid, a Gaussian function. This study uses the sigmoid function as the transfer function for hidden and output layer neurons.

$$f = \frac{1}{1 + e^{-x}} \quad (8)$$

To enhance the accuracy of estimated results, the FWA is used to optimize the network weights. The main idea is to convert the weight matrices of the ANN into individuals of population-based optimization algorithms. This study chooses the mean squared error as a network error function shown in Eq. (9) and the objective of the FWA is to minimize this function.

$$E(w(t)) = \frac{1}{T} \cdot \sum_{t=1}^T \sum_{k=1}^K (P_k - A_k)^2 \quad (9)$$

where  $E(w(t))$  is the error at the  $t^{th}$  iteration,  $w(t)$  is the vector of weights of the connections at the  $t^{th}$  iteration,  $P_k$  and  $A_k$  are predicted and actual values of the effort of the  $k^{th}$  output node.  $K$  is the number of output nodes and  $T$  is the number of patterns in the training dataset.

#### B. Proposed Architecture of FNN

This study employs a multilayer perceptron neural network to estimate the completion time for an agile project. A MLP is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. A MLP includes multiple layers of nodes in a directed graph, with each layer fully connected to the next one. In general, the architecture of the MLP might have many hidden layers and each hidden layer can consist of many neurons. However, many studies and experimental results also indicate that one hidden layer is sufficient for most of the forecasting problems as [23], [24], and [25]. Therefore, this work utilizes the architecture of the MLP with one hidden layer.

Another difficult task when choosing good parameters for the ANN is the number of hidden neurons. Setting a suitable architecture of the ANN for a specific problem is an essential issue because the network topology directly influences its computational complexity and generalization ability. Too much hidden neurons will drive the ANN to the over-fitting in

which the ANN performs well on training data and poorly on data it has not seen. We conducted the experiments with the number of hidden neurons varying from 4 to 60. It can be seen that the value of hidden neurons which is more than 40 will give a low error rate with regard to the training data set, but the testing error is quite high meanwhile if the number of hidden neurons is 27 then the training error is very low and the testing error is lowest. In our study, we therefore choose 27 neurons for the hidden layer.

As for the input layer, there are two neurons being the total number of story points and the project final velocity. The output of the ANN has only one neuron being the completion time of the project.

#### C. Fireworks Algorithm

Fireworks algorithm inspired by observing fireworks explosion is proposed by Tan and Zhu [12] for global optimization of complex functions. In this algorithm, two kinds of search processes are employed as well as generating mechanisms for keeping the diversity of sparks. When a firework explodes, a shower of sparks will be created around the firework. The explosion process of a firework might be considered as a local search around a specific point. In order to seek a point  $x_j$  such that  $f(x_j) = y$ , ‘fireworks’ are constantly set off in potential space until one ‘spark’ target or reaching a point that is relatively close to the point  $x_j$ .

In this paper, a location presents a possible set of optimized weights and biases for the ANN. In the FWA, for each generation of the explosion,  $n$  locations are chosen. After the explosion, the locations of sparks are obtained and assessed. The algorithm stops when the optimal location is found. Otherwise,  $n$  other locations are selected from the current sparks and fireworks for the next generation of the explosion. It can be seen that the success of the FWA relies greatly on a good design of the explosion process and an appropriate approach for choosing locations.

##### 1) Design of Fireworks explosion

Through inspecting fireworks display, two specific kinds of the fireworks explosion behavior are found. A good firework explosion creates numerous sparks which centralize the explosion center. On the other hand, a bad firework explosion generates a few sparks which scatter around the explosion center. Two these behaviors are shown in Fig. 1.

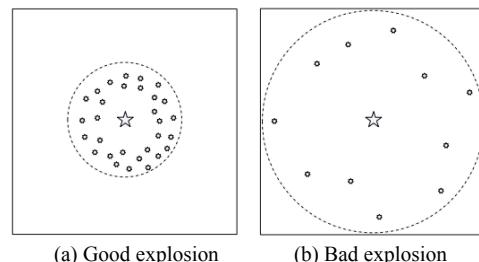


Fig. 1. Two kinds of fireworks explosion

From the viewpoint of a search algorithm, a good firework means that the firework might be near to the optimal location. Therefore, it is proper to use more sparks to search the local area around the firework. In contrast, a bad firework shows

that the firework can be far from the optimal location. In this case, the search radius should be larger. In the FWA, a good firework generates more sparks and the explosion amplitude is smaller when compared to the bad one.

**Number of Sparks:** Suppose that the FWA is designed for finding the optimal solution of a general optimization problem as follows:

$$\text{Minimize } f(x) \in R, x_{\min} \leq x \leq x_{\max} \quad (10)$$

where  $x = [x_1, x_2, \dots, x_D]$  is a location in the potential space,  $f(x)$  is an objective function and in this paper  $f(x) = E(w(t))$ ,  $x_{\min}$  and  $x_{\max}$  denote the bounds of the potential space,  $D$  is the dimensionality of the location  $x$ . Let  $HN$  be the number of hidden neurons, the value of  $D$  can be computed by using Eq. (11) whose details are shown in Table III.

$$D = IW\{1,1\} + b\{1,1\} + OW\{2,1\} + b\{2,1\} \quad (11)$$

TABLE III. PARAMETERS FOR SPECIFYING THE SIZE OF INDIVIDUALS

Value	Symbol	Description
$2 \cdot HN$	$IW\{1, 1\}$	Weights of the connections from the input layer to the hidden layer
$HN$	$b\{1, 1\}$	Biases of neurons in the hidden layer
$HN \cdot 1$	$OW\{2, 1\}$	Weights of the connections between the output layer and the hidden layer
1	$b\{2, 1\}$	Biases of output neurons

As mentioned above, the number of hidden neurons in this study is 27, so  $HN = 27$ .

Next, the number of sparks provided by each firework  $x_i$  is determined as Eq. (12).

$$s_i = t \times \frac{y_{\max} - f(x_i) + \delta}{\sum_{i=1}^n (y_{\max} - f(x_i)) + \delta} \quad (12)$$

where  $t$  is a parameter controlling the total number of sparks created by  $n$  fireworks,  $y_{\max} = \max(f(x_i))$  ( $i = 1, 2, \dots, n$ ) is the maximum or worst value of the objective function among  $n$  fireworks, and  $\delta$  denotes the smallest constant in the computer, is employed to avoid zero-division-error.

In order to avoid overwhelming impacts of gorgeous fireworks, bounds are defined for  $s_i$ , which is described in Eq. (13).

$$\hat{s}_i = \begin{cases} \text{round}(a \times t) & \text{if } s_i < a \cdot t \\ \text{round}(b \times t) & \text{if } s_i > b \cdot t, \quad a < b < 1 \\ \text{round}(s_i) & \text{otherwise} \end{cases} \quad (13)$$

where  $a$  and  $b$  are const parameters.

**Amplitude of Explosion:** In contrast to the design of sparks number, the amplitude of a good firework explosion is smaller than that of a bad one. The amplitude of the explosion for each firework is specified as Eq. (14).

$$A_i = \hat{A} \cdot \frac{f(x_i) - y_{\min} + \delta}{\sum_{i=1}^n (f(x_i) - y_{\min}) + \delta} \quad (14)$$

where  $\hat{A}$  is the maximum explosion amplitude, and  $y_{\min} = \min(f(x_i))$  ( $i = 1, 2, \dots, n$ ) is the minimum or best value of the objective function among the  $n$  fireworks.

**Generating Sparks:** In the explosion process, sparks might meet the influences of explosion from random  $z$  dimensions. In the FWA, the number of the affected directions is randomly

obtained from Eq. (15).

$$z = \text{round}(D \times \text{rand}(0, 1)) \quad (15)$$

where  $D$  is the dimensionality of the location  $x$ , and  $\text{rand}(0, 1)$  is a random number distributed uniform in the range of  $[0, 1]$ .

The location of a spark of the firework  $x_i$  is obtained by using Algorithm 1.

**Algorithm 1.** Find out the location of a spark

```

Initialize the location of the spark:  $\tilde{x}_j = x_i$ 
 $z = \text{round}(D \times \text{rand}(0, 1))$ 
Randomly choose  $z$  dimensions of  $x_j$ 
Compute the displacement:  $d = A_i \times \text{rand}(-1, 1)$ ;
for each dimension  $\tilde{x}_k^j \in \{\text{pre-selected } z \text{ dimensions of } \tilde{x}_j\}$  do
     $\tilde{x}_k^j = \tilde{x}_k^j + d$ ;
    if  $\tilde{x}_k^j < x_k^{\min}$  or  $\tilde{x}_k^j > x_k^{\max}$  then
        map  $\tilde{x}_k^j$  to the potential space:  $\tilde{x}_k^j = x_k^{\min} + |\tilde{x}_k^j| \% (x_k^{\max} - x_k^{\min})$ ;
    end if
end for

```

To maintain the diversity of sparks, there is another way of generating sparks called Gaussian explosion, which is presented in Algorithm 2. A function  $Gaussian(1, 1)$ , which is a Gaussian distribution with mean 1 and standard deviation 1, is utilized to define the coefficient of the explosion.  $\hat{m}$  sparks of this type are created in each explosion generation.

**Algorithm 2.** Find out the location of a specific spark

```

Initialize the location of the spark:  $\tilde{x}_j = x_i$ ;
 $z = \text{round}(D \times \text{rand}(0, 1))$ ;
Randomly choose  $z$  dimensions of  $\tilde{x}_j$ ;
Compute the coefficient of Gaussian explosion:  $g = Gaussian(1, 1)$ ;
for each dimension  $\tilde{x}_k^j \in \{\text{pre-selected } z \text{ dimensions of } \tilde{x}_j\}$  do
     $\tilde{x}_k^j = \tilde{x}_k^j \cdot g$ ;
    if  $\tilde{x}_k^j < x_k^{\min}$  or  $\tilde{x}_k^j > x_k^{\max}$  then
        map  $\tilde{x}_k^j$  to the potential space:  $\tilde{x}_k^j = x_k^{\min} + |\tilde{x}_k^j| \% (x_k^{\max} - x_k^{\min})$ ;
    end if
end for

```

## 2) Selection of locations

At the beginning of each explosion generation,  $n$  locations will be selected for the fireworks explosion. In the FWA, the current best location  $x^*$  in the current generation is always retained for the next explosion generation. After that,  $n - 1$  locations are chosen based on their distance to other locations in order to maintain the diversity of sparks. The general distance between a location  $x_i$  and other locations are defined as Eq. (16).

$$R(x_i) = \sum_{j \in C} d(x_i, x_j) = \sum_{j \in C} \|x_i - x_j\| \quad (16)$$

where  $C$  is the set of all current locations of both fireworks and sparks.

Then the selection probability of a location  $x_i$  is defined as Eq. (17).

$$p(x_i) = \frac{R(x_i)}{\sum_{j \in C} R(x_j)} \quad (17)$$

When assessing the distance, any distance measure might be used involving Euclidean distance, Manhattan distance, Angle-based distance, etc. In this paper, Euclidean distance is employed.

Algorithm 3 shows the overview of the FWA. During each

explosion generation, two kinds of sparks are generated respectively as presented in Algorithm 1 and Algorithm 2. In the first kind, the explosion amplitude and the number of sparks rely on the quality of the corresponding firework. In contrast, the second type is produced using a Gaussian explosion process, which carries out seeking in a local Gaussian space around a firework.

**Algorithm 3.** Framework of the FWA

```

Randomly choose  $n$  locations for fireworks;
while stopping criteria is not met do
    Set off  $n$  fireworks respectively at the  $n$  locations:
    for each firework  $x_i$  do
        Compute the number of sparks that the firework produces:  $\hat{s}_i$  by
        using Eq. (12) and Eq. (13);
        Find out locations of  $\hat{s}_i$  sparks of the firework  $x_i$  using
        Algorithm 1;
    end for
    for  $k = 1 \rightarrow \hat{m}$  do
        Randomly choose a firework  $x_j$ ;
        Create a specific spark for the firework by using Algorithm 2;
    end for
    Choose the best location and keep it for the next explosion generation;
    Randomly select  $n - 1$  locations from the two types of sparks and the
    current fireworks according to the probability given in Eq. (17);
end while
```

#### D. Training ANN using Fireworks Algorithm and Levenberg-Marquardt

The FWA has a strong capability of finding the global optimized result and the LM algorithm [26] has a strong ability to seek the local optimized solution. This paper proposes the combination of FWA with LM to train the ANN. The key idea of this method is that the FWA is employed at the beginning stage of searching for the optimum. Then, the training process is continued with the LM algorithm. In the first stage, the FWA finishes its training process; LM algorithm then begins training with the optimal weights of FWA algorithm with 1000 epochs more. The LM algorithm interpolates between the Newton method and gradient descent approach where it approximates the error of the network with a second order expression.

The diagram of hybrid FWA-LM algorithm is shown in Fig. 2.

## V. EXPERIMENTS

### A. Evaluation criteria

The performance of the proposed approach is assessed by using criteria as below:

- The Mean Squared Error (MSE) is computed by the following equation:

$$MSE = \frac{1}{T} \cdot \sum_{i=1}^T (A_i - P_i)^2 \quad (18)$$

where  $A_i$  and  $P_i$  are actual and predicted effort values of the  $i^{th}$  test data respectively.

- The Mean Magnitude of Relative Error (MMRE) is the average percentage of the absolute values of the relative errors over an entire data set. Given  $T$  tests, the MMRE is calculated as Eq. (19).

$$MMRE = \frac{100}{T} \cdot \sum_{i=1}^T \frac{|P_i - A_i|}{A_i} \quad (19)$$

- $PRED(N)$  indicates the average percentage of estimates that were within  $N$  percent of the actual values. Given  $T$  tests, then:

$$PRED(N) = \frac{100}{T} \cdot \sum_{i=1}^T \begin{cases} 1 & \text{if } \frac{|P_i - A_i|}{A_i} < \frac{N}{100} \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

For example,  $PRED(25) = 80\%$  means that 80% of the estimates are within 25 percent of the actual.

- The squared correlation coefficient ( $R^2$ ), also known as the coefficient of determination is computed as Eq. (21)

$$R^2 = 1 - \frac{\sum_{i=1}^T (A_i - P_i)^2}{\sum_{i=1}^T (A_i - \bar{A})^2} \quad (21)$$

where  $\bar{A}$  is the mean of actual effort values.

The higher the values of  $R^2$  and  $PRED(N)$  are, the better the values of estimated results are. In contrast, the lower the values of MSE and MMRE are, the more accurate the values of estimated results are.

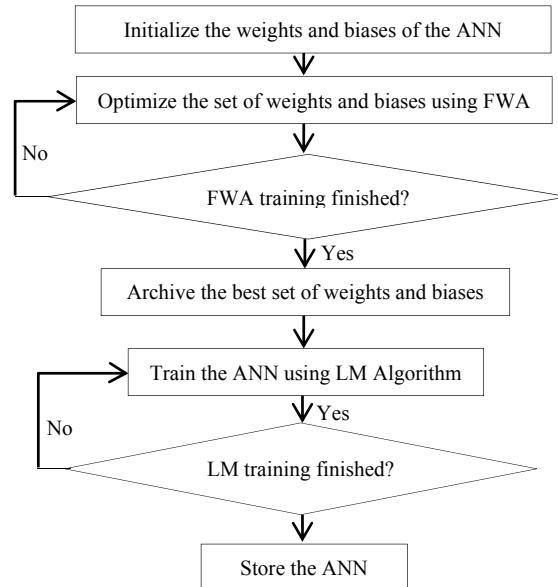


Fig. 2. The diagram of the hybrid FWA-LM Algorithm

### B. Data sets

The proposed method is validated using the data set of twenty-one projects developed by six software companies as presented in Zia's work [21]. The data set is three-dimensional. The first dimension is the number story points required to accomplish the project, the second one shows the velocity of the project, and the third presents the actual effort required to complete that project. Zia *et al.* used this data for predicting effort using the linear regression analysis. In this study, the first fifteen projects were utilized to train the ANN and the others were used for testing the proposed approach.

### C. Evaluating the performance of Fireworks Algorithm

In order to evaluate the effectiveness of the FWA, the

combination of LM with the other algorithms including directed artificial bee colony (DABC) algorithm [27], which is an improved version of artificial bee colony (ABC) algorithm, Teaching-Learning based optimization (TLBO) [27], Teaching-learning based artificial bee colony (TLBABC) [29] are compared to the proposed algorithm. LM algorithm

without using optimization algorithms is also used to compare to the hybrid FWA-LM approach. In this case, the number of epochs for the training process using only LM is equal to the sum of the number of cycles of the optimization algorithm and the number of epochs of the LM algorithm. Table IV shows the average results of the experiment after 10 execution times.

TABLE IV. THE EXPERIMENTAL RESULTS OF ALGORITHMS

<b>Effort</b>	<b>Velocity</b>	<b>Time</b>	<b>FWA-LM</b>	<b>TLBO-LM</b>	<b>DABC-LM</b>	<b>TLBABC-LM</b>	<b>LM</b>
289	2.8	112	108.8	88.0	99.0	98.0	94.8
113	2.8	39	41.6	41.9	43.4	43.1	45.9
141	2.8	52	53.1	52.6	54.3	54.1	55.8
213	2.8	80	81.2	75.2	80.1	80.3	80.8
237	2.7	56	54.3	52.5	54.5	54.4	54.3
91	2.7	35	35.5	34.9	36.2	36.1	40.4
<b>MSE</b>		<b>3.7983</b>	103.3450	32.8917	36.8467	65.0967	
<b>MMRE (%)</b>		<b>2.9339</b>	7.0923	5.5907	5.5710	9.9702	
<b>PRED(7.19)</b>		<b>100</b>	66.6667	66.6667	66.6667	33.3333	
<b>R<sup>2</sup></b>		<b>0.9946</b>	0.8530	0.9532	0.9476	0.9074	

From Table IV, it can be seen that the hybrid FWA-LM algorithm gave the best results on all evaluation criteria compared to the TLBO-LM, DABC-LM, LBOABC-LM and LM algorithms. The value of PRED(7.19) is chosen according to the results of Zia's work [21]. It also finds that the testing errors using FWA-LM are relatively small. This experiment pointed out that the proposed method showed the effort estimation results quite accurate and the FWA is an effective algorithm to optimize the set of weights and biases of the ANN. It is also clear that if we only use the LM algorithm for the training process then the accuracy of estimated results is fairly low. This indicates the important role of optimization algorithms in enhancing the precision of the ANN.

#### D. Comparison of proposed approach with the other types of Neural Networks

In [21], Zia *et al.* used the regression method to estimate the effort of agile projects and results were computed on twenty-first projects. This study uses fifteen projects for the training process and six other projects for testing. In order to ensure comparability, we only report the results of Zia's work on six projects of our testing set. In [22], Panda *et al.* used different types of neural networks such as GRNN, PNN, GMDH Polynomial Neural Network and Cascade-Correlation Neural Network. The performance of our proposed method will be compared with these studies. Table V presents the comparison of obtained results using different types of ANN.

TABLE V. COMPARISON OF RESULTS USING DIFFERENT TYPES OF ANN

<b>Name of metric</b>	<b>R<sup>2</sup></b>	<b>MMRE</b>	<b>PRED(%)</b>
General Regression Neural Networks	0.7125	0.3581	85.9182
Probabilistic Neural Networks	0.6614	1.5776	87.6561
GMDH Polynomial Neural Network	0.6259	0.1563	89.6689
Cascade Correlation Neural Network	0.9303	0.1486	94.7649
Zia's Regression	0.9661	0.6634	50
Our method	<b>0.9946</b>	<b>0.0293</b>	<b>100</b>

The obtained results indicated that our method

outperformed all kinds of ANN in Panda's work and the regression method of Zia. The experimental results pointed out that our proposed approach is a promising orientation in improving the accuracy of ANN and giving the right effort estimation results for agile projects.

#### VI. THREATS TO VALIDITY

This paper proposed how to apply ANN trained by using Fireworks algorithm in cooperation with LM for estimating the effort of the agile software development project.

Threats to construct validity insist on the way that the effort estimation models are defined. In this paper, the proposed model assumes that the value of initial project velocity is given by taking from the past projects developed by the same team in similar working conditions. However, when a team is new, the company will not have any past record of it. In that case, no obvious assignment to initial project velocity can be allocated. To solve this issue, we can use the average velocity values of all the teams working in similar conditions with the same size of the project and assign them to initial project velocity and then these data are utilized to train the ANN.

In our study, threats to external validity are related to the generalization to the other types of dataset. In this work, records of twenty-one projects developed by six software houses from the work of Zia *et al.* [16] are used without information with regard to the type of projects taken for research. To increase the persuasiveness, data covering all categories of software developed by agile methods should be collected and experimented.

The threat to internal validity might be caused by how the empirical study was conducted. Because of a small amount of dataset, the testing data is quite small in size with only six projects being used for testing. Therefore, the optimal precision of the performance of proposed method might not be guaranteed. Moreover, the ANN training approaches in this paper using FWA have the initial population being usually randomly generated, so the experiments might deliver various results. To cope with this problem, we carried out each experiment 10 times, and we followed rigorous statistical

procedures to evaluate the obtained results. Ten runs were a rule-of-thumb limit proposed by Ali *et al.* [17].

## VII. CONCLUSION

The story point method is one of the software effort estimation techniques that can be employed for agile software projects. In this study, the total number of story points and the project velocity are utilized to train the ANN to give the effort involved in developing an agile software product. The accuracy of estimation results is enhanced by using the FWA and LM to optimize weights and biases of the ANN. The experimental results indicated that the FWA is an effective algorithm to improve the accuracy of the ANN in comparison with the other algorithms such as DABC, TLBO, and TLBABC. The proposed ANN also outperformed the other types of ANN in other studies like the General Regression, Probabilistic, GMDH Polynomial, and Cascade Correlation neural networks. This study is also expanded by using other machine learning techniques such as support vector machine, random forest, stochastic gradient boosting based on the story point approach.

## REFERENCES

- [1] B. Boehm, B. Clark, E. Horowitz, C. Westland, R. Madachy, R. Selby, "Cost Models for Future Software Life Cycle Processes: COCOMO 2.0," *Annals of Software Engineering*, vol. 1, no. 1, pp. 57-94, 1995.
- [2] L. H. Putnam, "A General Empirical Solution to the Macro Software Sizing and Estimating Problem," *IEEE Transaction on Software Engineering*, vol. SE-4, no. 4, pp. 345-361, 1978.
- [3] F. R. Freiman, R. D. Park, "PRICE software model-Version 3, An overview," in *Proc. of IEEE-PINY Workshop on quantitative Software Models*, 1979, pp. 32-41.
- [4] S. M. Satapathy, B. P. Acharya, S. K. Rath, "Class Point Approach for Software Effort Estimation Using Stochastic Gradient Boosting Technique," *ACM SIGSOFT Software Engineering Notes*, pp. 1-6, 2014.
- [5] C. Blum, K. Socha, "Training feed-forward neural networks with ant colony optimization: an application to pattern classification," in *Proc. of the Fifth International Conference on Hybrid Intelligent Systems*, 2005, pp. 233-238.
- [6] R. Sexton, R. Dorsey, and J. Johnson, "Optimization of neural networks: A comparative analysis of the genetic algorithm and simulated annealing," *European Journal of Operational Research*, vol. 114, pp. 589-601, 1999.
- [7] R. Sexton, B. Alidaee, R. Dorsey, J. Johnson, "Global optimization for artificial neural networks: a tabu search application," *European Journal of Operational Research*, vol. 106, pp. 570-584, 1998.
- [8] C. Ozturk, D. Karaboga, "Hybrid Artificial Bee Colony algorithm for neural network training," in *IEEE Congress of Evolutionary Computation*, 2011, pp. 84-88.
- [9] T. Back, H. P. Schwefel, "An overview of evolutionary algorithms," *Evolutionary Computation*, vol. 1, no. 1, pp. 1-23, 1993.
- [10] D. Karaboga, C. Ozturk, "Neural networks training by artificial bee colony algorithm on pattern classification," *Neural Network World*, vol. 19, no. 3, pp. 279-292, 2009.
- [11] S. Mirjalili, S. Z. M Hashim, H. M. Sardroodi, "Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm," *Applied Mathematics and Computation*, vol. 218, no. 22, pp. 11125-11137, 2012.
- [12] Y. Tan, Y. Zhu, "Fireworks Algorithm for Optimization," in *Proc. of the First International Conference Advances in Swarm Intelligence*, 2010, pp. 355-364.
- [13] S. Keaveney, K. Conboy, "Cost estimation in agile development projects," in *Proceedings of the Fourteenth European Conference on Information Systems*, Göteborg, Sweden, 2006, pp. 183-197.
- [14] E. Coelho, A. Basu, "Effort Estimation in Agile Software Development using Story Points," *International Journal of Applied Information Systems*, vol. 3, no. 7, pp. 7-10, 2012.
- [15] P. Abrahamsson, J. Koskela, "Extreme Programming: A Survey of Empirical Data from a Controlled Case Study," in *Proceedings of International Symposium on Empirical Software Engineering*, 2004, pp. 73-82.
- [16] I. Hussain, L. Kosseim, O. Ormandjieva, "Approximation of COSMIC functional size to support early effort estimation in Agile," *Data & Knowledge Engineering*, vol. 85, pp. 2-14, 2013.
- [17] R. Popli, N. Chauhan, "Cost and effort estimation in agile software development," in *International Conference on Optimization, Reliability, and Information Technology*, 2014, pp. 57-61.
- [18] A. E. D. Hamouda, "Using Agile Story Points as an Estimation Technique in CMMI Organizations," in *Agile Conference (AGILE)*, Kissimmee, Finland, 2014, pp. 16-23.
- [19] A. L. Oliveira, "Estimation of software project effort with support vector regression," *Neurocomputing*, vol. 69, no. 13, pp. 1749-1753, 2006.
- [20] S. M. Satapathy, A. Panda, S. K. Rath, "Story Point Approach based Agile Software Effort Estimation using Various SVR Kernel Methods," in *International Conference on Software Engineering and Knowledge Engineering*, Vancouver, Canada, 2014, pp. 304-307.
- [21] Z. K. Zia, S. K. Tipu, S. K. Zia, "An Effort Estimation Model for Agile Software Development," *Advances in Computer Science and its Applications*, vol. 2, no. 1, pp. 314-324, 2012.
- [22] A. Panda, S. M. Satapathy, S. K. Rath, "Empirical Validation of Neural Network Models for Agile Software Effort Estimation based on Story Points," *Procedia Computer Science*, vol. 25, no. 2015, pp. 772-781, 2015.
- [23] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303-314, 1989.
- [24] A. Omidi, E. Nourani, M. Jalili, "Forecasting stock prices using financial data mining and Neural Network," in *Proc. of the 3rd International Conference on Computer Research and Development*, 2011, pp. 242-246.
- [25] A. Adebiyi, C. Ayo, M. O. Adebiyi, S. Otokiti, "Stock Price Prediction using Neural Network with Hybridized Market Indicators," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 3, no. 1, pp. 1-9, 2012.
- [26] M. Hagan, M. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989 - 993, 1994.
- [27] M. S. Kiran, O. Findik, "A directed artificial bee colony algorithm," *Applied Soft Computing*, vol. 26, pp. 454-462, 2015.
- [28] R. V. Rao, V. Patel, D. P. Vakharia, "Teaching-Learning-Based Optimization: An optimization method for continuous non-linear large scale problems," *Information Sciences*, vol. 183, pp. 1-15, 2012.
- [29] T. T. Khuat, M. H. Le, "Applying Teaching-Learning to Artificial Bee Colony for Parameter Optimization of Software Effort Estimation Model," *Journal of Engineering Science and Technology*, (Accepted), 2016.
- [30] S. Ali, L. Briand, H. Hemmati, R. Panesar-Walaweege, "A systematic review of the application and empirical investigation of search-based test case generation," *IEEE Transactions on Software Engineering*, vol. 36, no. 6, pp. 742-762, 2010.



**Thanh Tung Khuat** completed the B.S degree in Software Engineering from University of Science and Technology, Danang, Vietnam, in 2014. Currently, he is participating in the research team at DATIC Laboratory, University of Science and Technology, Danang. His research interests include software engineering, software testing, evolutionary computation, Intelligent Optimization Techniques and Applications in software engineering.



**My Hanh Le** is currently a lecturer of the Information Technology Faculty, University of Science and Technology, Danang, Vietnam. She gained M.Sc. degree in 2004 and Ph.D. degree in Computer Science from The University of Danang in 2016. Her research interests are about software testing, nature-inspired algorithms and more generally application of heuristic techniques to problems in software engineering.

# Critical evaluation of Maintainability Parameters using code metrics

Bhawana Mathur  
Dept. of CS&E  
JECRC University  
Jaipur, India

Manju Kaushik  
Dept. of CS&E  
JECRC University  
Jaipur, India

**Abstract**— Software maintenance is a noteworthy feature of software development life cycle, hence earlier approximation of work for maintainability plays a vibrant role. The C sharp small programs are programmed in console applications like a reverse number & check if it a palindrome, check whether given string is a palindrome or not, and so many. The 40 programs on Visual Studio 2012 are compiled and analyze the code metrics. After analysis code metrics parameters like MI, DIT, LOC, class coupling and cyclomatic complexity results are found. Existing approaches for maintainability estimation are the correlation between code metrics like maintainability index, cyclomatic complexity, Depth in Inheritance, class coupling, Line of Code in the experiments. On the off chance that the coefficient quality is in the negative shift, before which it implies the relationship between the variables is adversely associated, or as one worth increases, the different declines ,like Depth of Inheritance between cyclomatic complexity , Depth of Inheritance between class coupling, Lines of Code between Depth of Inheritance, Maintainability Index between Lines of Code. This paper likewise gives various understanding to the viable utilization of Maintainability Index. To reiterate, stay to remark the source codes yet don't put a lot of trust in remarks to enhance maintainability.

**Keywords-** *Code Metrics; Maintenance,; Maintainability Index; Lines of Code; Halstead Volume; Cyclomatic Complexity; Depth of Inheritance; Class coupling; smells; Lines of Code (LOC).*

## I. INTRODUCTION

In an evolving domain, programming is additionally inclined to software maintainability to adjust programming support is one of the key procedures of the software life cycle. The explanation behind the product updates is to keep programming operation, avert and revise flaws in the product and enhance the usefulness of the product. Support alludes to the alterations made to programming frameworks after their underlying discharge. It is unrealistic to build up a product framework that does not require support since change is the automatic nature of programming frameworks [1].

Maintenance is characterized by the IEEE as "the procedure of changing a product framework or part later conveyance to right blames, show signs of improvement execution or else different Characteristics, adjust to a modified domain". The maintainability is firmly identified. With programming support in light of the fact that the ease with which the maintenance of the framework is performed hence maintainability. There have been numerous endeavors to evaluate the maintainability of a product framework. The generally utilized programming metric

which measures the practicality is known as maintainability index (MI).

The maintainability index was given by Oman and Hagemeister and was made out of various software metrics which is the source of software maintainability. It comprises of a polynomial expression and results in a number demonstrating the general framework maintainability. Khan et.al. in their book characterized maintainability index as a gathering of software metrics to be specific McCabe's Cyclomatic many-sided quality (cc), Halstead's volume (v) and lines of code (LOC) that influence maintainability of the product. Maintainability Index might change from new code added to the current source code because of bugs altering or other remedial activities. As indicated by Coleman a Maintainability Index estimation of 85 shows that the product is exceptionally maintainable, an estimation of 85 and 65 moderate maintainability and a quality underneath 65 demonstrates that the framework is hard to keep up [2].The maintainability index (MI) is a generally known as an estimation of maintainability. Run of the mill estimations of maintainability index ranges from 200 to-100s. Higher maintainability index is the reflection of better maintainability. The issue of foreseeing the maintainability of programming has been composed on how maintainability can be anticipated by utilizing different instruments and procedures at the season of planning by reason software design metrics.

Studies have been carrying out to establish a link between object-oriented software metrics and its maintainability. They have also found that these metrics can be used as interpreters of maintenance effort. Exact expectation of software maintainability can be helpful due to the accompanying reasons:

- It ventures supervisors interestingly the profitability and expenses alongside undertakings.*
- It provides managers with information to use beneficial resources in the best way.*

## II. LITERATURE SURVEY

This section intends to highlights a review of the literature on the use of software metrics and maintainability index [2][3][4][5].

TABLE I.  
MAINTAINABILITY

Author	Technique	Work done
Coleman, D., Ash, D., Lowther, B. and Oman, P. 1994	Hierarchical multidimensional assessment models, Polynomial regression models, an aggregate complexity, Principal components analysis, Factor analysis.	They have connected measurements based software maintainability models to 11 industrial programming frameworks and utilized the outcomes for actuality finding and choices.
Welker, K. D. 2001	It comprises of a polynomial expression.	As it was first proposed by Oman and Hagemeister, the MI is contained of weighted Halstead measurements (exertion or volume), McCabe's Cyclomatic Complexity, lines of code (LOC), and then a number of remarks. Two mathematical statements were exhibited: one that considered remarks and another one that did not.
Ganpati1, A., Kalia .A., Singh, H., 2012	The software metrics were intentional utilizing Resource Standard Metrics (RSM) instrument and Crystal Flow device.	In this study, the MI of four furthermore regular OSS specifically Apache, Mozilla Firefox, MySQL and FileZilla for fifty progressive discharges was experimentally inspected. The MI as far as software metrics viz. Lines of Code (LOC), Cyclomatic Complexity (CC), and Halstead Volume (V) was figured for all the fifty progressive adaptations of four OSS.

### III. RESEARCH QUESTIONS

RQ1: What is the use of Minimum and Maximum Value of Maintainability Index?

Sol: A high value means better maintainability and minimum value means lower maintainability.

RQ2: What is the use of this analysis in these experiments?

Sol: So, many researchers work on these fields but we find out better maintainability of object-oriented software like c sharp and the analyses the code metrics.

RQ3: Which research methodology is followed in this paper and why?

Sol: We use Descriptive Study and Correlation Research Methodology used. For Metrics, that is strongly associated with other, we choose only one of them for further consideration.

### IV. THE GOAL OF THE STUDY

Analysis Code Metrics through object-oriented (C Sharp) Software with the help of Code Metrics to find obtainable better maintainability through various parameters like Maintainability index, LOC, Cyclomatic Complexity, Depth of Inheritance, Class Coupling etc. An unprejudiced approach of this study is to assess maintainability index for comparatively object-oriented C# Experiments and to find out optimizing results with the help of Visual Studio Code Metrics. To evaluate consequences, establish through code metrics and to enhance maintainability, assessing code quality in an object-oriented programming.

Identify applicable sponsor/s here. (sponsors)

TABLE II.  
CODE METRICS AND EXPERIMENT

Projects	Maintainability Index	Cyclomatic Complexity	Depth of Inheritance	Coupling Class	Lines of Code
E1	15	12	1	2	35
E2	93	7	3	3	11
E3	91	6	2	3	11
E4	91	8	2	4	16
E5	84	8	2	3	18
E6	74	8	1	7	6
E7	67	7	1	12	13
E8	66	14	1	11	11
E9	85	29	1	12	37
E10	72	6	1	8	8
E11	72	6	1	8	8
E12	74	20	1	12	41
E13	72	13	1	2	28
E14	74	2	1	2	9
E15	77	4	1	1	7
E16	68	6	1	3	20
E17	68	4	1	2	12
E18	66	4	1	2	14
E19	68	4	1	2	14
E20	71	5	1	8	15
E21	90	16	1	4	20
E22	86	8	1	3	19
E23	71	2	1	2	10
E24	77	2	1	3	6
E25	68	9	1	2	20
E26	67	5	1	2	14
E27	78	2	1	1	6
E28	60	6	1	8	23
E29	68	3	1	1	18
E30	72	3	1	2	10
E31	65	5	1	1	17
E32	82	4	1	3	9
E33	91	9	1	4	8
E34	91	10	1	4	17
E35	91	10	1	4	17
E36	91	9	1	8	14
E37	85	5	1	1	11
E38	87	8	1	5	16
E39	89	9	1	6	10
E40	90	11	1	9	16

### V. RESEARCH METHODOLOGY

The source code of C# 40 Experiments runs in Visual Studio 2012. Discover Code Metrics like parameters Maintainability Index, Depth of Inheritance, Cyclomatic Complexity, Class Coupling and Lines of Code. Descriptive Study and Correlation

Research Methodology is used here. For Metrics, that is strongly associated with other, for further consideration.

#### A. Descriptive Methodology

A measure of variation (for dispersion) describes the spread or deviation of the individual values around the central value of a set of data. Therefore, difference sets of data may have the same central value but differ greatly in terms of variation or dispersion.

##### 1) Different measures of variation

There are following five different measures of variation :

- a) Range
- b) Quartile deviation (or semi-interquartile range)
- c) Mean deviation
- d) Standard deviation
- e) Variance

#### B. Correlation

Literally, correlation means an association of two or more facts. In statistics, the correlation may be defined as 'the tendency of simultaneous variation between two variables'. The distribution involving two variables are called bivariate distribution and the distribution involving more than two variables are called multivariate distribution. In statistics, the degree of correlation between two or more variables is studied. Correlation can be defined according to the following points:

- 1) Measures the relative strength of the linear relationship between two variables.
- 2) Unit-less.
- 3) Ranges between -1 and 1.
- 4) The closer to -1, the stronger the negative linear relationship.
- 5) The closer to 1, the stronger the positive linear relationship.
- 6) The closer to 0, the weaker any positive linear relationship

## VI. SOURCE CODE MEASUREMENT

Run of the mill software metrics is the span of the code (measured in Lines of Code and number of statements) and the code complexity (measured through unpredictability figures like the Cyclomatic Complexity). According to parameters, a result was concluded by code metrics. An arrangement of such cutoff points from various effective system composing rules characterizes a programming standard, i.e., any correlation between these qualities can prompt an agent perspective of the nature of the experimented projects.

The metrics considered are among the comprehensive reports and utilized as a part of the information and are as following:

#### A. A Number of lines of code (LOC)

It measures the physical size of the system code, barring clear lines and remarks.

#### B. Cyclomatic Complexity $V(g)$

It was proposed by McCabe, these metrics check the quantity of control flow graph of a project constituent. Cyclomatic complexity worth relies on the quantity of branches created by contingent proclamations (if-then-else). It activities have the structural complexity of the part.

#### C. Maintainability Index (MI)

MI measures maintainability by focusing on the size, the intricacy, and the self-distinction of the code considered. A new maintainability index is obtained after experimentation of the current source code and a product, "new code" is formed. However, since MI depends on its features which are considered as qualities of software. It is similarly autonomous without a doubt; the span of these progressions might be used to analyze frameworks of disparate size. The coefficients of the process for Maintainability Index have been adjusted [8]. On different programming, frameworks kept up by Hewlett-Packard. Maintainability Index defenders confirmed this type of the MI perfect for the mathematical statement. In most cases, modern estimated delicate product frameworks are applied. High MI values show high Maintainability.

#### D. Coupling

In 1974, Stevens et al. initially characterized Coupling in the setting of organized advancement as "the measure of the quality of affiliations built up by an interface particle from one module to another". Coupling is a measure of association of two items. For instance, objects A and B are coupled if a strategy for questioning A calls a technique or gets to a variable in item B. Classes are coupled when techniques report in one class reported strategies or characteristics of alternate classes.

#### E. The Depth of Inheritance Tree (DIT)

The profundity of a class of the legacy progressive system is characterized as the most extreme length of the class hub to the root/guardian of the class order tree and is planned by progenitor classes. In cases including various legacy, the DIT is the most extreme length of the hub to the base of the tree [7]

## VII. ANALYSIS, IMPLEMENTATION, AND VALIDATION

In this paper, show the objective to build object-oriented software.

TABLE III. DESCRIPTIVE ANALYSIS OF CODE METRICS

	Maintainability Index	Cyclomatic Complexity	Depth of Inheritance	Coupling Class	Lines of Code
Mean	76.18	7.73	1.13	4.50	15.38
Standard Error	2.21	0.83	0.06	0.53	1.29
Median	74.00	6.50	1.00	3.00	14.00
Mode	91.00	6.00	1.00	2.00	11.00
Standard Deviation	13.95	5.25	0.40	3.37	8.15
Sample Variance	194.71	27.59	0.16	11.33	66.50
Kurtosis	8.32	6.25	12.57	-0.06	2.66
Skewness	-2.04	2.10	3.48	1.04	1.58
Range	78.00	27.00	2.00	11.00	35.00
Minimum	15.00	2.00	1.00	1.00	6.00
Maximum	93.00	29.00	3.00	12.00	41.00
Sum	3047.00	309.00	45.00	180.00	615.00
Count	40.00	40.00	40.00	40.00	40.00
Confidence Level (95.0%)	4.46	1.68	0.13	1.08	2.61

TABLE IV. CORRELATION BETWEEN CODE METRICS

	Maintainability Index	Cyclomatic Complexity	Depth of Inheritance	Coupling Class	Lines of Code
Maintainability Index	1.00				
Cyclomatic Complexity	0.11	1.00			
Depth of Inheritance	0.32	-0.03	1.00		
Coupling Class	0.09	0.60	-0.12	1.00	
Lines of Code	-0.30	0.73	-0.08	0.28	1.00

Co-variation between two variables in opposite direction are negatively correlated with Depth of Inheritance and Cyclomatic Complexity are -0.03, Coupling Class and Depth in Inheritance are -0.12, Lines of code and Maintainability Index are -0.30, as well as Lines of codes and Depth in Inheritance, are -0.08. In the other way, the correlation between the two variables in

which can be expressed by a straight line is called linear correlation. In perfect linear correlation, the amount of change in one variable bears a constant ratio to the amount of change in the other.

TABLE V. CORRELATION BETWEEN MAINTAINABILITY INDEX AND CYCLOMATIC COMPLEXITY

	Maintainability Index	Cyclomatic Complexity
Maintainability Index	1	
Cyclomatic Complexity	0.114	1

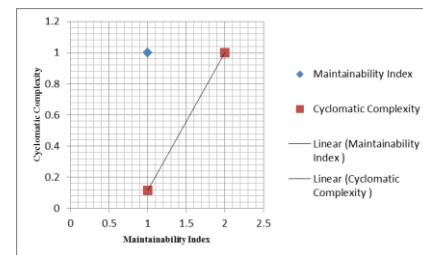


Figure 1. Correlation between Maintainability Index and Cyclomatic Complexity

## VIII. SOFTWARE MAINTAINABILITY METRICS HELP IDENTIFY PROBLEM AREAS

Software maintainability needs more workout for the engineer and hence it belongs to the development of software lifecycle. A forty programming has done four sorts of maintenance on new arrangements or upgrades: remedial, versatile, perfective, and deterrent. These moves have made supplementary time to finish if the code is difficult to oversee in any case. A PC program with these capacities requires expanded programming: poor code quality, undetected vulnerabilities, source code imperfections, inordinate specialized multifaceted nature, vast frameworks, defective reported frameworks, over the top dead code. Asset needs to keep on developing as situations turn out to be continuously intricate and applications are immediately created. Software maintainability metrics give a modest, and exact way to deal with identifying conceivable reasons for unmaintainable frameworks. Thus, researcher association gets vision into change regions and can screen basic frameworks in each part its application. The issue of foreseeing the maintainability of programming is largely recognized in the business and much has been composed on how maintainability can be evaluated by utilizing diverse devices. Procedures of the phases of planning with the assistance of software design metrics concentrates on how to manage and find out the firmness of connection between Object Oriented software metrics and its maintainability. They have moreover built up that these metrics can be appropriate as indicators of maintenance effort. Exact

expectation of software maintainability can be productive in view of the accompanying rationale:

- 1) *It helps project managers in contrast to the productivity and costs of software.*
- 2) *It gives managers with information for emphatic designing for the use of valuable resources.*

## IX. DISCUSSION

Utilizing the Maintainability Index to survey a source code and is this way distinguish and measure maintainability. It is a powerful approach. The Maintainability Index gives one little point of view of the exceptionally complex issues of software maintenance. The Maintainability Index gives a fabulous manual for direct human examination. This paper gives various updates for generalized utilization of Maintainability Index. To recap, keep on remarking the source codes, however, do not place an excess of confidence in remarks to enhance maintainability. Keep on measuring maintainability utilizing Maintainability Index without translating the outcomes in a vacuum. Know about the confinements of target measurements, for example, MI. While changing advancements it requires evolving measurements. In designing, maintainability it should be kept in mind that deficiencies should overcome in future with regards that it benefit as much as possible from proficiency, dependability, and security, meet new necessities, make future upkeep less demanding, or deal with a changed situation. While designing it should be kept in mind that defective segments should be recovered in spite of substituting newer segments. Utilizing the MI to survey source code and recognizing and measure maintainability is a viable methodology. The maintainability index gives one little point of view of the profoundly complex issues of software maintenance. The MI gives a magnificent manual for direct human examination [6][7][8].

## X. RESULT

The Maintainability Index of Experiments 40 was practical over statistical tests. The results show that two variables co-varying in the same direction are positively correlated i.e. a positive correlation between cyclomatic complexity and maintainability index, Depth of Inheritance and maintainability index, class coupling and maintainability index ,class coupling and cyclomatic complexity lines of codes and cyclomatic complexity and lines of codes and class coupling of code metrics. Similarly, Co-variation between the two variables in opposite direction is negatively correlated. The increase in one variable results in a corresponding decrease in the other .For example, increase in lines of codes results in a corresponding decrease in maintainability index, depth in inheritance and cyclomatic complexity ,class coupling and depth in inheritance as well as lines of codes in code metrics. Maximum the maintainability Index is 93 of experiment 2. Minimum Maintainability Index is 15 of experiment 1 after Statistical analysis of a set of 40 programs in c sharp object-oriented programming. Utilizing the Maintainability Index to evaluate a source code and consequently recognize and measure Maintainability is a

compelling methodology. The Maintainability Index gives one little viewpoint of the exceptionally complex issues of Software maintenance.

## XI. CONCLUSION AND FUTURE SCOPE

The Maintainability Index gives a brilliant manual for direct human examination. This paper gives various considerate to the pragmatic utilization of Maintainability Index.

During working in this area of investigation, a lot of scope for future work has been observed. There is need of further empirical investigations to find maintainability index. One technique for testing this objective is by controlled test and investigation. Programming designing endeavors the cost, build dependability, and expansion heartiness and in addition to other things .The objective of this analysis is to grow the establishments of software designing so those work with programming can settle on savvy decisions when fabricating and keep up Systems. Similar Experiments can be carried out for maintainability index. There is in need of further empirical investigations and found Maintainability Index. In this work, Maintainability Index of software is computed based of Cyclomatic Complexity, Lines of Code (LOC), Depth of Inheritance, and Class coupling and their correlation coefficient.

## ACKNOWLEDGEMENT

I am also deeply indebted to JECRC University Foundation who has given me this opportunity to fulfill Ph.D. work. We want to thank JECRC University Foundation for providing their excellent tool freely and their kind helps me for the tool use.

## REFERENCES

- [1] Malhotra R. and Chug A.(2014), “A Metric Suite for Predicting Software Maintainability in Data Intensive Applications”, Kim H.K.,AO S., Amouzegar M.A.(eds.), *Transactions on Engineering Technologies* , Springer Netherlands, pp. 161-175.
- [2] Kumar L. and Rath S.K.(2015), “Neuro–Genetic Approach for Predicting Maintainability Using Chidamber and Kemerer Software Metrics Suite” , Herwig U., Phayung M., Sirapat B., (eds.),*In Recent Advances in Information and Communication Technology*, Springer International Publishing, 361,pp. 31-40
- [3] Coleman D., Ash D., Lowther B. and Oman P. (1994), “Using metrics to evaluate software system maintainability, “*Computer*”, 27(8), pp.44-49.
- [4] Welker K.D., (2001), “The software maintainability index revisited”, *CrossTalk* , 14, pp.18-21.
- [5] Ganpati A., Kalia A. and Singh H.(2012), “A Comparative Study of Maintainability Index of Open Source Software”, *Int. J. Emerg. Technol. Adv. Eng.* , 2(10), pp.228-230.
- [6] Rivero C.R., Hernandez I., Ruiz D. and Corchuelo R.(2013), “Benchmarking data exchange among semantic-web ontologies”, *IEEE Transactions on Knowledge and Data Engineering*, 25(9), pp.1997-2009.

- [7] Coleman D., Lowther B. and Oman P. (1995), "The application of software maintainability models in industrial software systems", *Journal of Systems and Software*, 29(1), pp.3-16.
- [8] Henry S., Humphrey M. and Lewis J. (1990), "Evaluation of the maintainability of object-oriented software", Liu Y.W.(eds.),10<sup>th</sup> Conference on Computer and Communication Systems, *TENCON'90*, Hong Kong, Sept 24-27, 1990. IEEE, pp. 404-409.

#### AUTHORS PROFILE

Authors Profile

Bhawana Mathur

Bhawana Mathur ,Research Scholar, Dept. of Computer Science and Engineering at JECRC University, Jaipur, Rajasthan, India. She has completed her MCA from IGNOU.

Manju Kaushik

Dr. Manju Kaushik, Associate Professor, Dept. of Computer Science and Engineering at JECRC University, Jaipur, Rajasthan, India. She has completed her Ph.D. from Mohan Lal Sukhadia University, Udaipur, Rajasthan, India.

# An efficient $(n, n)$ - threshold secret sharing Scheme using on FAPKC4 and hash function

Ali Saeidi Rashkolia,  
Department of Mathematics,  
Graduate University of Advanced Technology,  
Kerman, Iran,

Mohammad Mahdi Zahedi  
Department of Mathematics,  
Graduate University of Advanced Technology,  
Kerman, Iran,

Masoud Hadian Dehkordi  
School of Mathematics,  
Iran University of Science and Technology,  
Tehran, Iran,

**Abstract**—The main purpose of this paper is to give a  $(n, n)$  - thresholdsecret sharing scheme based on the inversion of weakly invertible finite automata. It is varifiable, practical it does not face with time-spending computation such as "discrete logarithm" moreover both combiner and participants can investigate the validity of exchanged data.

Security can be reduced to the generalization of finite automata public key cryptosystem FAPKC4, because the secret is encrypted by using it. This is a strong property since the FAPKC4 is believed to be secure.

**Keywords-component; finite automaton, secret sharing, weakly invertible, weak inverse, hash function, public key cryposystem.**

## I. INTRODUCTION

A secret sharing scheme (SSS) allows one to split a secret  $s$  into different pieces, called shares, which are given to the set of participants  $P$  such that only certain qualified subsets of participants can recover the secret using their respective shares [9]. The collection of those qualified sets of participants is called the access structure corresponding to  $s$ . Secret sharing plays an important role in protecting secret information from becoming lost, destroyed, or falling into the wrong hands [2],[3],[4]. Blakley [1] and Shamir [5], in 1979, independently, came out with a scheme known as the  $(t, n)$  threshold secret sharing scheme. In recent years, secret sharing schemes have found many applications in diverse areas such as access control systems, e-voting schemes and digital cash protocols, to name a few. SSS is said to be verifiable if the participants can check the correctness of their shares given by the dealer

and the reconstructed secret given by the combiner, and the combiner can check whether the participants have submitted the correct shares or not. In 2008, Wei et al. [8] proposed a renewable secret sharing scheme for general access structure. A renewable secret sharing scheme allows new secrets to be added. In addition, the participant set and the access structure can be changed dynamically without updating any participant's share.

Automata theory is a mathematical theory to investigate behavior, structure and their relationship to discrete and digital systems. It can be considered as a natural model of cryptosystems. Since up to now, in studying the cryptoststem based on automata, the invertibility of finite automata has a main role, for example in studying the FAPKC4 in [7]. All the cryptosystems based on invertibility theory of finite automaton, in which their securities depends on the difficulties of inversion of nonlinear finite automata. For more information about the invertibility of finite automata, the reader may be referred to [6]. The proposed scheme is a verifiable, renewable, secret sharing scheme where each secret can be reconstructed independently. The uses of inversion of the weakly invertible finite automata also a one-way collision resistant hash function. In the next section we introduce the necessarily preliminaries about automata theory. In section III, is devoted to main idea. The security of the scheme is checked in section IV, and finaly section V is devoted to the conclusion of the paper.

## II. PRELIMINARIES

### A. Finite Automata and Compound of tow automata

**Definition 1.** [6] A finite automata is a quintuple  $(X, Y, S, \delta, \lambda)$ , where:

- $X$  is a nonempty finite set called the input alphabet of the finite automaton;
- $Y$  is a nonempty finite set called the output alphabet of the finite automaton;
- $S$  is a nonempty finite set called the set of states of the finite automaton;
- $\delta$  is a function from  $S \times X$  to  $S$  called the state transition function of the finite automaton;
- $\lambda$  is a function from  $S \times X$  to  $Y$  called the output function.

To simplify the notions we use  $\|_{i=1}^k x_i$  instead of the word  $x_1 x_2 \dots x_k$ . Let  $A = (X, Y, S, \delta, \lambda)$  be a finite automaton. The state transition function  $\delta$  and the output function  $\lambda$  can be extended to words, i.e. elements of  $X^*$  recursively, as follows:  $\delta(s, \varepsilon) = s$ ,  $\delta(s, \|_{i=0}^n x_i) = \delta(\delta(s, x_0), \|_{i=1}^n x_i)$ ,  $\lambda(s, \varepsilon) = \varepsilon$ ,  $\lambda(s, \|_{i=0}^n x_i) = \lambda(s, x_0) \lambda(\delta(s, x_0), \|_{i=1}^n x_i)$ , where  $s \in S$ ,  $n \in N$  and  $\|_{i=0}^n x_i \in X^{n+1}$ . In an analogous way,  $\lambda$  may be extended to  $X^\omega$ .

**Definition 2.** Suppose that  $f$  is a single-valued mapping from  $S^{p+1} \times X^{r+1}$  to  $Y$ , and  $g$  is a single-valued mapping from  $S^{p+1} \times X^{r+1}$  to  $S$ . We define the finite automaton  $M(X, Y, S^{p+1} \times X^r, \delta, \lambda)$  by

$$\begin{aligned} \lambda((s_i, \dots, s_{i-p}, x_{i-1}, \dots, x_{i-r}), x_i) &= y_i, \\ \delta((s_i, \dots, s_{i-p}, x_{i-1}, \dots, x_{i-r}), x_i) &= (s_{i+1}, \dots, s_{i+1-p}, x_i, \dots, x_{i+1-r}), \end{aligned} \quad (1)$$

where

$$\begin{aligned} y_i &= f(s_i, \dots, s_{i-p}, x_i, \dots, x_{i-r}), s_{i+1} = \\ g(s_i, \dots, s_{i-p}, x_i, \dots, x_{i-r}). \end{aligned} \quad (2)$$

**Definition 3.** Suppose that  $f_t^*$  is a single-valued mapping from  $X^r \times S^{p+1} \times Y^t$  to  $X$ , and  $g$  is a single-valued mapping definded by Definition 2. We define the finite automaton  $M^* = (Y, X, X^r \times S^{p+1} \times Y^t, \delta^*, \lambda^*)$  by

$$\lambda^*((x_{i-1}, \dots, x_{i-r}, s_i, \dots, s_{i-p}, y_{i-1}, \dots, y_{i-t}), y_i) = x_i,$$

$$\begin{aligned} \delta^*((x_{i-1}, \dots, x_{i-r}, s_i, \dots, s_{i-p}, y_{i-1}, \dots, y_{i-t}), y_i) &= \\ = (x_i, \dots, x_{i+1-r}, s_{i+1}, \dots, s_{i+1-p}, y_i, \dots, y_{i+1-t}). \end{aligned} \quad (3)$$

where,

$$\begin{aligned} x_i &= f_t^*(x_{i-1}, \dots, x_{i-r}, s_i, \dots, s_{i-p}, y_i, \dots, y_{i-t}), s_{i+1} = \\ g(s_i, \dots, s_{i-p}, x_i, \dots, x_{i-r}). \end{aligned} \quad (4)$$

**Definition 4.** Suppose that  $\varphi$  is a single-valued mapping from  $Y^k \times W^{n+1} \times Z^{h+1}$  to  $Y$  and  $\psi$  is a single-valued mapping from  $Y^k \times W^{n+1} \times Z^{h+1}$  to  $W$ . We define the finite automaton  $M' = (Z, Y, Y^k \times W^{n+1} \times Z^h, \delta', \lambda')$  by  $\lambda'((y_{i-1}, \dots, y_{i-k}, w_i, \dots, w_{i-n}, z_{i-1}, \dots, z_{i-h}), z_i) = y_i$ ,

$$\begin{aligned} \delta'((y_{i-1}, \dots, y_{i-k}, w_i, \dots, w_{i-n}, z_{i-1}, \dots, z_{i-h}), z_i) &= \\ = (y_i, \dots, y_{i+1-k}, w_{i+l}, \dots, w_{i+1-n}, z_i, \dots, z_{i+1-h}), \end{aligned} \quad (5)$$

where

$$y_i = \varphi(y_{i-1}, \dots, y_{i-k}, w_i, \dots, w_{i-n}, z_i, \dots, z_{i-h}),$$

$$w_{i+1} = \psi(y_{i-1}, \dots, y_{i-k}, w_i, \dots, w_{i-n}, z_i, \dots, z_{i-h}). \quad (6)$$

**Definition 5.** Suppose that  $\varphi_t^*$  is a single-valued mapping from  $Z^h \times W^{n+1} \times Y^{t'+k+1}$  to  $Z$  and  $\psi$  is a single-valued mapping definded by Definition 4. We define the finite automaton  $M'^* = (Y, Z, Z^h \times W^{n+1} \times Y^{t'+k}, \delta'^*, \lambda'^*)$  by

$$\begin{aligned} \lambda'^*((z_{i-1}, \dots, z_{i-h}, w_i, \dots, w_{i-n}, y_{i-1}, \dots, y_{i-t'-k}), y_i) &= z_i, \\ \delta'^*((z_{i-1}, \dots, z_{i-h}, w_i, \dots, w_{i-n}, y_{i-1}, \dots, y_{i-t'-k}), y_i) &= \\ = (z_i, \dots, z_{i+1-h}, w_{i+1}, \dots, w_{i+1-n}, y_i, \dots, y_{i+1-t'-k}), \end{aligned} \quad (7)$$

where

$$\begin{aligned} z_i &= \varphi_t^*(z_{i-1}, \dots, z_{i-h}, w_i, \dots, w_{i-n}, y_i, \dots, y_{i-t'-k}), \\ w_{i+1} &= \psi(y_{i-t'-1}, \dots, y_{i-t'-k}, w_i, \dots, w_{i-n}, z_i, \dots, z_{i-h}). \end{aligned} \quad (8)$$

**Definition 6.** Suppose that  $\varphi_t^*$  is a single-valued mapping definded by Definition 5. and  $\psi$  is a single-valued mapping definded by Definition 4. We difine the compound finite automaton of  $M$  and  $M'^*$  by

$$\begin{aligned} C'(M, M'^*) &= (X, Z, Z^h \times W^{n+1} \times S^{t'+k+p+1} \times \\ &\quad X^{t'+k+r}, \delta'', \lambda'') \quad \text{where} \\ \lambda''((z_{i-1}, \dots, z_{i-h}, w_i, \dots, w_{i-n}, s_i, \dots, s_{i-t'-k-p}, \\ &\quad x_{i-1}, \dots, x_{i-t'-k-r}), x_i) = z_i, \\ \delta''((z_{i-1}, \dots, z_{i-h}, w_i, \dots, w_{i-n}, s_i, \dots, s_{i-t'-k-p}, \\ &\quad x_{i-1}, \dots, x_{i-t'-k-r}), x_i) \\ = (z_i, \dots, z_{i+1-h}, w_{i+1}, \dots, w_{i+1-n}, s_{i+1}, \dots, s_{i+1-t'-k-p}, \\ &\quad x_i, \dots, x_{i+1-t'-k-r}), \end{aligned}$$

and

$$\begin{aligned} \varphi'^*(z_{i-1}, \dots, z_{i-h}, w_i, \dots, w_{i-n}, f(s_i, \dots, s_{i-p}, x_i, \dots, x_{i-r}), \dots, \\ f(s_{i-t'-k}, \dots, s_{i-t'-k-p}, x_{i-t'-k}, \dots, x_{i-t'-k-r})), \end{aligned}$$

$$\begin{aligned} w_{i+1} &= \psi(f(s_{i-t'-1}, \dots, s_{i-t'-1-p}, x_{i-t'-1}, \dots, x_{i-t'-1-r}), \dots, \\ f(s_{i-t'-1-k}, \dots, s_{i-t'-1-k-p}, x_{i-t'-1-k}, \dots, x_{i-t'-1-k-r}), w_i, \dots, w_{i-n}, \\ &\quad z_{i-h}), \end{aligned}$$

$$s_{i+1} = g(s_i, \dots, s_{i-p}, x_i, \dots, x_{i-r}). \quad (9)$$

**Definition 7.** Two finite automaton  $M$  and  $M^*$  satisfy  $P_{IC1}(M, M^*)$ , if the following condition is hold: for any state  $s_0 = (s_0, \dots, s_{-p}, x_{-1}, \dots, x_{-r})$  of  $M$  and any  $x_0, x_1, \dots \in X$ , if  $\|_{i=0}^\infty y_i = \lambda(s_0, \|_{i=0}^\infty x_i)$  then  $\|_{i=0}^\infty y_i = \lambda^*(s_t^*, \|_{i=t}^\infty y_i)$  where  $s_t^* = (x_{-1}, \dots, x_{-r}, s_0, \dots, s_{-p}, y_{t-1}, \dots, y_0)$ .

**Definition 8.** Suppose that  $\psi$  is a single-valued mapping defined by Definition 4. Two finite automaton  $M'$  and  $M'^*$  satisfy the condition  $P_{IC2}(M'^*, M')$ , if the following condition is hold: for any state

$$\begin{aligned} s'_0^* &= (z_{-1}, \dots, z_{-h}, w_0, \dots, w_{-n}, y_{-1}, \dots, y_{-t'-k}) \text{ of } M'^* \text{ and} \\ \text{any } y_0, y_1, \dots \in Y &\quad \text{if } \|_{i=0}^\infty z_i = \lambda'^*(s'_0^*, \|_{i=0}^\infty y_i) \text{ then} \\ \|_{i=0}^\infty y_i &= \lambda'(s'_{t'}, \|_{i=t'}^\infty z_i) \end{aligned}$$

where  $s'_{t'} = (y_{-1}, \dots, y_{-k}, w_{t'}, \dots, w_{t'-n}, z_{t'-1}, \dots, z_{t'-h})$  and

$$w_{i+1} = \psi(y_{i-t'-1}, \dots, y_{i-t'-k}, w_i, \dots, w_{i-n}, z_i, \dots, z_{i-h}), i = 0, \dots, t' - 1. \quad (10)$$

Now we present the public key cryptosystem which uses finite automaton for encryption. Tao and Chen proved the following theorem in [7].

**Theorem 1. (FAPKC4)** Assume that  $M^*$  and  $M$  satisfy  $P_{IC1}(M, M^*)$  and  $M'$  and  $M'^*$  satisfy  $P_{IC2}(M'^*, M')$ . For any state  $s'' =$

$(z_{-1}, \dots, z_{-h}, w_0, \dots, w_{-n}, s_0, \dots, s_{-t'-k-p}, x_{-1}, \dots, x_{-t'-k-r})$  of  $C'(M, M'^*)$  and any  $x_0, x_1, \dots \in X$ , if  $\lambda''(s'_0, ||_{i=0}^\infty x_i) = ||_{i=0}^\infty z_i$ , and  $||_{i=0}^\infty y_i = \lambda'(s'_t, z_{t'} z_{t'+1} \dots)$ , then  $\lambda^*(s_t^*, ||_{i=t}^\infty y_i) = ||_{i=0}^\infty x_i$ , where  $s_t'' = (x_{-1}, \dots, z_{-r}, s_0, \dots, s_{-p}, y_{t-1}, \dots, y_0)$ ,  $s'_t = (y_{-1}, \dots, y_{-k}, w_{t'}, \dots, w_{t'-n}, z_{t'-1}, \dots, z_{t'-h})$ ,  $y_i = f(s_i, \dots, s_{i-p}, x_i, \dots, x_{i-r})$ ,  $i = -t' - k, -t' - k + 1, \dots, -1$ , and  $w_1, \dots, w_{t'}$  are computed by (10).

### III. SECRET SHARING

#### A. Initialization and construction phase

Let  $X, Y$  be the column vector spaces over  $GF(q)$  of dimension  $l$ . We show the players of the scheme with  $P_\alpha$ ,  $0 \leq \alpha \leq n - 1$  and the dealer of the scheme with  $D$ . The dealer computes the shares of the secret and distributes them to all the players in the scheme. Our scheme is a  $(n, n)$ -threshold scheme, so in order to recover the secret, the  $n$  players have to participate with their shares. Let the secret be  $x_0, x_1, \dots, x_n$ . The dealer should share it among  $n$  players. The dealer executes the following procedure:

(a) Construct a finite automaton  $M_\alpha = (X, Y, S^{p_\alpha+1} \times X^{r_\alpha}, \delta_\alpha, \lambda_\alpha)$  defined by (1) and (2) and a finite automaton  $M_\alpha^* = (Y, X, X^{r_\alpha} \times S^{p_\alpha+1} \times Y^{t_\alpha}, \delta_\alpha^*, \lambda_\alpha^*)$  defined by (3) and (4), for  $\alpha = 0, 1, \dots, n - 1$ , so that  $M_\alpha$  and  $M_\alpha^*$  satisfy conditions  $P_{IC1}$ .

(b) Construct a finite automaton

$$M'_\alpha = (Z, Y, Y^{k_\alpha} \times W^{n_\alpha+1} \times Z^{h_\alpha}, \delta'_\alpha, \lambda'_\alpha)$$

using (5) and (6) and construct a finite automaton

$$M'^*_\alpha = (Y, Z, Z^{h_\alpha} \times W^{n_\alpha+1} \times Y^{t'_\alpha+k_\alpha}, \delta'^*_\alpha, \lambda'^*_\alpha)$$

using (7) and (8) for  $\alpha = 0, 1, \dots, n - 1$ , so that  $M'^*_\alpha$  and  $M'_\alpha$  satisfy  $P_{IC2}$ .

(c) Construct the finite automaton

$$C'(M_\alpha, M'^*_\alpha) = (X, Z, Z^{h_\alpha} \times W^{n_\alpha+1} \times S^{t'_\alpha+k_\alpha+p_\alpha+1} \times X^{t'_\alpha+k_\alpha+r_\alpha}, \delta''_\alpha, \lambda''_\alpha)$$

using definition 2.6 for  $\alpha = 0, 1, \dots, n - 1$ .

(d) Choose arbitrary state

$$s''_{\alpha,e} = (z_{-1,\alpha}, z_{-2,\alpha}, \dots, z_{-h_\alpha,\alpha}, w_{0,\alpha}, \dots, w_{-n_\alpha,\alpha}, s_{0,\alpha}, \dots, s_{-t'_\alpha-k_\alpha-p_\alpha,\alpha}, x_{-1,\alpha}, \dots, x_{-t'_\alpha-k_\alpha-r_\alpha,\alpha})$$

of  $C'(M_\alpha, M'^*_\alpha)$  and compute  $y_{i,\alpha} = f_\alpha(s_{i,\alpha}, \dots, s_{i-p_\alpha,\alpha}, x_{i,\alpha}, \dots, x_{i-r_\alpha,\alpha})$  where  $i = -t'_\alpha - k_\alpha, \dots, -1$ . Take  $s'_{\alpha,out,d} = (y_{-1,\alpha}, \dots, y_{-t'_\alpha-k_\alpha,\alpha})$ , for  $\alpha = 0, 1, \dots, n - 1$ .

(e) Compute and publish the ciphertext from following procedure:

$S := \text{secret}$ .

for  $\alpha = 0$  to  $n - 1$  do

• Add  $t_\alpha + t'_\alpha$  digits, say  $x_{b+1}, \dots, x_{b+t_\alpha+t'_\alpha}$ , to the  $S$ .

• Compute the ciphertext  $z_0, z_1, \dots, z_{z_{b+t_\alpha+t'_\alpha}}$  as follows:

$$||_{i=0}^{b+t_\alpha+t'_\alpha} z_i = \lambda''_\alpha(s''_{\alpha,e}, ||_{i=0}^{b+t_\alpha+t'_\alpha} x_i).$$

where  $\lambda''_\alpha$  is the output function of  $C'(M_\alpha, M'^*_\alpha)$ .

•  $S := z_0, z_1, \dots, z_{b+t_\alpha+t'_\alpha}$ .

end do .

(f) Distribute  $\alpha$  (the priority),  $M_\alpha^*$ ,  $M'^*_\alpha$ ,  $s'^*_{\alpha,out,d}$ ,  $s''_{\alpha,e}$ ,  $S$  and  $t_\alpha, t'_\alpha$  to  $P_\alpha$  for  $\alpha = 0, 1, \dots, n - 1$ .

#### B. Recovery phase

To recover the secret, we need to all the shares. In fact, only  $n$  players can recover the secret  $S$ . Any  $p_\alpha$  has two automaton  $M_\alpha^*$  and  $M'^*_\alpha$  for  $\alpha = 1, \dots, n$ . We call the input of  $M_\alpha^*$  as the input of  $P_\alpha$  and the output of  $M'^*_\alpha$  as the output of  $P_\alpha$ . The recovery with verification process is as following procedure:

i) Sort all the players by their priority.

ii) For  $\alpha = n - 1$  down to 0 do

1)  $S :=$  the decryption of  $S$  by  $P_\alpha$  using  $C'(M'^*_{\alpha-1}, M'^*_\alpha)$ .

2)  $S :=$  remove  $t_{\alpha-1} + t'_{\alpha-1}$  digits from  $S$ .

end do.

### IV. SECURITY OF THE SCHEME

We check the security of the scheme with respect to the shares and the secrets.

**Theorem 2.** Only  $n$  participants in the  $(n, n)$  - threshold secret sharing scheme, which introduced in section III, can retrieve the secret.

*Proof.* It is obvious that  $n$  participants can retrieve the secret using Theorem 1. Now suppose that less than  $n$  participants want to retrieve the secret. Since the secret encrypted  $n$  times by FAPC4 and FAPC4 is secure, thus less than  $n$  participants can not retrieve the secret.

- If one of participants be absent in  $P = \{p_i\}_{i=1}^r$  and the other be prepared, and want to access the secret  $s$ , by above theorem this is impossible.

- The dealer D; considers a one - way, suitable secure hash function H such that its the domain including of the column vector space X, Y and publishes the hash function H, and the

values  $H(m)$ ,  $H(s_i)$ , for  $i = 0, 1 \dots, n$  where  $s_i$  is the secret share of  $p_i$ ,  $m$  is the secret.

- Let the group of participants  $P = \{p_i\}_{i=1}^r$  submit their shares to the combiner to get  $m$ . Then the combiner can check whether particular participant has given his or her secret share  $s_i$  correctly or not, by verifying it with the corresponding public values  $H(s_i)$ .
- The participants in  $P$  can check whether the combiner is giving them back the correct secret  $m$  or not, by verifying it with the public value  $H(m)$ .
- An adversary can attempt to achieve the participants share by using public values of  $H(s_i)$  and Therefore the adversary needs to the pre-image of the hash function  $H$  for accessing the privately shares allocated to participants where the calculation is difficult. Hence shares are secure under a suitable secure collision resistant one-way hash function.

#### IV. CONCLUSIONS AND FUTURE WORKS

In this paper, we have presented a secret sharing scheme based on automata. The next characteristics of its applies a one-way collision resistant hash function. moreover operations like modular multiplication, exponentiation and discrete logarithm are not used, thereby reducing the computational cost of the scheme to quite a large extent. Also the proposed scheme is practical, efficient and secure against notorious conspiracy attacks.

#### ACKNOWLEDGMENT

The authors are highly grateful to the Department of Mathematics, Graduate University of Advanced Technology for providing an excellent research environment..

#### REFERENCES

- [1] G. R. Blakley "Safeguarding cryptographic keys", *The National Computer Conference 1979*, AFIPS, Vol. 48, 1979, pp. 313-317.
- [2] Y. F. Chen, Y. K. Chan, C. C. Huang, M. H. Tsai and Y. P. Chu, "A multiple-level visual secret-sharing scheme without image size expansion", INFORM SCIENCES, Vol. 177, 2007, pp. 4696-4710.
- [3] M. H. Dehkordi, Y. Farzaneh, "A New Verifiable Multi-secret Sharing Scheme Realizing Adversary Structure", WIRELESS PERS COMMUN, Vol. 82, 2015, pp. 1749-1758.
- [4] M. H. Dehkordi, S. Mashhadi, "An efficient threshold verifiable multi-secret sharing", COMP STAND INTER, Vol. 30, 2008, pp. 187-190.
- [5] A. Shamir, "How to share a secret", COMMUN ACM, Vol. 22, 1979, pp. 612-613.
- [6] R. Tao, "Finite Automata and Application to Cryptography", TSINGHUA University Press, Springer, 2008.
- [7] R. Tao, S. Chen, The generalization of public key cryptosystem FAPKC4, CHINESE SCI BULL, Vol. 44, 1999, 784-790.
- [8] Y. Wei, P. Zhong and G. Xiong, "A multi-stage secret sharing scheme with general access structures", In *Wireless Communications, Networking and Mobile Computing (WiCom)*, 4th International Conference on, IEEE, 2008, pp. 1-4.
- [9] A. Saeidi, M. M. Zahedi and A. Nakhaei Amroodi, "A new secret sharing based on finite automaton public key cryptosystem", J THEOR PHYS & CRYPTOGR, Vol. 6, 2014.

#### AUTHORS PROFILE

**Ali Saeidi Rashkolia** received his M.Sc. degree in Mathematics from Islamic Azad University Kerman Branch, Iran, in 2002. He is currently a Ph.D. student in the department of applied mathematics at Graduate University of Advanced Technology (KGUT), Kerman, Iran. His research interests include Cryptography and Automata.

**Mohammad Mahdi Zahedi** received his Ph.D. degree in Mathematics from Bahonar University, Iran, in 1990. He is currently a professor of mathematics at the Department of Mathematical Sciences in Iran Graduate University of Advanced Technology (KGUT), Kerman, Iran. His research interests include Fuzzy Systems, Algebra and Automata.

**Masoud Hadian Dehkordi** received his Ph.D. degree in Mathematics from Loughborough University, UK, in 1998. He is currently a professor of mathematics at the school of Mathematical Sciences in Iran University of Science and Technology (IUST), Tehran, Iran. His research interests include Number Theory, Cryptography and other related topics.

# Trust and Risk Based Approach for the Management of Dynamic Break-Glass Access in the Cloud Federation Environments

Manoj V. Thomas<sup>1</sup>, K. Chandrasekaran<sup>2</sup>

*Department of Computer Science and Engineering,*

*National Institute of Technology Karnataka*

*Surathkal, Mangalore, Karnataka, India*

<sup>1</sup>manojkurissinkal@gmail.com

<sup>2</sup>kchnitk@gmail.com

**Abstract**—Personal Health Records (PHRs) are highly sensitive; and hence proper access control mechanisms need to be enforced in dealing with access requests involving such data. With the emergence of the inter-cloud computing, the PHR service providers can combine different services from multiple Cloud Service Providers (CSPs) into a single service or application for advantages such as better quality of health care and reduced health care cost. In this combined service delivery model, patients' data are stored in the CSPs in cloud federation, and hence the effective access control mechanism should be enforced by the CSPs. During emergency situations, availability of the healthcare data is more important than confidentiality, and hence relevant medical data should be made available to the concerned people irrespective of the employed access control model. But, how to identify the legitimate access request is an issue to be solved in this domain. In this paper, we are proposing a trust and risk-based mechanism for finding the legitimacy of the emergency access requests in the cloud federation environment. The proposed mechanism calculates the risk involved in the access request and takes a suitable access decision by calculating the trust value of the user. The workflow of the proposed approach is also discussed. We have implemented the proposed approach using the CloudSim toolkit, and the analysis of the results is also given. The analysis shows that the proposed approach is efficient in dealing with the break-glass access requests in the cloud federation environment.

**Index Terms**—authorization; break-glass; cloud federation; emergency; PHR; risk; trust.

## I. INTRODUCTION

The widespread acceptance of Cloud Computing has contributed to the design and development of Cloud Federation. Cloud Federation is an association of different Cloud Service Providers. In the standard Cloud Computing model, a client gets the services from a single Cloud Service Provider. At the same time, in the cloud federation the client can get the services combined from different CSPs in the federation. The CSPs in the federation can share the cloud infrastructure between them in order to have better resource utilization and improved QoS to the cloud consumers. Normally, there will be Service Level Agreements (SLAs) between the CSPs in a Cloud Federation regarding the details of the services agreed

among them. Thus, Cloud Federation helps a CSP to have a collection or pool of resources from different CSPs, and this aggregation of resources can take place at different service levels of the Cloud Computing stack. Thus, Cloud Federation at the SaaS level enables a CSP to combine different software services offered by other CSPs in the federation into a single service or application and deliver it to its cloud users.

### A. Need for Dynamic Break-Glass Mechanism in the Cloud Federation Environment

Electronic Medical Records controlled and managed by the patients are known as Personal Health Records (PHRs). By utilizing the cloud based health care applications, the various users such as patients, doctors, nurses, other medical professionals etc. can access the medical data of the patients anytime, anywhere. Since the PHR data are treated as highly sensitive, proper access control mechanisms need to be enforced in dealing with access requests involving PHR data, in order to permit only the authorized users to access the data. During emergency situations, availability of the healthcare data is more important than confidentiality, and hence relevant medical data should be made available to the concerned people irrespective of the employed access control model. The break-glass concept was introduced in [1], and it is the way to extend a person's access rights in exceptional situations. Since the modeling of all emergency situations is difficult to achieve, it is possible that the personal information of patients are misused. In the health care domain, there is the possibility of some users trying to access the health data of patients beyond their access rights for making undue advantages.

By using the multi-cloud based health care services, the quality of the health care given to patients can be improved, while reducing the overall health care cost. In order to enhance the cloud provider's service capabilities, new technologies such as cloud mashups were introduced [2]. Cloud mashups in the health care domain combine different services from multiple cloud providers into a single service or application. This service composition helps the CSPs offer more efficient

services and functionalities to clients at lower costs. Even though, the researchers have been working regarding the security of the PHR data, one of the issues which have not got a perfect solution is how to handle access request to PHR data during emergency situations, when the patients' information is stored in a multi-cloud or cloud federation environment.

### B. Role of Trust in the Cloud Federation Environments

Human beings trust others depending upon the environment or contexts, and this trust values change from time to time. According to Azzedin and Maheswaran [3], trust is defined as: "trust is the firm belief in the competence of an entity to act as expected such that the firm belief is not a fixed value associated with the entity but rather it is subject to entity's behaviour and applies only within a specific context at a given time". One entity can trust another entity in a system also based on the reputation of that particular entity. In this way, the reputation of an entity can be effectively used for building the trust [4][5]. For an entity, there can be either direct or indirect experience with another entity. Direct experience shows that the entities have had some direct interactions between them in the past, and how one entity learns about the behaviours of the other entity using this interactions. Indirect experience of an entity is developed based on the recommendations given by other trusted members in the community. In a multi-cloud environment, it requires the association among multiple clouds, and the effective establishment and management of trust among the various CSPs and also between cloud users and CSPs is of paramount importance [6].

Even though cloud federation offers various advantages, establishment of trust among the partners in the federation is a challenging issue [7][8]. Researchers have been working on various trust models in the Cloud Computing domain that evaluates the trust of various CSPs [9][10]. Majority of these trust models focused on evaluating and managing the trust between cloud users and the CSPs. Very few of the proposed trust models focuses on effective trust management in the Inter-Cloud domain and hence, the present Cloud Federation scenario requires effective trust management approaches. Although it has been discussed that efficient resource allocation and utilization requires a high degree of trust values [11], to the best of our knowledge, the issue of solving the break-glass access management, taking trust into consideration in a cloud federation environment has not been addressed in a satisfactory manner.

### C. Risk Management in the Access Control

In traditional access control models, every access request is evaluated based on the pre-established policies. In dynamic access control systems, every access request is analyzed dynamically, considering not only the security policies, but the context, attributes of the entities and also the risks involved in granting the access request. The risk of allowing a process in a system is defined as the potential damage that can happen due to that process, and it is calculated as a product of the probability of occurrence of an undesired event and its impact

on the system [12]. Before taking an access decision, risk-based access control systems conduct a risk analysis of the access requests made, and a numeric value is assigned to the risk. Then, depending on the risk threshold maintained in the system, access is either permitted or rejected.

Thus, the major contributions of this paper are:

- Design and implementation of a trust-based break-glass mechanism in the cloud federation environment, for dealing with the emergency access requests of PHR users.
- Design and implementation of an approach to evaluate the risk involved in an emergency access request.
- Design and implementation of an approach to calculate the local and recommended trust of a PHR user.
- Implementation of the proposed approach using the CloudSim toolkit.
- Discussion of the results obtained highlighting the advantages and the disadvantages of the proposed approach.

The rest of the paper is organized as follows: Section II describes the work done in the area of PHR management and the break-glass access management in the cloud-based health care environments, highlighting the merits and demerits of various approaches. Section III and IV present the proposed approach of trust and risk based management of break-glass access in the cloud federation environments. Section V presents the results and analysis of the proposed approach, and Section VI discusses the pros and cons of the approach. Finally, Section VII concludes the paper with pointers to the future works.

## II. LITERATURE REVIEW

In this section, we discuss the relevant research activities in the area of cloud-based PHR management and the break-glass access control, analyzing the works carried out by the researchers.

In [13][14][15][16], the authors discuss the emergency access management of PHR information. These works use the predefined emergency staff concept. Based on the policy enforced by the PHR owner, the emergency staff can access the various categories of PHR information. In this work, threshold cryptosystem is used in which a trusted group of people selected by the PHR owner grants the access rights to emergency staff when the PHR owner is unable to grant the permission during emergencies. But, how the genuineness of the access request will be practically verified is not specified. Also, the various issues such as how to decide the members in the emergency team, and also how to make sure that the trusted group of people are online when the break-glass access is requested are not discussed in this paper. Also, is every member in the emergency staff trusted equally by the PHR owners, for granting access to the PHR data during emergencies, is another issue to be solved. But, during emergency situations, practically it can be somebody else other than the predefined medical staff, who would be attending to a patient. Hence, if we restrict the break glass access to a few selected people, it may not always be useful. In our proposed work, every member of the medical staff can request the emergency (break glass) access to the patients' data. The eligibility of access is

determined dynamically when an access request is made, by calculating the risk value of emergency access, and also the trust value of the user requesting the emergency access.

There are many research works dealing with the PHR management in the cloud environment using various encryption techniques such as Attribute-Based Encryption (ABE) [17, 18, 19, 20] and its variants such as CP-ABE [18, 21], KP-ABE [22] etc. But, the effective management of emergency access requests by the PHR users is to be incorporated with them to meet the real-life emergency situations.

In [18], the authors used ABE for encrypting the PHR data of patients and the concept of emergency department (ED) is used to provide the break-glass access. A medical staff who wants to access the patient's health data in an emergency situation, has to contact the ED which will give the required key after proper authentication of the user and also verifying the emergency situations. After the emergency access, the patient also revokes the emergency key through the ED. This work also uses the emergency staff concept whose access requests need to be verified by the ED. Again, how to verify the legitimacy of the access request is not included in this work. The authors in [23][24] also discuss the use of emergency department (ED) for dealing with the break-glass access in the cloud environment. However, the location of the ED is not specified in this work. In their work, no risk-calculation of the access request is considered. Also, the trust values of the CSPs and the PHR users is not considered in this work. In this work, the ED has to verify an emergency situation before the emergency access keys are given to a PHR user. But, how to verify the emergency situation is not discussed in this work. In our work, we are proposing an approach for verifying the legitimacy of emergency access requests in the cloud federation environment. We are not using the emergency staff concept; instead all members of the medical staff with the PHR service provider are allowed the break glass access to the patients' data after the associated risk and the trust calculations.

The authors in [25] discuss the emergency access to PHRs in a distributed environment using Digital Rights Management (DRM) techniques. But, how to decide the legitimacy of the break glass access request by a user is not discussed. In [26], the authors discuss the PHR management system in which the PHR files are organized in a hierarchical manner to make the key distribution efficient. However, the issue of break glass access management is not included in this work. In [27], the authors propose a proxy encryption based access control mechanism in a multi-user environment in which every access operation involves a proxy server. But, this approach does not offer a fine-grained access control and a proper break-glass mechanism. In [28], the authors discuss a PHR management system wherein the personal health data encrypted by the patients can be decrypted when legitimate access requests from the medical staff are made. Here also, how to verify the legitimacy of the access requests is not discussed. In [29], the authors discuss an architecture for sharing health data in a multi-cloud environment. In this architecture, the

health data of the patients are distributed as fragments across different cloud service providers, and it uses the concept of secret sharing to assemble the required health data upon request. However, in this multi-cloud architecture, no break glass access is implemented.

Based on the literature survey, we can see that there is no effective approach that calculates the legitimacy of the break-glass access requests in the cloud federation environment. Hence, in this paper, we are proposing a method that determines the legitimacy of the break-glass access requests by calculating the risk involved in the access requests, and also by calculating the local and recommended trust values of the user requesting the PHR data of patients.

### III. TRUST AND RISK BASED BREAK-GLASS ACCESS MANAGEMENT IN THE CLOUD FEDERATION

In medical applications, sometimes exceptional or special access requests need to be handled during emergency situations. The medical records of the patients contain private and sensitive information. Generally, these data cannot be accessed by all medical professionals in a hospital other than the consulting doctor or the ones explicitly permitted by the PHR owner. But, in the case of an emergency, in order to save a patient's life, a PHR user, such as a nurse needs to be permitted to access the PHR data of the patient, if the consulting doctor is unavailable at that time. If the access control system does not have this required break-glass mechanism, in this case, either the nurse may not be able to perform the emergency service, or a doctor's access rights may be given to the nurse, which may result in some misuse. Also, practically it is not possible for a patient to predict and plan in advance which specific person(s) will request his PHR data during emergencies. Hence, there should be a flexible access control mechanism that deals with break-glass access requests for PHR data of a patient.

#### A. Novelty of the Approach

In the future, multi-cloud based health care services using technologies such as cloud mashups will be widely utilized, as it offers healthcare with improved quality and reduced overall cost. From the literature review carried out, we could understand that the emergency access management in the multi-cloud based health care domain requires efficient solutions. The novelty of the proposed approach can be summarized as:

- To the best of our knowledge, this is the first work that discusses the issue of break-glass access management in the cloud federation environment considering the trust values of various CSPs and PHR users.
- This work does not use the emergency staff concept because during emergency situations, practically it can be somebody else other than the predefined medical staff, who would be attending to a patient. Also, the emergency staff concept has various disadvantages such as how to decide the members in the emergency team, and also how to make sure that the trusted group of people are online when needed.

- This work does not use the concept of emergency department (ED) as it has various issues such where the ED is located, and also how the ED verifies the emergency situation.
- This work does not use the threshold-based crypto system because in this case, the genuineness of the access request should be verified by the trusted people before they give their share of key, and how practically it is verified is an issue.
- In this work, all members of the medical staff with the PHR service provider are allowed the break glass access to the patients' data after the associated risk and the trust calculations. The eligibility of a medical staff for the emergency access is determined dynamically when an access request is made, by calculating the risk value of emergency access, and also the trust value of the user requesting the emergency access.

#### B. PHR Management in the Cloud Federation

The context discussed in this paper is that of a cloud federation based healthcare scenario as shown in the Fig. 1. In this case, it is assumed that the health care provider, such as hospitals aggregate the services from more than one CSP, and the combined services or the application is used by the PHR owners for the storage and processing of their health data. Also, it is assumed that the CSPs whose services are aggregated are part of a cloud federation. Now, the combined service is accessed by different users such as doctors, nurses, lab staff etc. As shown in the figure, three CSPs (Cloud Service Provider-A, B and C) are part of the cloud federation and the health care services such as Cloud Service-1 and Cloud Service-2 from the CSP-B and CSP-C respectively are combined and used by the Health Care Provider. This combined service is then used by the various PHR users (User-1, User-2, ..., User-N) such as doctors, nurses, lab staff etc. as shown in the figure. Our proposed access control mechanism needs to be

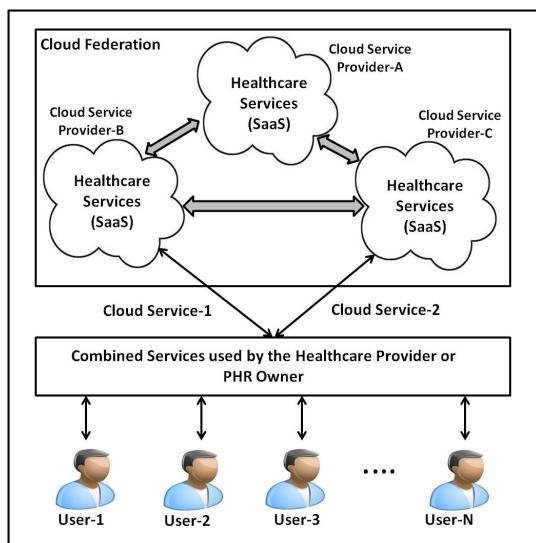


Fig. 1. PHR Management in the Cloud Federation

implemented and enforced by the CSPs in the federation. The PHR owners have to use the required encryption mechanism to ensure the fine-grained access control of their personal health data stored in the cloud servers. In our proposed approach, we do not deal with the encryption mechanisms to be used for protecting the health data of the patients. Our approach deals with emergency access requests and takes a decision as to whether the request should be permitted or not considering the various parameters. There is no 'emergency staff' concept used in this work; and the requests of different medical staff are analysed dynamically considering the risk value of the access request made, and also the trust value of the user requesting the break-glass access.

#### C. Access Control Framework

The overall view of the proposed access control framework with the break-glass management is shown in the Fig. 2. In this model, a PHR user makes the request to access the PHR data stored with any of the CSPs in the federation, and this access request is handled by the access control module with the break-glass mechanism. The various functional components in this framework are:

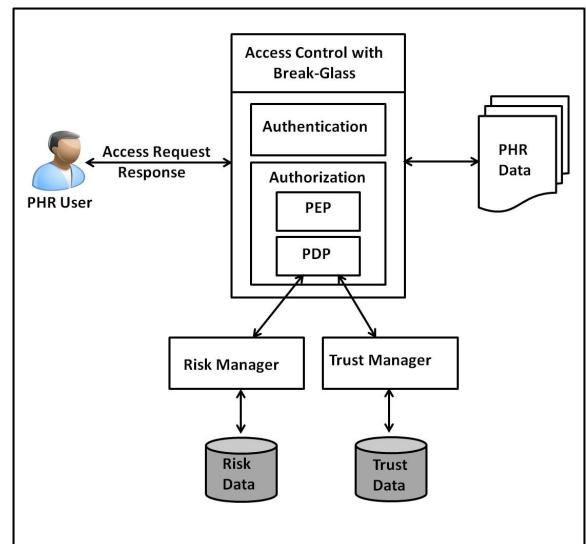


Fig. 2. Overview of the Access Control Framework with Break-Glass Mechanism

1) **Authentication:** Before allowing a break-glass access to the PHR data of patients, the user who wants to access the data should be authenticated. Every user has a username-password pair which is encrypted using AES-256 algorithm and this is used for the authentication of the user.

2) **Authorization:** This component verifies the access rights of the requesting user and takes a decision as to whether the access request should be permitted or not. This component has two modules, PEP (Policy Enforcement Point) and PDP (Policy Decision Point). The PEP contacts the PDP for access decision, and implements the access decision taken by the PDP. Whenever a user makes a break-glass access request, the PDP contacts the Risk Manager module for calculating

the risk value of allowing the access request. It also contacts the Trust Manager module for calculating the trust value of the requesting user in the cloud federation environment.

#### D. Workflow of the Proposed Approach

The workflow of the proposed approach for the management of break-glass access requests is shown in the Fig. 3. When

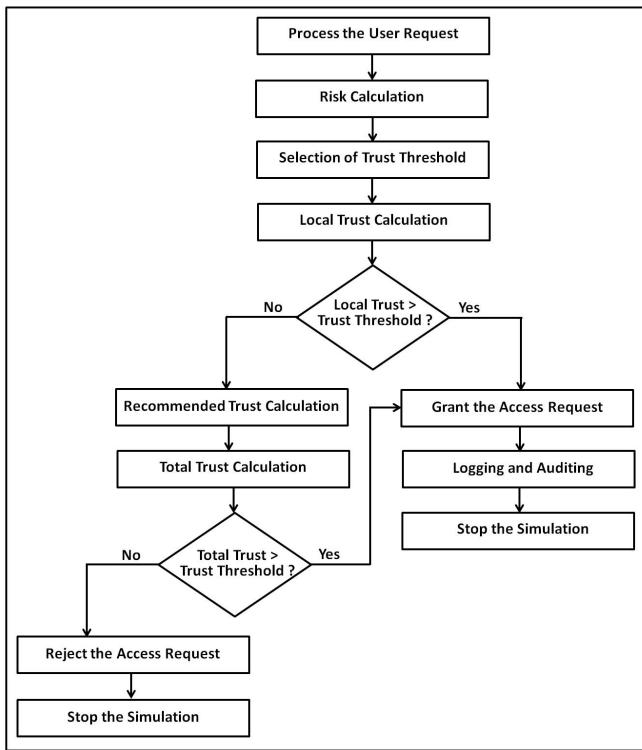


Fig. 3. Workflow of the proposed Approach for Break-Glass Access Management

a CSP receives the break-glass access request from a user, it carries out the authentication and the authorization processes. If the authentication is failed, the access request is rejected and the message is communicated to the requesting user. If the authentication of the user is successful, then the access rights of that particular user for the requested PHR data are checked to verify if the user has been given explicit access rights by the data owner himself. If the requested access rights are allowed by the PHR owner, then the access to the required PHR data is permitted. If there is no explicit authorization given by the data owner for the requesting user, then the proposed break-glass access mechanism is executed to decide whether the requested access needs to be permitted or not.

As shown in the figure, when an access request is made by a user, the risk value corresponding to that access request is calculated. Depending on the calculated risk value, the trust threshold is selected. Then, the local trust calculation of the requesting user is performed. If the local trust value of the user requesting the break-glass access is greater than the trust threshold selected, then the access request is permitted. The details of the permitted access are logged and then audited

for determining whether the break-glass access was genuine or not. Also, the database is updated accordingly to record the details of the permitted access and the modified trust value of the user who made use of the break-glass access, into the corresponding tables in the database.

If the local trust value of the requesting user is less than the trust threshold, then the recommended trust is calculated as explained in the next section. Then, the CSP calculates the total trust value of the user considering both the local and the recommended trust values. If the total trust value is greater than the threshold value, then the break-glass access request of the user is permitted. Then, the details of the permitted access are logged and audited for determining whether the break-glass access was genuine or not, and the database is updated accordingly to reflect the details of the access request permitted and the modified trust value of the user in the required tables in the database. If the total trust value of the user is still less than the trust threshold, the requested break-glass access request is rejected.

### IV. PROPOSED APPROACH FOR THE MANAGEMENT OF BREAK GLASS ACCESS REQUESTS IN THE CLOUD FEDERATION

The various functional components in the proposed approach for dealing with the break glass access requests are shown in the Fig. 4. They are discussed below:

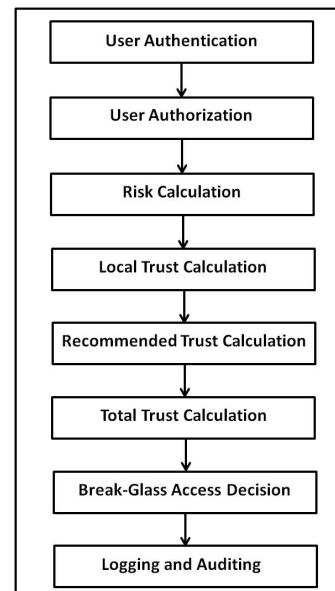


Fig. 4. Proposed Approach for the Break-Glass Mechanism

#### A. User Authentication

This component verifies the identity of the person trying to access the health data of the patient. The identity details of the PHR users are stored with the CSP hosting the application or service. The passwords of the PHR users are encrypted using the Symmetric Encryption scheme, AES-256, and this serves the purpose of securing the stored passwords against various attacks.

### B. Authorization

This component verifies the access rights of the user trying to access the patient's information. In our implementation, a user requesting emergency access to a patient's data is allowed to do so in two cases. In the first case, a user is allowed access when he has the break-glass access privilege granted explicitly by the PHR owner. This may be the case such as the user trying to access the patient's health information is the consulting doctor of the patient or a medical professional allowed by the PHR owner. In the second case, the user has no explicit access rights. Hence, the emergency access requests are handled by the proposed trust-based approach, and a suitable access decision is taken accordingly considering the various parameters as discussed later.

In order to implement the first case, we have considered a parameter 'Degree-of-Bias' in our implementation which can be set to any value of -1, 1 or 0 by the PHR owner. For a particular user and a health file, the value of -1 for the 'Degree-of-Bias' parameter means that the PHR owner has not given permission to the specific PHR user in order to access the specific health file of the patient under any circumstances including emergency situations. A value of 1 for the parameter indicates that the specific PHR user is allowed to access the health data of the patient under all circumstances. A value of 0 for the parameter means that the PHR owner has not given any explicit preferences for the user. Thus, if the requesting user is not permitted to access the patients' data according to the normal access policies of the system, the proposed approach takes a suitable decision as to whether the access request should be permitted or not considering the risk and the trust values associated with the request.

### C. Risk Calculation

In order to calculate the risk value of an access request, the following three factors are considered in our work.

1) *The impact factor (Factor-1):* This factor calculates the impact of allowing the access request made by the user in terms of possible violation of confidentiality, integrity and the availability of the data. In our implementation, in order to show the working of our approach, we have considered four access operations such as read, write or update, download and delete operations. Thus, the impact factor is calculated as:  $\sum_{i=1}^3 P_i * I_i / 3$ , where  $P_i$  ( $i=1, 2, 3$ ) corresponds to the probability of having a violation of the Confidentiality (C), Integrity (I) and the Availability (A) respectively and  $I_i$  ( $i=1, 2, 3$ ) takes value of 1 if there is a corresponding violation of C, I or A because of the access operation, and  $I_i$  takes a value of 0 if there is no violation of C, I or A respectively. The initial probability for an access request to violate C, I and A of the PHR data is considered equal and taken as 0.33

2) *The sensitivity factor (Factor-2):* This factor calculates the sensitivity of the data being accessed. In our work, we have considered five types of reports of patients containing details such as their basic information, allergy details, lab and scan reports etc. Depending on the sensitivity of the information contained in these reports, each of these reports is assigned

a sensitivity value ranging from 1 (lowest) to 5 (highest). So, whenever, an access request is received for a report of a patient, the sensitivity factor is calculated as:

$$\text{Sensitivity} = (\text{Assigned Sensitivity} / \text{Maximum Sensitivity})$$

3) *The probability of malicious access factor (Factor-3):*

This factor shows the probability of the current access request being malicious. The proposed system, after auditing, records all the malicious accesses to the data being protected. Thus, the total number of malicious accesses made to the data of a particular patient by a specific user is available. Hence, the probability of malicious access to a particular data is calculated as:

$$\text{The probability of malicious access} = (\text{Total number of Malicious Access} / \text{Total number of Access Permitted})$$

Hence, the value of the Total Risk of the current access request is calculated as:

$$\text{Total Risk} = (\text{Factor-1} + \text{Factor-2} + \text{Factor-3})/3.$$

Depending on the value of the total risk calculated corresponding to an access request, the current trust threshold is dynamically selected by a CSP in the federation. Now, the local trust value of the requesting user is calculated.

### D. Local Trust Calculation of the User

The calculation of local trust of the user requesting break-glass access involves the following parameters:

1) *Probability of Success:* This parameter considers the number of successful break-glass attempts made by the user in the past. In our implementation, we record every break-glass access permitted, and the process of auditing can be used to verify the break-glass access as successful or not. Hence, this parameter is calculated as: Probability of Success= $(x/y)$ , where  $x$  is the total number of successful break-glass access permitted and  $y$  is the total number of break-glass access requested by that specific user.

2) *Degree of Association:* For calculating the local trust of the user, the total period of association of the user showing how long he has been associating with the healthcare provider is taken into account by considering the date and time of joining of the user (Doctor, Nurse, Lab Technician etc.) with the hospital or the health care centre. Based on the date and time of joining, the Degree of Association is given a value  $x$  for the user, where  $x \in [0, 1]$ .

3) *History of Interaction:* This shows the lead of the number of successful break-glass accesses over the number of unsuccessful break-glass accesses made by a user in the past. It is calculated as: History of Interaction= $(x-y)/z$ , where  $x$  is the total number of successful break-glass accesses permitted,  $y$  is the total number of unsuccessful break-glass accesses made and  $z$  is the total number of break-glass accesses permitted by the specific user.

4) *Existing Trust:* This shows the existing trust value of a user before the current trust value is calculated. As this factor indicates, a user with a higher existing trust value is expected to have a more positive impact on the calculation of the current trust value than a user having a lower existing trust value.

5) *Access Level*: In our work, we have considered three categories of users such as doctors, nurses and lab technicians, and each category of users is assigned a numeric value ranging from 1 to 3 showing the access level associated with them. Thus, the category of doctors is given the access level of 3, nurses given the access level of 2, and the category of lab technicians is given the access level of 1. Hence, the Access Level factor of the requesting user is calculated as: Access Level=(Assigned Level/Maximum Level)

6) *Access Right*: This parameter shows the access right value of the user requesting the break-glass access. In our work, we have considered four access operations such as read, download, write and delete with respect to the PHR data of the patients, and each operation is given a numeric weightage such as read=1, download=2, write=3 and delete=4. Hence, this parameter is calculated as: Access Right=(Assigned Right/Maximum Right)

7) *Permitted Factor*: When a user is making a break-glass access request, the Permitted Factor is considered to show how many break-glass access requests made by the user were permitted in the past. Hence, this parameter is calculated as: Permitted Factor=(x/y), where x is the total number of break-glass access requests permitted and, y is the total number of break-glass access requests made by the user.

8) *Genuine Factor*: This factor shows the ratio of the genuine break-glass accesses made by a specific user to the total number of break-glass accesses permitted by the user in the past. Hence, this is calculated as: Genuine Factor=(x/y), where x is the total number of genuine break-glass accesses made, and y is the total number of break-glass access requests permitted by the user.

Thus, the local trust value of the requesting user is calculated as:

**Trust Value=(Probability of Success + Degree of Association + History of Interaction + Existing Trust + Access Level + Access Right + Permitted Factor + Genuine Factor)/8**

9) *Trust Decay Factor of the User*: In the cloud federation domain, the trust value of a user is considered to be dynamic and the calculated trust value decays over time. Hence, we have considered the Trust Decay Factor while calculating the trust value of the requesting user in the federation. This decay factor is selected depending on when the requesting user had the last transaction with any of the CSPs in the federation. The decay factor is adjusted in such a way that the trust value gets decremented more when the date and time of the last transaction of the user with a CSP becomes older. In our implementation, this decay factor is represented as  $1/x$ , where  $x \in [1, 2]$ , depending on the date and time of the last transaction. In our prototype simulation, the parameter x takes values 1.1, 1.2, 1.4, 1.6, 1.8 and 2 corresponding to six ranges of the elapsed time since the last transaction, such as less than one month, 1-3 month(s), 3-6 months, 6-9 months, 9-12 months and greater than one year respectively.

Hence, the final value of the Local Trust of the user is calculated as:

#### **Local Trust Value=Trust Value X Trust Decay Factor**

If the local trust value is greater than the trust threshold, the break-glass access is permitted. Otherwise, the recommended trust of the requesting user is calculated.

#### *E. Recommended Trust Calculation of the User*

In order to calculate the recommended trust of the requesting user, initially, the trusted CSPs in the federation are identified.

1) *Selection of Trusted CSPs*: When a CSP gets an emergency access request, the CSP calculates the trust value of other CSPs in the federation to identify the trusted CSPs, and from these trusted CSPs, the feedback of the requesting user is collected. In order to select the trusted CSPs, a CSP considers the parameters such as Probability of Success, History of Interaction, Existing Trust, Degree of Association and QoS Value.

The Probability of Success of the requesting CSP with any other CSP in the federation is calculated as: **Probability of Success**=(x/y), where x is the total number of successful transactions and y is the total number of transactions initiated with that CSP. History of Interaction shows the lead of the number of successful transactions over the number of unsuccessful transactions with a particular CSP. It is calculated as: **History of Interaction**=(x - y)/z, where x is the total number of successful transactions, y is the total number of unsuccessful transactions and z is the total number of transactions by a specific CSP with another CSP in the federation. **Existing Trust** shows the existing trust value of a CSP towards another CSP in the federation, before the current trust value is calculated. For calculating the trust of a CSP in the federation, the total period of association of the CSP with the federation is taken into account, by considering the date and time of joining of the CSP with the Cloud Federation. Based on the date and time of joining the federation, the **Degree of Association** is given a value x for the CSP, where  $x \in [0, 1]$ .

While calculating the trust value of a CSP in the federation, the QoS parameters are considered separately to distinguish one CSP from another in the federation. Hence, this calculation involves the following factors:

**Availability Factor**: Calculated as (x/y), where x is the total number of times the service from a specific CSP was available when requested and, y is the total number of service requests made to that CSP.

**Reliability Factor**: Calculated as (x/y), where x is the total number of times the service was reliable and, y is the total number of times the service was available from that CSP.

**Confidentiality Factor**: Calculated as (x/y), where x is the total number of times the confidentiality was intact with the service from a CSP and, y is the total number of times service was available from that CSP.

**Integrity Factor**: Calculated as (x/y), where x is the total number of times the integrity was intact with the service from the CSP and, y is the total number of times service was available from that CSP.

**Response Time Factor**: Calculated as (x/y), where x is

the total number of times the response time was within the promised limit and,  $y$  is the total number of times service was available from that CSP.

Hence, the final QoS Value corresponding to a CSP in the federation is calculated as:

$$\text{QoS Value} = (\text{Availability Factor} + \text{Reliability Factor} + \text{Confidentiality Factor} + \text{Integrity Factor} + \text{Response Time Factor})/5$$

Thus, the Trust Value of the CSP is calculated as:

$$\text{Trust Value} = (\text{Probability of Success} + \text{History of Interaction} + \text{Degree of Association} + \text{Existing Trust} + \text{QoS Value})/5$$

While calculating the QoS values, we have considered five factors such as availability, reliability, confidentiality, integrity and response time factors. In our prototype simulation, just to show the working of our approach, we have given equal weights to all the parameters. In real time cloud federation environment, it will vary from one CSP to another depending on their business objectives.

2) *Trust Decay Factor of the CSP*: In the cloud federation domain, the trust value of a CSP also is considered to be dynamic and the calculated trust value decays over time. This decay factor is selected depending on when a CSP had the last transaction with any other CSP in the federation. In our implementation, this decay factor is represented as  $1/x$ , where  $x \in [1, 2]$ , depending on the date and time of the last transaction. We have selected the decay factor as  $1/x$  to show the variation in the trust value of a CSP, where  $x$  depends on the time elapsed since the last transaction of the requesting CSP with any other CSP in the federation. In our prototype simulation, the parameter  $x$  takes values 1.1, 1.2, 1.4, 1.6, 1.8 and 2 corresponding to six ranges of the elapsed time since the last transaction, such as less than one month, 1-3 month(s), 3-6 months, 6-9 months, 9-12 months and greater than one year respectively. In real time implementation, the parameter  $x$  is also CSP-specific. Practically, different CSPs can use different values for  $x$  for the same time period. It also depends on how long the cloud federation has been in existence, and also how long the requesting CSP has been a member of this federation. Accordingly, a CSP in the federation can decide the value of  $x$ .

Hence, the Total Trust Value of a CSP in the federation is calculated as:

$$\text{Total Trust Value} = \text{Trust Value} \times \text{Trust Decay Factor}$$

After calculating the trust values of all the possible CSPs, those CSPs with trust values greater than a specific threshold are selected into the list of trusted CSPs, and the recommended trust is calculated considering their feedbacks.

3) *Recommended Trust Calculation of the User*: The CSP contacts the trusted CSPs and each of the trusted CSPs calculates its current trust value of the user specified, and communicates that trust value to the CSP that asked for it. The CSP then aggregates the trust values collected from the trusted CSPs to calculate the final recommended trust of the requesting user in the federation. After calculating the final recommended trust value, the CSP calculate the total trust

value of the requesting user as:

$$\text{Total Trust Value} = (\text{Local Trust} + \text{Recommended Trust})/2$$

Based on this total trust value of the requesting user, the CSP decides to either accept or reject the break-glass access request from that user.

In our simulation, total trust value of a user is calculated as the average of the local trust and the recommended trust values. Local trust value is based on own experience of working with a particular user, and the recommended trust is based on the feedback from other trusted CSPs. In our implementation, in order to calculate the recommended trust of a PHR user, initially the trusted CSPs are selected. Here also for selecting the trusted CSPs, the trust-threshold used is CSP-specific. Generally, it can be reasonably high (0.7 in our case). Then, the feedback regarding a specific user is collected from these trusted CSPs. That is the reason why we have given equal weightage to local trust and the recommended trust. Again, in the real time cloud federation implementation, a CSP can use different weights such as 0.6 for the local trust value and 0.4 for the recommended trust value. In our prototype simulation, just to show the working of the proposed mechanism, we have used equal weights (0.5) for both the local trust and the recommended trust values.

#### F. Logging and Auditing

In our implementation, before the access request is permitted, a warning message will be sent to the user requesting the break-glass access. After every permitted break-glass access, the details of the access request such as the User-ID, date and time of access, and the specific resource accessed etc. are encrypted and entered into a log file. Also, auditing is used to verify whether the break-glass access was genuine or not. The auditing of the permitted break-glass access can be performed by either the consulting doctor of the patient or the administrator of the combined multi-cloud health care application. The result of the audit is recorded in the corresponding table in the database, which could be used for taking the break-glass access decision for the same user in the future. In our work, we assume that a break-glass access can be practically audited within 24 hours. Since we are taking the access decision considering the current trust value of the user, no user will be allowed to have break-glass accesses multiple times in a row, for the same patient, if the previous break glass access is not audited by the system.

#### G. Trust Update of the User

After a break-glass access and the subsequent auditing, if it is found that the access was genuine, the trust value of the user who performed the break-glass access is updated as:

$$\text{New Trust} = \text{Old Trust} + [(x/y) * \text{Old Trust}] / 10, \text{ where } x \text{ is the number of genuine break-glass accesses made by the user, and } y \text{ is the number of total break-glass access permitted by the user.}$$

Also, after a break-glass access and the subsequent auditing, if it is found that the access was not genuine, the trust value

of the user who performed the break-glass access is updated as:

**New Trust=Old Trust-[ $(x/y) * Old\ Trust$ ]**, where  $x$  is the number of not-genuine break-glass accesses made by the user, and  $y$  is the number of total break-glass access permitted by the user.

## V. RESULTS AND ANALYSIS

### A. Experimental Setup

In our work, we have considered 25 CSPs in the simulated cloud federation environment. We have carried out the simulation experiments on a system with Intel (R) Core (TM) i7-3770, CPU 3.40 GHz, 8.00 GB RAM and 32-bit Operating System (Ubuntu 14.04). Softwares used for the implementation include CloudSim-3.0.3, Eclipse IDE version 3.8, MySQL Workbench Community (GPL) for Linux/Unix version 6.0.8 and Java version 1.7.0\_55.

### B. Break-Glass Mechanism in the Cloud Federation

We assume that the PHR service provider combines the healthcare services from the CSPs in the federation. We also assume that there are SLAs among the CSPs in the federation who offer cloud based health care services, to share the trust values of users among them. In order to show the working of the prototype, we have considered 1500 users in three categories such as doctors, nurses and lab staff who make requests to access the PHR data of 500 patients. We have considered 5 types of reports/files of patients, and 4 types of access operations such as read, download, write and delete operations. In the example shown in this paper, we have considered the scenario when User-4 (doctor) requests write operation on File-3 of Patient-1 stored at CSP-1. In this case, the PHR owner has not given explicit write permission to the requested file.

Thus, when CSP-1 receives the access request, it calculates the risk-value of the break-glass access request made. The Fig. 5 shows the Risk Calculation of the break-glass access request made by User-4 requesting write operation on File-3 of Patient-1. In our work, we have considered 3 risk factors, and they are calculated as explained in the Section 4.3. For this implementation, we have used the sample database and as shown in the Fig. 5, the Risk Factor-1 is calculated as 0.33, the Risk Factor-2 is calculated as 0.6 and the Risk Factor-3 is calculated as 0.011. The final risk value corresponding to the access request made is calculated as the average of the above three factors and in our case, it is 0.314.

```
--Risk Calculation-----
Risk Factor-1 = 0.330
Risk Factor-2 = 0.600
Risk Factor-3 = 0.011
Final Risk Value = 0.314
```

Fig. 5. Risk Calculation

The Fig. 6 shows the Risk-Trust Table maintained by CSP-1 where the user has requested break-glass access. This table shows the trust threshold to be selected corresponding to a

particular risk score. In our case, since the calculated risk score is 0.314 (as shown in the Fig. 5), the selected trust threshold is 0.6.

-- Risk Trust Table -----		
Risk Value	Trust Threshold	
0.0-0.2	0.55	
0.2-0.4	0.6	
0.4-0.6	0.7	
0.6-0.8	0.75	
0.8-1.0	0.8	
Selected Trust Threshold = 0.6		

Fig. 6. Risk-Trust Table

The Fig. 7 shows the calculation of the local trust of the user by CSP-1. The various parameters considered are Probability of Success, Degree of Association, History of Interaction, Existing Trust, Access Level, Access Right, Permitted Factor and Genuine Factor, and they are calculated as explained earlier in Section 4.4. Thus, as shown in the figure, the calculated values of the various parameters are 0.848, 1.0, 0.728, 0.616, 1.0, 0.4, 0.967 and 0.876 respectively. Now, the average of all these eight factors is multiplied by the calculated Trust Decay factor (0.714) to get the final local trust value (0.575) of the user. Since the final local trust of the user is less than the trust threshold (0.6) maintained by the system, the CSP-1 calculates the recommended trust of the user.

```
--Calculation of Local Trust of the User ---
1. Probability of Success = 0.848
2. Degree of Association = 1.000
3. History of Interaction = 0.728
4. Existing Trust = 0.616
5. Access Level = 1.0
6. Access Right = 0.400
7. Permitted Factor = 0.967
8. Genuine Factor = 0.876
-----
Trust Decay Factor = 0.714
Final Local Trust = 0.575
Local Trust < Trust Threshold
```

Fig. 7. Calculation of Local Trust

In order to calculate the recommended trust of the requesting user, CSP-1 identifies the trusted CSPs. The Fig. 8 shows the CSP-Trust table generated by CSP-1 to which the user has made the break-glass access request. This table shows the current trust values of CSP-1 towards every other CSP in the federation, along with the associated Trust Decay Factor. These trust values are calculated by considering the parameters Probability of Success, Degree of Association, History of Interaction, Existing Trust and the QoS values as explained in the section 4.5. From this table, all the CSPs whose calculated trust values are above the `CSP_Trust_Threshold` (0.7) are selected to the set of Trusted CSPs.

The Fig. 9 shows the Trusted CSPs of CSP-1, along with their corresponding trust values. From the figure, it is seen that the total number of CSPs in the set of trusted CSPs is 15. Now, CSP-1 asks for the recommendation from these trusted CSPs regarding the requesting user.

---Calculation of Recommended Trust---			
-- CSP_Trust Table --			
CSP_ID	Trust_Value	Decay_Factor	Total_Trust_Value
2	0.886	0.833	0.739
3	0.772	0.833	0.643
4	0.935	0.833	0.779
5	0.804	0.833	0.670
6	0.905	0.909	0.823
7	0.907	0.833	0.756
8	0.775	0.833	0.646
9	0.874	0.833	0.728
10	0.866	0.833	0.721
11	0.680	0.833	0.566
12	0.870	0.833	0.725
13	0.845	0.833	0.704
14	0.874	0.833	0.728
15	0.633	0.833	0.528
16	0.866	0.833	0.721
17	0.821	0.833	0.684
18	0.792	0.833	0.660
19	0.931	0.909	0.847
20	0.859	0.909	0.781
21	0.791	0.833	0.659
22	0.895	0.833	0.746
23	0.889	0.833	0.741
24	0.882	0.833	0.735
25	0.810	0.833	0.675

Fig. 8. CSP-Trust Table

Trusted CSPs	
CSP_ID	Trust_Value
2	0.739
4	0.779
6	0.823
7	0.756
9	0.728
10	0.721
12	0.725
13	0.704
14	0.728
16	0.721
19	0.847
20	0.781
22	0.746
23	0.741
24	0.735

CSP\_Trust\_Threshold Value = 0.7  
Total Number of CSPs above the CSP\_Trust\_Threshold Value:15

Fig. 9. Trusted CSPs

The Fig. 10 shows the Recommended Trust Table of CSP-1, and as shown in the figure, the number of CSPs responded is 12. Other CSPs may not have any data to calculate the trust value of the requesting user. This table shows the trust value of every responded CSP and the corresponding trust value of the user as returned by it. Also, as shown in the figure, these two values are multiplied to get the recommended value of the user. The average value of all the recommended values returned by the responded CSPs is calculated to get the total recommended value of the user. As shown in the figure, the total value of the recommended trust of the requesting user is calculated as 0.643.

Now, the average of the total local trust (0.575 as shown in the Fig. 7) and the total recommended trust (0.643 as shown in the Fig. 10) values are taken to get the total trust value of the user. Hence the total trust of the user is calculated as 0.609. In our work, while calculating the total trust value of

---Recommended_Trust Table---				
CSP_ID	Trust_Value	Returned_Trust_Value	Recommended_Trust	
7	0.756	0.845	0.639	
2	0.739	0.831	0.614	
9	0.728	0.843	0.614	
6	0.823	0.856	0.705	
20	0.781	0.842	0.658	
10	0.721	0.875	0.631	
4	0.779	0.850	0.662	
13	0.704	0.877	0.617	
16	0.721	0.839	0.605	
14	0.728	0.880	0.641	
12	0.725	0.855	0.620	
19	0.847	0.842	0.713	

No. of CSPs Responded : 12  
Total Recommended Trust = 0.643

Fig. 10. Recommended Trust Table

the user, equal weightage (0.5) is given to both the local trust and the recommended trust as the recommendations are taken from the trusted CSPs of CSP-1.

Now, since the total trust value (0.609) of the requesting user is greater than the trust threshold (0.6), the break-glass access request is permitted with a warning message as shown in the Fig. 11. In our work, the details of the break-glass access are logged into a file after encrypting with AES-256 algorithm.

Total Trust Value = 0.609

Trust Value > Trust Threshold  
You are going to access a Report Which you are not authorized to.  
This access details will be logged in and communicated to the Admin.  
Do you want to proceed?

Y  
Access Permitted  
Time Duration = 92111

Fig. 11. Access Decision

### C. Results and Analysis

In order to test and validate the proposed approach in the Cloud Federation environment, we have implemented the Cloud Federation of 25 CSPs using the CloudSim toolkit [30]. Sample database is created and used as the database for testing our algorithm. We have considered the break-glass access request of a user in such a way that there is no explicit authorization given by the PHR owner for the specific user, and hence the proposed approach is executed to take the suitable access decision.

The Fig. 12 shows the number of accepted break-glass access requests of 5 users, and how the 100 break-glass requests of these five users are treated in our work. The figure shows the number of accepted break-glass requests of the users in three cases. The first case shows the number of requests accepted considering the local trust of the user alone. The second case indicates the number of requests accepted considering the local and the recommended trust of the requesting user. The third case shows the number of requests rejected even after considering the local and the recommended trust of the user. As shown in the figure, out of 100 access requests made by User-1, 48 times the access requests were accepted using

the local trust alone, and 20 times the access requests were accepted using both the local and the recommended trust of the user. The figure also shows that, 32 times the break-glass access requests of User-1 were rejected due to insufficient trust even after considering the local and the recommended trust of the requesting user. As shown in the figure, the corresponding data for User-2, User-3, User-4 and User-5 are (32, 32, 36), (40, 40, 20), (36, 20, 24) and (40, 24, 36) respectively.

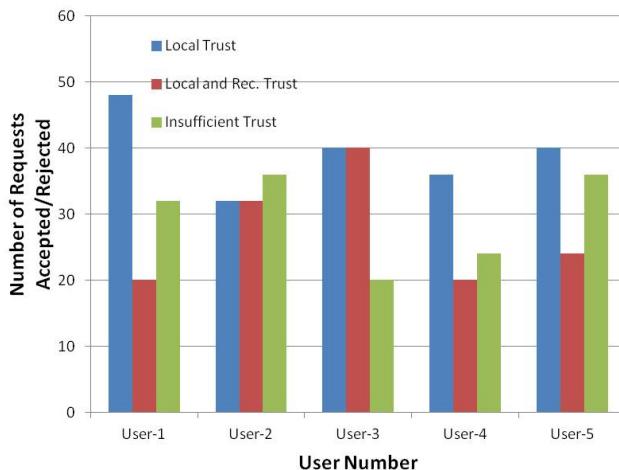


Fig. 12. Analysis of the Accepted/Rejected Requests

The results show that the total number of accepted break-glass requests of a user increases with the recommendation from the trusted CSPs in the federation, and hence the overall efficiency of the cloud-based healthcare services. Since we have not used any real time data in our simulation, we assume that the permitted break-glass accesses were genuine, as in real life cases, the auditing of a break-glass access proves the genuineness of the access made.

The Fig. 13 shows the average time taken for the break-glass access decision for 5 users in our simulation. The figure shows the service decision time taken for the users in two cases. The first case shows the time taken for the break-glass access decision considering the local trust of the user alone. The second case shows the time taken considering both the local and the recommended trust values of the requesting user. As shown in the figure, the average time taken for the break-glass access decision considering the local trust of the User-1 alone is 4317 ms and the average time taken considering the local and the recommended trust values of User-1 is 5475 ms. The corresponding data for User-2, User-3, User-4 and User-5 are (2776, 5161), (2867, 5217), (4648, 5827) and (2675, 5672) respectively. As shown in the figure, even though the calculation of the recommended trust of a user takes longer compared to the calculation of the local trust alone, the efficiency of the cloud-based healthcare services is improved as more break-glass access requests are accepted by the cloud based healthcare service.

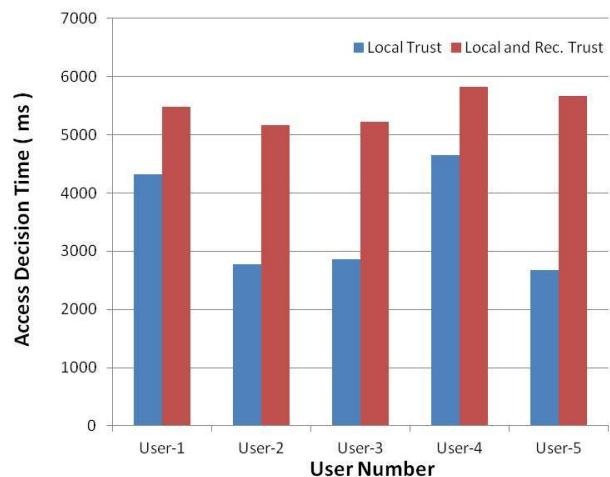


Fig. 13. Analysis of the Access Decision Time

## VI. PROS AND CONS OF THE APPROACH

The major advantage of the proposed approach of break-glass management in the cloud federation environment is that it helps the CSP to take emergency access decision effectively in the cloud federation environment. It helps a CSP in the federation to identify the PHR user requesting access to a patient's data as trustworthy or not. The proposed approach calculates the risk value of the access request made, and then the local and the recommended trust values of the requesting user. Then, based on the calculated trust value, the proposed trust based approach helps to effectively decide whether the emergency access request should be permitted or not in the cloud federation environment. Thus, the approach improves the performance, responsiveness and the efficiency of the healthcare services delivered by the CSPs in the federation. In the proposed approach, we consider the trusted CSPs of any CSP to get the recommendation of a requesting user in the federation. Here, we have assumed that the specific CSP has a good transaction history with the trusted CSPs in the federation. Thus, our approach helps to take the access decisions efficiently during emergency situations, ensuring timely and efficient service to the clients. As far as we know, this is the first work that employs the trust-based approach for the management of dynamic break-glass requests in the cloud federation environment. Since there are no similar works available that deals with the management of dynamic break-glass requests in the cloud federation domain, we were not able to compare our approach with other approaches.

## VII. CONCLUSIONS AND FUTURE WORK

This paper presents a novel trust-based approach for the management of dynamic break-glass access requests in the Cloud Federation environment. It shows that by calculating the dynamic trust of the requesting user in the federation, break-glass access requests can be effectively managed. The proposed break-glass access mechanism calculates the risk

value of the access request, and the local and the recommended trust values of the requesting user to evaluate the degree of the trustworthiness of the user requesting emergency access to the personal sensitive health data of the patients. The proposed approach was validated using the CloudSim toolkit. The analysis of the obtained results shows the effectiveness of the proposed approach. In our implementation, we have used the sample database created for testing the approach. As a future work, we plan to implement the proposed approach in an OpenNebula cloud environment using real time data. Also in our future work, we plan to incorporate the session adaptive risk and the trust management into the proposed approach. Thus, considering the importance of the emerging cloud federation and the cloud-based healthcare services, the proposed approach is relevant and efficient in dealing with the emergency access requests of users where the sensitive health information has to be protected, at the same time made available during emergency situations.

## REFERENCES

- [1] Break-glass: An approach to granting emergency access to healthcare systems. White paper, Joint NEMA/COCIR/JIRA Security and Privacy Committee, 2004.
- [2] Chandrasekhar, S., M. Singh, G. E. Tingjian, R. Krishnan, Gail Joon Ahn, and Elisa Bertino, "Collaboration in Multicloud computing environments: Framework and security issues", IEEE Transactions on Cloud Computing, vol. 46, no 2, 2013, pp. 76-84.
- [3] F. Azzedin and M. Maheswaran, "Towards trust-aware resource management in grid computing systems", in Proc. of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid, Washington, USA, 2002, p. 452.
- [4] Vijayakumar, V. R. S. D., R. S. D. Wahida Banu, and Jemal H. Abawajy, "An efficient approach based on trust and reputation for secured selection of grid resources", International journal of parallel, emergent and distributed systems 27, no. 1, pp. 1-17, 2012.
- [5] Jøsang, Audun, Roslan Ismail, and Colin Boyd, "A survey of trust and reputation systems for online service provision", Decision support systems, vol. 43, no. 2, pp. 618-644, 2007.
- [6] Abawajy, Jemal, "Establishing trust in hybrid cloud computing environments." In Proc. of 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2011, pp. 118-125.
- [7] Govil, Saumitra Baleshwar, Karthik Thyagarajan, Karthikeyan Srinivasan, Vijay Kumar Chaurasiya, and Santanu Das, "An approach to identify the optimal cloud in cloud federation", International Journal of Cloud Computing and Services Science (IJ-CLOSER), vol. 1, no. 1, pp. 35-44, 2012.
- [8] Sánchez, Rosa, Florina Almenares, Patricia Arias, Díaz-Sánchez, Daniel and Marín, Andrés, "Enhancing privacy and dynamic federation in IdM for consumer cloud computing", IEEE Transactions on Consumer Electronics, vol. 58, no. 1, pp. 95-103, 2012.
- [9] W. Li and L. Ping, "Trust model to enhance security and interoperability of cloud environment", In Cloud Computing, Springer, pp. 69-79, 2009.
- [10] M. Ahmed and Y. Xiang, "Trust ticket deployment: a notion of a data owner's trust in cloud computing". In Proc. of 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2011, pp. 111-117.
- [11] Vijayakumar, V., and R. S. D. W. Banu, "Security for resource selection in grid computing based on trust and reputation responsiveness", International Journal of Computer Science and Network Security, vol. 8, no. 11, pp. 107-115, 2008.
- [12] N. N. Diep, S. Lee, Y.-K. Lee, and H. Lee, "Contextual risk-based access control," In Security and Management, pp. 406-412, 2007.
- [13] Thummavet, P., and S. Vasupongayya, "A novel personal health record system for handling emergency situations." In Proc. International Conference of Computer Science and Engineering (ICSEC), 2013, pp. 266-271. IEEE, 2013.
- [14] M. N. Huda, S. Yamada, and N. Sonehara, "Privacy-aware access to patient-controlled personal health records in emergency situations," Pervasive Computing Technologies for Healthcare, PervasiveHealth2009, 2009, pp. 1-6.
- [15] J. Sun, X. Zhu, C. Zhang, and Y. Fang, "HCPP: Cryptography based secure EHR system for patient privacy and emergency healthcare", In Proc. of Distributed Computing Systems, ICDCS, 2011, pp. 373-382.
- [16] Tong, Yue, Jinyuan Sun, Sherman SM Chow, and Pan Li, "Towards auditable cloud-assisted access of encrypted health data." In Proc. of IEEE Conference on Communications and Network Security (CNS), 2013, pp. 514-519.
- [17] Brucker, Achim D., Helmut Petritsch, and Stefan G. Weber, "Attribute-based encryption with break-glass." In Information Security Theory and Practices, Security and Privacy of Pervasive Systems and Smart Devices, Springer Berlin Heidelberg, pp. 237-244, 2010.
- [18] Li, Ming, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption", IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 1, pp. 131-143, 2013.
- [19] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proc. of CCS '06, 2006, pp. 89-98.
- [20] Bethencourt, J., Sahai, A., Waters, B., "Ciphertext-policy attribute-based encryption." In Proc. of IEEE Symposium on Security and Privacy, IEEE Computer Society Press, Los Alamos, 2007, pp. 321-334.
- [21] At Risk of Exposure - in the Push for Electronic Medical Records, Concern Is Growing About How Well Privacy Can Be Safeguarded. [Online]. Available: <http://articles.latimes.com/2006/jun/26/health/he-privacy26>, 2006.
- [22] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," In Proc. of 13th ACM conference on Computer and Communications Security, CCS, 2006, pp. 89-98.
- [23] Singh, Ramkinker, Mohan K., and Vipra Gupta, "Dynamic Federation in Identity Management for Securing and Sharing Personal Health Records in a Patient-centric Model in Cloud", International Journal of Engineering & Technology (0975-4024), vol. 5, no. 3, 2013, pp. 2201-2209.
- [24] Li, Ming, Shucheng Yu, Kui Ren, and Wenjing Lou, "Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings", In Security and Privacy in Communication Networks, Springer Berlin Heidelberg, pp. 89-106, 2010.
- [25] Künzi, Julien, Paul Koster, and Milan Petkovic, "Emergency access to protected health records", In MIE, pp. 705-709, 2009.
- [26] J. Benaloh, M. Chase, E. Horvitz, and K. Lauter, "Patient controlled encryption: ensuring privacy of electronic medical records," in Proc. of the 2009 ACM workshop on Cloud computing security, CCSW, 2009, pp. 103-114.
- [27] C. Dong, G. Russello, and N. Dulay, "Shared and searchable encrypted data for untrusted servers," in Proc. of DBSec'08, 2008, pp. 127-143.
- [28] Sun, Jinyuan, Xiaoyan Zhu, Chi Zhang, and Yuguang Fang, "Hcgp: Cryptography based secure ehr system for patient privacy and emergency healthcare." In Proc. of 31st IEEE International Conference on Distributed Computing Systems (ICDCS), 2011, pp. 373-382.
- [29] Ermakova, Tatiana, and Benjamin Fabian, "Secret sharing for health data in multi-provider clouds." In Proc. of 15th IEEE Conference on Business Informatics (CBI), 2013, pp. 93-100.
- [30] Calheiros, Rodrigo N., Rajiv Ranjan, Anton Beloglazov, César AF De Rose, and Rajkumar Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", Software: Practice and Experience, vol.41, no. 1, pp. 23-50, 2011.

# CROSS SLOT MICROSTRIP PATCH ANTENNA WITH DUAL POLARIZATION

Nazia Hasan

Student ECE Deptt., UTU Dehradun  
Uttarakhand Technical University, Dehradun  
Dehradun, India

Dr.S.C.Gupta

ECE Deptt., DIT Dehradun  
Dehradun Institute of Technology Dehradun  
Dehradun, India

**Abstract** — A single-feed circular microstrip patch antenna having reconfigurable polarization capability is proposed. This proposed antenna has a very simple structure; two slots are created at an angle of 45 degree and 135 degree in the shape of X at the centre of patch antenna, and one Micro Electromechanical switch is inserted at the centre of the created slot to alter the polarization of antenna. When switch is in ON position, the polarization will be linear and if switch is OFF, polarization will be circular. Polarization will be confirmed with the help of axial ratio plot. Microstrip feed line is used in this structure.

**Keywords-***Circular Polarization, microstrip patch, Xshape slot, MEM switch*

## I. INTRODUCTION

Microstrip patch antennas have always been a very attractive choice for the researchers, because of its various advantages such as low volume, light weight, low cost, ease of construction, conformal pattern, compatibility with incorporated circuits and so many others. Mostly microstrip patch antennas are designed for linear polarization, but in some applications such as satellite communication circular polarization is required because of its insensitivity to transmitter or receiver orientation. Polarization diversity [1] may be a great issue in effectively addressing the multipath-fading effects in recent wireless communication systems [2].

A microstrip patch antenna is one of the most widely used radiator which is used to generate circular polarization. Circular polarized functions and polarization diversity are becoming main design concerns for practical applications of microstrip patch antennas.

In this paper a dual polarized antenna is designed, which can provide secure communication by switching the polarization. This antenna may be very helpful in military applications also.

In the work presented in this paper a microstrip patch antenna is designed at 3.6 GHz that is linearly polarized [3]. Duroid substrate is used in this design and the dielectric constant is 2.2[4 ].Two crossed slots in the shape of 'X' created at the centre of patch antenna, as a result of which the linear polarization of

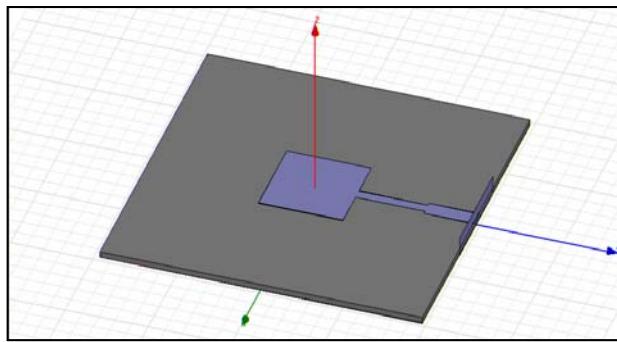
the antenna is converted to circular polarization. In this way circular polarization is achieved [5]. This antenna is converted to reconfigurable antenna by inserting a MEM switch at the center of x shape slot. The cross slot etched antenna shows two types of polarization, the polarization is linear at switched on position and the polarization is circular at switched off position [6]. In this way the cross slot etched antenna provides dual characteristics.

Table 1 shows the design parameters of basic microstrip patch antenna without any slot [7].

## II. ANTENNA DESIGN

S.No.	Parameter	Value
1.	Designed Frequency	3.6 GHz
2.	Dielectric constant	Duroid (tm) 2.2
3.	Width of patch	22.58 mm
4.	Length of patch	22.58 mm
5	Type of feed	Stripline feed
6.	Width of quarter wave transformer	0.582 mm
7.	Length of QWT	20.8325 mm
8.	Width of 50 Ω TL	4.84 mm
9.	Length of 50 Ω TL	15 mm
10.	Polarization	Linear

**Table 1. Various dimensions of Patch antenna**



**Fig.1 Microstrip Patch Antenna at 3.6 GHz**

The patch antenna is designed on the foundation of transmission line model (TLM) and expressions used to calculate parameters like length, width of patch, etc. are given by the authors in [8] which are shown below:

$$W = \frac{c}{2f} \sqrt{\frac{2}{\epsilon_r + 1}} \quad (1)$$

where  $\epsilon_r$  is the substrate dielectric constant, while W is width of the patch and H is height of the substrate. The patch we have used in this model is square patch, means length and width of the patch are same. The dimensions of the patch are extended to account the fringing effect; the extension of length is given by formula,

$$\Delta L = 0.412h \frac{(\epsilon_{eff} + 0.3).(W/h + 0.264)}{(\epsilon_{eff} - 0.258).(W/h + 0.8)} \quad (2)$$

Since the length has been extended by  $\Delta$  on each side of the patch, effective length is given by,

$$L_{eff} = \frac{c}{2f \sqrt{\epsilon_{eff}}} \quad (3)$$

Patch resonant length L is given by,

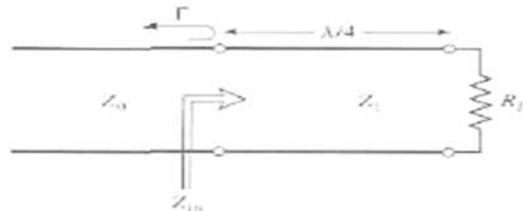
$$L = L_{eff} - 2\Delta \quad (4)$$

Using the values given by the TLM approximation, parameters for the antenna were calculated for 3.6 GHz. The dielectric substrate Duroid ( $\epsilon_r = 2.2$ ) with height  $h = 1.57$  mm is used.

Quarter wave transformer is a very useful and practical circuit that is used for impedance matching and it also provides a simple transmission line circuit that further demonstrates the properties of standing waves on a mismatched line.

When it is desired to match a load resistance  $R_L$  with a feed line of characteristic impedance  $Z_0$ , a piece of lossless transmission line of characteristic impedance  $Z_1$  and  $\lambda/4$  length is used to connect them, as to make the reflection coefficient  $\Gamma=0$  looking into the  $\lambda/4$  matching section. The matching impedance  $Z_1$  is given by,

$$Z_1 = \sqrt{Z_0 * R_L} \quad (5)$$



**Fig.2 quarter wave matching transformer**

The length of the transformer is  $\lambda/4$ , which gives length 20.8325. Width of the conductor is given by the formula

$$\frac{w}{h} = \frac{8 \exp(A)}{\exp(2A) - 2} \quad (6)$$

for  $Z_0(\epsilon_r)^{1/2} > 89.91$ , that is  $A > 1.52$  and for  $Z_0(\epsilon_r)^{1/2} \leq 89.91$ , that is  $A \leq 1.52$

$$\frac{w}{h} = \frac{2}{\pi} \left\{ B - 1 - \ln(2B-1) + \frac{\epsilon_r - 1}{2\epsilon_r} \left[ \ln(B-1) + 0.39 - \frac{0.61}{\epsilon_r} \right] \right\} \quad (7)$$

where,

$$A = \frac{Z_0}{60} \left\{ \frac{\epsilon_r + 1}{2} \right\}^{1/2} + \frac{\epsilon_r - 1}{\epsilon_r + 1} \left\{ 0.23 + \frac{0.11}{\epsilon_r} \right\} \quad (8)$$

The width of 50-ohm transmission line is calculated by eq. (7) giving a value of 4.84mm.

The dimensions of the ground plane were taken according to the lengths of patch, the quarter wave transformer and the 50-Ohm transmission line. So the length of the ground plane, we calculated is given by;

$$L_g = L + 2*(L_{QWT} + L_{TL}) \quad (9)$$

where L is the length of the patch,  $L_{QWT}$  is the length of the quarter wave transformer;  $L_{TL}$  is the length of the 50-Ohm transmission line.

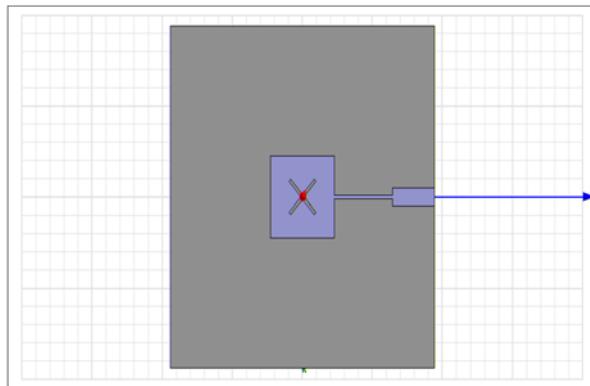
The Patch antenna and the ground plane are made of copper sheet; thickness of copper sheet is taken to be 0.1 mm. The antenna is fed RF signals ranging from 1 to 5 GHz, at the port with the help of a waveport as

shown in fig.1 The antenna without slot has linear polarization. In next step we change the polarization of antenna from linear to circular. For this purpose, a slot in the shape of X is created at the center of the patch. The dimensions of the slots being calculated according to the following equation

$$\text{Length } a = \text{length of patch} / 10 \quad (10)$$

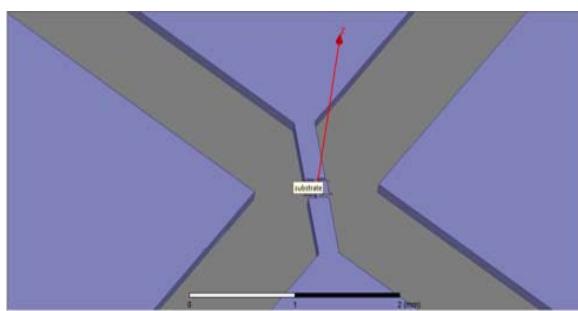
$$\text{Width } b = a / 10 \quad (11)$$

Calculated slot length is 8.6 mm & Slot width is 0.86 mm.



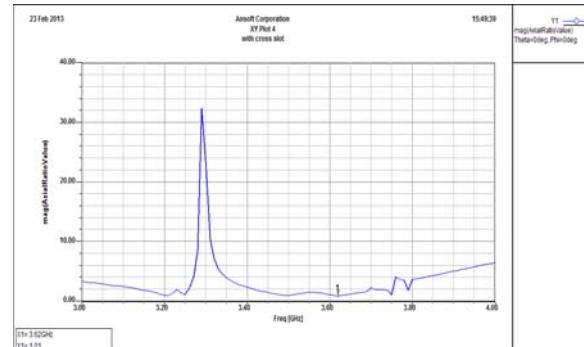
**Fig.2 cross slot antenna**

This cross slot antenna gives circular polarization. If a switch is inserted at the center of this slot, then in Switch On position the polarization will be linear and if the switch is in OFF position the polarization will be circular[9]. The cross slot antenna with switch ON is shown in fig.3 given below. The design of Switch is taken from [10].



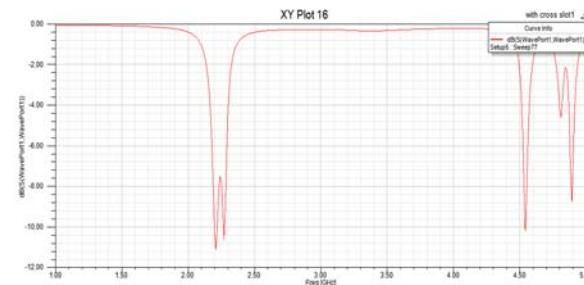
**Fig.3 cross slot antenna (switch ON)**

The polarization of Cross slot antenna can be observed with the help of axial ratio plot. The value of axial ratio from Fig 4 is found to be greater than 1 that confirms linear polarization.



**Fig.4 Axial ratio of cross slot antenna (switch ON)**

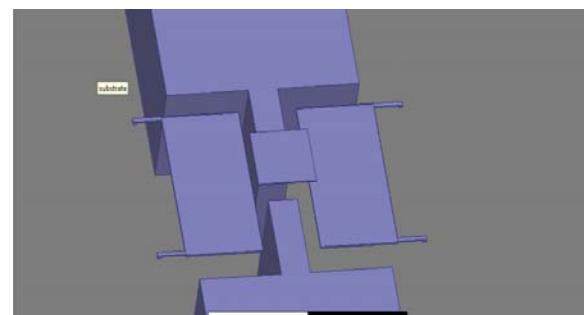
The s-parameter plot of cross slot antenna (switch ON) is presented in Fig 5 showing the resonance at 4.5 GHZ.



**Fig.5 S11 plot of cross slot antenna (switch ON)**

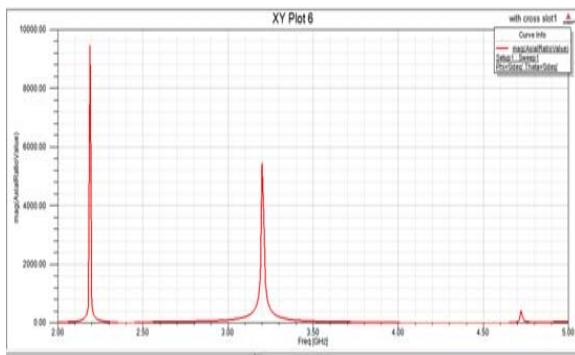
When the switch is turned OFF the polarization changes from linear to circular.

The design of cross slot etched microstrip patch antenna (switch OFF) is shown in figure given below.



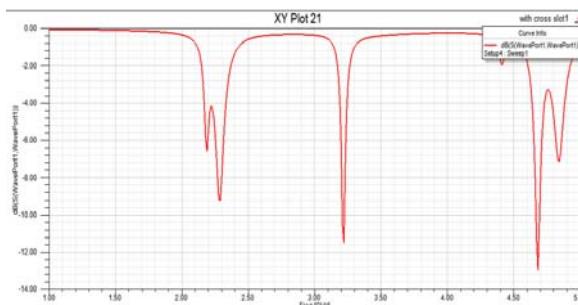
**Fig.6 cross slot antenna with switch OFF**

The circular polarization can be observed by plotting axial ratio. This axial ratio plot is showing circular polarization because the axial ratio magnitude is 1 at the frequency 3.3 GHz.



**Fig.7 Axial ratio of cross slot antenna (switch OFF)**

The s-parameter plot of cross slot antenna in switch OFF position showing the resonance at 3.3 GHZ



**Fig.8 S11 plot of cross slot antenna (switch OFF)**

### III. RESULTS & ANALYSIS

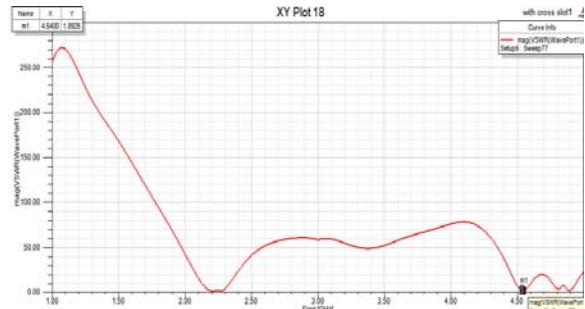
#### A. Software and settings

This Circularly and linearly polarized antennas were simulated on EM solver Ansoft HFSS (High Frequency Structure Simulator), the radiations were measured by taking the help of infinite sphere setup in far field. The RF Signals were setup from 1 GHz to 5 GHz, simulated adaptively.

#### B. Analysis of results

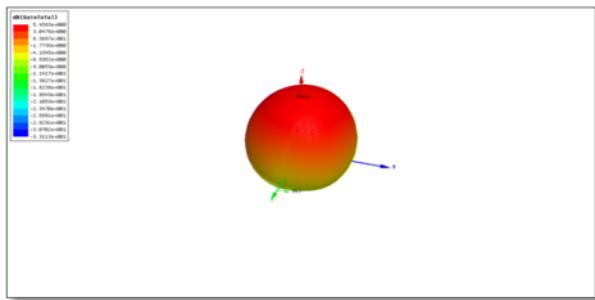
The simulation results are shown in figures 9 to 12, it can be observed that the cross slot antenna with switch OFF is providing circular polarization and cross slot antenna with switch ON is showing linear polarization

Cross slot with switch ON is showing resonance at frequency 2.4 and 4.5 GHz. Figure 7 shows the axial ratio for cross slot antenna (OFF), as desired its magnitude is 1.01, which shows that this cross slot etched microstrip patch antenna giving circular polarization.



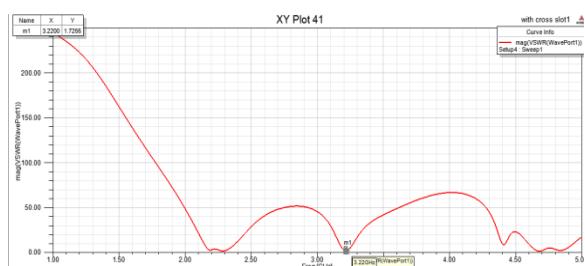
**Fig.9 VSWR of cross slot antenna (switch ON)**

This cross slot etched microstrip patch antenna provides circular polarization. Gain of cross slot antenna with switch ON is shown in figure 10. Figure 10 is showing the gain 5.54 dB.

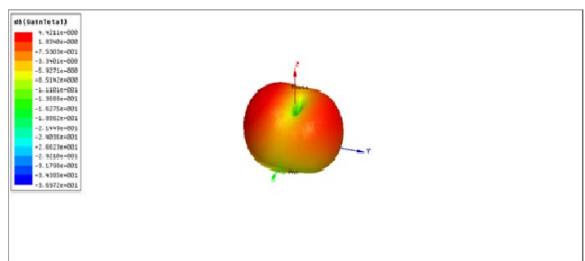


**Fig.10 Gain of cross slot antenna (switch ON)**

The cross slot antenna provides wide bandwidth. The VSWR of cross slot antenna with switch OFF is shown in figure given below.



**Fig.11 VSWR of cross slot antenna (switch OFF)**



**Fig.12 Gain of cross slot antenna (switch OFF)**

#### IV. CONCLUSION

The design of cross slot etched polarization reconfigurable antenna using MEM switch is presented in this paper. This designed cross slot Reconfigurable antenna changing the polarization using turning ON or OFF the switch.

The designed antenna because of switching polarization is suitable for use in defense applications, surveillance, countermeasures and communication.

#### ACKNOWLEDGEMENT

The authors are very much thankful to DEAL (Defense Electronics Applications Laboratory), Dehradun for granting permission to use their laboratory to make use of Ansoft HFSS.

#### REFERENCES

1. Xing-Peng Mao; Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; Department of Information Engineering, Harbin Institute of Technology (Weihai), Weihai, Shandong, P.R. China 264209. "On Polarization Diversity in Mobile Communications" 2006 international conference on communication technology, IEEE, pp. 1-4.
2. Dmitri B Strukov and Konstantin K Likharev "CMOL FPGA: A RECONFIGURABLE ARCHITECTURE FOR HYBRID DIGITAL CIRCUITS WITH TWO TERMINAL NANODEVICES" Published 19 April 2005, IOP Publishing Ltd, Nanotechnology, volume 16, No.6.
3. Hall P.S., "Microstrip linear array with polarization control", IEEE Proc., Vol. 130, pp. 215-224.
4. Symeon Nikolaou, Ramanan Bairavasubramanian, Cesar Lugo, Ileana arrasquillo, Dane C. Thompson, George E. Ponchak, Manos M. Tentzeris, "Pattern and Frequency reconfigurable Annular Slot Antenna Using PIN Diodes" IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 54, NO. 2, FEBRUARY 2006.
5. Nazia Hasan, UTU Dehradun, Dr.S.C.Gupta, DIT Dehradun "A DUAL BAND MICROSTRIP PATCH ANTENNA WITH CIRCULAR POLARIZATION" Conferences on Advances in Communication and Control system (CAC2S), 2013.
6. Eko Tjipto Rahardjo; Dept. of Electr. Eng., Univ. Indonesia, Depok, Indonesia; Fitri Yuli Zulkifli;
7. Basari; Desriansyah Yudha Herwanto "Circularly polarized microstrip antenna array for UAV application" Antennas & Propagation (ISAP), 2013, Proceedings of the International Symposium on (Volume:02 ), IEEE, pp. 870-872.
8. Nazia Hasan, UTU Dehradun, Dr.S.C.Gupta, DIT Dehradun "POLARIZATION TUNABLE MICROSTRIP PATCH ANTENNA USING RF-MEMS" IEEE 3rd International Advance Computing Conference (IACC), 2013.
9. "Antenna Theory, Analysis and Design", 2<sup>nd</sup> Edition, Balanis C. A., 1976. Wiley-Interscience, John Wiley and Sons, Inc., Publications, pp. 771.
10. Y. J. Sung, T. U. Jang, and Y.-S. Kim" A Reconfigurable Microstrip Antenna for Switchable Polarization" IEEE MICROWAVE AND WIRELESS COMPONENTS LETTERS, VOL. 14, NO. 11, NOVEMBER 2004.
11. \*Nitin, \*\*Vipul Sharma, \*\*\*S.S. Pattnaik, \*\*Tanuj Garg \*\*\*S. Devi and \*\*Sameer Kaul "DESIGN AND SIMULATION OF A LOW ACTUATION VOLTAGE WIDEBAND RF MEMS SWITCH ON HFSS" Journal of Natural & Physical Sciences, Vol. 23 (1-2) (2009) 85-89

# Layer Based Log Analysis for Enhancing Security of Enterprise Datacenter

Samuel Getachew Tadesse<sup>1</sup>

Department of Computer Science  
Haramaya University  
Haramaya, Ethiopia

Dejene Ejigu Dedefa<sup>2</sup>

Department of Computer Science  
Addis Ababa University  
Addis Ababa, Ethiopia

**Abstract**—The paper explores how log analysis is key for enhancing network security of enterprises. Now a days the issues of security becomes great concern because of the interconnection among organizations with WWW. Routine log analysis is beneficial for identifying security incidents, policy violations, fraudulent activity, and operational problems. Security is a means for assuring health and help to identify attacks. Enterprises must perform log analysis to discover different attacks by considering heterogeneous log records. We used multilevel log analysis to identify attacks found at different layers of data center through scrutinizing log events of various network devices, applications and others. Thus, to discover different attacks considering heterogeneous log records are basis for analysis. In our work log records were organized together into common format and analyzed based on their features. In central engine clustering and correlation are core of log analyzer that work together with attack knowledge base to identify attacks. Clustering algorithms such as Expectation Maximization, K-means were used to determine the number of clusters and filter events based on filtering threshold respectively. On the other hands, correlation finds a relationship or association among log events and generates new attack definitions. Finally, we evaluated log analyzer prototype of the proposed system and obtained an encouraging result with average precision of SOM#34 and AAU is 84.37 and 90.01 respectively. Further study and implementation of log analysis can significantly enhance data center security of enterprises. Generally, this paper demonstrates the application of log analysis for enhancing security of enterprise data center and our proposed solution will be discussed.

**Keywords**— *Log File; Log Analysis; Layered approach; Attack Identification; Data Center; Network Security*

## 1. INTRODUCTION

In fact a number of enterprises are more interconnected with the advent networking technologies and large amounts of information relayed on these infrastructures. With this respect, Network plays vital role for the establishment of communication. However, security is left aside and becomes big issue for enterprises. The structure of network itself provides suitable condition for intruders to create security events. Since attacks are unstable in their nature that lead not to bring one size fit solution, in which organizations secure their network using their own means.

Log files are set of records collected from log generating devices. They are considered as ideal source of information for security management [3, 4, 5]. By collecting and analyzing log files, security professionals can determine loopholes in their

network and accordingly execute proactive mitigation strategies. In general, our work was aimed at making detail analysis to produce log analyzer having layered data center security approach for organizations aligned with their security policies, procedures, and standards.

In this paper, previously conducted researches in areas related to the concept of log analysis to build an enhanced data center security is presented and deeper discussion on various proposed log analysis techniques is made. Also, specific works in relation to log analyzer having different approaches will be elaborated in the next sections which are related and relevant in terms of their objectives.

## 2. LOG FILE ANALYSIS

Currently security gets more attention in many organizations than ever before. This is due to the growth of Internet and dynamic nature of emerging attacks towards an organizations data center [1]. When organizations ensure security in their business strategy, then the confidentiality, integrity and availability of data will be assured in the data center. Security requirement has a direct relationship with the growth of a data center. It plays fundamental role in contributing towards the development of organizations.

Therefore, network devices generated information (log file) is considered as a means to identify, detect, analyze and take a remedy action accordingly. This enables administrators to easily handle the monitoring activity of the entire data center infrastructure with minimal data loss, time, effort and other expenses.

Recently, the expansion of Internet leads many organizations to be victim of various attacks type and create channel for easy dissemination across organizations' data centers freely. Security becomes great point of interest in which it is accomplished through the process of log file analysis. In such circumstances several research dimensions are conducted towards the data center for guaranteeing security at the required level. The works done so far can be taken as inputs and used to bring a newly proposed solution aimed to enhance network security.

Log files are rich source of information and have been analyzed in the past for a variety of purposes and reasons, such as system maintenance, software testing and validation, forensic analysis and for anomaly detection. The following section will briefly discuss works which has been done related to our work. Here we have categorized the works which are

done so far based on purpose for the usage of log files as discussed below.

### 3. RELATED WORKS

#### 3.1 Log Analysis as a Security Aid

In [6] the collected log files and use an information retrieval open source tools to index log files' fields and search for patterns of suspected behaviors, which may indicate intrusions. The aim of their work was to use the tool for indexing and searching for attack like patterns on log files. The application indexes every occurrence of specific strings (e.g. denial of access, wrong credential, and so on). After that, the system tries to find events within similar occurrences comparing with the data searched for. Fuzzy searches have been used to detect same attack like patterns such as brute force attack.

However, the application is very limited to analyze few number of log files type with known log formats and patterns. Hence, multiple types of log need to be considered and also correlating such logs from myriad of resources is necessary. And also, the system didnot use a well prepared log data by incorporating a preprocessing task. An attack knowledge base which is dynamically updated must be constructed in order to easily handle the analysis process.

In [2] proposed an automated forensic diagnosis system to reconstruct the attacks actions after a security incident has been occurred. Their system analyzes a set of log files created by the different applications running in the network. The system is composed of four modules: event collection, event pre-processing, event correlation, and attack graph generator; all of them working on victim system log files to recreate in an automated fashion and come up with the attacker actions represented by attack scenario. First, event collection module gathers the log files in their original format then, the events pre-processing module adjusted timestamp of log files, normalizes the attributes of the log files and saves them in a repository (event container). Later, the event correlation phase proceed by: first, atomic attack definitions from an attack knowledge base are used to find specific attack actions, and second, the attack actions found are then correlated to build attack scenario describing complex multi-steps attacks. Finally, actions are represented through graphs to facilitate the interpretation to the end-user.

However, the proposed system did not assume that the attack knowledge base will be outdated due to variable nature of attacks and attackers' actions. And also, in preprocessing stage the log record may be incomplete for unknown reason or has different log format but the system did not put any metrics to clean a data. This can reduce the detection accuracy of the system. In attack knowledge base port is left as an attribute which is important parameter like IP address to have a better insight about intruders activity from log analysis.

#### 3.2 Log Analysis using Data Mining

In [7] propose a system that parse/isolate logs from various sources and then cluster the logs using data mining tool (WEKA). The framework first collects unlabelled heterogeneous logs, then parse each raw log individually and isolate log entries when necessary. Secondly, process of

clustering of log entries before filtering. Thirdly, again parse the clustered logs to make it visible for filtering. Later on, the process of filtering proceeds to filter the clustered events. Finally, the system combines the filtered events attribute values which are exactly alike.

However, the proposed work lacks to create common log format through log normalization in a preprocessing module for identification of log in its proprietary log format. In addition, it is better to construct an updated Attack Knowledge Base and compare each filtered event against the knowledge base. Additionally, in their system finding association (correlation) among log records is not considered for attack detection.

In [8] develop Unsupervised Heterogeneous Anomaly Detection system (UHAD) which scrutinizes heterogeneous logs, without using a trained model on traffic behavior or knowledge of anomalies, and uses a two step strategy in clustering normal and abnormal events. They introduce new algorithm for filtering, by which the filtering threshold is calculated based on the volume of log events and the number of log events clusters. First, the component log preprocessing was used to extract the data from the logs with additional functions, such as isolator and timestamp synchronizer. Secondly, the event clustering component separates abnormal events from the normal ones using various logs and also finds possible number of clusters (K) to group the events using expectation maximization algorithm. Thirdly, filtering clustered events component remove the normal events (noise) whilst retaining the abnormal events for further processing. Later on, aggregation of filtered in events component combines the redundant events thereby reducing the events in the filtered log. Then, transferring events component extract the features from various aggregated logs as stated in Generic Format (GF) to store in Generic Format Log (GFL). Finally, the system detects anomalous events by analyzing features such as IP address analysis and port number analysis.

However, the system lacks preparing log files through pre processing to produce a better detection results. It also lacks building attack knowledge base to extract atomic attack definitions for increased anomaly detection accuracy. In their system the concept of correlation among log events is not taken into consideration.

In [13] discusses a data mining tools Simple Log Clustering Tool (SLCT) and LogHound that were designed for assisting system management to extract knowledge from event logs. The automation of event log analysis is an important research concept in network and system management. In order to tackle such problem they propose a data mining technique to obtain knowledge about events. SLCT is basically employs clustering algorithm for analyzing textual event logs where each log record represents certain event. On the other hand, LogHound employs a frequent item set mining algorithm for discovering frequent patterns from event logs.

However, their research did not consider anomaly detection methods as part of their work and the effectiveness of combing SLCT and LogHound, in order to build an event log anomaly detection system. Another drawback is their proposed system did not include more preprocessing techniques to produce efficient log analysis system.

### 3.3 Log Analyzer in real time

In [44] real time log analyzer system was proposed begin with collecting logs data from devices in the network into central server by removing garbage data and make correlation between them into one common table. Then, it converts them into one common format. Customized learning algorithms such as association rule, tf-idf, k-means clustering, and decision tree were used to analyze and interpret data to get important information from log data. Finally, the system converts again into the graphical formats for easy understanding. The system uses adaptive learning algorithm to process the data stream and data model that changes along with time and to flush out the old model when the model is too old. In general, the system target was to detect the abnormal activities using combination of signature-based and learning algorithms based techniques. Their proposed architecture of log analyzer system is shown below in figure 3.1.

Even though, the work come up with low false positive rate but still it lacks further refinement to process the log files collected with more algorithms to gain a better log analyzer. During conversion of log data to common format they left how to extract important features from log file which have direct relationship with detection accuracy of the log analyzer. The implementation of tf-idf algorithm is less important for detecting attacks for large log corpus since it results irrelevant detection results.

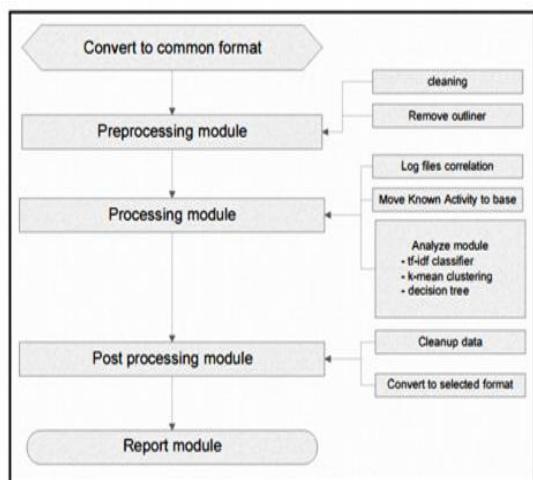


Figure 3.1 Real-time Log Analyzer system architecture

### 3.4 Log Analysis for Security Management

In [9] the authors developed a method for using the multi-agent technology to collect and analyze log data generated by network devices and security devices, and then generating a fixed format data structure and build log collection and analysis systems by incorporating preprocessing operation. The system analyzes the status of the network and information security, and created the centralized storage to provide services for the later research. At first, the log collection agent collects the data of all the network devices, security devices, applications systems and host system, then format these log data into pre-defined log format and store in the log central storage. And then, filter it according to the rules defined in the rule module. Users can have flexibly to adjust the parameters

of filter and filter out the data in which the user may be interested from the many log information according to the actual needs.

However, the proposed system is rule based in which currently emerging attacks do not go along with the defined rule which create security hole in the organization. The absence of structured and up-to-date attack knowledgebase and absence of log correlation bring bottleneck for analysis of log files. The preprocessing module task is not well known for making the system more robust. In addition, in their system the type of activities carried out in log analysis stage is not well stated.

In [10] proposed an approach for receiving, storing and administrating log events. They presented a secure audit log management system focusing on security, flexibility, performance, and portability. Furthermore, they come up with a design solution that allows organizations in distributed environments to send audit log transactions from different local networks to one centralized server in a secure way. They proposed system which has the capability to analyze logs from different log sources such as, firewalls, IDS, servers, and clients. The proposed system consists of one centralized server located on a secured location connected to the inner parts of the network of the supported system. To collect those log events syslog and Simple Network Management Protocol (SNMP) protocols are used. Then, the agents read local logs and transfer the information with the syslog or/and SNMP protocol to the log server. The security audit logs can also be transmitted over the network to the system using standard User Datagram Protocol (UDP), syslog protocol or Transmission Control Protocol (TCP). In such way the system collects log information from all types of clients, servers, firewalls and network equipment. Moreover, the log server is able to detect activation and deactivation of nodes and network equipment on the supervised network, i.e. the network where logs are collected from.

However, the system lacks works related to the preparation and implementation of standardized log formats for heterogeneous log data, as well as normalizing of logs in standardized way. In addition, the issue concerned with the approach or technique used for detecting anomalous events from the log file is not considered.

In [11] develop a model composed by a set of agents in order to collect, filter, normalize, and to correlate events coming from diverse devices. The model provides a capability for analyst in the evidence search process of a forensic investigation. Agents collect log files and send them to an event container. Afterwards, events are filtered by reducing the number of events which are not related with the attack and then normalize to standardize the information of logs. Once all logs are in the same place and format, the correlation engine processes the events and generates a diagnosis of how the system was penetrated. Correlation analysis assigns relationship between multiple events related directly or indirectly with the system violations. In other words, correlating events help at identifying the attacker actions by analyzing events of diverse applications all together.

However, the proposed system limited to correlation (association of events) and exclude log mining algorithms for

multiplespectral identification of attack to obtain better attack detection results. In addition, the system did not use fine tuned filtered event from event container which means if the event contains incomplete information then the system consider as an attack which is not true.

In [14] authors bring a new method for correlating intrusion logs with the main processes of intrusion detection system. It is based on a centralized log correlation system which is composed of six components: data provider, preprocessor, analyzer, manager and controller, responder, and evaluator. In the proposed system, data provider collects data from network logs audited data file (off-line mode) or live network logs (on-line mode) and sends text data to the processor component. Then, the preprocessor converts text data into numeric one and if necessary converts numeric data into binary or normalized form, and sends them to a Self Organizing Map (SOM) neural net based analyzer. In preprocessor, after extracting features from each record, each feature is converted from text or symbolic form into numerical form. In the next step preprocessing convert data into binary, or normalized and scaled form. For normalizing feature values, a statistical analysis is performed on the values of each feature, based on the existing data from dataset and then acceptable maximum value for each feature is determined. The analyzer uses data either for training and testing its SOM neural net or for analyzing and detecting intrusions/attacks. An IDS evaluator component provides a facility for reporting true detection rate, true type detection rate, false positive detection rate, false negative detection rate, and other criteria such as detection rate of three attacks categories, to evaluate their log correlation in the intrusion detection system.

However, the proposed system uses statistical approach that results less detection accuracy comparing to data mining learning schema and limits correlation process to be less effective to identify more intrusions/attacks in the IDS log. The proposed system uses approach for log data collection in online or offline mode in similar way but parameters must be set for proper identification of the log data mode. In the preprocessing component data cleaning is not included which helps to obtain more features.

In [12] a prototype system is developed and implemented based on relational algebra to build the chain of evidence. It is used to preprocess the real generated data from logs and classify the suspicious user based on decision tree. The proposed work describes the nature of the event information and the extent to which it is correlated such event information despite its heterogeneous nature and origins. First, the system begins by extracting log files of the web server and firewall and stored in the central location. In this stage, the data are transformed in a suitable format for conducting effective analysis. Secondly, the chain of evidence analyzer takes the firewall log and web log from the centralized log. It applies the rule based correlation by URLs and Time techniques and creates the training data set. Later on, a decision tree is constructed from the resultant training data set by applying decision tree algorithm which helps to take a proper a decision in suspicious users.

However, the parameters URL and time is not enough for building correlation system and the system did not have a

organized way to analyze log files. This means that it simply takes a garbage collected log records as it is without preprocessing and no common log format which leads a system to be less efficient to identify a malicious users. In addition, the system is limited to use decision tree algorithms in which applying more log mining algorithms gives better detection accuracy. The following figure shows the proposed log analyzer framework.

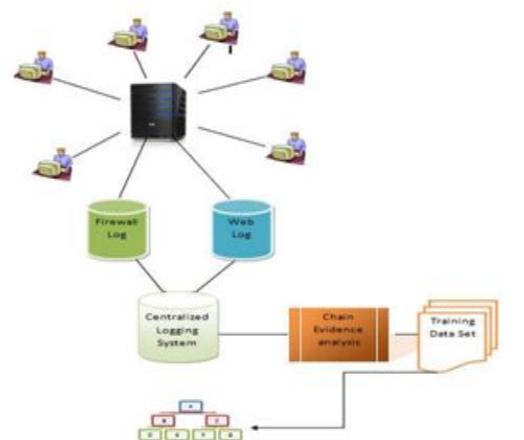


Figure 3.1 Log Analyzer Frameworks

In [15] propose a system used to analyze intersections of log files that come from different applications and firewalls installed on one computer, and intersections resulting from log files coming from different computers. And also, it is concerned with the issues involving large scale log processing which helps to analyze log records. They have used firewalls' log files coming from web server and from regular desktop computer (in both cases coming from the same period of time), and web server's access-log file from the same time period. During the initial preprocessing stage they have removed from all logs entries that were related to intranet, and they have left only those entries that came from outside of their LAN.

However, in the proposed system log preprocessing stage entries are selected based on the source they originated (i.e. can be from intranet or outside LAN) which has no importance to identify attacks in the network. The system did not put any technique or systematic method for detecting attacks from log records.

### 3.5 Build Log Management Architecture

In [12] authors aim to suggest log management architecture with more common functions that are used by vendors. They proposed log management architecture having collection server which is the first module for collecting received logs from various log generator devices such as firewalls, NIDS, operating systems, application systems, etc. Then, log generators send logs by transmitting protocols like syslog, IDMEF, CEE, CEF and SNMP. Thus, collection server must

be able to understand all log formats. After studying various SIEM vendor architectures on log management the most important functionalities are considered as follows: Normalization, filtering, reduction, rotation, time synchronization, aggregation and integrity check. Finally, storage server keeps logs for forensic, auditing and off line analysis. In addition, they consider log security in their architecture.

However, the proposed architecture functionality is not evaluated and tested. And also the architecture did not include log pre-processing and analysis component which are core in log analysis.

In [16] proposes a defense in depth network security architecture and applies the data mining technologies to analyze the alerts collected from distributed intrusion detection and prevention systems (IDS/IPS). The key component of the Global Policy Server (GPS) is the security information management (SIM) which consists of an online detecting phase and an offline training phase. The system consists of four main components in the online detecting phase: First, the online data miner, which classifies the records in active database to detect attacks. Then, the rules tuner which runs the machine learning algorithm tunes the parameters of rules accordingly. Later, the GLS, which receives logs from LPSs stores them into an active database. Finally, policy dispatcher waits for the commands from the online miner.

However, the experimental results demonstrate the proposed work is highly effective only for detecting the DDOS attacks which is not for other attack. It also did not show if we take a shorter time interval between the events it is difficult to suggest about occurrence of the false alarm rate. Also, the model only uses classifier as mining technique.

#### 4. PROPOSED SYSTEM ARCHITECTURE

In fact, the development of system is determined based on the composition of many subcomponents (parts) of the entire system. Hence, the integration and interoperability of those components will produce the expected system so that its objective will be met.

Our model of log file analyzer with layer based data center security consists of eight major components, including :- Log Files Repository(LFR), Log File Pre-processor(LFPP), Network Security Information Manager(NSIM), Attack knowledge Base(ANB), Central Engine(CE), Action Center or Remediation Area, and Audit Reporter as shown in figure 4.1. Log file collector is simply concerned with gathering heterogeneous log files generated by those leveled devices in the data center network and put in central repository (LFR) for preprocessing which is second component of the system.

The aim of the log file pre-processor is to adjust or prepare the various input log files considering different metrics and feed to the processing unit called central engine. Log parsing, log cleaning, log normalization, and log aggregation are the major activities conducted in this component. Once this part is completed, then the preprocessed log files are stored in NSIM component and ready for later usage by central engine.

Network Security Information Manager is a module that works as a central repository for the entire raw of log data as

well as controls the integrity, accessibility and confidentiality of those pre-processed log files received from log file preprocessing component. Attack knowledge base is part of the system which consists of atomic attack definition received from the central engine and provides required atomic attack to the central engine.

The central engine component is the heart (core) of the system which is responsible to perform the overall log file processing. To achieve this the subcomponents include clustering and correlation or association module which comprise their own components. On the other hand, an action center or a remediation section is a component that deals with providing reasonable response towards emerging incident to the administrator through the user interface. It incorporates alert production, notification through SMS, generating report, and visualizations. Finally, the audit reporter module is a repository for the generated information from action center component for long term usage. Therefore, our research is mainly concerned with enhancing data center security trends of organizations through building log file analyzer (i.e. intended to identify and inform the state of the data center within certain period of time). The following section will briefly elaborate the concept related to our proposed architectural design for log files analysis using layered approach of data center security.

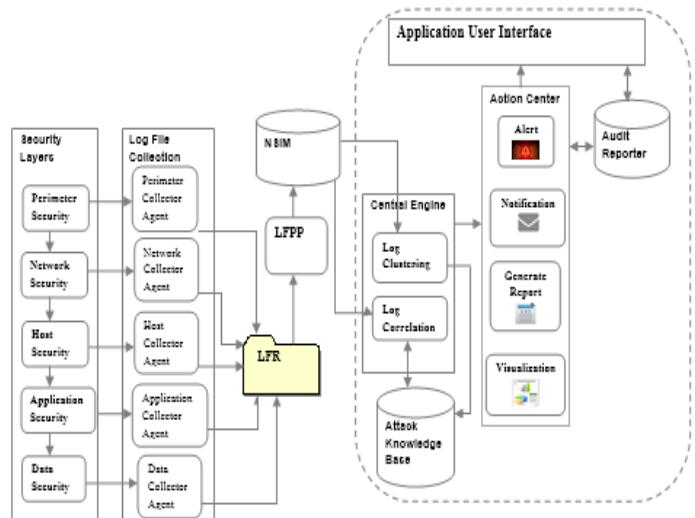


Figure Proposed Log Analyzer System Architecture

#### 5. RESULTS AND DISCUSSION

The clusters produced by clustering algorithms were calculated using weka experimenter with 10 fold cross validation and 10 iterations to allow every part of the log record to be tested. True Positives (TP), True Negatives (TN), False Positives (FP) were measured to calculate the precision and False Positive Rate (FPR) of clusters. Precision is calculated by dividing true identified log by total number of clustered logs ( $\text{precision} = \text{TP} / \text{Total number of clustered log events}$ ). And False Positive Rate is calculated by dividing incorrectly identified log events to a given cluster by the sum of incorrectly identified log events to a given cluster and correctly clustered log events into other cluster ( $\text{FPR} = \text{FP} / (\text{FP} + \text{TP})$ ).

(FP + TN)). The evaluation of clustered log events for SOTM#34 and AAU is summarized below in Table 1 and Table 2 respectively.

**Table 1 Evaluation of Clustered Log Events for SOTM#34**

Log File	clustered log records	Evaluation				
		TP	FP	TN	Precision (%)	FPR (%)
access_log	3,554	3,281	73	97	92.31	42.94
error_log	3,692	3450	192	44	93.44	81.35
ssl_err or_log	374	298	76	21	79.6	78.35
Iptable syslog	179, 752	168,269	1103 4	110 2	93.61	90.91
Snorts yslog	69,039	58,480	1055 9	527	84.70	95.24
Mail_1 og	1,172	794	378	80	67.74	82.53
Messa ges	1,166	935	531	204	80.18	72.24
Secure	1,587	1,449	138	36	91.30	79.31

**Table 2 Evaluation of Clustered Log Events for AAU**

Log File	clustered log records	Evaluation				
		TP	FP	TN	Precision (%)	FPR (%)
acce ss_lo g	15868324	1467623 1	892093	9578	92.48	98.93
error _log	4194204	3672201	322003	5697	87.55	98.26

## 6. CONCLUSION AND FUTURE WORKS

Following results of this research the process of finding attacks out of the whole log entries can be easily conducted and possible solution can be determined. We presented a log analyzer architecture organized in a layer for enhancing security of data center. The system analyzes a set of heterogeneous log files by collecting log data, preprocess them, build central engine for analysis and taking remedial measure through action center are the processes of the proposed system. The central engine module is heart of the entire system performing log processing or analysis. It was used for finding security holes in bidirectional fashion using both clustering and correlation techniques. In order to validate the usability of our system we used a real network based log records of both SOTM#34 and AAU data center devices. Based on those we found several attack actions which leads us for construction of an attack scenarios. In general, the

following are some of potential future works for continuation of our work.

- Apply more log analysis or mining approaches to obtain useful knowledge and reduce false positive and false negative results.
- Take more log records to make the work generic.
- Understands users' behavior from analyzed logs.
- Create intelligent attack knowledge base for the sake of forensics, auditing and others.

## ACKNOWLEDGMENT

I would like to express my deepest gratitude to my advisor Dr. Dejene Ejigu in which the work would not be possible without his motivation, enthusiasm, continuous as well as constructive supervision, and encouragements. I want to extend my appreciation to Mr. Zelalem Assefa manager of EthERNET and other members for providing necessary information related to research. Finally, thanks to my colleagues, families, teachers, friends and others who have contributed in one or another ways for successful accomplishment of this work.

## REFERENCES

- [1] Bagchi, K., and Udo G., "An analysis of the growth of computer and Internet security breaches", Communications of the Association for Information Systems, Issue12, pp. 684-700, 2003.
- [2] Herreras, J., Gomez, R., "Log Analysis Towards an Automated Forensic Diagnosis System", Availability, Reliability, and Security, 2010. ARES '10 International Conference on, pp.659-664, 15-18 Feb, 2010.
- [3] Pingchuan Ma, "Log Analysis-Based Intrusion Detection via Unsupervised Learning", Unpublished Master's Thesis, Master of Science in School of Informatics, University of Edinburgh, UK, 2003.
- [4] Deepak Upadhyaya and Shubha Jain, "Model for Intrusion Detection System with Data Mining", International Journal of Advanced Research in Computer Engineering and Technology, pp. 145-148, Volume 1, Issue 4, June 2012.
- [5] Prashant Achari, Susanta Adhikary, Jayashree Madugundu, and Mungara Jitendranath "Knowledge and Rule Based Learning Engine to Analyze the Logs for Troubleshooting", International Journal for Scientific International Journal for Scientific Research and Development, Vol. 2, Issue 04, 2014
- [6] Leite, Jorge Pinto, "Analysis of log files as a security aid", In Information Systems and Technologies (CISTI), 6th Iberian Conference, pp. 1-6. IEEE, 2011.Sorot Panichprecha, "Abstracting and Correlating Heterogeneous Events to Detect Complex Scenarios", unpublished PhD Thesis, Queensland University of Technology, Brisbane, Australia, March 2009.Research and Development, Vol. 2, Issue 04, 2014.
- [7] Asif-Iqbal, H., Nur Izura Udzir, Ramlan Mahmod, and Abdul Azim Abd Ghani. "Filtering events using clustering in heterogeneous security logs", Information Technology Journal 10, No. 4, pp.798-806, 2011.
- [8] Ghani Abdul, "An unsupervised heterogeneous log-based framework for anomaly detection", Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.
- [9] Shengyan Shi, Shen Xiaoliu, Zhao Jianbao, and Ma Xinke, "Research on System Logs Collection and Analysis Model of the Network and Information Security System by Using Multi-agent Technology", In Multimedia Information Networking and Security (MINES), Fourth International Conference , pp. 23-26. IEEE, 2012.
- [10] Söderström, Olof, and Esmiralda Moradian. "Secure Audit Log Management", Procedia Computer Science 22, pp. 1249-1258, 2013.
- [11] Herreras Jorge, and Roberto Gomez, "A log correlation model to support the evidence search process in a forensic investigation", In

- Systematic Approaches to Digital Forensic Engineering, SADFE 2007.  
Second International Workshop on, pp.31-42. IEEE, 2007.
- [12] Madani, Afsaneh, Saed Rezayi, and Hossein Gharaee, "Log management comprehensive architecture in Security Operation Center (SOC)", In Computational Aspects of Social Networks (CASON), International Conference, pp. 284-289. IEEE, 2011.
- [13] Vaarandi, Risto, "Mining event logs with slct and loghound", In Network Operations and Management Symposium,NOMS . IEEE, pp. 1071-1074, 2008.
- [14] Sayed Omid Azarkash and Saeed Shiry Ghidary, "Logs Correlation: Current Approaches, Promising Directions, and Future Policies", Vol. 2(5), pp.4413-4322, Journal of Basic and Applied Scientific Research, 2012.
- [15] Kowalski, K., Beheshti M., "Analysis of Log Files Intersections for Security Enhancement", Information Technology: New Generations, ITNG 2006. Third International Conference, pp.452-457, 10-12 April 2006.
- [16] Nen-Fu Huang, Chia-Nan Kao, Hsien-Wei Hun, Gin-Yuan Jai, Chia-Lin Lin, "Apply Data Mining to Defense-in-Depth Network Security System", Proceedings of the 19<sup>th</sup> International Conference on Advanced Information Networking and Applications (AINA), 2005.

#### AUTHORS PROFILE

1. Name: Samuel Getachew

- **Academic Rank:** Lecturer
- **Institution:** Haramaya university, College of Computing and Informatics, Department of Computer Science,Haramaya, Ethiopia
- **Email** – [samuelgetachew34@gmail.com](mailto:samuelgetachew34@gmail.com)
- **Mobile Phone** - +251-9112887176

2. Name: Dejene Ejigu (Ph.D.)

- **Academic Rank:** Assistant Professor
- **Institution:** Addis Ababa University, Head of Information Technology Doctoral program, Member of computer science department.
- Advisor of Master's Thesis and gave lecture for Selected topics in computer science and Computer Security course
- **Email** - [ejigud@yahoo.com](mailto:ejigud@yahoo.com)
- **Mobile Phone** - +251-911982031
- **Office Phone** - +251-255530392
- **Fax**-+251-255530325

# Solving Bi-objective Two-Dimensional Rectangle Packing Problem using Binary Cuckoo Search

Amandeep Kaur Virk

Assistant Professor, Sri Guru Granth Sahib World University, Fatehgarh Sahib, India. Ph.D. Research Scholar, University College of Engineering, Punjabi, University, Patiala, India.

Dr. Kawaljeet Singh

Director, University Computer Centre, Punjabi University, Patiala, India

**Abstract:** The work presented here optimizes the rectangular packing problem in which rectangular items are packed on a rectangular stock sheet. Our objective is to maximize the utilization of the rectangular sheet and to minimize the number of non-guillotine cuts required to cut various pieces. Binary version of cuckoo search algorithm has been used to solve this discrete problem. A series of computational experiments have been conducted to evaluate the performance of the new cuckoo search metaheuristic technique. It appears from the computational analysis that the cuckoo search algorithm is able to give good solutions.

**Keywords:** Nesting Problem, Cuckoo Search, Multi-objective optimization, Non-guillotine cutting.

## 1. INTRODUCTION

Nesting (Cutting and packing) problem is a combinatorial optimization problem that finds wide applicability in industries related to leather [17], wood [11], sheet metal cutting [19], paper [12,20], automobiles and ship building [3]. This problem deals in finding an optimum arrangement for multiple items on a large containing area such that minimum material is wasted, as optimum design layouts can lead to considerable savings in raw material. The cutting stock problem is a NP-hard optimization problem [4,6,8]. The two-dimensional non-guillotine stock cutting problem consists of packing or cutting rectangular items of predetermined sizes onto a rectangular stock sheet which is bigger in size than the items. Non-guillotine cutting means that the cuts may not go from one end to the other end of stock sheet [7]. The primary aim of this problem is to pack the items in such an arrangement that maximum utilization (utilization factor) or minimum wastage (trim loss) of the stock sheet is achieved. Due to many real world applications of packing and cutting problem, consideration of other criteria like profitability, number of cuts, setup cost etc. are also important to cover all aspects of production process. This study takes number of cuts (non-guillotine) as the second objective for cutting process because optimization of number of cuts is crucial in determining the tool life and cost of the cutting process.

Evolutionary algorithms have been used over the years to solve complex combinatorial problems. One of the most recent evolutionary approach developed by Xin-She Yang and Suash Deb in 2010 [18] is Cuckoo Search (CS) Algorithm. Cuckoo Search algorithm was inspired by the behavior of cuckoo bird which uses the nests of other host birds to lay its eggs. The host bird may discover the alien egg and destroy it or may abandon the nest. Levy flights are used by cuckoos to search a new nest for laying their eggs.

This study uses Binary Cuckoo Search (BCS) algorithm proposed by Gherboudj et al. in 2012 [5] to solve the two-dimensional non-guillotine rectangle packing problem.

The outline of this paper is as follows: Section II describes the rectangle packing problem with multiple objectives. Cuckoo Search algorithm is discussed in Section III. Section IV covers the experimental results and discussions.

## 2. RECTANGLE PACKING PROBLEM

The two-dimensional non-guillotine rectangle packing problem consists of a large rectangular stock sheet of given length L and width W and an order list of small rectangular items  $i$  of specified length  $l_i$  and width  $w_i$ ,  $i = 1, 2, 3, \dots, n$ , to be cut from stock sheet. The items are allowed to rotate by 90°. The cuts are non-guillotine which means that the cuts may not go from one end of the stock sheet to another end [7]. The topology of cutting and packing problems given by Wäscher et. al in 2007 [16] classifies the two dimensional rectangle packing problem as two-dimensional rectangular single large object placement problem (2D-SLOPP).

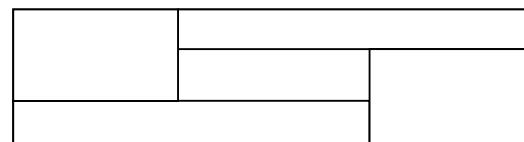


Fig. 1: Non-guillotine cuts

The primary objective of this problem is to find a layout of items on stock sheet which maximizes the utilization of the stock sheet. This paper considers a second objective which is minimizing the number of cuts required to maximize the tool life.

The multi-objective goal of this problem can be stated as:

$$\text{Maximize } \sum_{i=1}^n \frac{l_i w_i}{N(LW)} \quad 0 \leq i \leq n \quad (1)$$

$$\text{Minimize } \sum_{j=1}^N \sum_{i=1}^n C_i P_{ij} \quad 0 \leq i \leq n, 0 \leq j \leq N \quad (2)$$

where

L is the length of stock sheet,

W is the width of stock sheet,

$l_i$  is the length of  $i^{th}$  piece,

$w_i$  is the width of  $i^{th}$  piece, for  $i = 1, 2, 3, \dots, n$ .

N is total number of available stock sheets

$C_i$  is the number of cuts required to cut  $i^{th}$  item

$P_{ij}$  is the number of pieces if  $i^{th}$  shape on  $j^{th}$  stock sheet.

### a) Maximizing utilization factor

Utilization factor is defined as the amount of stock sheet space used to place rectangular items. Our aim is to place rectangles in an optimum arrangement so that maximum space is utilized. Many heuristics have been proposed in literature for optimum placement of objects. The heuristic used in this paper is Best Fit Decreasing (BFD) heuristic. The items are sorted in decreasing order in BFD and placed at a position where they best fit by leaving minimum wastage. The placement strategy used to place rectangles on stock sheet is constructive approach (CAA) with maximum adjacency [13,15]. The strategy starts by placing the first item at the bottom left of the stock sheet. The next piece can be placed adjacent to the already placed piece (horizontally or vertically). The best position for next piece is selected on the basis of maximum adjacency i.e. the position which shares maximum common boundary with the already placed piece is selected.

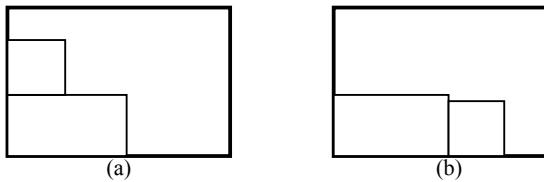


Fig. 2: (a) Second item placed on top of first piece (b) Second item placed along with first piece

Figure 2 above shows two placing positions for second rectangle. The ideal position is shown in (b) as the two rectangles share maximum common boundary (adjacency).

### b) Minimizing the number of cuts

Optimization or efficient use of cutting equipment can be achieved by minimizing or reducing the number of independent cuts required by a packing arrangement. Cuts are simply defined as the number of dissimilar edges within a packing arrangement. We take into account non-guillotine cuts.

The required number of cuts is counted using the following method [14]. The total number of cuts required by a single rectangle is four as it has four edges. For each rectangle, all edges of the rectangle are checked whether they lie on the edge of mother sheet or not. If an edge lies on mother sheet edge, then the number of cuts for each such occurrence is reduced by one as sheet edge does not require a cut. The figure 3 below shows a rectangle with two edges lying on the boundary of the sheet. So, the number of cuts required to cut this rectangle is reduced to two.

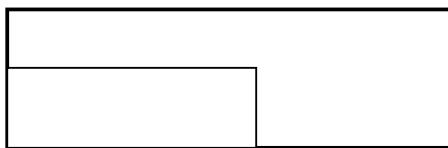


Fig. 3: Rectangle with edges on the stock sheet boundary

Next we check the alignment of each rectangle with all other rectangles placed on the stock sheet. A rectangle can share a common edge with other rectangles horizontally or vertically. There are two possible cases:

Case1: When rectangles are aligned or associated sidewise (fully or partially).

Case 2: When rectangles are aligned or associated one above the other (fully or partially).

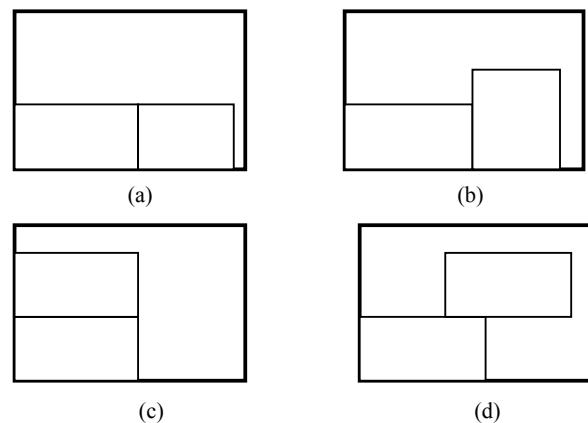


Fig. 4: (a) Fully sidewise aligned rectangles (b) Partially sidewise aligned rectangles (c) Fully top wise aligned rectangles (d) partially top wise aligned rectangles

For both cases, if rectangles touch each other, then the number of cuts is decreased by one.

## 3. CUCKOO SEARCH ALGORITHM

Cuckoo Search algorithm is one of the recent and simple nature inspired optimization algorithm based on the behavior of cuckoo bird. Cuckoos cleverly hack the nests of other birds to lay their own eggs. Cuckoos are under constant threat that their eggs might be discovered by the host birds. If discovered, their eggs are either destroyed by host bird or the host bird abandons the nest. The three idealized rules proposed by Yang and Deb in 2009 [18] for cuckoo search algorithm are:

1. Each cuckoo can lay one egg at a time in a randomly chosen nest.
2. High quality solutions or eggs will be carried over to next generation.
3. The total number of host nests is fixed and a host bird can discover the alien cuckoo egg with a probability  $p_a \in [0, 1]$ . If the egg is discovered, the host bird can either destroy the egg or abandon the nest to build a completely new nest at a new location.

The new nests at new location are built using levy flight (eq. 3). Levy flight provides a random walk where random step is drawn from levy distribution (eq. 4) which has infinite variance with an infinite mean. The steps form a random walk process with a power-law step-length distribution with a heavy tail (Yang and Deb, 2010).

$$x_i^{t+1} = x_i^t + \alpha + \text{Levy}(\lambda) \quad (3)$$

$$\text{Levy} \sim u = t^{-\lambda} \quad (4)$$

Where  $\alpha$  is the step size related to the scale of the problem of interest. In general, we take  $\alpha = O(1)$ . The product  $+$  means entry wise multiplication similar to those used in particle swarm optimization [18].

Levy Flight prevents the problem of being trapped into a local optimum by obtaining the best solution so far. The basic steps of cuckoo search algorithm are given in figure 5.

---

**Objective function**  $f(x)$ ,  $x = (x_1, \dots, x_d)^T$ ;

**Initial a population** of  $N$  host nests  $x_i$  ( $i = 1, 2, \dots, N$ );

**while** ( $t < \text{MaxGeneration}$ ) or (stop criterion);

- Get a cuckoo (say  $i$ ) randomly by Lévy flights;

- Evaluate its quality/fitness  $F_i$ ;
  - Choose a nest among  $n$  (say  $j$ ) randomly;
  - if ( $F_i > F_j$ ),
    - Replace  $j$  by the new solution;
    - end
  - Abandon a fraction ( $pa$ ) of worse nests
  - build new ones at new locations via Lévy flights;
  - Keep the best solutions (or nests with quality solutions);
  - Rank the solutions and find the current best;
- 
- end while**

Fig. 5: Cuckoo Search Algorithm

The cuckoo search algorithm works for continuous optimization problems as the solution to a problem is a set of real numbers. Nesting problems are discrete binary optimization problems which accept binary values as solution [2]. Gherbouj et al. in 2012 [5] proposed a binary version of cuckoo search algorithm (BCS). This paper uses the proposed BCS to solve the bi-objective rectangle packing problem. It works as follows:

#### 1. Initial population

A set of initial nests is generated randomly which represents the current location of cuckoo eggs. The fitness function of each nest is calculated. The initial population is made diverse by making it a mix of both good and bad solutions.

#### 2. Levy Flights

New nests are built using levy flight which provides randomization by using levy distribution (eq. 3 and 4). Levy flights provide global exploration.

#### 3. Binary solution representation

The solution built using levy flight is converted into a binary solution. To get a discrete solution, we generate flipping values for the solution using sigmoid function as follows:

$$S(x_i) = \frac{1}{1 + e^{-x_i}} \quad (5)$$

Where  $S(x_i)$  is the flipping chance of solution bit  $x_i$ .

For each solution, a random number  $\gamma$  is generated from the interval  $[0,1]$ . The random number is compared with flipping value. If the random number generated is lower than the flipping value of  $x_i$ , then  $x_i$  takes value 1. Otherwise  $x_i$  takes the value 0. So, the parts with value 1 are selected for new solution and the ones with value 0 are not selected.

#### 4. Evaluate Fitness

A nest from initial population is chosen at random and its fitness is compared with the newly generated binary solution. The random nest is replaced with the current nest if the fitness value of latter is better. Next a fraction of worst nests ( $pa$ ) is abandoned so that best quality solutions move to next iteration. At the end, the global best solution is updated and

the whole process is repeated until maximum number of iterations is reached.

---

**Input:** Real solution representation  $x_i$

For ( $i = 1$  to (problem size))

{

$$S(x_i) = \frac{1}{1 + e^{-x_i}}$$

If (random number  $\gamma < S(x_i)$ )

$x_i = 1$ ;

Otherwise

$x_i = 0$ ;

}

---

Fig. 6: Binary solution representation

## 4. Experimental Results And Discussion

The cuckoo search algorithm was implemented in MATLAB 7.3. We consider ten different classes of randomly generated problems, in which, Berkey and Wang [1] proposed the first six classes and Martello and Vigo [10] introduced the last four classes. The classes are listed in table 1 below.

Classes VII to X are based on the following types of items defined on the basis of the length  $L$  and width  $W$  of the stock sheet.

Type 1:  $w_j$  uniformly random in  $[2/3W, W]$ ;  $h_j$  uniformly random in  $[1, 1/2H]$

Type 2:  $w_j$  uniformly random in  $[1, 1/2W]$ ;  $h_j$  uniformly random in  $[2/3H, H]$

Type 3:  $w_j$  uniformly random in  $[1/2W, W]$ ;  $h_j$  uniformly random in  $[1/2H, H]$

Type 4:  $w_j$  uniformly random in  $[1, 1/2W]$ ;  $h_j$  uniformly random in  $[1, 1/2H]$ .

For each class five instances are generated with number of items: 20, 40, 60, 80 and 100. The problem instances are given by Lodi et al. [9] and are available publicly ([www.or.deis.unibo.it/research.html](http://www.or.deis.unibo.it/research.html)).

Table 2 below summarizes the computational results for all the instances of each class. The first two columns of the table give the class number and number of items respectively. Next, the multiple objectives, number of cuts required and utilization factor are shown. The last column shows the execution time (in seconds) required to solve the various problem instances. The used cuckoo search metaheuristic is able to give us good overall solutions.

TABLE 1  
Problem Set

Class	Stock Sheet (LxW)	Item ( $l_i$ and $w_i$ )
I	10x10	uniformly random in $[1, 10]$
II	30x30	uniformly random in $[1, 10]$

III	40x40	uniformly random in [1,35]
IV	100x100	uniformly random in [1,35]
V	100x100	uniformly random in [1,100]
VI	300x300	uniformly random in [1,100]
VII	100x100	Type1 with probability 70%, Type 2,3,4 with probability 10% each
VIII	100x100	Type2 with probability 70%, Type 1,3,4 with probability 10% each
IX	100x100	Type 3 with probability 70%, Type 1,2,4 with probability 10% each
X	100x100	Type 4 with probability 70%, Type 1,2,3 with probability 10% each

TABLE 2  
Computational Results

Class	Number of items	Number of cuts	Utilization Factor (%)	CPU Time (in seconds)
Class I	20	27	95.12	1.78
	40	82	95.72	9.04
	60	81	95.16	10.82
	80	110	96.03	28.37
	100	96	95.87	19.89
Class II	20	40	82	3.35
	40	84	89.72	15.67
	60	129	92.07	32.48
	80	162	88.33	58.40
	100	164	91.38	68.02
Class III	20	44	85.42	2.42
	40	112	91.02	14.71
	60	92	86.69	13.17
	80	197	83.47	36.68
	100	219	87.18	51.26
Class IV	20	40	88.71	3.82
	40	86	84.84	15.87
	60	115	87.30	21.46
	80	181	84.07	42.84
	100	223	82.34	83.71
Class V	20	41	84.43	2.12
	40	109	89.33	10.95
	60	109	87.05	10.95
	80	95	89.66	13.73
	100	199	91.95	42.37
Class VI	20	45	69.67	3.54
	40	144	82.14	27.87
	60	145	84.42	37.70
	80	159	83.51	59.34
	100	203	82.92	82.87
Class VII	20	47	87.42	3.32
	40	67	89.6	6.48
	60	109	85.51	13.23
	80	162	84.23	20.43
	100	173	86.87	30.29
Class VIII	20	52	89.97	3.109
	40	132	168	90.70
	60	189		96.23

	80	248	92.31	42.87
	100	220	93.66	50.51
Class IX	20	54	79.50	2.06
	40	111	85.10	7.093
	60	166	86.93	16.68
	80	179	89.33	26.87
	100	241	93.06	36.20
Class X	20	37	87.51	1.68
	40	56	92.65	4.75
	60	80	90.71	7.98
	80	182	89.35	29.53
	100	164	91.43	36.37

## 7. CONCLUSION

This paper addressed multi-objective rectangle packing problem in which rectangular items are to be packed on a rectangular stock sheet such that utilization factor is maximized and the number of cuts required is minimized. Binary cuckoo search algorithm has been used to solve the problem using set of instances taken from literature. The experimental results demonstrate that the algorithm provides good quality solutions. More new metaheuristic techniques can be used to solve the problem.

## REFERENCES

- [1] Berkey J.O., Wang P.Y., "Two-dimensional finite bin packing algorithms," *Journal of the Operational Research Society*, vol. 38, pp.423-429,1987.
- [2] Carravilla, M. A. , Ribeiro, C and Oliveira, J. F., "Solving nesting problems with non-convex polygons by constraint logic programming," *International Transactions in Operational Research*, Wiley Online Library, vol. 10, No. 6, pp 651–663,2003.
- [3] Cheok, B.T. and Nee, A.Y.C., " Algorithms for Nesting of Ship/Offshore Structural Plates," *Advances in Design Automation*,vol. 32, No. 2, pp. 221-226,1991.
- [4] Garey, M. and Johnson, D., "Computers and intractability, a guide to the theory of NP- Completeness," *W.H. Freeman and Company, San Francisco*,1979.
- [5] Gherboudj, A., Layeb, A., and Chikhi,S., "Solving 0-1 knapsack problems by a discrete binary version of cuckoo search algorithm," *International Journal of Bio-Inspired Computation*, vol. 4, No. 4, pp. 229-236,2012.
- [6] Hopper, E. and Turton, B.C.H., "A Review of the application of metaheuristic algorithms to 2D strip packing problems," *Artificial Intelligence Review*,vol. 16, pp. 257–300,2001a.
- [7] Leung, T.W., Yung, C.H., Troutt, M.D., " Applications of genetic search and simulated annealing to the two-dimensional non-guillotine cutting stock problem," *Computers and Industrial Engineering*, vol.40, pp. 201–214,2001.
- [8] Liu, H.Y. and He, Y.J., "Algorithm for 2D irregular-shaped nesting problem based on the NFP algorithm and lowest-gravity-center principle," *Journal of Zhejiang University – Science*,vol. 7, No. 4, pp. 570-576,2006.
- [9] Lodi,A.,Martello,S.,Vigo,D., "Heuristic and meta heuristic approaches for a class of two dimensional bin packing problems," *INFORMS Journal of Computing*,vol.11, pp.345–357,1999b.
- [10] Martello S., Vigo D., " Exact solution of the two-dimensional finite bin packing problem," *Management Science* ,vol. 44, pp.388-399,1998.
- [11] Reinaldo, M. and Luciano, B., "Optimising the cutting of wood fibre plates in the hardboard industry," *European Journal of Operational Research*, vol. 183, pp. 1405-1420,2007.
- [12] Selow, R., Junior, F.N., Heitor, S. and Lopes, H.S., "Genetic Algorithms for the Nesting Problem in the Packing Industry," *The International Multi Conference of Engineers and Computer Scientists (IMECS)*, pp. 1-6,2007.
- [13] Terashima-Marín, H., Ross, P., Farias-Zárate, C. J., López-169 Camacho, E., & Valenzuela-Rendón, M., "Generalized Hyper-
- [14] Heuristics for Solving 2D Regular and Irregular Packing Problems," *Annals of Operations Research*, vol. 179, No. 1, pp. 369-392,2010.
- [15] Tiwari,S and Chakraborti, N., "Multi-objective optimization of a two-dimensional cutting problem using genetic algorithms," *Journal of Materials Processing Technology*, vol. 173, pp. 384-393,2006.
- [16] Uday, A., Goodman, E. D., & Debnath, A. A., "Nesting of irregular shapes using feature matching and parallel genetic algorithms," *Genetic and evolutionary computation conference late breaking papers, San Francisco, California, USA*, pp. 429–434,2001.
- [17] Wascher, G., Haasner, H., Schumann, H., " An improved typology of cutting and packing problems," *European Journal of Operational Research*, vol. 183, pp. 1109–1130,2007.
- [18] Yang H.H. and Lin C.L., "On genetic algorithms for shoe making nesting," *A Taiwan case Expert Systems with Applications*,vol. 36, No. 2, pp. 1134-1141,2009.
- [19] Yang, X.S. and Deb, S., "Engineering Optimisation by Cuckoo Search, " *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 1, No. 4, pp. 330–343,2010.
- [20] Yaodong, C. and Xiaoxia, S., "Applying parallelogrammic strips for cutting circles from stainless steel rolls," *Journal of Materials Processing Technology*, vol. 205, pp. 138-145,2008.
- [21] Yaodong, C. and Yiping, L., "Heuristic algorithm for a cutting stock problem in the steel bridge construction," *Computers and Operations Research*, vol. 36, pp. 612 – 622,2009.

# Optimising Mobile Adhoc Energy demands with Probabilistic Max Drift and Longevity Scheme for realizing Green Campus status in Higher Education Institutions

Kesava Rao Alla <sup>#1</sup>, Soong Der Chen <sup>2</sup>

*Department of Graphics and Multimedia, College of Information Technology,  
University Tenaga Nasional, Malaysia.*

<sup>1</sup> alla248@yahoo.com

<sup>2</sup> chensoong@uniten.edu.my

**Abstract—**Green computing is an approach of optimizing the usage of Computer systems without compromising on system output and performance. As each hardware and software component contributes to the overall system energy requirement, this research is about presenting an investigation on minimizing the energy demand over a network in a Higher Education Institution (HEI) which can contribute towards achieving and maintaining a green campus environment. The principal contribution of this research is an amplification method that uses Probabilistic Max Drift in reducing Mobile Ad hoc energy demands and improving the durability. Comparison of performance of amplified networks was simulated using Java with their initial layouts. Furthermore, extended probabilistic method is added to Max Drift Scheme, and the effects are assessed by comparing on network lifetime when combined with network amplification. This system uses bi-connectivity directly to improve network lifetime, and also it introduces the network maintenance improvement to promote green environment. The results show that the energy consumption was reduced to a significant level of 17% when tested for one of the HEI, which thus plays a key role in fulfilling the green computing requirements and provides a pathway to realising the green campus. With these findings, it is envisaged that this system with less network resource usage could very well be applicable for any other HEI or any other environment with a demand for higher volumes of network communication resources.

**Index Terms—**Adhoc , Mobile , Energy Demands , Green Computing , Institution or University

## I. INTRODUCTION

Mobile Ad hoc grids are most popular network communication. However, wide spread of such deployment is yet to be established. Existing mobile ad hoc networks generally follow the structure in which dynamic mobile nodes are linked to an enormous ad hoc network. These dynamic mobile nodes have an ability to communicate over wireless links, and they connect devices with wireless connection by acting as a gateway to reach the larger network in a bottom-up fashion fabricated grids. In general, ad hoc network requires fewer grids and connects many nodes in a decentralized fashion. Such networks are very elastic and can be rapidly deployed together with many wireless networks, such as sensor networks, networks of unmanned vehicles in the army command and

control. Energy efficiency is important in wireless nodes, since energy wasted at the wireless sensors are non-replaceable in different situations. So, preservation of energy is one of the key points in extending lifetime of network. A wireless node often spends a lot of energy in transportation [1]. The Maximum Drift problems are used to find the possible maximum drift in a single-source and single-sink network. In this paper, the implementation and operation of Probabilistic Max Drift Algorithm in Programming language Java is presented. Personal Digital Assistants (PDA) are used in the test platform since PDAs are less expensive and light weight with high mobility. The remaining part of the paper is organized as follows. Section 2 explains the related works. Section 3 describes the detailed implementation and its description of Probabilistic Max Drift Algorithm. Section 4 presents the experimental results with detailed discussion and comparative analysis of the proposed design with existing designs. Conclusion and further research is presented in section 5.

## II. RELATED WORK

Modern day Higher Education Institutions (HEI) are densely populated tech-savvy communities with students, academic and supporting staff and researchers residing in and around campus. These communities are actively using many mobile devices such as laptops, PDAs, mobile phones, smart phones, smart cameras, gaming devices, etc., round the clock [2]. Many of these devices massively consume data transmission through wireless networks for browsing, downloading academic content and using it across multiple platforms, mailing, downloading and sharing music, using social networks, chatting and distributing videos in high definition formats on the web. In this scenario, it is obvious that the HEI's are having bigger issues of expanding its wireless routers. Even if the routers are increased, the energy consumption is a major issue. In order to resolve these issues, researchers are relying on a technology called green computing for better energy consumption rates [3]. To create a green campus, it is necessary to apply green computing as one of the components that start with software

and hardware design, installation, maintenance and disposal of the equipment after the shelf life. The key factor behind the implementation is, utilizing all these systems with optimum usage and with lesser energy consumption as much as possible. Ad hoc networks can be used as a feature of green computing. Ad hoc is a strategy and a vision to bring industry relevant mobility solutions while shortening deployment and reducing the burden on IT resources. The proposed method offers a rounded approach to mobility that combines applications, devices, and the network as the platform. This is to support HEI's networking infrastructure mobility, and the unified wireless network provides flexibility to achieve their requirement within the specified budget [4].

### III. METHODOLOGY

This proposed method and architecture offers HEIs huge mobile network requirements for the entire campus area with effective mobility beyond just providing an wireless connection. Whenever mobile ad-hoc network is configured, the supreme drift was calculated using Probabilistic Max Drift Algorithm and energy demand intended by using Probabilistic Max Drift Reduced Energy Demand method. Based on calculated vigor demand, the maximum drift route was calculated using the Algorithm.

#### A. Probability Max Drift

In the proposed mobile ad hoc network, graph data structure was used to organize the network. A graph is a collection of nodes and edges, which connects different nodes. A node is represented with a Java object. Each edge is associated with source node, sink node, and a boolean data set that indicates a directed or undirected edge.. In the directed boundaries, the source and sink nodes are equal. In an undirected graph, all the edges are strictly undirectional. Node information can be obtained from a graph specification. Adding or removing of nodes can be done by mutators. If any extra additional node or edge occurs, then that action has no effect. HashMap (Java object) is used to represent graph objects[5]. Each node in the graph is a key, mapped to a set of its boundaries together. For example, each edge in the set is indicated by a node A, then A is one of the nodes of that edge. This is the representation of an invariant state. The opposite node of a newly added edge is also taken as a key of the HashMap. The adjacency set of the newly added edge contains a corresponding edge. This invariant enforced by all the mutators of an undirected graph, and a number of valuable accessors provided by this graph are in the interface. One such observation can take an unmodifiable view of a graph. Whether a given node or edge is limited in the graph or not can be checked using this graph interface [6]. In a conceptual basis augmentation, algorithm implementation is parallel. The implementation was done in Java, and Java has built-in geometric awt package, due to the advantage of which feature Java was selected for implementation. The awt package is useful for performing unions and intersections on regions [7]. The simple mobile ad hoc network amplification flow diagram is shown in Figure 1.

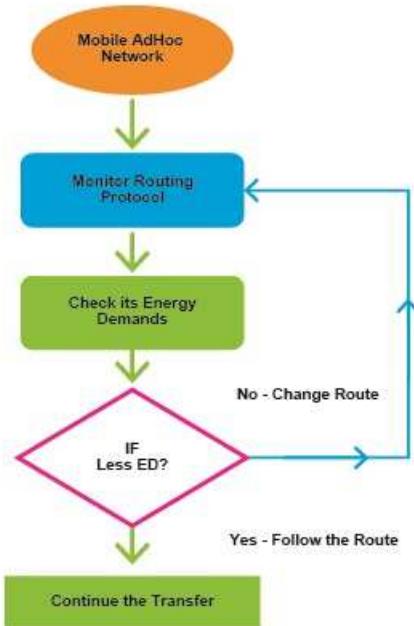


Fig. 1: Amplification Scheme for Mobile Adhoc Network

#### B. Probabilistic Maximum-Drift Algorithm

After configuring the mobile ad-hoc network , the supreme drift was calculated using Probabilistic Max Drift Algorithm. The Probabilistic Max Drift algorithm is used to calculate the maximum drift with the principal of vigor demand. Connected weighted digraph with n apexes numbered as 1 to n and set of boundaries E is represented as the transportation of mobile ad-hoc network with the following properties: The apex without an entering edge is called source, which is represented as 1[8]. The apex without leaving edge is called as sink and represented as n. Edge capacity is represented with the weight  $u_{ij}$  of each directed edge (i, j ). The starting and ending node of a network is taken as source and sink. All the other pieces considered as points in which the drift can be bypassed[9]. In apex, the total value of the incoming must be equal to the total value of the physical energy. This condition is the requirement for the drift conservation. For any intermediate apex I, the amount of energy sent via the edge (i, j ) is represented as  $x_{ij}$  and the drift-conservation requirement can be expressed as shown in the Equation 1.

$$\sum_{j:(j,i) \in E} x_{ji} = \sum_{j:(i,j) \in E} x_{ij} \text{ for } i=1,2,\dots,n-1 \quad (1)$$

Figure 2 shows the mobile ad-hoc network architecture example. The total amount of left and right-hand sides expresses the total in drift and out drift as entering and leaving apex i, respectively. The total out drift from the source is equal to the total in drift into the sink is called the value of the drift,

which is shown in Equation 2.

$$\sum_{j:(1,j) \in E} x_{1j} = \sum_{j:(j,n) \in E} x_{jn} \quad (2)$$

To find a drift-augmenting route for a drift  $x$ , the trails from

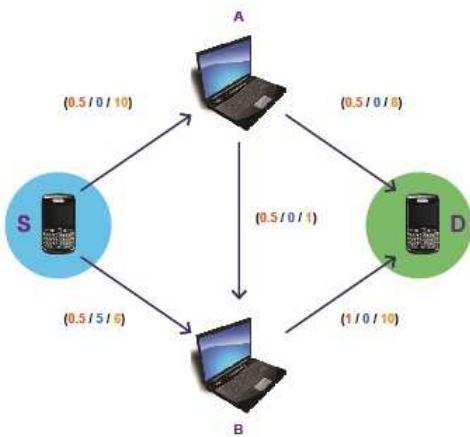


Fig. 2: Example of Mobile Ad-hoc network Architecture

source to sink in the underlying undirected chart with any two consecutive vertices  $i, j$  are needed to be either linked by a directed edge from  $i$  to  $j$  with some positive unused volume  $X_{ij} = u_{ij} - A_{ij}$  (so that it can increase the drift through that edge by up to  $A_{ij}$  units), or linked by a directed edge from  $j$  to  $i$  with some positive drift  $X_{ji}$  (so that it can decrease the drift through that edge by up to  $A_{ji}$  units).

Boundaries of the first type are called forward boundaries, since their tail is listed before their head in the apex list  $1 \rightarrow \dots i \rightarrow j \dots \rightarrow m$ . Boundaries of the second type are called negative directed boundaries because their tail is listed after their head in the route list  $1 \rightarrow \dots i \leftarrow j \dots \rightarrow m$ . The first route  $1 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 5 \rightarrow 6$  is represented as  $(1, 4), (4, 3), (2, 5)$ , and  $(5, 6)$ , which are the positive directed boundaries, and  $(3, 2)$  is the backward edge. For a given drift-augmenting route, let  $X$  be the least of all the unused capacities  $X_{ij}$  of its forward boundaries and all the drifts  $A_{ji}$  of its negative directed boundaries. It is easy to see that if we upsurge the current drift by  $r$  on each forward edge and reduced it by this amount on each negative directed edge, a feasible result could be obtained[10].

#### Algorithm 1 Probabilistic Intelligent Max Drift Longitivity Algorithm

//Implements the Probabilistic Intelligent Max Drift Longitivity Algorithm

//Input: Single source and Single sink network with positive integer capacities  $u_{ij}$  on its boundaries  $(i, j)$

//Output: A maximum drift  $A$

$A_{ij} = 0$  to every edge  $(i, j)$  in the network //Initialization

//Start from source and assign name for source as  $\infty$ ,

//insert the source to the empty priority queue PQ

while (!IsEmpty(PQ))

{

I = GetFront(PQ)

DeleteFront(PQ))

for (i=1 to m) // Positive Directed Edge

for (j=1 to m)

{ if j is unnamed

$X_{ij} \leftarrow u_{ij} - A_{ij}$

if  $X_{ij} > 0$

$n_j \leftarrow \min\{l_i, X_{ij}\}$ ;

}

label j with  $n_j$ ,  $i+$

InsertRear(PQ,j))

for (j=1 to m) // Negative Directed Edge

{

if j is unnamed

if  $A_{ij} > 0$

$n_j \leftarrow \min\{n_j, A_{ji}\}$ ; name j with  $n_j$ ,  $i-$

} InsertRear(PQ,j)) If sink is reached and named

// augmentation begin from back to front

$j \leftarrow m$  //start from the sink,  $j \neq 1$  //the source hasn't been reached

If the second name of apex j is  $i+$

$A_{ij} \leftarrow A_{ij} + n_m$

else

//the second name of apex j is  $i-$

$A_{ji} \leftarrow A_{ji} - n_m$

$j \leftarrow i$ ;  $i \leftarrow$  the apex indicated  $i$ 's second name

reinitialize all apex names and start from source

reinitialize PQ with the source

return A

// the current drift is maximum

here are four possible combinations of positive directed edge and negative directed boundaries incident to apex I as illustrated below.

$\xrightarrow{+r} i \xrightarrow{+r}, \xrightarrow{-r} i \xleftarrow{-r}, \xleftarrow{-r} i \xrightarrow{+r}, \xleftarrow{-r} i \xleftarrow{-r}$

For each of them, the drift-conservation requirement for apex  $i$  will grip after the drift adjustments designated above the

edge arrows. Since  $r$  is the least among all the positive unused volumes on the forward boundaries, all the positive currents on the backward limits the drift-augmenting route and the new movement will satisfy the volume restraints. Finally, adding  $r$  to the drift on the first edge of the supplementing route will increase the value of the drift by  $r$ . Under the assumption that all the edge dimensions are integers,  $r$  will be a positive integer too. Hence, the drift value raises at least by 1 on each iteration of the augmenting-route method. As the value of a maximum drift is restricted above (e.g., by the sum of the capacities of the source boundaries), the augmenting-route way has to stop after a finite number of iterations. The final movement always turns out to be maximal, irrespective of a sequence of supplementing routes.

An application of this algorithm was tested with different configuration setups to achieve maximum drift for optimised Mobile Adhoc network which is shown here in Figures 3a to 3d and 4a to 4c. Figure 3a shows the sample ad-hoc network. In that, the connection link was assigned with probability of selecting particular route and current drift with maximum capacity. In figure 3b, the mobile nodes were assigned with labels. In that, the first label indicates minimum possible drift and the second label indicates drift from particular node. From this configuration, augment route was identified, but maximum drift was not achieved. In Figure 3c configurations, all the labels were erased and the next possible argument route was identified. In Figure 3d, the second argument route was identified, but not maximum drift. Figure 4a, again depicted the erased reflection within the labels and continued to search for further optimised route with maximum drift. In Figure 4b, labels were reassigned and next argument route was identified. In Figure 4c configurations, maximum drift of network route was achieved.

### C. Experiment

In this research, sixteen node testbeds were created and tested in multi-cast experiments. For evaluating the experimental results, a new longevity method was used and compared with the existing methods. Further, vigor utilization in a stationary grid scenario was studied and verified for the accuracy of the grids in topology and membership active scenarios. During the evaluation period, all the grid plans were operated on 2.4 GHz bandwidth and communicated at the capacity of 2 mb/s with the power of 1mw. The WaveLan devices were operated in an ad hoc mode.

For implementing the optimised mobile network, two different systems were used, the details of the systems are as follows :

### D. System Setup

- 1) System Specification : Modern Ad hoc grid setup are Dell i7 laptops with intel core processor and HP Proliant server connected to standard wireless grid.
- 2) Software Specifications: Godern grid was developed on fedora 21 linux kernal.

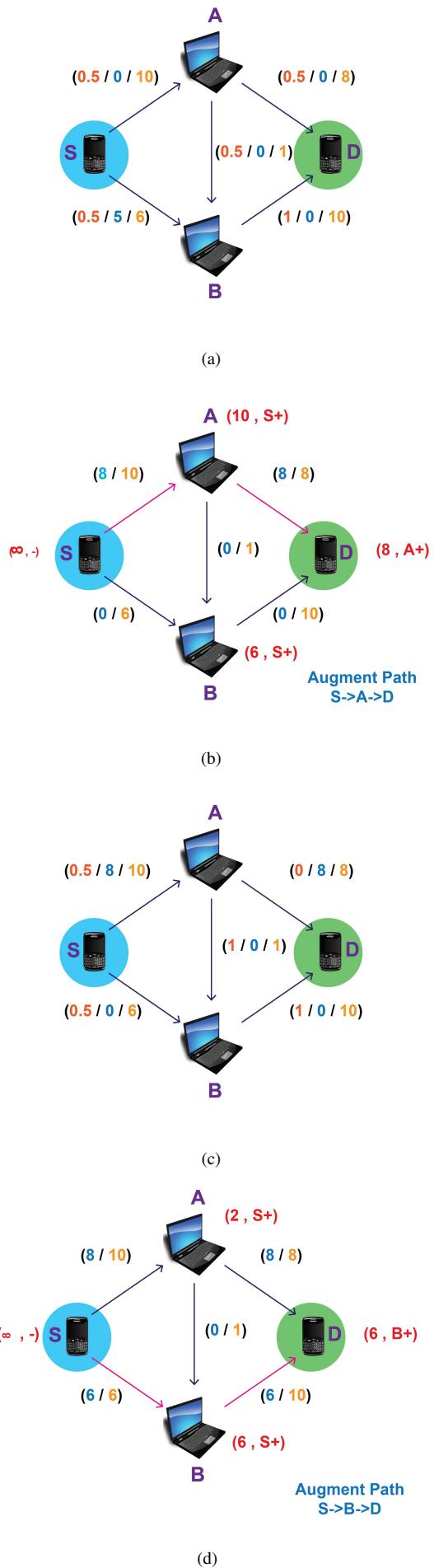


Fig. 3: Proposed Mobile Adhoc Method Implementation setup

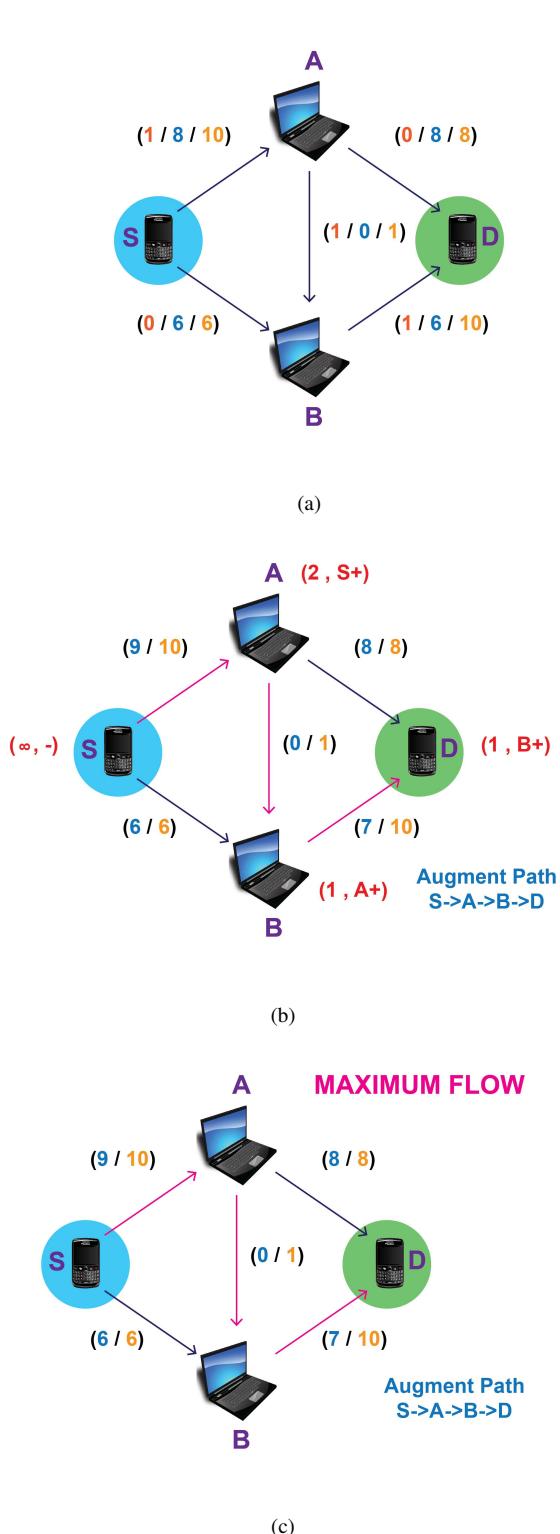


Fig. 4: Proposed Mobile Adhoc Method Implementation setup II

#### IV. RESULTS AND DISCUSSIONS

Figure 5 shows the MANET implementation configuration based on vigor demand, In this, max drift route constructed on demand as standard routine.

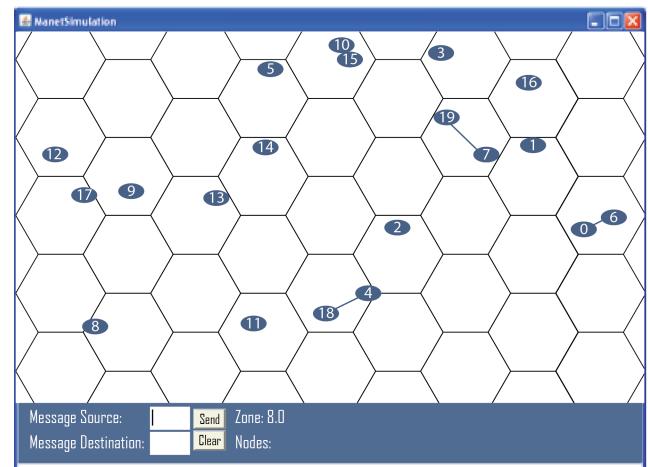


Fig. 5: MANET Energy Demand Java Implementation - Simulation run setup

Figure 6 shows the new energy consumption of two nodes with different routing conditions. It was observed that as the time progressed, node E consumed less energy compared to the other nodes with different routing conditions.

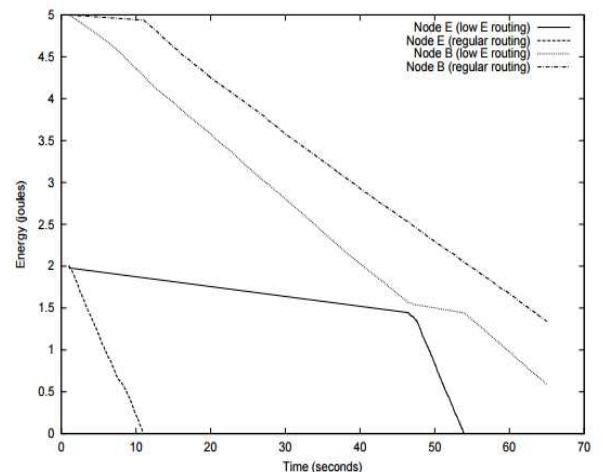


Fig. 6: New Energy Consumption

Figure 7 shows the energy rate required by longevity compared to the existing method. It was noticed that the energy rate required for the longevity method was less compared to the existing Augmenting Max-flow min-cut Algorithm method. It was also observed that, as the nodes increase the energy rate relatively reduced markedly. Hence, it was concluded that, the requirement of green campus could be achieved with the MANET configuration.

## V. CONCLUSION AND FURTHER RESEARCH

In this paper, Probabilistic Max drift algorithm is applied in a dynamic manner for achieving the energy efficiency of ad-hoc networks. The maximum drift algorithm discovers whenever a new node joins/relivies from grid and the grid drift is routed consequently. Peer to peer communication of Mobile Ad hoc network was performed using a dynamically altering topology. Mobile Ad hoc infrastructure was simulated using Java with group mobility, group communication and terrain blockage representations. The obtained results show that the energy consumption reduced by 13 to 24% when tested for an HEI. Also, it is further noticed that the overall throughput increased by 17% compared to the existing methods. Hence, it is concluded that the proposed system can be applicable as one of the key component to any HEI to realise their green campus status.

The research by Asgarali Bouyer11, is focused on cloud computing for an HEI which is more towards overall system based solution. In this paper, an algorithm is applied to reduce the energy requirement, and its efficiency has been proved without any compromise on the network demand. This finding would raise interest for further research on studying on other components that are contributing to the energy requirements including the wired network such as fibre optic connections. Logical and Programming solution for addressing the energy demands in an ICT system is more handy and can lead to the production of new hardware equipment.

## REFERENCES

- [1] P. Basu and J. Redi. Movement control algorithms for realization of fault-tolerant ad hoc robot networks. *IEEE Network*, 18(4),36-44, 2004.
- [2] Vaduvur Bharghavan, Alan Demers, Scott Shenker, and Lixia Zhang. MACAW: a media access protocol for wireless LAN's. In *SIGCOMM '94: Proceedings of the conference on Communications architectures, protocols and applications*, 212-225, 1994.
- [3] M. Maleki, K. Dantu, and M. Pedram. Lifetime prediction routing in mobile ad hoc networks. *IEEE Wireless Communications and Networking*, 2, 1185- 1190, 2003.
- [4] I. Stojmenovic and X. Lin. Power-aware localized routing in wireless networks. *IEEE Transactions on Parallel and Distributed Systems*, 12:1122 - 1133, 2001.
- [5] Robert Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146-160, 1972.
- [6] C.-K. Toh. Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks. *IEEE Communications Magazine*, 39(6), 138-147, June 2001.
- [7] Wei Ye, John Heidemann, and Deborah Estrin. Medium access control with coordinated adaptive sleeping for wireless sensor networks. *IEEE Transactions on Networking*, 12(3):493-506, 2004.
- [8] Sung-Ju Lee, Mario Gerla, and Ching-Chuan Chiang. Ondemand multicast routing protocol. *IEEE WCNC*, 12981302, 1999.
- [9] Sagar Sanghani, Timothy X Brown, Shweta Bhandare, and Sheetalkumar Doshi. Ewant: The emulated wireless ad hoc network testbed. In *IEEE WCNC2003*, volume 3, 1844 1849, 2003.
- [10] Lusheng Ji, Mary Ishibashi, and M. Scott Corson. An approach to mobile ad hoc network protocol kernel design. *IEEE WCNC*, 13031307, 1999.
- [11] Asgarali. B, Bahman.(2014) A The Necessity Of Using Cloud Computing In Educational System. Elsevier Computer Standards Interfaces.

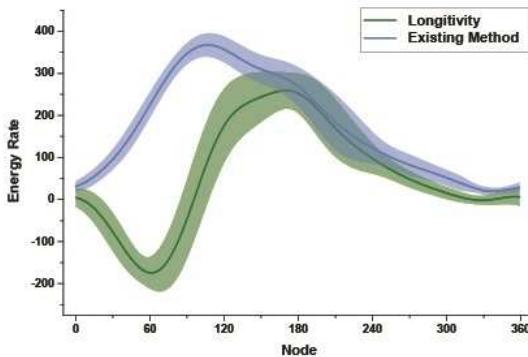


Fig. 7: Comparison of Energy Demand for green Campus

Figure 8 shows the time complexity comparison of the proposed algorithm with the existing method. The average execution time required for existing method was 47.66 ms and with the proposed longevity method was only 26.83 ms. Hence, it was observed that as the number of nodes increased the proposed longevity algorithm execution time requirement was not increased randomly as compared to the existing Augmenting Max-flow min-cut Algorithm method. Further, from the obtained results it can be concluded that the longevity algorithm provides higher efficiency.

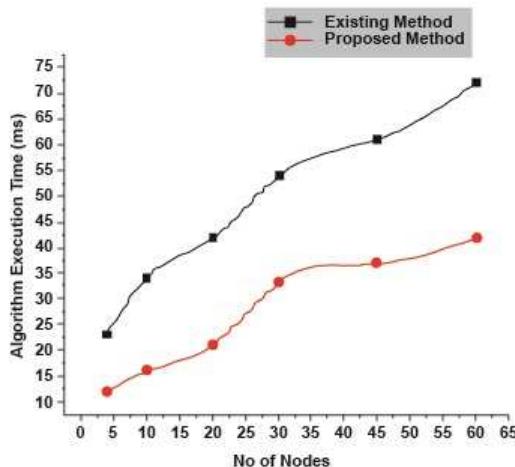


Fig. 8: Time Complexity Comparison

# NeTMids: Neighbor Node Trust Management Based Anomaly Intrusion Detection System for Wireless Sensor Networks

Syed Muhammad Sajjad<sup>1</sup>, Muhammad Yousaf<sup>2</sup>

*Riphah Institute of Systems Engineering,  
Riphah International University,  
Islamabad, Pakistan.*

<sup>1</sup> muhammad.sajjad@riu.edu.pk

<sup>2</sup> muhammad.yousaf@riu.edu.pk

**Abstract**—Timely detection of anomalous activities in Wireless Sensor Network is critical for the smooth working of the network. This paper presents an intrusion detection technique based on the calculation of trust of the neighboring nodes. In the proposed IDS, each node observes the trust level of its neighboring nodes. Based on these trust values, neighboring nodes may be declared as trustworthy, risky or malicious. Trustworthy nodes are recommended to the forwarding engine for packet forwarding purposes. The proposed scheme successfully detects Hello Flood Attack, Jamming Attack and Selective Forwarding Attack by analyzing the network statistics and malicious node behavior. The simulation results show that network performs better when neighbor node trust management based anomaly detection technique is in place.

**Index Terms**—Wireless Sensor Network, Intrusion Detection System, Trust management, Risk, Trusted Node

## I. INTRODUCTION

Wireless Sensor Network (WSN) is an emerging notion and has gained enormous diligence of the research community due to increasing modernization of the technology. WSN is a self-organized network of large number of low power and low cost sensor nodes [1]. These sensor nodes are light-weight and movable devices having capabilities of sensing, communicating and processing the information to the targeted users. They have limited transmission range and communicate directly with nodes lying within its transmission range. Communication with a far end node is performed via intermediate nodes. Sensor networks are susceptible of exterior and interior outbreaks[2] [3] [4].

Sensor nodes often lack ability of dealing a tough attacker owing to its resource constrained nature. In this case secondary level of defense often called intrusion detection system is required [5] [6] [7]. Exploitation efforts by the attacker can be detected with the help of intrusion detection system. The confidence and faith of a node in the ability, consistency and trustworthiness of other nodes is termed as trust [8]. Trust based on direct observation of a node is also called direct trust or first-hand information. A node's observation and opinion about other nodes based on their earlier performances in an explicit perspective on a certain period of time is termed as

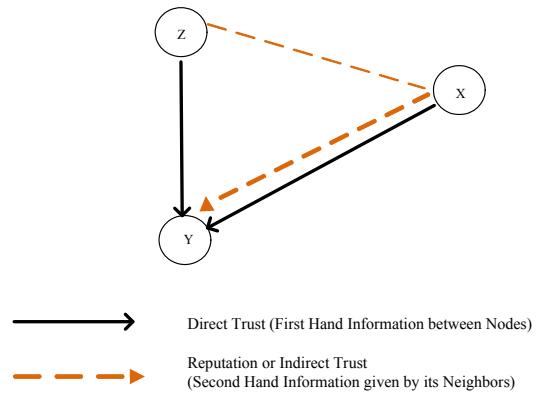


Fig. 1. Direct and Indirect Trust

reputation [9] [10] [11]. Reputation is also called indirect trust or second hand information [12] [13] [14]. In figure 1 a node's direct and indirect trust (reputation) is illustrated. Direct observation of node 'Y' about node 'X' is direct trust of node 'Y' on node 'X'. The trust values of Node 'Z' given to node 'Y' by its neighboring node 'X' is called indirect trust or second hand information or reputation. This paper elaborates a neighbor node trust calculation and evaluation based anomaly intrusion detection technique. Remaining of the paper is structured as follows: Section II covers the related work. In section III, we present the adversary model and describe the attacks which the proposed NeTMids successfully detects. Section IV provides the details of the proposed NeTMids. It presents discussion about the system model, its components and blocks of the proposed NeTMids. Results are discussed in section V. Finally, section VI concludes the paper.

## II. RELATED WORK

The idea of trust computation based intrusion detection systems originates with the design of an IDS by Wang et al. [15] for mobile ad hoc networks (MANETs) based on trust variations and chain of evidence. The assessment of the network node is carried out periodically. A trust assessment

and reputation interchangeability based intrusion detection method is offered by Ebinger et al. [16]. The combination of reputation, trust and confidence with trustworthiness causes an improvement in the detection of intrusion [17][18]. Various trust management mechanisms [19][20][21] have also been presented for WSN. The primary objectives of these techniques include security of systems and reliability of the information [22][23][24]. A trust based IDS is proposed by [25] for cluster WSN. Cluster head (CH) performs the trust calculation and evaluation of nodes present in the cluster. Honesty (social trust) and supportiveness as well as energy consumptions (quality of service trust) are the assessment metrics used by the authors for the identification of malicious activity. Base station evaluates the trust level of cluster head (CH). Fuzzy logic and theory in combination of evidence theory based IDS is presented in [26]. Node's behavior is observed and malicious nodes are identified by the validation of normal process. An IDS for the localization and detection of the anomalies in WSN is presented by [27]. The decision about the adversary is achieved by taking inference from the calculation and observation of the specially designated measurement nodes. Malicious node detection based on the neighbor node calculation is carried out in [28][29][30]. In [28] information fabrication attack is detected. Spatial correlation is used in order to detect anomalous activities in neighboring nodes. In [29] statistical distribution and high computational complexity of the nodes are the disadvantages of IDS. In [30] though, the cooperation between nodes makes this IDS robust, the main drawback is overhead due to communication. WSN requires a flexible, light weight and an effective IDS for the identification of internal malicious nodes. Therefore, a lightweight IDS is required. We present, in this paper, a lightweight neighbor node trust calculation and evaluation based anomaly intrusion detection technique.

### III. ADVERSARY MODEL AND ATTACKS

Based on the security requirements, attacks on network are categorized as interruption, interception, modification and fabrication. Interruption involves breaking the availability of the target network. An attack aiming to compromise the confidentiality of network is often termed as interception. The integrity of the message is endangered in modification attack. In Fabrication, attack is intended to threaten the authentication. Mote class attacker possesses the ability of jamming the entire network. A laptop class Attacker have extra proficient central processing unit (CPU), radio transmitter having high power, efficient battery and sensitive antenna. Thus, they are far more destructive than Mote-class Attacker. An attacker that originates a sharp increase in Denial of Service (DoS) and causes greater depletion of network resource by means of false data injection to the network is often called external attacker. Contrarily, a node which has attained network access and can perform malicious activity is termed as insider attacker. In this work, we assume an attacker having the capabilities of performing Jamming, Hello flood and Selective Forwarding Attacks.

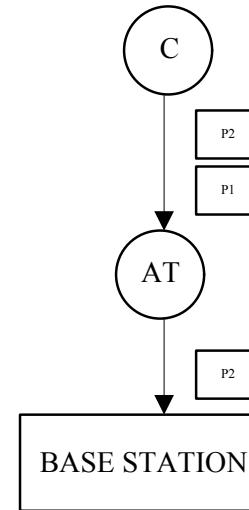


Fig. 2. Selective Forwarding Attack by Node 'AT'

#### A. Jamming Attack

The intentional interference performed on physical reception and transmission in any communication is termed as Jamming. Basic aim of this attack is Denial of Service (DoS) attack. Node considers its communication medium busy all the time. Jamming is performed by an attacker called jammer [31]. There are four types of Jammers [32]. If motes are unable to find medium idle due to the continuous transmission of radio signal by an adversary, the attacker is termed as constant jammer. Regular injection of packets is performed by deceptive jammer. A random jammer arbitrarily switches between jamming and sleeping so as to save its energy. A reactive jammer prefers jamming when it feels there is a communication process on the channel.

#### B. Selective Forwarding Attack

In multi-hop networks, there is a prime assumption that participating nodes will correctly and honestly forward the packets to the required destination. A malicious node may

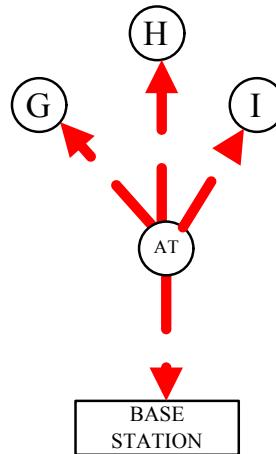


Fig. 3. HELLO Flood Attack by Node 'AT'

not forward any message and simply drop them. This attack is a special attack in which selective packets are dropped. Consequently, the attack may be classified as a special form of Black-hole attack wherein the malicious node drops all the packets. The behavior of the malicious node causing black-hole attack may be detected and notified by the neighboring nodes in the whole network. In Selective Forwarding Attack, the malicious node selectively drops packets so as to minimize suspiciousness of the detecting node. This attack has worse impact when occurred in nodes lying near to base station. In this way, wireless sensor network conveys wrong information about the observed environment. In figure 2, an attacker node 'AT' selectively drops packet P1.

### C. HELLO Flood Attack

Many protocols use HELLO message for route discovery purpose. Nodes receiving this HELLO message assume that this message is from a node of normal communication range. A laptop-class attacker having high transmission power can convince all the other nodes that he is their neighboring node, as a result of which all the neighboring nodes will respond to the HELLO message and will waste their energy. Nodes consider that the attacker is within one hop communication range. In case of any low cost route advertisement by the attacker, nodes forward their messages towards the attacker. In figure 3, the attacker node 'AT' floods all his neighbor nodes with HELLO packet.

## IV. THE PROPOSED NETMIDS

Block diagram of the proposed intrusion detection system is shown in figure 4 and detailed flow chart of the same is shown in figure 5. The proposed intrusion detection system has a trust manager, which manage the direct and indirect trust (reputation) of a node. The behavior classifier classifies the behavior of the node as attacker, trustworthy and risky based on the trust values and calculation obtained from the trust manager. In case of the trustworthy behavior, the observed node is recommended to the forwarding engine for packet forwarding. When behavior of the observed node is identified as risky, its risk factor is evaluated and updated. If the observing node is willing to take risk, it recommends the observed node having risky behavior to the forwarding engine for forwarding. This status of the observed node is saved in the recommendation data base. If the observing node does not want to take risk, it stores the risk factor of the observed node in recommendation data base. In case of attack behavior, the attack classifier distinguishes attack pattern based on the calculation described in the following subsections. The observed node is declined for forwarding purpose. The status of the observed nodes is saved in the recommendation data base.

### A. System Model

In the proposed IDS, a node  $y_0$  calculates the level of trust of its neighboring nodes. The neighbors of  $y_0$  is a set of nodes having one hop contact with node  $y_0$  and are represented

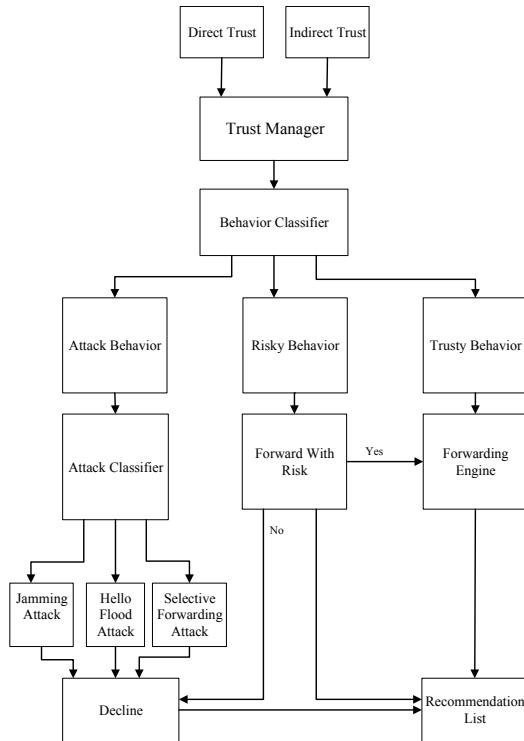


Fig. 4. Block Diagram of Proposed IDS

as  $N_b(y_0)=\{y_1, \dots, y_n\}$ . Every node  $y_i$  possesses different attributes. The set of attributes of node  $y(i)$  can be denoted as  $A_{y_0} = \{a_1, \dots, a_n\}$ . The activity of the node  $y_i$  is observed by the sensor node  $y_0$  by observing its individual attributes. If node  $y_0$  observes its neighboring nodes  $N_i(y_0)=\{y_1, \dots, y_i\}$  it stores the set of the corresponding attribute vectors  $A_{N_b(y_0)} = \{A_{y_1}, \dots, A_{y_i}\}$ .

More precisely the attributes of any node include Received Signal Strength, Packet Sending Rate, Control Packet Gen-

TABLE I  
MATHEMATICAL NOTATIONS

Notation	Description
$N_b(y)$	Set of neighbor nodes of node $y_0$
$RSS(y)$	Received signal strength of node $y_i$
$PGR(y)$	Control Packet generation rate of node $y_i$
$PRR(y)$	Packet receiving rate of node $y_i$
$PLR(y)$	Packets delivery ratio of node $y_i$
$PDR(y)$	Packet dropping rate of node $y_i$
$PFR(y)$	Packet forwarding rate of node $y_i$
$PFO(y)$	Packet forwarding ratio
$TR$	Required trust of node $y_i$
$T_C(y)$	Current trust of node $y_i$
$T_D(y)$	Direct trust of node $y_i$
$T_I(y)$	Indirect trust about the node $y_i$
$R_F(y)$	Risk Factor

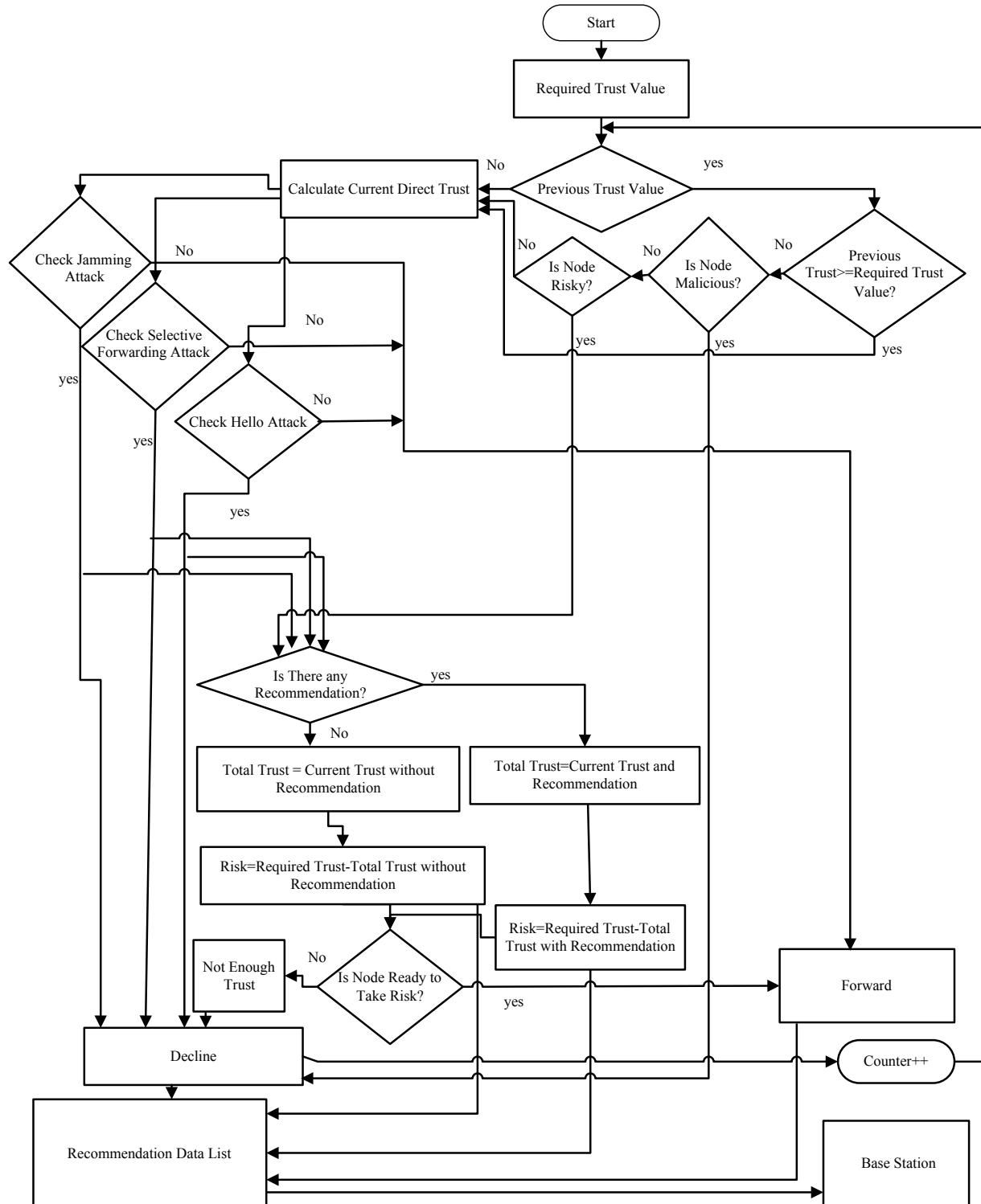


Fig. 5. Flow Chart of Proposed IDS

erating Rate, Packets Delivery Ratio, Packet Dropping Rate, Packet Forwarding Rate and Packet Acknowledgment Rate. The amount of power received in any radio signal is termed as Received Signal Strength. The Received Signal Strength of the node  $y_i$  observed by the node  $y_0$  is represented as  $RSS(y)$ . A

node  $y_i$  is considered malicious if its received signal strength is higher than the average of received signal strength of its neighbors. In this case the node  $y_0$  is considered to have undergone jamming attack. Control Packet Generation Rate is the number of control packets generated in a specific interval

of time.  $CPG(y)$  is the Packet Generation Rate of node  $y_i$  monitored by the node  $y_0$ . A node is considered malicious if it generates high number of control packets than the average of control packets generated by its neighbors. In this case, the node  $y_0$  is considered to have undergone a Hello Flood Attack. Packet Receiving Rate is the total number of packets received in a specific period of time.  $PRR(y)$  is the Packet Receiving Rate of node  $y_i$  monitored by the node  $y_0$ . The ratio of the packets successfully conveyed to the designation node in comparison to the number of packets sent by a sender is called Packets Delivery Ratio.  $PLR(y)$  is the Packets Delivery Ratio of mote  $y_i$  monitored by the mote  $y_0$ . If packets are directed to certain node via some intermediate nodes and the intermediate nodes drop some packets then the ratio between packet sent to intermediate node and the packets conveyed by the intermediate node to the destination node is termed as Packet Dropping Rate.  $PDR(y)$  is the Packet Dropping Rate of node  $y_i$  monitored by the node  $y_0$ . In a multi-hop scenario, a node forwards packets of its neighbors. The rate of packet receiving and its subsequent forwarding to its destination by a node is termed as Packet Forwarding Rate.  $PFR(y)$  is the Packet Forwarding Rate of node  $y_i$  monitored by the node  $y_0$ . A node  $y_i$  is said to be involved in Selective Forwarding Attack, if its packets forwarding rate is much less than the packets forwarding rate of its neighbors.

#### B. Detection of Jamming Attack

Let  $RSS(y)$  is the Received Signal Strength of node  $y_i$  observed by node  $y_0$  during time interval  $t$ . The total Received Signal Strength of node  $y_i$  observed by node  $y_0$  at any arbitrary time interval  $t$  is the direct trust of the observing node. Mathematically

$$TD_t(y_{0i}) = \frac{1}{RSS_t(y_{0i})} \quad (1)$$

Reputation is the values of the Received Signal Strength of node  $y_i$  given by its neighboring nodes to the observing node  $y_0$ . The net reputation is the average of these values. Mathematically

$$TI_t(y_{01}) = \frac{1}{avg[RSS_t(y_{21}), RSS_t(y_{31}), \dots, RSS_t(y_{n1})]} \quad (2)$$

$$TI_t(y_{01}) = \frac{1}{(1/n) \sum RSS_t(y_{n1})} \quad (3)$$

Current Trust value is then given by

$$TC_t(y_{01}) = \frac{1}{average[TD_t(y_{01}), TI_t(y_{01})]} \quad (4)$$

The Required Trust value is given in the following equation.

$$TR_t(y_0) = \frac{1}{(1/i) \sum TC_t(y_{0i}) + k} \quad (5)$$

K is constant. At any instant  $i$  if the current trust of a node is less then the required trust value, node is suffering from jamming Attack. Mathematically

$$TC_t(y_{01}) < TR_t(y_0) \quad (6)$$

#### C. Detection of Selective Forwarding Attack

Packets may be dropped due to congestion in the network or by the malicious node intentionally. The packets forwarded successfully is the ratio of the packet forwarding rate  $PFR(y)$  and packet receiving rate  $PRR(y)$ . The packet forwarding ratio of node  $y_i$  at any instant  $t$  observed by node  $y_0$  is given as  $PFO(y_i) = \frac{PFR(y)}{PRR(y)}$ . Selective Forwarding Attack is detected using multi-path packet forward and multi-hop activity sharing using packet sequence number. The number of packets received by a node from path one is compared with the packets received via path two. The total number of packets sent by the originator is learned from sequence number. The total packet forwarding ratio of node  $y_1$  observed by node  $y_0$  at any arbitrary time interval  $t$  is the current direct trust of the observing node. Mathematically  $TD_t(y_{01}) = PFO(y_1)$ . Reputation is the values of the packets forwarding ratio of node  $y_i$  given by its neighboring nodes to the observing node. The net reputation is the average of all these values. Mathematically

$$TI_t(y_{01}) = (1/n) \sum TD_t(y_{n1}) \quad (7)$$

The current trust value is then given by

$$TC_t(y_{01}) = avg[TD_t(y_{01}), TI_t(y_{01})] \quad (8)$$

The Required Trust value is given in the following equation.

$$TR_t(y_0) = (1/i) \sum TC_t(y_{0i}) + k \quad (9)$$

K is constant. At any instant  $i$  if the current trust of a node is less then the required trust value, node is suffering from jamming Attack. Mathematically

$$TC_t(y_{01}) < TR_t(y_0) \quad (10)$$

#### D. Detection of HELLO Flood Attack

let  $PGR(y)$  is the total control packets generating rate of node  $y_i$  observed by node  $y_0$  during time interval  $t$ . The total control packets generating rate of node  $y_i$  observed by node  $y_0$  at any arbitrary time interval  $t$  is the direct trust of the observing node. Mathematically  $TD_t(y_i) = \frac{1}{PGR_t(y_i)}$ . Reputation is the values of the total control packets generating rate of node  $y_i$  given by its neighboring nodes to the observing node  $y_0$ . The net reputation is the average of these values. Mathematically  $TI_t(y_{01}) = \frac{1}{avg[PGR_t(y_{21}), PGR_t(y_{31}), \dots, PGR_t(y_{n1})]}$

$$TI_t(y_{01}) = \frac{1}{(1/n) \sum PGR_t(y_{n1})} \quad (11)$$

Current Trust value is then given by

$$TC_t(y_{01}) = \frac{1}{average[TD_t(y_{01}), TI_t(y_{01})]} \quad (12)$$

The Required Trust value is given in the following equation.

$$TR_t(y_0) = \frac{1}{(1/i) \sum TC_t(y_{0i}) + k} \quad (13)$$

TABLE II  
SIMULATION PARAMETERS

Network Size	200mX200m
Nodes	60,80,100,120,140,160,180,200
Duration	5000 Rounds
Rx Power	50mW
Tx Power	50mW
Packet Size	800 bits
Initial Energy	0.5 Joules
Malicious Nodes	5,10,15
Transmission Radius	10 m
Location	Random Distributed

K is constant. At any instant  $i$  if the current trust of a node is less than the required trust value, node is suffering from jamming Attack. Mathematically

$$TC_t(y_{0i}) < TR_t(y_0) \quad (14)$$

#### E. Detection of Trustworthy (Good) Nodes

A node is said to be trustworthy if its current direct trust value is greater than required trust value.

#### F. Risky Factor

The value of risk is given as

$$RF_t(y_{0i}) = TR_t(y_0) - TC_t(y_{0i}) \quad (15)$$

#### G. Storage of Node Status's For Future Use (Reputation)

Recommendation Data Base stores the status of the node. On the basis of calculation, a node may be found malicious, trustworthy or risky. These statistics are used in the future interaction with the nodes. A trustworthy node is recommended for interaction, a malicious node is declined, while decision about packet forwarding through risky node is made if the node intending to send data is willing to take risk. In that case, a node having lowest risk is selected for forwarding purposes.

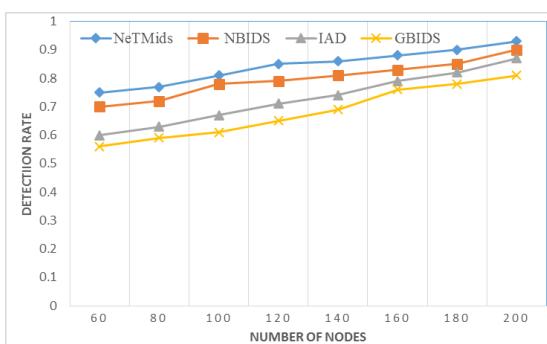


Fig. 6. Detection rate per 100 rounds along varying network size

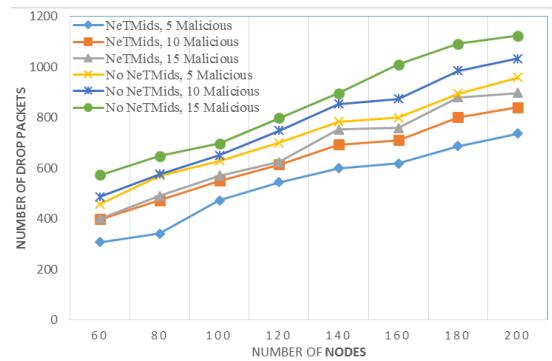


Fig. 7. Packets drops per 100 rounds along varying network size

#### H. Forwarding Decision Based On Node's Status

After the successful determination of the node status as malicious, trustworthy or risky, decision about the packet forwarding through any neighbor node is taken by the packet sending node. The criteria for packet forwarding is, the selection of safest path rather than selecting shortest path. If there isn't any good node in neighbor list for packet forwarding, node having minimum risk is selected for packet forwarding. Mathematically

$$\text{Min}[RF_t(y_{0i})] \quad (16)$$

#### V. RESULTS AND DISCUSSION

The proposed intrusion detection system is implemented using MATLAB. Nodes are randomly deployed in an area of 200 x 200 square meters. Simulation parameters are shown in the table 2.

Simulations are performed for network size of 60, 80, 100, 120, 140, 160, 180, and 200 nodes. For each network size, per 100 round results are discussed for the following four performance metrics.

- Detection Rate
- Packets Drop Ratio (During its Movement to Next Hop)
- Successful Packet Delivery Ratio to Base Station
- Number of Dead Nodes (Network Life Time)

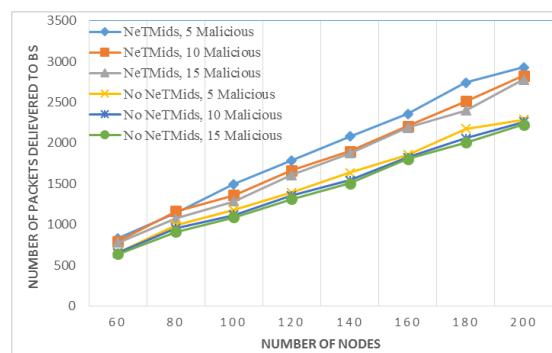


Fig. 8. Packets delivered to base station per 100 rounds along varying network size

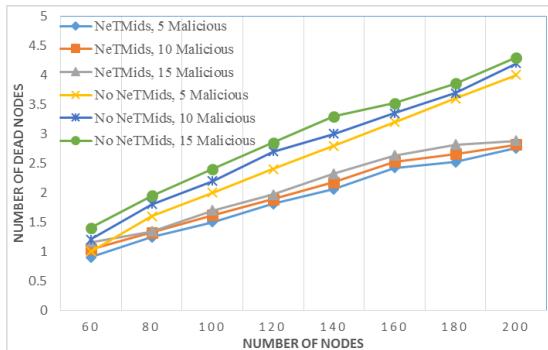


Fig. 9. Numbers of dead nodes per 100 rounds along varying network size

#### A. Detection Rate

The detection rate of the proposed NeTMids is compared with the detection rate of [26][27] and [28]. Figure 6 shows that the detection rate of the proposed NeTMids is better than the detection rate of [26][27][28], due to the fact that the proposed IDS distinguishes observed nodes as trustworthy, risky and malicious based on their trust. Also, observing nodes do not solely depend on the direct observation of node's reputation but it takes into consideration their current trust.

#### B. Packets Drop Ratio (During Its Movement to Next Hop)

Packets dropping behavior of nodes is observed with and without NeTMids per hundred rounds in the presence of five, ten and fifteen malicious nodes respectively. As shown in figure 7, when NeTMids is deployed, the packets dropping rate is minimal owing to the IDS property of timely detecting and reporting the packet dropping nodes. It is notable that the number of packets drop increases with the increase in number of malicious nodes in the network.

#### C. Packet Delivery Ratio

The number of packets successfully delivered to base station are investigated per hundred rounds along network size of 60, 80, 100, 120, 140, 160, 180 and 200 nodes. As shown in figure 8, packet delivery increases with the increase in network size. Also, with the proposed NeTMids and in the presence of five, ten and fifteen malicious nodes, the packets delivery is better than the packets delivered to base station in the absence of the proposed NeTMids. The reason is based on observed trust values and subsequent determination of observing node behaviour, malicious nodes are reported and not recommended for forwarding packets.

#### D. Network Lifetime (Detection Of Dead Nodes)

Due to different kind of attacks on network, specially jamming attack exhaust the energy resources of the network. Timely detection of these attacks and the subsequent exclusion of the attacker node is compulsory for the smooth functioning of network. Network life time increases with the proposed NeTMids as the nodes originating energy consuming attacks are identified timely and are not used for forwarding purposes as shown in Figure 9.

## VI. CONCLUSION

This paper proposed an intrusion detection technique for WSN based on the principle that nodes observe the activities of their neighboring nodes and report the anomalous behaviour. Based on these observations and reports, a node can be declared as trustworthy, risky or malicious. The proposed NeTMids solution successfully detects hello flood, jamming and selective forwarding attacks in WSN. The proposed intrusion detection system can be further extended to the IPv6 connected wireless sensor networks which can further be extended to the Internet of Things (IoTs). Simulation results show that detection rate of the proposed IDS is better as compared to the earlier proposed solutions. Packet drop ratio with the proposed IDS is less as compared to the earlier IDS due to the fact that malicious nodes are timely identified and removed from the network, consequently resulting in an increased packets delivery ratio. Network life time is also extended with the proposed IDS as malicious nodes, originating hello flood attack and causing battery drainage, are timely identified and removed from the network.

## REFERENCES

- [1] C. Y. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.
- [2] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures," *Ad hoc networks*, vol. 1, no. 2, pp. 293–315, 2003.
- [3] D. Martins and H. Guyennet, "Wireless sensor network attacks and security mechanisms: A short survey," in *Network-Based Information Systems (NBiS), 2010 13th International Conference on*. IEEE, 2010, pp. 313–320.
- [4] K. Xing, S. S. R. Srinivasan, M. Jose, J. Li, X. Cheng *et al.*, "Attacks and countermeasures in sensor networks: a survey," in *Network security*. Springer, 2010, pp. 251–272.
- [5] K. Ioannis, T. Dimitriou, and F. C. Freiling, "Towards intrusion detection in wireless sensor networks," in *Proc. of the 13th European Wireless Conference*, 2007, pp. 1–10.
- [6] A. H. Farooqi and F. A. Khan, "Intrusion detection systems for wireless sensor networks: A survey," in *Communication and networking*. Springer, 2009, pp. 234–241.
- [7] Z. S. Bojkovic, B. M. Bakmaz, and M. R. Bakmaz, "Security issues in wireless sensor networks," *International journal of Communications*, vol. 2, no. 1, pp. 106–115, 2008.
- [8] M. Momani, "Bayesian methods for modelling and management of trust in wireless sensor networks," Ph.D. dissertation, University of Technology, Sydney, 2008.
- [9] F. Azzedin and M. Maheswaran, "Evolving and managing trust in grid computing systems," in *Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on*, vol. 3. IEEE, 2002, pp. 1424–1429.
- [10] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in p2p networks," in *Proceedings of the 12th international conference on World Wide Web*. ACM, 2003, pp. 640–651.
- [11] X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: a survey," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 2, pp. 52–73, 2009.
- [12] R. Feng, X. Xu, X. Zhou, and J. Wan, "A trust evaluation algorithm for wireless sensor networks based on node behaviors and ds evidence theory," *Sensors*, vol. 11, no. 2, pp. 1345–1360, 2011.
- [13] H. Deng, Y. Yang, G. Jin, R. Xu, and W. Shi, "Building a trust-aware dynamic routing solution for wireless sensor networks," in *2010 IEEE Globecom Workshops*. IEEE, 2010, pp. 153–157.

- [14] N. Poolsappasit and S. Madria, "A secure data aggregation based trust management approach for dealing with untrustworthy motes in sensor network," in *2011 International Conference on Parallel Processing*. IEEE, 2011, pp. 138–147.
- [15] F. Wang, C. Huang, J. Zhao, and C. Rong, "Idmtm: A novel intrusion detection mechanism based on trust model for ad hoc networks," in *22nd International Conference on Advanced Information Networking and Applications (aina 2008)*. IEEE, 2008, pp. 978–984.
- [16] P. Ebinger and N. Bißmeyer, "Terec: Trust evaluation and reputation exchange for cooperative intrusion detection in manets," in *Communication Networks and Services Research Conference, 2009. CNSR'09. Seventh Annual*. IEEE, 2009, pp. 378–385.
- [17] S. M. Sajjad, S. H. Bouk, and M. Yousaf, "Neighbor node trust based intrusion detection system for wsn," *Procedia Computer Science*, vol. 63, pp. 183–188, 2015.
- [18] S. M. Sajjad and M. Yousaf, "Security analysis of ieee 802.15. 4 mac in the context of internet of things (iot)," in *Information Assurance and Cyber Security (CIACS), 2014 Conference on*. IEEE, 2014, pp. 9–14.
- [19] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava, "Reputation-based framework for high integrity sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 4, no. 3, p. 15, 2008.
- [20] K. Liu, N. Abu-Ghazaleh, and K.-D. Kang, "Location verification and trust management for resilient geographic routing," *Journal of Parallel and Distributed Computing*, vol. 67, no. 2, pp. 215–228, 2007.
- [21] R. A. Shaikh, H. Jameel, B. J. d'Auriol, H. Lee, S. Lee, and Y.-J. Song, "Group-based trust management scheme for clustered wireless sensor networks," *IEEE transactions on parallel and distributed systems*, vol. 20, no. 11, pp. 1698–1712, 2009.
- [22] U. E. Tahta, S. Sen, and A. B. Can, "Gentrust: A genetic trust management model for peer-to-peer systems," *Applied Soft Computing*, vol. 34, pp. 693–704, 2015.
- [23] A. Ahmed, K. A. Bakar, M. I. Channa, K. Haseeb, and A. W. Khan, "A trust aware routing protocol for energy constrained wireless sensor network," *Telecommunication Systems*, vol. 61, no. 1, pp. 123–140, 2016.
- [24] S. Che, R. Feng, X. Liang, and X. Wang, "A lightweight trust management based on bayesian and entropy for wireless sensor networks," *Security and Communication Networks*, vol. 8, no. 2, pp. 168–175, 2015.
- [25] F. Bao, R. Chen, M. Chang, and J.-H. Cho, "Trust-based intrusion detection in wireless sensor networks," in *2011 IEEE International Conference on Communications (ICC)*. IEEE, 2011, pp. 1–6.
- [26] R. Wu, X. Deng, R. Lu, and X. Shen, "Trust-based anomaly detection in wireless sensor networks," in *2012 1st IEEE International Conference on Communications in China (ICCC)*. IEEE, 2012, pp. 203–207.
- [27] S. Zheng and J. S. Baras, "Trust-assisted anomaly detection and localization in wireless sensor networks," in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*. IEEE, 2011, pp. 386–394.
- [28] A. Stetsko, L. Folkman, and V. Matyas, "Neighbor-based intrusion detection for wireless sensor networks," in *Wireless and Mobile Communications (ICWMC), 2010 6th International Conference on*. IEEE, 2010, pp. 420–425.
- [29] F. Liu, X. Cheng, and D. Chen, "Insider attacker detection in wireless sensor networks," in *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*. IEEE, 2007, pp. 1937–1945.
- [30] Y. WANG and G. LI, "A group-based intrusion detection scheme in wireless sensor networks," *CHINESE JOURNAL OF SENSORS AND ACTUATORS*, pp. 879–881, 2009.
- [31] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2005, pp. 46–57.
- [32] W. Xu, K. Ma, W. Trappe, and Y. Zhang, "Jamming sensor networks: attack and defense strategies," *IEEE network*, vol. 20, no. 3, pp. 41–47, 2006.

# NOVEL HYBRID IMAGE ENCRYPTION (64-Bit) BASED ON RUBIK CUBE AND BLOCK CIPHER

*Jasdeep Singh Chauhan*

*Student*

*Dept. of CSE Rayat Bahra Campus  
Ropar*

*Amanpal Singh Rayat*

*Asst. Professor*

*Dept. of CSE, Rayat Bahra Campus  
Ropar*

**Abstract:** Cryptographic Encryption is a method, for the protection of useful information so that only those for whom it is intended can read and process it. Numerous applications are there which require the rapid and strong security against the unauthorized users. For example, securing Military related information, securing sensitive online transactions, securing online transmission of data for real time applications like stock market apps, electronic mails or data transmission of social applications and online personal photograph albums like applications demand for the high security as these are stored and transmitted throughout the internet. The image Encryption is one of the techniques used for alteration of the images into faint form so that the image cannot be seen by the prohibited person. In this paper we explore the Novel Hybrid technique to encrypt image by following the concept of Rubik Cube encryption phenomenon (stream cipher) and combine it with block cipher.

**Index Terms:** - Novel Hybrid, Encryption, Decryption, Rubik Cube, Symmetric key cryptography, Secure Force Algorithm, secret key.

## I. INTRODUCTION

The use of multimedia data such as digital images, videos, audios etc. is increasing with the continuing growth in information technology. Such applications are like video conversation, online photograph albums, imaging systems and so forth. Nowadays, these applications play a vital role towards the various aspects of our daily life, including learning, business and for personal usage.

Digital data Image encryption algorithms try to convert original data image to another form of image that is hard to understand and recognize. One can say in other word, to keep the image confidential between authorized users, it is essential that nobody (un-authorized user) could get to know the content without a key for decryption. Moreover, special and reliable security in Storage and transmission of digital images are needed in many applications, such as cable-TV, online personal photograph album, medical imaging systems, military image communications and confidential video conferences, etc. In order to fulfill such a task, many image encryption methods have been proposed.

Image Encryption is generally the term used for translation of an original figure to the ciphered mode when converting the original figure from its ciphered

form is known as Image Decryption. However, there are many schemes included for enciphering the data includes both textual data and digital data. On the basis of way of encryption carried out, there are two types of cryptographic encryptions:

**Secret key cryptography:** Secret key cryptography is the set of steps in cryptography to carry out encryption with single key called secret key. The whole process of encryption is based on just one key. In this type of cryptographic encryption, both the sender and receiver are known to same secret key. The sender uses this key to encrypt the input data and receiver on the other hand uses the same key for decryption. This is also known as symmetric key encryption.

**Public key cryptography:** In this type of cryptographic encryption there are two keys to carry out the whole process. The first key, also called public key, is used by sender to encrypt the message and is done by easy set of steps whereas another key is used for decryption is known to receiver only. This type of encryption is also known as asymmetric encryption.

The use of cryptographic techniques plays the great role in carrying out transmission of data over network with security. The cryptographic techniques need some set of rules called algorithms that are publically defined yet considered to be secure. Now a day almost every kind of data is transmitted over the internet that includes both textual and digital data, so security plays great role in that to ensure information security and safety. There are many algorithms available for textual data encryption as textual data encryption is far easy and fast as compared to image or video data encryption as in case of multimedia data the bits are more dependent to each other and hard to process it with textual encryption.

The security of images or one can say digital data is the major concern of this paper. Traditional image encryption algorithms such as private key encryption algorithms (DES and AES), public key algorithms such as RSA (Rivest Shamir Adleman), and the family of elliptic-curve-based encryption algorithms (ECC), as well as the international data encryption algorithm (IDEA), may not be the desirable candidates for image encryption to better extent, particularly for fast and real-time communication applications. In recent years, several encryption schemes have been proposed [1–12]. These encryption schemes can be classified into different categories such as value transformation [1–4], pixels position permutation [5–8], and chaotic systems [9–12].

## II. PROBLEM FORMULATION

The proposed work is basically a cryptographic security algorithm in which image is encrypted to a form that is not understandable and recognizable unless and until it is decrypted with security. The focus is mainly on to major concepts in the whole process of encryption of image. The algorithm focuses on encryption level as good as Secure Force algorithm provides for textual data and on the other hand with less time consumed in encryption and decryption process. The comparison of SF with AES algorithm on FPGA platform is mentioned in [22].

The basic motivation to the proposed approach is to improve the fast and secure communication of multimedia data over internet transmission. For security of data many algorithms are developed but not every algorithm works smoothly when it comes to the point of fast transmission with high security. In this proposed algorithm the collective work of Secure Force algorithm which is a light weight and less calculative algorithm is combined with Rubik Cube algorithm to get better result in the form of good encryption with more security yet with better speed.

### III. NOVEL HYBRID IMAGE ENCRYPTION BASED ON RUBIK CUBE AND BLOCK CIPHER.

With the improvement in the today internet communication technology, numerous applications are there with various encounters that includes power utilization, security problems, scalability and design simulation problems. There are networks called Wireless Sensor Networks which claims an algorithm to provide high and trustworthy security with small power utilization as of low resources available to them. The proposed Novel Hybrid algorithm is one of the nominees for fast and secure algorithms. The design of proposed algorithm delivers low and simple complexity architecture. To safeguard energy efficient execution, it is suggested to lower the number of encryption rounds [15] [16]. In Novel Hybrid algorithm each encryption round encompasses six modest arithmetic operations on 64-bit data to certify security followed by scrambling of data. These steps create the satisfactory amount of confusion and diffusion in data to challenge various types of attacks. The key expansion process, implemented at the decoder, involves the same very light weight process followed by descrambling and simple 16-bit diffusion and reform the original image and it contains simple mathematical operations such as multiplication, addition, rotation, permutation and other operations like transportation is there to produce keys for the encryption process. However, the keys generated must be transmitted securely to the encoder side for the encryption process. The process of Novel Hybrid algorithm consists of 4 major blocks. The detailed description of logical operations used of the proposed algorithm can be found in [1]. **Key Expansion Block:** Key expansion is the very first and primary method that is used to generate altered keys for encryption and decryption. There are different operations that are performed in order to create confusion and diffusion for more security. This whole process is to reduce the possibility of weak key as well as to increase the key strength and to ensure energy efficient implementation, it is suggested to lower the number of encryption rounds [13]. The round keys (Kr) are derived from

the input cipher key by means of the key schedule. The whole process consists of two components: key expansion (where key is expanded to perform operations on it) and key reduction (where expanded key undergoes logical operations to make final key). The key expansion performs dividing 64-bit key into 16-bits keys to operate logical operation on them where as key reduction phase includes performing operation on 16-bits keys blocks to convert them to final 16-bit key in order to make it able for 64-bit data encryption. The block diagram of the key expansion block is shown in fig. 1. **Key Management Protocol:** The key management protocol is there to accomplish transmission of key over the network to decoder end to ensure the safety of key. As the whole process of encryption and decryption depends on the same key so key security is must for the algorithm security. **Encryption Block:** The process of encryption is initiated with the key expansion block where keys generated are securely received by the encoder side through the communication channel which is assumed to be secured by following any protocol rules. Then process of encryption is carried out with simple operations with XOR, OR, AND, XNOR, left shift (LS), substitution (S boxes) and swapping operations, are performed to create high level confusion and diffusion. The detailed block diagram of the encryption block is shown in fig. 2.

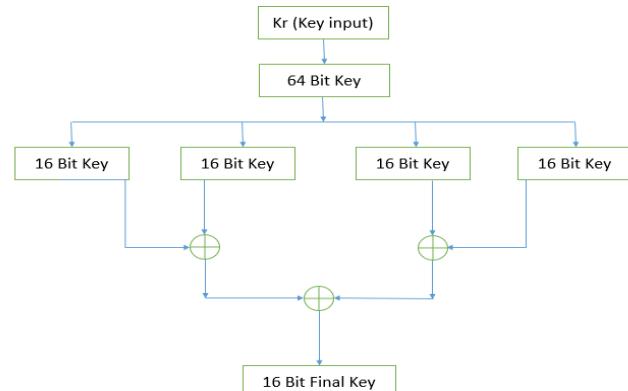


Fig. 1. KEY EXPANSION BLOCK.

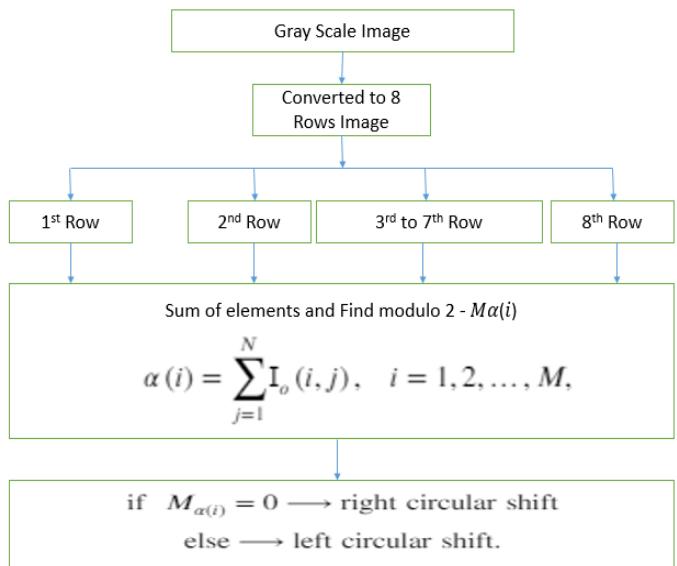


Fig. 2. SCRAMBLING OF IMAGE.

encryption. The large number of replication loops, the better is the encryption but on the same end the greater consumption of resources and more time consuming.

From encryption the encrypted data is passed to the decrypted block which process the encrypted block with the reverse encryption process which also contains one decryption round, and it uses the same key K1 to apply the process of decryption on the encrypted message to get it back to a plain text. To examine the overall performance of the specific block of each decryption, encryption and key expansion block is executed separately. The thorough explanation of the sub-modules of Novel hybrid will be given in their particular sections.

### Key Expansion Block

Key expansion is the foremost process that is used to produce keys for the process of decryption and encryption cycles to carry out. Different operations are executed counting various logical operations in order to create diffusion and confusion. This process is to lessen the opportunity of generation of feeble keys as well as to upsurge the strength of key. The round key (K1), that is derived from the input cipher key (Kr), whose length is same as length of image matrix, by means of the key schedule. The process involves of breaking 64-Bit key into 16 bit blocks of keys and taking XOR of first two sets and last two sets and then take XOR of the final two sets to produce final 16-bit key. Here the key generated is of size 16-bit whereas the block size of image to be encrypted is of 64-bit, for that we need to convert image matrix into rows of 8 count and columns count may vary that when converted to binary led to 64-bit column count.

### Encryption Block

We have key Kr to start the procedure of encryption. The very first step of encryption contains the scrambling of image as needed to create diffusion and confusion of the original data image. For that the block of image is considered and processed row wise. Each row of the image is processed one after the another (serial wise). The scrambling of image is carried out by taking the submission of all elements of row, on which processing is done and take modulo 2 of the submission. The submission of image is carried out as:

$$\alpha(i) = \sum_{j=1}^N I_o(i, j), \quad i = 1, 2, \dots, M,$$

Where I represent the image matrix and (i,j) represents the image matrix values and j varies from 1 to length of image matrix row represented by N. Then Modular 2 is calculated for a(i) and is processed by the following mentioned equation:

if  $M_{\alpha(i)} = 0 \rightarrow$  right circular shift  
else  $\rightarrow$  left circular shift.

where  $M_{\alpha(i)}$  represents the modular 2 of image matrix row elements.

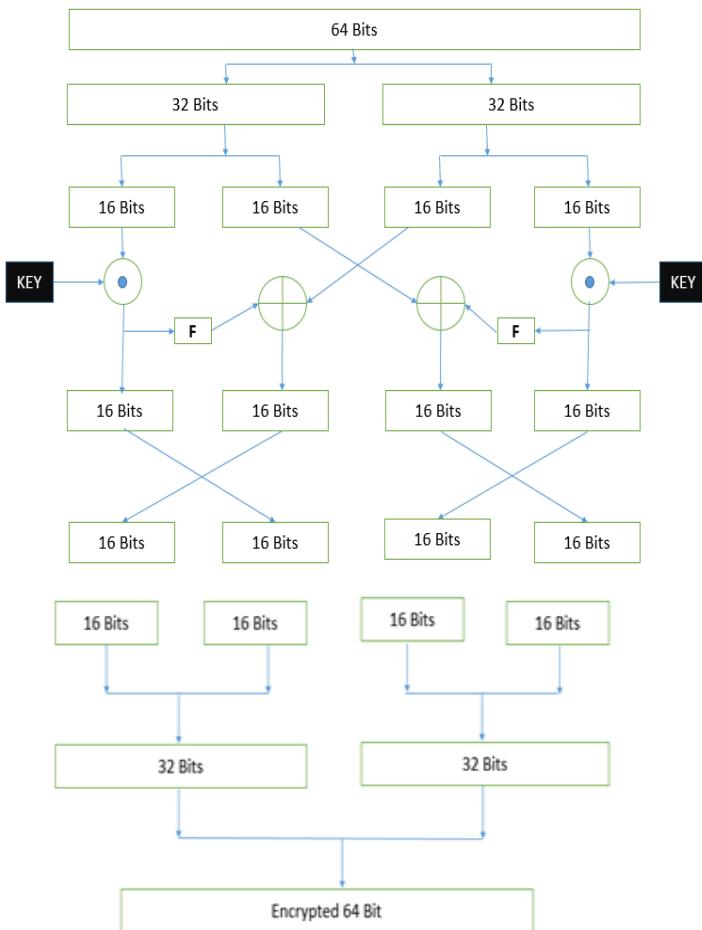


Fig. 3. ENCRYPTION BLOCK.

**Decryption Block:** In the process of decryption the whole process of encryption is executed but in opposite manner with same key used in encryption process.

The execution of encryption process on hardware level holds many of the choices that are used in execution process including loop unrolling that may be full, partial and etc., pipelining, substitution box designs, data path width optimization etc. Though it is good to use but the complete parallel loop unroll architecture is chosen only at those locations where high throughput is required.

The Novel hybrid algorithm consists of three units; the key generation module, image input encryption module, and the encrypted image decryption module. The proposed algorithm termed Novel hybrid chains all the three modules together to form a whole single-unit. First, a key is passed and it is converted into 64-bit binary key then the generated key authorizes through the key generator which generates Final Key of 16 bits to encrypt 64-bits block of image in encryption phase. The generated key act as a input to the encryption block along with the plain text from the binary image and converts it into a cipher text of 64-bits by applying the encryption process. In that round key K1 will be used and in the final step, the encrypted message generated by the same key used. The same procedure of encryption is repeated again and again for high end recursive

If calculated modular 2 is zero, then the row is shifter towards right side by unit present in Kr key array and on the other hand if it is not zero then it is shifted to left by the units present in Kr key location.

The above mentioned steps are repeated for all rows of image matrix as represented to create confusion and diffusion.

The below image represents the output after scrambling is done on input image.

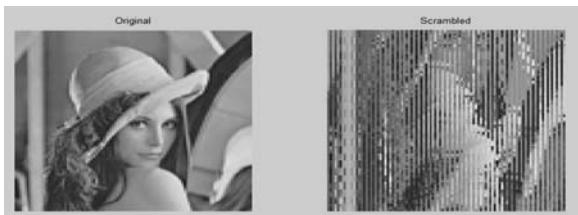


Fig. 4. SCRAMBLED IMAGE.

Once the key expansion block generates the key as per needed to accomplish block cipher encryption on image with each block of 64-bits, the scrambled image is converted into binary form in order to encrypt it accordingly. In binary form image block contains 64 columns and n number of rows. The encryption of image is carried out row wise with each row size of 64-bits, four times the size of key in binary format. The process of encryption executes by breaking the binary scrambled image 64-bits data into 16-bits four blocks separately. As we have key of size 16-bits that is same size of each block created for image data. The very first and the fourth block of image data is operated with XNOR operation with input key and the output of these blocks are operated with highly permuted and substituted box to create a lot of confusion and operated with XNOR operation with second and third block of input image data to produce another second and third block of encrypted form. The output four blocks are swapped and interchanged with first by second and third by fourth to create more confusion. The four blocks are combined to make 64-bits encrypted data of image block. These steps are performed repeatedly over all the rows of binary image data to produce encrypted image data. The output of encrypted seems as follow:

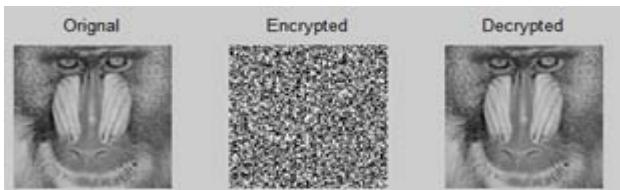


Fig. 5. BABOON IMAGE ENCRYPTION AND DECRYPTION BY PROPOSED ALGORITHM

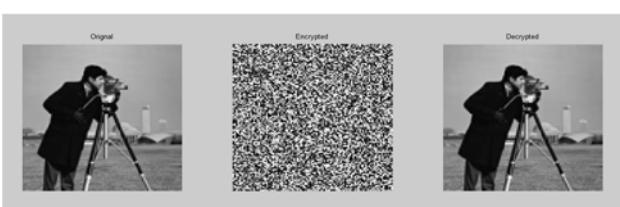


Fig. 6. CAMERA-MAN IMAGE ENCRYPTION AND DECRYPTION BY PROPOSED ALGORITHM.

#### IV. EXPERIMENTAL RESULTS AND TEST CASES

The encryption power of the proposed algorithm is calculated by the universal defined tests. These tests include:

**Visual Testing:** The resolution of visual testing is to highlight the existence of resemblances or one can say correlation, association and relation between plain-image and shuffled and (or) ciphered image i.e., if the scrambled and (or) ciphered image does not or does contain any features of the plain-image.

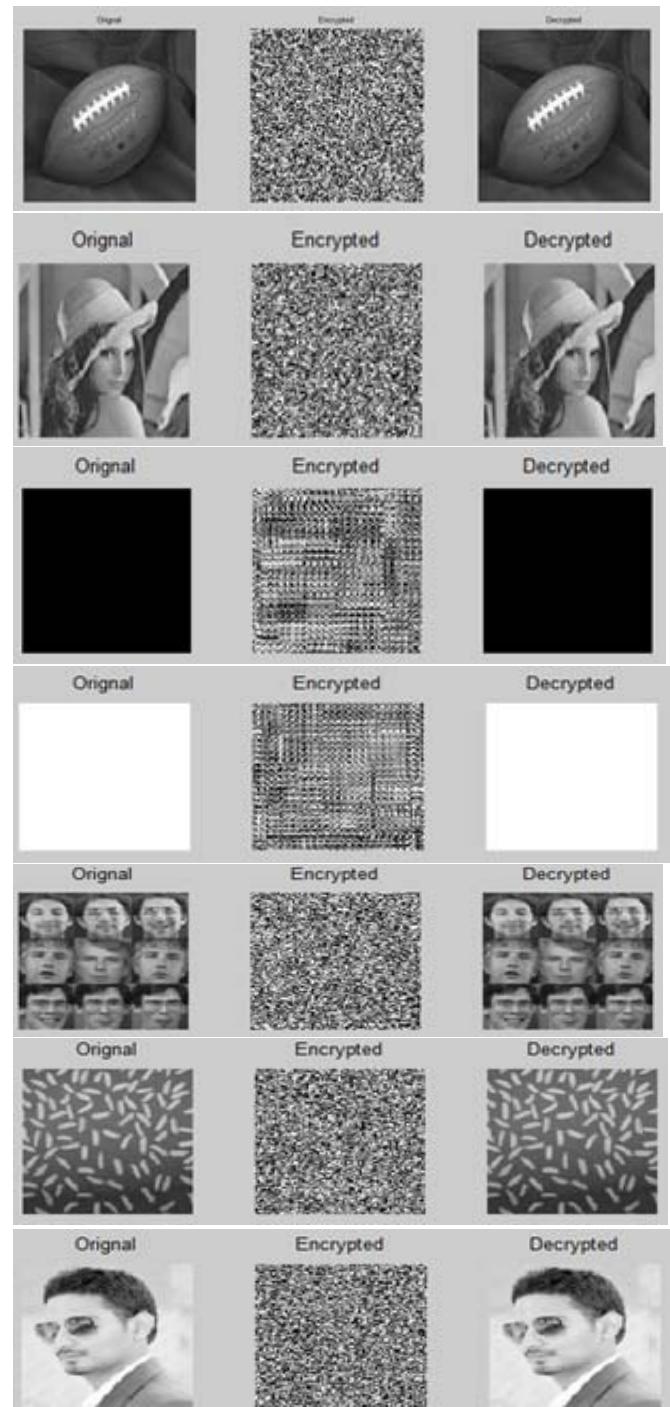


Fig. 7. DIFFERENT IMAGES SHOWS ENCRYPTION AND DECRYPTION BY PROPOSED ALGORITHM TO DEPICT VISUAL TESTING. INCLUDES IMAGES (FOOTBALL, LENA, BLACK, WHITE, ORAL FACES, RICE, BOY).

From the above encrypted images, it is clearly represented that encrypted images are very different from original image and gives no clue about original image as per seen visually.

#### Security Assessment check by Statistical Analysis Testing:

There are mainly two statistical analyses performed to showcase the confusion and diffusion of the encrypted image scuffling and shuffling or used ciphering algorithm named Histogram Analysis test and examination of the inter-relation or correlation coefficient between adjacent or together pixels. The very first of it is Histogram Test: This histogram test is used to perform analysis of pixels' distribution within an image, by representing those pixels' number relative to each intensity level.

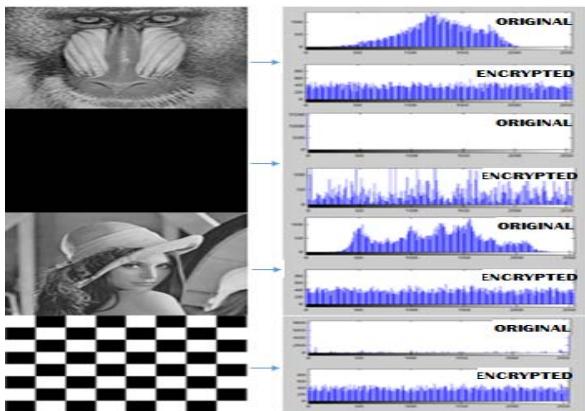


Fig. 8. DIFFERENT IMAGES SHOWS ORIGINAL AND ENCRYPTED IMAGES HISTOGRAM GRAPHS. INCLUDES IMAGES (BABOON, BLACK, LENA, CHECKED)

**PIXELS CORRELATION TEST:** Pixels correlation test is the renowned test that is generally performed in plain images to judge how toughly correlation exists between any subjectively chosen pixel with its adjacent pixels, that be vertically, diagonally or horizontally oriented. For the encrypted image with little dependency of pixels led to worthy encryption rather than produced image with high pixels' dependency. Mentioned table represents the pixel correlation values of encrypted and original images.

Table 1. PIXEL CORRELATION VALUES OF DIFFERENT IMAGES FOR BOTH ENCRYPTION AND DECRYPTION

IMAGE	ENCRYPTED	ORIGINAL
cameraman.tif	0.0105	0.9402
Football.jpg	-0.0074	0.9561
Lena.jpg	-0.0116	0.9379
Rice.tif	0.0162	0.851
ORLFace.jpg	0.0045	0.8724
Boy.jpg	-0.011	0.9456
baboon.jpg	-0.0008	0.8631
checkbox.jpg	-0.013	0.8963
black.jpg	0.0115	NaN
White.jpg	-0.017	NaN

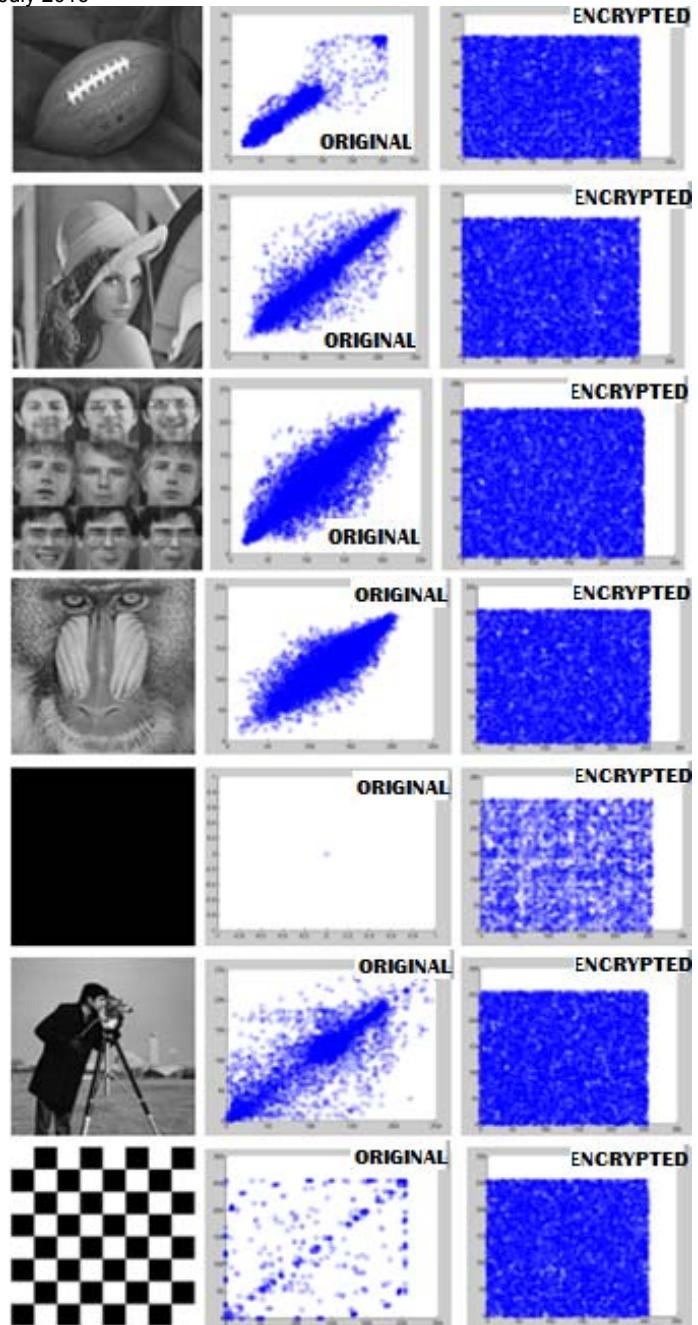


Fig. 9. DIFFERENT IMAGES SHOWS ORIGINAL AND ENCRYPTED IMAGES CORRELATION COEFFICIENT. INCLUDES IMAGES (FOOTBALL, LENA, ORAL FACES, BABOON, BLACK, CAMERA MAN, CHECKED).

**Security Assessment test by Differential Analysis:** The Differential analysis related to cryptography undertakes that the invader is able to craft small changes in the input plain image and judges the output that is the processed version of image. By doing that, scuffling, shuffling and (or) encryption key(s) and (or) the meaningful interrelationship between original images from that the original image can be generated. Therefore, a necessary property of the proposed modified algorithm is to be sensitive to the small changes in plain-image. Thus the attack by differential process can miss its efficiency led it to useless practically as if even one minor change in the original image led to significant change in its processed versions.

Another analysis test to ensure encryption goodness is analysis based on differential parameters. These includes two parameters – Number of Pixels Change Rate (NPCR) and Unified Average Changing Rate (UACI). To approach the performances of ideal image encryption algorithms, the value of NPCR must be large as possible i.e. nearer to 100 percent or can say close to unity, while UACI value generated must be around 33% to be more good. The more it is nearer to 33% the better is the encryption. Following table represents the NPCR and UACI values calculated for encryption process done on different images.

Table 2. NPCR AND UACI VALUES OF DIFFERENT IMAGES.

IMAGES	NPCR	UACI
<a href="#">cameraman.tif</a>	99.59	30.8527
<a href="#">Football.jpg</a>	99.61	31.316
<a href="#">Lena.jpg</a>	99.68	28.4551
<a href="#">Rice.tif</a>	99.49	27.9647
<a href="#">ORLFace.jpg</a>	99.71	28.7727
<a href="#">myimage.jpg</a>	99.58	35.1324
<a href="#">baboon.jpg</a>	99.6	26.9939
<a href="#">checkboard.jpg</a>	99.5	46.6261
<a href="#">black.jpg</a>	99.2	50.9327
<a href="#">White.jpg</a>	99.63	47.8185
Average	99.559	35.48648

**Security Analysis by Entropy Assessment:** Image entropy is a quantity which is used to describe the business of the image i.e. the amount of information which must be coded by the compression algorithm. The ideal value of entropy of image is calculated as 8-bits, for gray-scale images of 256 levels. Nearer the entropy better is the encryption held by algorithm. The proposed algorithm for the encryption of an image must give entropy nearer to 8-bits to make it more efficient as in practice, the resulted entropy is smaller than the ideal one. As we know the smaller is the entropy, the greater the degree of predictability i.e. the poor security led to threatens encryption system's security. Computing the entropy value of the encrypted image, which is very close to the ideal value of 8 (more accurately 7.9992, much closer than values resulted under different algorithms), we can say that the proposed encryption algorithm is highly robust against entropy attacks. For the proposed algorithm the entropy is calculated for five loops as follow:

Table 3. ENTROPY ANALYSIS OF DIFFERENT IMAGES FOR FIVE CYCLES.

PHOTO	1st Cycle	2nd Cycle	3rd Cycle	4th Cycle	5th Cycle	Original Image
<a href="#">cameraman.tif</a>	7.9547	7.9777	7.9782	7.9808	7.9827	7.0443
<a href="#">Football.jpg</a>	7.945	7.9792	7.9786	7.9815	7.9816	6.6442
<a href="#">Lena.jpg</a>	7.9781	7.9834	7.9819	7.9825	7.9805	7.3983
<a href="#">Rice.tif</a>	7.9559	7.9776	7.9823	7.9815	7.9828	6.9329
<a href="#">ORLFace.jpg</a>	7.9784	7.9827	7.9816	7.9811	7.9836	7.4688
<a href="#">myimage.jpg</a>	7.9565	7.9804	7.9804	7.9787	7.9835	7.0565
<a href="#">baboon.jpg</a>	7.9731	7.9815	7.9786	7.9786	7.9793	7.1071
<a href="#">checkboard.jpg</a>	6.7664	7.6072	7.9115	7.9686	7.9815	2.9337
<a href="#">black.jpg</a>	4.2006	5.2175	6.3779	7.2614	7.6435	0
<a href="#">White.jpg</a>	4.0059	5.0328	6.097	6.9633	7.5363	0
Average By Cycle	7.07146	7.372	7.6248	7.8058	7.9035	5.25858

**Key Sensibility Test:** Proposed Encryption algorithms must have great sensibility to encryption key that is the key used in the whole encryption process: this simply means that any small change in the key should lead to a huge change in the encrypted, or decrypted, image data. I performed two tests to illustrate the sensibility of key of our scheme. The very first one displays the impact of a key alteration in the process of image encryption. In this the same image is encrypted by original key and by single bit change in the key. The encrypted images must be far different from each other. If they are far different from each other than proposed algorithm is key sensible on the other hand if not than it is not. Another test includes decryption of encrypted image with original key and with another key just with 1-bit difference. The decrypted images produced must be far different from each other for the good key sensibility of encryption algorithm. To calculate the difference between the produced images with one-bit difference we have to calculate the NPCR (Number of Pixels Change Rate) and UACI (Unified Average Changing Intensity) of the encrypted images. In the proposed algorithm following table represents the output of NPCR and UACI produced as follows:

Table 4. NPCR AND UACI VALUES FOR DIFFERENT IMAGES FOR ONE-BIT KEY DIFFERENCE.

IMAGE	NPCR	UACI
cameraman.tif	99.58	32.8154
Football.jpg	99.12	33.0303
Lena.jpg	99.59	33.4188
Rice.tif	99.77	33.3683
ORLFace.jpg	99.67	33.0593
myimage.jpg	99.15	33.6345
baboon.jpg	99.52	33.3772
checkboard.jpg	99.66	33.8985
black.jpg	99.13	33.3443
White.jpg	99	32.43
Average	99.419	33.23766

The above stated table depicts the NPCR and UACI values of encryption images with original key and key with one-bit difference. Another test contains decryption of image with original key and key with one-bit difference. The following figures represents the proposed algorithm is key sensitive and even one-bit change in key do not produce image nearer to original image.

The below mentioned figures contains ‘Original’ image that is the original image and Decrypted image that is decrypted by original key and Decrypted1 and Decrypted2 are produced by key with one-bit difference. It depicts that in proposed algorithm even a bit difference in key do not bring original back from encrypted image.

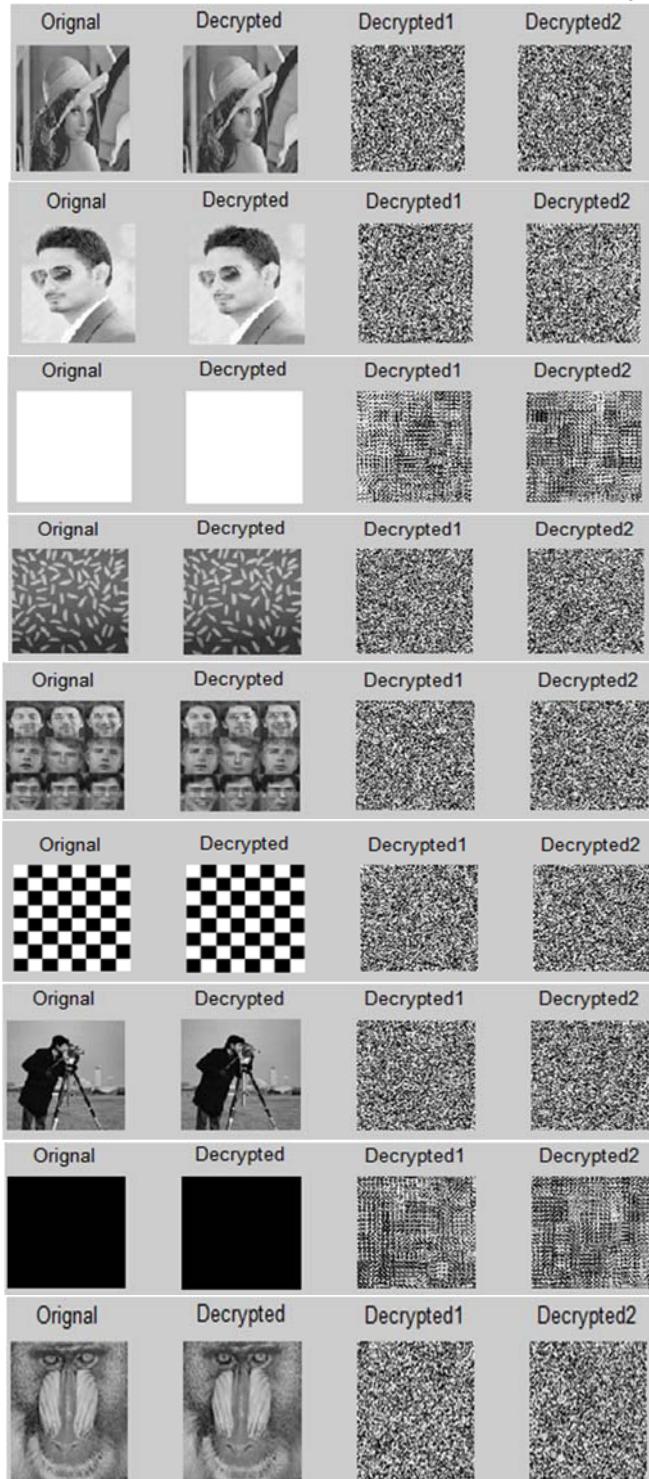


Fig. 9. DIFFERENT IMAGES SHOWS ORIGINAL AND DECRYPTED VISUAL ANALYSIS PROCESSED BY ORIGINAL KEY AND TWO ANOTHER KEYS WITH ONE-BIT DIFFERENCE. INCLUDES IMAGES (LENA, BOY, WHITE, RICE, ORAL FACES, CHECKED, BOY IMAGE ENCRYPTION WITH SECURE FORCE, RUBIK CUBE [14] AND PROPOSED ALGORITHM.

**Encryption Time:** Encryption time depicts the time taken by algorithm during encryption process. Encryption time plays important role in designing the algorithm. Lower the encryption time better the algorithm encryption and vice versa. The encryption time for proposed algorithm for different images and for different cycles figured out below:

Table 5. ANALYSIS OF ENCRYPTION TIME FOR DIFFERENT IMAGES FOR DIFFERENT CYCLES.

PHOTO	1st Cycle	2nd Cycle	3rd Cycle	4th Cycle	5th Cycle	AVERAGE TIME
cameraman.tif	0.4567	0.4643	0.4651	0.4649	0.4676	3.86382
Football.jpg	0.4586	0.4589	0.4608	0.4647	0.4574	0.46008
Lena.jpg	0.4643	0.4553	0.4652	0.4563	0.4594	0.4601
Rice.tif	0.4859	0.4601	0.4637	0.6515	0.5649	0.52522
ORLFace.jpg	0.4635	0.4525	0.4579	0.4556	0.4653	0.45896
myimage.jpg	0.4542	0.4604	0.4478	0.4475	0.4472	0.45142
baboon.jpg	0.468	0.4708	0.4606	0.4642	0.4609	0.4649
checkboard.jpg	0.4742	0.4642	0.4489	0.4566	0.4683	0.46244
black.jpg	0.4443	0.4624	0.4642	0.4662	0.4673	0.46088
White.jpg	0.4529	0.4489	0.4593	0.4587	0.4598	0.45592
AVERAGE TIME BY	0.46226	0.45978	0.45935	0.47862	0.4718	0.466364

## V. RESULT COMPARISON

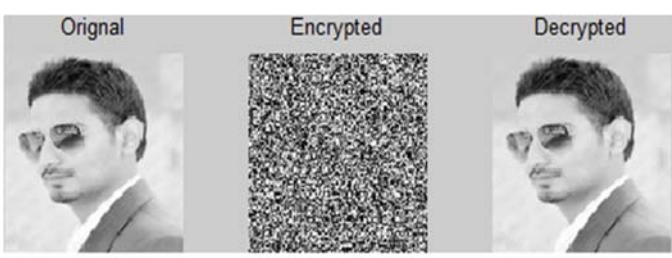
The evaluation of SF algorithm was agreed out on certain well known constraints used by numerous authors [17], [18], [19], [20] and [21]. Although SF shows good results and its performance is comparable to other algorithms [14], [22], [23], [24] and [25] in terms of computation cost, but it is still not as claimed. The following table represents the comparison of proposed algorithm with Rubik Cube and Secure Force Algorithm on different parameters:



RUBIK CUBE ALGORITHM



SECURE FORCE ALGORITHM



PROPOSED ALGORITHM

Fig. 10. BOY IMAGE ENCRYPTION WITH SECURE FORCE, RUBIK CUBE [14] AND PROPOSED ALGORITHM.

Table 6. COMPARISON OF VARIOUS PARAMETERS FOR SECURE FORCE, RUBIK CUBE AND PROPOSED ALGORITHM.

COMPARISON	PROPOSED ALGO	RUBIK CUBE	SECURE FORCE
ENTROPY (ORIGINAL IMAGE)	7.0565	7.0565	7.0565
ENTROPY (FINAL IMAGE)	7.9821	7.6256	7.9813
NPCR	99.66	99.38	99.62
UACI	34.8005	24.7476	34.9138
ENCRYPTION TIME (Per Cycle)	0.5842	0.0792	2.9558

From the above demonstrated table it is clear that the Entropy, NPCR and UACI is better than both and more nearer to secure force algorithm whereas the execution encryption time for Rubik cube is still better than both but as its Entropy, NPCR and UACI is far less than proposed algorithm and Secure Force, the execution time for proposed algorithm is considered better without compromising the security and speed of execution of algorithm.

## VI. CONCLUSION

In this paper I implement Novel Hybrid image encryption algorithm by applying cryptographic and logical functions on the input image. This paper focuses of processing images with less encryption consuming time without compromising the encryption security. The output decrypted image is same as that of original image without any loss of information. The encryption process managed in this algorithm is 64-bit encryption algorithm. Another advantage of this algorithm is the whole security depends on one small size key, which is easy to manage and easy to transmit.

## VII. REFERENCES

1. Z. Liu, L. Xu, C. Lin, J. Dai, and S. Liu, "Image encryption scheme by using iterative random phase encoding in gyrator transform domains," Optics and Lasers in Engineering, vol. 49, no. 4, pp. 542–546, 2011.
2. Q. Guo, Z. Liu, and S. Liu, "Color image encryption by using Arnold and discrete fractional random transforms in IHS space," Optics and Lasers in Engineering, vol. 48, no. 12, pp. 1174–1181, 2010.
3. Z. Liu, H. Chen, T. Liu, et al., "Image encryption by using gyrator transform and Arnold transform," Journal of Electronic Imaging, vol. 2, no. 4, pp. 345–351, 1993.
4. R. Tao, X. Y. Meng, and Y. Wang, "Image encryption with multiorders of fractional fourier transforms," IEEE Transactions on Information Forensics and Security, vol. 5, no. 4, pp. 734–738, 2010.
5. R. Zunino, "Fractal circuit layout for spatial decorrelation of images," Electronics Letters, vol. 34, no. 20, pp. 1929–1930, 1998.
6. G. Zhang and Q. Liu, "A novel image encryption method based on total shuffling scheme," Optics Communications, vol. 284, no. 12, pp. 2775–2780, 2011.
7. X.-Y. Zhao and G. Chen, "Ergodic matrix in image encryption," in Proceedings of the 2nd International Conference on Image and Graphics, vol. 4875, pp. 394–401, August 2002.
8. Z.-L. Zhu, W. Zhang, K.-W. Wong, and H. Yu, "A chaos-based symmetric image encryption scheme using a bit-level permutation," Information Sciences, vol. 181, no. 6, pp. 1171–1186, 2011.
9. C. K. Huang and H. H. Nien, "Multi chaotic systems based pixel shuffle for image encryption," Optics Communications, vol. 282, no. 11, pp. 2123–2127, 2009.
10. G. Chen, Y. Mao, and C. K. Chui, "A symmetric image encryption scheme based on 3D chaotic cat maps," Chaos, Solitons and Fractals, vol. 21, no. 3, pp. 749–761, 2004.
11. X. Y. Wang, L. Yang, R. Liu, and A. Kadir, "A chaotic image encryption algorithm based on perceptron model," Nonlinear Dynamics, vol. 62, no. 3, pp. 615–621, 2010.
12. Y. Wang, K. W. Wong, X. Liao, and G. Chen, "A new chaos-based fast image encryption algorithm," Applied Soft Computing Journal, vol. 11, no. 1, pp. 514–522, 2011.
13. R. Chandramouli, S. Bapatla, and K. P. Subbalakshmi, "Battery power-aware encryption", ACM Transactions on Information and System Security, Vol. 9, No. 2, May 2006, pp. 162–18.
14. National Institute of Standards and Technology, "SkipJack and KEA algorithm specifications (Version 2.0)," May 1998. [A. Perrig, R. Szewczyk, J. Tygar, V. Wen, and D. Culler, "Spins: security protocols for sensor networks". ACM/Kluwer Wireless Networks, 8(5):521–534, 2002.]
15. C. Karlof, N. Sastry, and D. Wagner, "TinySec: A link layer security architecture for wireless sensor networks." In Proc. of the 2nd ACM SenSys, 2004.
16. Kumar, A., & Tiwari, M. N. (2012). "Effective implementation and avalanche effect of AES. International Journal of Security", Privacy and Trust Management (IJSPTM), 1(3/4),
17. Shrivkumar, S., & Umamaheswari, G. (2011, July), "Performance Comparison of Advanced Encryption Standard (AES) and AES key dependent S-box-Simulation using MATLAB." In Process Automation, Control and Computing (PACC), 2011 International Conference on (pp. 1-6). IEEE.
18. Zeghid, M., Machhout, M., Khriji, L., Baganne, A., & Tourki, R. (2007), "A modified AES based algorithm for image encryption." International Journal of Computer Science and engineering, 1(1), 70-75.

19. Elminaam, D. S. A., Abdual-Kader, H. M., & Hadhoud, M. M. (2010), "Evaluating The Performance of Symmetric Encryption Algorithms." IJ Network Security, 10(3), 216-222.
20. Yoon, J. W., & Kim, H. (2010). "An image encryption scheme with a pseudorandom permutation based on chaotic maps." Communications in Nonlinear Science and Numerical Simulation, 15(12), 3998-4006.
21. Khan, Shuaat, M. Sohail Ibrahim, Haseeb Amjad and Mansoor Ebrahim, "FPGA Implementation of Secure Force (64-bit); Full loopunroll architecture of Secure Force: a low complexity encryption algorithm," unpublished. Matsui M, "Linear cryptanalysis method for DES cipher. Advances in Cryptology-EuroCrypt'93". Berlin: Springer- Verlag, 1994. 386-397.31-35.
22. D. Hong, J. Sung, S. Hong, J. Lim and S. Lee "HIGHT: A new block cipher suitable for low-resource device", Cryptographic Hardware and Embedded Systems Vol. 4249, 2006, pp. 46-59.
23. R.L. pavan, M.J.B. Robshaw, R.Sidney, and Y.L. Yin. "The RC6 Block Cipher". Ver 1.1, August 1998.
24. S.I. Huang, and S. Shieh, "SEA: Secure encrypted data aggregation in mobile WSNs", International Conference on Computational Intelligence and Security, IEEE, 2007.
25. Yoon, J. W., & Kim, H. (2010). An image encryption scheme with a pseudorandom permutation based on chaotic maps. Communications in Nonlinear Science and Numerical Simulation, 15(12), 3998-4006.
26. MATLAB, The MathWorks, Inc

### VIII. AUTHOR's PROFILE

**Jasdeep Singh Chauhan.** Born in 1991 Received Bachelor of Technology degree in 2013 from Lovely Professional University, chehra Punjab, India. He has done his Master of Technology in 2016 from Rayat Bahra Campus, Ropar, Rail Majra, Punjab, India.

# E-Learning Systems Risks and their Security

Kassid Asmaa, El kamoun Najib ; STIC Laboratory, Chouaib Doukkali University

**Abstract—** the security of Information Systems is a major challenge for all organizations today, because people can only use a system if they trust it. Especially when they are using open and distributed environment like E-learning platforms, as e-learning increases in popularity and reach, the need to understand security concepts will also increase

The goal of this research is to identify some key security issues that must be taken into consideration in developing and using an E-learning platform. In order to do it, this paper examines the basic concepts of security in computing, and some characteristics of E-learning platforms that introduce new threats and ways to attack, we will also discuss some security aspects of one of the most popular E-learning systems: Moodle.

**Index Terms**—Security requirements, E-learning platform, Security in E-learning platform.

## I. INTRODUCTION

With the development of Information Technology, E-learning is developing rapidly, it's become an important gadget in future learning trends, since it's an inexpensive way to help people acquire knowledge and skills in different domains while living every day's ordinary life. Worth knowing that E-learning is a flexible term used to describe the newest method of teaching throughout the online internet technology , but, Internet is insecure and illegal activities are on the rise with the passage of each day, Unfortunately, We have seen considerable effort being put into development of the content and infrastructure for the e-learning system, yet there is hardly any effort being put into these system for making them secure. In many cases security is considered as a technology that increases the complexity of processes and makes everyone's life harder. However, we must to take into account that people only use a system if they trust it. Thus security is a primordial technology that each system must integrate. Considering though the scenario of a purely interactive e-learning application constructed over heterogeneous, distributed and open architectures, the potential threats to security cannot be neglected [1].

When the e-Learning appeared there 20 years, it consisted of a text on the screen for remote access to educational resources and with isolated learning environment similar to reading a book. The e-Learning was not effective or popular enough among learners, but now with Web 2.0 technologies e-Learning has become more interactive and rich media content. These new technologies have a different vision of education, by providing tools for different learning by students; its offers the ability to share material in all kinds of formats such as videos, slideshows, word documents and PDFs. Conducting live online classes and communicating with professors via chat and message forums is also an option available to users . There are several benefits to e-learning:

- **Rentability and saves time** by reducing the time taken away from the office, removing displacement costs, online learning helps you to save money and increase workplace productivity.
- **Disponibility anytime, anywhere** by in comparison with traditional learning, significantly reduces the time needed to locate information. It also offers access to online resources, databases, periodicals, journals and other material you wouldn't normally have access to from a library.
- **Discretionary** by allowing each individual to tackle the subject at their own pace, with interactive tasks being set in place to ensure a thorough understanding throughout each module, because many persons cannot feels comfortable learning in a large group.

Due to the new trends in development of educational systems, the security management of e-learning systems have attracted more and more the attention of researchers and web applications developers. Because the marketing of e-learning is continuing to grow, with a continuous student demand for online learning, the need to understand it and also to understand the security issues associated with it will also increase.

During this research were used several publications, books and some of the top online security sites to highlight the actual security issues encountered today in the online learning environment.

The goal of this research is to identify the security issues faced by E-learning platforms and organizing the literature in this area of knowledge. This leads us, first, to an analysis of the security

in computing, because E-learning systems, as any computer system, face these basic issues. The remainder of this paper is organized as follows. Section 2 describes why we need to add security to a computer system and what are the aspects that need to be analyzed to build a security system. In this section are identified the concepts involving information security, such as the threats, the countermeasures, and the security goals. In Section 3 we describe, in a conceptual way, the E-learning systems, and the particular characteristics that make this kind of system different from the conventional systems and motivate us to do research in this area and explains common threats, followed by different risks faced by the participants and necessity of risk analysis in E-Learning. In Sections 4 and 5, we answer the main question of our research: how to achieve security to our E-learning system. Section 4 presents the main works in the literature and Section 5 do a detailed exam of the advantages and disadvantages of those works. Also we examined various security aspects of the most popular open source e-learning platform. Section 6 summarizes the content of the paper and discusses the future of the research in this field.

## II. SECURITY IN COMPUTING

Computing has become part of everyday life. So in computing individuals as well as groups and organizations are concerned about security, it is a term widely used today and its meaning is heavily dependent on its context, even in computing, the term security is found in various terminologies, such as data security, information security, network security, and computer security. All these terms are interrelated and we will consider them the same thing. Security in computing involves several concepts that sometimes are controversial by the literature. It's requires precise and formalized procedures.

### A. The need of security

The information system and the sensitive data of an organization are its most important capital, which should be protected against unauthorized access. The security experts of information systems are faced with security challenges. They must both, have a comprehensive approach to information system and knowledge of the operating system and installed applications.

Several technical possibilities are offered today with the spread of the Internet connection, the intensive use of communication systems and applications and complex architectures. These opportunities also introduce risks such cases where users violate the integrity of the system by behaving in a reckless manner due to generally involuntary manipulations that might promote the danger, this category of users are causing the majority of problems related to the security of an information system, also there is malicious people who manage to seep by exploiting a vulnerability in the system to capture sensitive information which they are not supposed to have access, in addition malicious programs like viruses, Trojans are developed to harm a system, or even modify or collect data for to be reused for malicious purposes.

For these reasons securing computer systems is a real and vital work, which consists in ensuring the hardware and software resources of an organization which cannot be, in any case, used outside a planned context.

### B. Security goals

The main goal for building a secure system is to remain dependable in the face of malice, error or mischance. As a discipline, security in computing focuses on tools, processes, and methods to design, and implement trustworthy systems [1]. So, all information security measures of any system try to address at least one of three goals cited and explained in the following table:

TABLE I  
SECURITY GOALS

Goal	Definition
Confidentiality	Is equivalent to privacy; measures undertaken to ensure confidentiality are designed to prevent sensitive information from reaching the wrong people, while making sure that the right people can in fact get it: Access must be restricted to those authorized to view the data in question[3].
Integrity	Integrity involves maintaining the consistency, accuracy, and trustworthiness of data over its entire life cycle. Data must not be changed in transit, and steps must be taken to ensure that data cannot be altered by unauthorized people. These measures include file permissions and user access controls. Version control maybe used to prevent erroneous changes or accidental deletion by authorized users becoming a problem.
Availability	Availability is best ensured by rigorously maintaining all hardware, performing hardware repairs immediately when needed and maintaining a correctly functioning operating system environment that is free of software conflicts. It's also important to keep current with all necessary system upgrades. Providing adequate communication bandwidth and preventing the occurrence of bottlenecks are equally important.

These goals form the confidentiality, integrity, availability also known as the *CIA triad*, are the basis of all security programs. Information security professionals who create policies and procedures must consider each goal when creating a plan to protect a computer system.

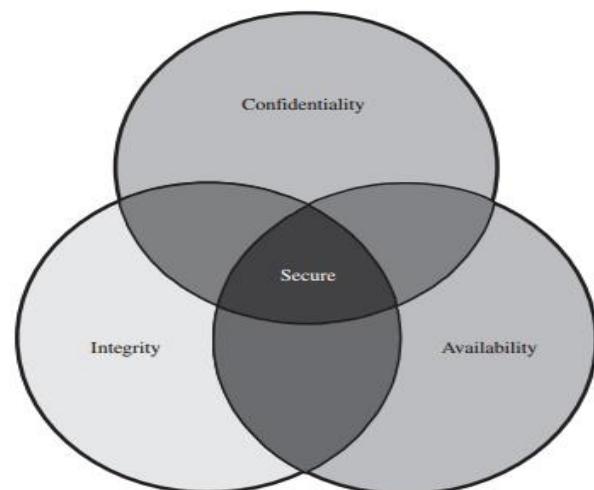


Fig. 1. Relationship between Confidentiality, Integrity, and Availability.

### C. Threats

A threat, in the context of computer security, refers to anything that has the potential to cause serious harm to a computer system. A threat is something that may or may not happen, but has the potential to cause serious damage. Threats can lead to attacks on computer systems, networks and more. They can put individuals' computer systems and business computers at risk, so vulnerabilities have to be fixed so that attackers cannot infiltrate the system and cause damage.

In the literature, threats are grouped in different ways according to their nature and objectives [3]. They are four main classes: interception or disclosure, interruption or disruption, modification and deception, and usurpation. Let's look closer at each of these types of threats in the following table:

TABLE II  
DIFFERENT THREATS IN COMPUTER SECURITY

Threats	Definition
Interception	'Disclosure 'This type of threat is well known and its idea is very simple: third party is listening a communication without authorization of the submitting parties. This type of failure, is accomplished when a malicious user gain access to a confidential asset: document, or a password.'
Interruption	The goal of interruption threats is to block legitimate users from getting services they can normally get from servers making the system lost, unavailable, or unusable. The most representative of this type of threat is the Denial of Service (DoS) and its basic function is force the target computer to process a large number of useless things, hoping to consume all its critical resources [4] Interruption can be achieved by overload of the communication channel. Availability controls may block this kind of threats
Deception	The main goal of this class of threats is deception or acceptance of false data, and occur when an unauthorized access and tampers with an asset. The attacker alters a message that contains data determining an action to be done, and the recipient, believing that this message is true, acts in accordance with the interests of the attacker. Another form of this attack occurs when an information is modified and an authorized user accept that information as correct and is mistaken.
Usurpation	Is a kind of fraud that results in the loss of personal data, such as passwords, user names, banking details or credit card numbers. The online identity theft is also sometimes called phishing.

These four classes of threats—interception, interruption, deception and usurpation—describe the kinds of problems we might encounter. In the next section, we will discuss some controls and countermeasures that will make our system secure and impenetrable.

### D. Countermeasures

To consider the controls or countermeasures that attempt to prevent exploiting a computing system's vulnerabilities, we begin by thinking about traditional ways to enhance physical security. Today we use a multipronged approach to protect our homes and offices. We may combine strong locks on the doors with a burglar alarm, reinforced windows, and even a nosy neighbor to keep an eye on our valuables. In each case, we

select one or more ways to deter an intruder or attacker, and we base our selection not only on the value of what we protect but also on the effort we think an attacker or intruder will expend to get inside [5].

Computer security has the same characteristics: we have many controls at our disposal. Some are easier than others to use or implement. Some are cheaper than others to use or implement. And some are more difficult than others for intruders to override. In this section, we present an overview of the controls available to us resumed in the following table:

TABLE III  
DIFFERENT COUNTERMEASURES IN COMPUTER SECURITY

Countermeasures	Definition
Cryptography or Encryption	<p>Is the formal name for the scrambling process. We take data in their normal, unscrambled state, called and transform them so that they are unintelligible to the outside observer; The scrambling process, which in the early days consisted of simple substitution of letters in a text, today, with the advent of computing, consisting in sophisticated algorithms and secret keys to transform data from a readable way to that which is unintelligible to those that do not have access to the data.</p> <p>Access controls is an important method of granting the three security principles of computing: confidentiality, integrity and availability, it is very important in the protection of the system, preventing unauthorized modification or disclosure of resources</p>
Software controls: access control	<p>Access control is concerned with restrictions on the actions of authenticated users. Authorization is a need of the system to determine if a legitimate user has the necessary rights and privileges to carry out a requested action within the system</p> <p>Numerous hardware devices have been created to assist in providing computer security. These devices include a variety of means, such as</p> <ul style="list-style-type: none"> <li>• hardware or smart card implementations of encryption</li> <li>• locks or cables limiting access or deterring theft</li> <li>• devices to verify users' identities</li> <li>• firewalls</li> <li>• intrusion detection systems</li> <li>• circuit boards that control access to storage media</li> </ul>
Hardware Controls	

## III. E-LEARNING PLATFORMS

### A. Overview

E-Learning systems have grown rapidly during the past years; they are diverse and widespread, with examples including Atutor, Moodle and OLAT.

E-Learning comprises both information and communication technologies. According to Rosenberg [4], there are three major criteria for eLearning:

- Updating, storing, exchanging information and its distribution;
- Distributing the information to the end user using available Internet technology;
- Targeting a wide field of education

They are large and dynamic with a variety of users and resources. The sharing of information, collaboration and interconnectivity are core elements of any e-Learning system [6]. Data must then be protected in order to maintain confidentiality, integrity and availability. Protecting against data manipulation, fraudulent user authentication and

compromises in confidentiality are important security issues in e-Learning. Meanwhile, e-Learning trends are demanding a greater level of interoperability for applications, learning environments and heterogeneous systems.

The authenticated E-Learning documents like learning materials, certificates, and question papers, lecture materials, mark sheets which are communicated from Manager to students and from Authors to teachers ... can be changed or modified, the educational assets can be also destroyed [7].

### B. Threats and risks in E-learning system

In this section we provide an overview of the most important cyber security risks that are relevant to Higher Education systems and distributed e-Learning systems.

In E-Learning system five significant participants are:

- Authors
- Managers
- Teachers
- Students

#### 1) Authors

As we know **Authors** can provide access to books, journal papers, for a wide range of students. They can also develop and implement the contents of these documents. As only registered Students can access those lecturer notes, assignments, class test paper, it is the Author's duty to protect against unauthorized use, modification and reuse of the data in different contexts related to E-Learning [8].

#### 2) Teachers

**The Discussions** are an essential component of teaching any course. One form of discussion can be through the online forum. An advantage of online forum discussions over oral discussions is that all written documents are stored electronically on a server, but the digital storage of contributions to a discussion constitutes a great risk for the privacy of Students as well as Teachers. Though in any teaching system maximum interaction can help Students as well as the Teachers to make their understanding clear .Only robust security mechanism can lead to this kind of interaction in the long run [9].

**The examination** system includes standardization of examination questions and list of questions possibly restrict the academic freedom of individual Teachers, so the relevant risk related to examination is directly associated with cheating ; also Teachers must be concerned about availability and non-repudiation of assessments, they must be aware of risk that Students receive the unaltered questions paper

#### 3) Students

Every Student must be aware of each and every document received from institute, Teachers or other Students. Because if intruders have edited the question papers or other important documents, he will have to face problems at the time of examination.

Storing login information: user ID and passwords, give a big chance to the attacker to prevent authorized learner from accessing the E-Learning server using many attacks.

Students are prompted to enter some confidential information to fake web sites which look like a real E-Learning website due to the phishing.

#### 4) Managers

A lot of risks in E-Learning platform involve inelegant people masquerading as Students and writing tests on behalf of enrolled Students and unauthorized help during the writing of online examination, so legal issues such as copyright, online testing, sending official documents ..., may be a big problem for those participants. In this case managers should take care of enrolment in a course and the cancellation of enrolment as and when required. Enrolment of one particular student in more than one course involves risk for the larger organization. There must be a plan for backups and recovery process test, if not it will be difficult to make the data up to date [10].

In General, e-university has to solve issues related to student authentication, unfair task performance, plagiarism, as well as the protection of the copyrighted material, placed on the web. So both the integrity of e-resources and smooth functioning of the educational computer systems must be protected.

### C. Classification of risks in E-learning by security goal

The following table resume some of these threats in the system classified by goals of security computing:

TABLE IV  
DIFFERENT THREATS IN E-LEARNING PLATFORMS

Threats	Authentication	Availability	Confidentiality	Integrity	Repudiation
An unauthorized party gaining access of the assets present	XXX		X		
An unauthorized party accessing and tempering with an asset used in E-learning.				XXX	
Denial of service : Prevention of legitimate access rights by disrupting traffic during the Transaction between users		XXX	X		
Person's denial of participation in any transaction of documents.		X			XXX
Insecure cryptographic storage; insecure direct object reference; information leakage and improper error handling		X	XXX		
Buffer overflow; cross site request forgery; cross site scripting; failure to restrict URL access; injection flaws; malicious file execution				XX	
Leakage of information by abusing communication channel		XX	XXX	X	

## IV. SECURING E-LEARNING SYSTEMS

In response to increasing threats, researchers have developed a number of countermeasures and solutions to improve security in E-learning. This section synthesize the related discussions in the literature

#### A. Security Protection Measures

E-learning became more user-centered and more secure with the help of new technologies:

##### 1) Cryptography

The purpose of confidentiality is to ensure that information and data are not disclosed to any unauthorized person or entity. One of the techniques in this aspect is cryptography. Cryptography plays a critical role when designing and implementing almost all kinds of electronic systems, different cryptographic tools (*JCrypTool* *Cryptool2*), are needed for the implementation of security in Internet based transactions [12]. Cryptography is an art of converting the data on the applications into incoherent or scrambled or in unintelligible format. It related to the study of mathematical algorithms related to aspects of information security such as confidentiality, data integrity, and data authentication Symmetric Key Encryption and Asymmetric Key Encryption are other two important encryption types [13].

##### 2) Digital Right Management

There are many reasons for wanting to manage the rights associated with intellectual property. Authors and artists wish to control what can be done with their creations, scholars wish to ensure that they receive proper attribution, commercial enterprises wish to support business models that involve licenses and fees, and consumers want an environment free of legal worries and unexpected costs. Although rights themselves are not technological in nature they are defined by laws, beliefs and practices technology can be used to transmit, verify, interpret and enforce rights as they apply to digital content and services. This is called digital rights management, or simply DRM.

In a distributed networked environment, multiple rights associated with multiple objects come into play as content and services are created, distributed, aggregated, disaggregated, stored, found, and used. This is particularly applicable to e-learning, where standards and technologies are being developed specifically to support the sharing and reuse of learning resources [14] ; DRM is of the major strategies to be implemented to reduce risks associated with E-Learning assets [15] is digital right management, E-Learning asset as services provided by E-Learning system such as learning resources, examination or assessment questions, Students' results, user profile, forum contents, Students' assignment and announcement in the E-Learning system. Digital Right Management makes the system safer for its contents.

##### 3) Distributed Firewall Solution

Distributed firewalls are host-resident security software applications that protect the enterprise network's servers and end-user machines against unwanted intrusion [16]; the difference between personal and distributed firewalls is the latter offer important advantages like central management, logging, and in some cases, access-control granularity. These features are necessary to implement corporate security policies in larger enterprises.

There are a lot of benefits/advantages of firewall includes [16]:

- Firewall protects hosts that are not within a topology boundary-topology independence
- Firewall provides protection against internal attacks.
- Firewall helps to eliminate single point of failure.
- They secure remote end-user machines.

-They secure critical servers on the network preventing intrusion by malicious code and "jailing" other such code by not letting the protected server be used as a launch pad for expanded attacks.

##### 4) Biometric Authentication

Among all authentication techniques like passwords, smart card, Digital signature and digital certificate, there is no guarantee that dishonest Students will keep their password secret. Password might be misused at the time of submission of assignment, receiving question papers, downloading of course materials, etc where biometric authenticity would give better security, Biometric computer authentication has an advantage it is based on something you are, which is not easily copied or stolen[17].

##### 5) Digital Watermarking

This solution allows an individual to add hidden copyright notices, audio, video, image signals. So multimedia database server of E-Learning system may be protected against unauthorized use by the way of digital watermarking. When also E-Learning information like question papers, important study materials, will invisible to the viewer, the chances of Hacking in this case will be less.

## V. RELATED WORK AND DISCUSSION

Meeting the security requirements in an e-learning system is an extremely complex problem because it is necessary to protect the content, services and the personal data not only for the external users, but also for the internal users, including system administrators. Many works are proposed in the literature try to give a satisfactory level of security taking into account the characteristic of security computing.

### A) Research works and discussion

In the literature, the indicated bibliography shows that there are many valuable scientific studies on e-learning, for instance[21,22,23,24].Although e-learning is a new concept, only several years old, it was granted a great deal of attention as revealed by the literature available on the topic. Mason and Rennie listed 180 key concepts subsumed within e-learning [25]. Unfortunately, the security aspects in e-learning appear extremely rarely, to achieve a good level of security, there are many important elements that must be taken into account, and this has been discussed in a good way and can be reached in [26]. In [27], proposals for Security of e-learning Systems and security requirements for Multi-agent systems have been discussed; Security case modeling has been taken into account with emphasis on use cases.

Security has already proved an important requirement for the success of MAS, so there are already some works in this research area cited in [28,29], showing the concern of multi-agent community with security.

Many solutions was discussed in many works for example: Authentication and authorization process is the first step that needs to be done to secure online learning environments. In [20] authors recommend the use of authentication procedure to

easily identify a legal user. This is meant to overcome the illegal usage of application. A system which is too heavily secured may be difficult access by the user. In order to put in equilibrium security and access. Graf [19] suggests protecting their intellectual property by increasing control of the person who owns the copyright on digital data for the entire period of its existence. He also proposes an approach to control the access to the information. in [29] the authors introduce a prototype architecture of 3 levels that highlights the security requirements of e-learning application and communication platform. Security requirements were modeled for access control, integrity and privacy through the case study of security use, the security mechanism of the multi agent platform FIPA-OS, PMA3 Multi agent platform was presented in this work.

One of the most developed security bases is "access control", it is an important method of grant the three security Principles of computing: confidentiality, integrity and availability, the control of how resources are accessed it is very important in the protection of the e-learning platforms, preventing unauthorized modification or disclosure of resources. Several access control models have been developed during last year's ,a relevant work is provided by Xiao et al. – an authorization mechanism based in RBAC model , the authors believe that using policies based on roles is possible to build a security architecture that automatically adapts to system changes. Credentials implement a notion of binary trust: the user has to produce a predetermined set of credentials like credit card numbers or proof of membership to certain groups to gain specific access privileges. The credential provides information about the rights, qualifications, responsibilities and other characteristics attributable to its bearer by one or more trusted authorities also it provides trust information about the authorities themselves. The integration of credential based access control with role-based access control make the security administration easy. Although credential based models solve the problem of access control in open systems to a great extent, but it still not enough to achieve a satisfy level in terms of given information about the behavior or action of the user between the time the credential was issued and its use, such information may play crucial parts in access control decisions, that why a lot of research has been done to improve the evaluation of trust on integrating the mechanism of history and context information (context awareness takes an important part which identifies the user's needs by analyzing the context information of user environment) of the user[32] .

Security is not always a technical problem; it is also a good organization. A general security policy respected by the users that contain not only a technical solution, but a combination of various methods, technologies, solutions and tools, figure 2 show an example of Elements of satisfactory level of security proposed for E-learning platforms [18].

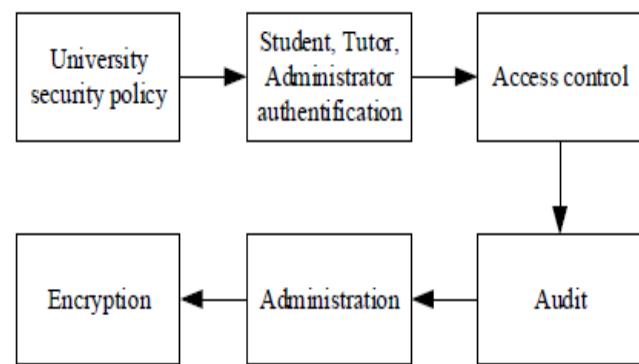


FIG.2. Components of satisfactory security in e-learning

### B) Moodle platform

Moodle became the most common used Learning Management Systems at that time. Moodle has the ability of tracking the learner's progress, and can be monitored by both teachers and learners. This fact implicitly includes both security and privacy threats and makes Moodle vulnerable system [30]. It is also known as a Course Management System (CMS), Virtual Learning comparison environment (VLE), and OSS e-learning platform which provides educators tools to create a course web site. It is used in 193 countries, with 400,000 registered users.

Moodle web page provides developer information, roadmap, coding guide and concurrent versioning system guide to access its source code and it has a long list of developers. It does not provide a formal model for future development [31].

#### 1) Moodle security issues

In this section we present a description of the most critical security flaws as discussed in literature. They are classified into four groups: authentication, availability, confidentiality and integrity attacks.

LMS is client/server web applications that, among rest, handle user requests coming from clients such as web browsers. To handle the user requests, they frequently require accessing security-critical resources (databases and files) at the server end [33].

The first design flaw of Moodle is related to the brute force attack. A brute force attack consists of trying every possible code, combination, or password until you find the right one. This type of attack may be performed to guess the password or user name. To guess the password, the user sends several requests to the web server with the blank cookie field so that the login failure count is reset to zero (Kumar & Kamlesh, 2011). To guess the user, a number of usernames are sent with an arbitrary password. Usually, if the response from the server is longer, the chances to guess the user are higher. To prevent this, Moodle added a *password policy system* (starting from version 1.9) which may be set up from: *Administration > Security > Site policies*. *This issue may be resolved also using a captcha system in the login page.*

Another security problem may occur when a session hijacking attack is used. Session hijacking is the act of taking control of a user session after successfully obtaining or generating an authentication session ID. Session hijacking

involves an attacker using captured, brute forced or reverse-engineered session IDs to seize control of a legitimate user's Web application session while that session is still in progress. The session is handled in Moodle by using two cookies: *MoodleSession* and *MoodleSessionTest* which are stored in the cookie hat is thrown on each HTTP demand within the header of the message. Acquiring a full HTTP request data with the cookie incorporated is easy because Moodle only uses SSL tunnels on the login service and a few administrative services. Because of that, most HTTP demand is done on plaintext that can be catch and easily decoded. After getting the cookie, the attacker can utilize this data on its own HTTP request, taking control of the target user session.

## 2) Moodle security overview

Here you see the most important things which help to secure Moodle platform:

Issue	Status	Description
Insecure dataroot	OK	Dataroot directory must not be accessible via the web.
Displaying of PHP errors	OK	Displaying of PHP errors disabled.
No authentication	OK	No authentication plugin is disabled.
Allow EMBED and OBJECT	OK	Unlimited object embedding is not allowed.
Enabled .swf media filter	OK	Flash media filter is not enabled.
Open user profiles	OK	Login is required before viewing user profiles.
Open to Google	OK	Search engine access is not enabled.
Password policy	Warning	Password policy not set.
Email change confirmation	OK	Confirmation of change of email address in user profile.
Writable config.php	Warning	PHP scripts may modify config.php.
XSS trusted users	Warning	RISK_XSS - found 31 users that have to be trusted.
Administrators	OK	Found 2 server administrator(s).
Backup of user data	Warning	Found 1 roles, 0 overrides and 5 users with the ability to backup user data.
Default role for all users	Critical	The default user role "Authenticated user" is incorrectly defined!
Guest role	OK	Guest role definition is OK.
Frontpage role	OK	Frontpage role definition is OK.

FIG.2. security overview in Moodle

- Insecure dataroot:** The dataroot is the directory where Moodle stores user files. It should not be directly accessible via the web.
- Insecure dataroot:** The dataroot is the directory where Moodle stores user files. It should not be directly accessible via the web.
- Register global:** register global is a PHP setting that must be disabled for Moodle to operate safely.
- Displaying of PHP errors:** If PHP is set to display errors, then anyone can enter a faulty URL causing PHP to give up valuable information about directory structures and so on.
- No authentication:** Use of the "no authentication" plugin can be dangerous, allowing people to access the site without authenticating.
- Allow EMBED and OBJECT:** Allowing ordinary users to embed Flash and other media in their texts (eg forum

posts) can be a problem because those rich media objects can be used to steal admin or teacher access, even if the media object is on another server.

- Enabled .swf media filter:** Even the flash media filter can be abused to include malicious flash files.
- Open user profiles :** User profiles should not be open to the web without authentication, both for privacy reasons and because spammers then have a platform to publish spam on your site.
- Open to Google:** Allowing Google to enter your site means that all the contents become available to the world. Don't use this unless it's a really public site.
- Password policy :** Using a password policy will force your users to use stronger passwords that are less susceptible to being cracked by a intruder.
- Password salt :** Setting a password salt greatly reduces the risk of password theft.
- Email change confirmation :** You should generally always force users to confirm email address changes via an extra step where a confirmation link is sent to the user.
- Writable config.php :** The config.php file must not be writeable by the web server process. If it is, then it is possible for another vulnerability to allow attackers to rewrite the Moodle code and display whatever they want.
- XSS trusted users :** Make sure that you trust all the people on this list: they are the ones with permissions to potentially write XSS exploits in forums etc.
- Administrators :** Review your administrator accounts and make sure you only have what you need.
- Backup of user data :** Make sure that only roles that need to backup user data can do so and that all users who have the capability are trusted.
- Default role for all users :** This checks that the registered user role is defined with sane permissions.
- Guest role:** This checks that the guest role is defined with sane permissions.
- FrontPage role:** This checks that the FrontPage user role is defined with sane permissions.

## VI. CONCLUSION

In order to support actual development of e-learning applications, the aspect of security should be considered primordially. In this paper we described some security aspects of e-learning platforms, we also reveals the prevalence of internal cyber-attack as well as a lack of proper IT policies and procedures in e-Learning systems, in light of their standard architecture and their specific security requirements. An examination of some security aspects of one of the most popular open-source e-learning systems: Moodle was given at the end. To conclude ; a system needs to implement security services such as authentication, encryption, access control, managing users and their permissions, as conclusion a secure learning platform should incorporate all the aspects of security and make most of the processes more transparent to the teacher and the student.

In future, the concept of m-learning will come in new electronically learning features, however new risks will also occur parallel with M-Learning.

## REFERENCES

- [1] Webber, C. G., Lima, M. D. F. W., Casa, M. E., & Ribeiro, A. M. (2007). Towards Secure e-Learning Applications: a Multiagent Platform. *Journal of Software*, 2(1), 60-69.
- [2] Anderson, J. R. (2001). Security engineering: A guide to building dependable distributed systems. John Wiley & Sons Inc.
- [3] Pfleegler, C. P., & Pfleegler, S. L. (2006). Security in computing (4th ed.). Prentice Hall Poggi, A., Rimassa, G., Tomaiuolo, M. (2001). Multi-user and security support for multi-agent systems. In Proceedings of WOA 2001, Modena, Italy (pp. 13-18).
- [4] Rosenberg M.J. (2011): E-learning strategies for delivering knowledge in the digital age, p.36. McGraw
- [5] Hill, New York. Harris, S. (2010). CISSP all-in-one exam guide (5th ed.). McGraw Hill.
- Heckman, C., Wobbrock, O. J. (1998). Liability for autonomous agent design. In Proceedings of the 2nd international conference on autonomous Agents, Minneapolis (pp. 392-399).
- [6] Rabai L. B. A. and Rjaibi N. (2012). Quantifying Security Threats for E-learning Systems.Education and e-Learning Innovations (ICEELI), 2012 International Conference, Tunis, Tunisia,July,2012
- [7] Zuev, V. I. (2012). E-learning security models. *Management Information Systems*, 7(2), 024-028.
- [8] Weippl Edgar R. "Security in E-Learning", Springer Publication ,2008
- [9] Ms. Ankita Chopra1 , Ms. Aakanksha Chopra2 –(2016), Application of Educational Data mining Techniques in E-Learning Systems with its Security Issues: A Case Study, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016
- [10] Chen Y. and He W. (2013). Security Risks and Protection in Online Learning: A Survey. *The International Review of Research in Open and Distance Learning*, 2013.
- [11] Sood S. K. (2012). Phishing Attacks: A Challenge Ahead. elearning papers, April 2012.
- [12] Martínez, T. S., Duráes, D. A., & Lucena, F. J. H. (2016). Management Conflicts in E-Learning Environment: Vulnerabilities in E-Learning Environments. *Interdisciplinary Perspectives on Contemporary Conflict Resolution*, 296.
- [13] Aakanksha Chopra (2013), "Comparative Analysis of Key Exchange Algorithms In Cryptography and its Implementation," *Journals of Innovations- IMS Noida*, vol. VIII, issue-2, Dec 2013, Print ISSN: 0974-7141, Online ISSN: 0976-1713.
- [14] Zamzuri Z. F. et al.(Eds.): "Computer Security Threats Towards the E-Learning System Assets" Communications in Computer and Information Science, 2011, Volume 180, Part 3, 335-345.
- [15] Nikhilesh Barik1 and Dr. Sunil Karforma2 (2012), "Risks and remedies in e-learning system"
- [16] Onyesolu, M. O., Ejiofor, V. E., Onyeizu, M. N., & Ugoh, D. (2013). Enhancing Security in a Distributed Examination Using Biometrics and Distributed Firewall System. *International Journal of Emerging Technology and Advanced Engineering*, 3(9), 65-70.
- [17] Sayed, B., Traore, I., Woungang, I., & Obaidat, M. S. (2013). Biometric authentication using mouse gesture dynamics. *IEEE Systems Journal*, 7(2), 262-274.
- [18] CorinaS avulescu , Zdzislaw Polkowski ; Deaconescu Ionu Cosmin ; Blidaru Cătălina Elena (2015), "Security in e-learning systems",ECAI 2015 - International Conference – 7th EditionElectronics, Computers and Artificial Intelligence 25 June -27 June, 2015, Bucharest, ROMÂNIA
- [19] Graf, F. (2009): Providing security for e-Learning. *Computers & Graphics*.
- [20] Najwa Hayaati Mohd Alwi and Ip-Shing Fan (2010), *Information Security in eLearning: A Discussion of Empirical Data on Information Security and eLearning*, Proceedings of the 5th International conference on e-learning 2010, Penang, Malaysia, 12-13 July 2010, Pages 282-290.
- [21] O'Neil, H. F., & Perez, R. S. (2013). *Web-based learning: Theory, research, and practice*. Routledge.
- [22] Bansal, P., & Sharma, D. M. (2015). DESIGNING ISSUES FOR E-LEARNING MODULES IN CLOUD PLATFORM. *International Journal of Information Technology & Computer Sciences Perspectives*, 2(3), 653-656.
- [23] Masud, M. A. H., & Huang, X. (2012). An e-learning system architecture based on cloud computing. *system*, 10(11).
- [24] Fallon, C., & Brown, S. (2016). *E-learning standards: a guide to purchasing, developing, and deploying standards-conformant e-learning*. CRC Press.
- [25] Mason, R., & Rennie, F. (2006). *Elearning: The key concepts*. Routledge.
- [26] Kambourakis G, Security and Privacy in m-Learning and Beyond: Challenges and state-of-the-art. *International Journal of u- and e- Service, Science and Technology*, Vol. 6, No. 3, June 2013.
- [27] S. H. Hasan, D. M. Alghazzawi, and A. Zafar "E-Learning systems and their Security" BRIS Journal of Adv. S & T (ISSN. 0971-9563) vol.2, no 3, pp. 83-92, 2014
- [28] Rodolfo Carneiro Cavalcante , Ig Ibert Bittencourt , Alan Pedro da Silva , Marlos Silva , Evandro Costa , Robério Santos, A survey of security in multi-agent systems, *Expert Systems with Applications: An International Journal*, v.39 n.5, p.4835-4846, April, 2012
- [29] Hasan,S. H., Alghazzawi, D. M., & Zafar, A. (2014). E-Learning systems and their Security. *BRIS Journal of Adv. S & T (ISSN. 0971-9563)* vol. 2, 83-92.
- [30] Muhsen, Z. F., Maaita, A., Odah, A., & Nsour, A. (2013). Moodle and e-learning Tools. *International Journal of Modern Education and Computer Science*, 5(6), 1.
- [31] M. Berry, An investigation of the effectiveness ofMoodle in primary education, in Deputy Head, 2005, Haslemere.
- [32] Asmaa.k, Najib.E (2016) , Towards a new access control model based on Trust-level for E-learning platform , *International Journal of Advanced Computer Science and Applications (IJACSA)*.
- [33] Sheo Kumar, Kamlesh Dutta(2011), investigation on security in lms moodle,International Journal of Information Technology and Knowledge Management January-June 2011, Volume 4, No. 1, pp. 233-238

# Classification of households after a traumatic shock, with the aid of Bayesian Networks: example of the post-electoral crisis in Côte D'Ivoire.

SAHA Kouassi Bernard<sup>1</sup>, BROU Konan Marcellin<sup>2</sup>, BABRI Michel<sup>3</sup>, OUMTANAGA Souleymane<sup>4</sup>

Institut National Polytechnique, Félix Houphouët-Boigny de Yamoussoukro (Côte d'Ivoire)

**Abstract:- Classification is a branch of multidimensional descriptive statistical analysis. This field of study has been the subject of several publication works. For the last couple's years, it is facing a renewal and a remarkable development with the multiplication of data .This situation requires, a deep analysis before the adoption of probabilistic model as suggested by the results. In this paper, we intend to study the social resilience and the vulnerability of urban populations' .Owing to the high concentration rate of population in big cities and the subsequent increase of modern plagues like rural exodus, galloping and blind urbanization with such corollaries as the creation of precarious districts and at times upper-crust in high-risks zones. So, within the framework of this study , we propose a deep analysis of data in general , the classification of Ivorian households according to their income , dwelling place after the shock of the social, political and the military crisis .This classification study should confirm or invalidate the opinion according to which the crisis was salutary to some people and a disaster for others, by causing a delay in the development of the country. Also through a modelling of the data collected on households made vulnerable by the post electoral crisis, in the form of Bayesian multidimensional models.**

**Index Terms:-** Bayesians networks, HIV-AIDS, Household, Resilience, Traumatic Shock, Post electoral crisis, Vulnerability.

## I. INTRODUCTION

**R**esilience is a concept that belongs to the physics of metals, transferred in social sciences, notably in economy and in psychology [1]. The literature on the concept is much diversified. Etymologically, it means to resist and to rebound in front of a significant and persistent adversity. Up to now, no consensus has been found as far as the definition of the concept is concerned because those offered are linked to cultural considerations and vary for that reason with societies and also time. So, to better explore the domain of the social resilience which is in full development in African countries after crises due to the changes of regimes or in the aftermath of natural

disasters and epidemics, the issue of resilience we will be the focus of our study in this paper. Several non-governmental organizations working in the health sector and in structures struggling against HIV-AIDS, are looking for tools which can help in reliable decision making, but do not have the necessary elements to reach their objectives. The discovery of the knowledge in databases and documents can help non-governmental organizations (NGOs) as well as professionals to transform their basic data into strategically information. Those organizations, which will take advantage of those techniques, will notice that they can help reduce their running costs while improving their quality and a quick decision making aid and a clear vision of the horizon. To reach this target we will make a state of the art on the definition of Community base-resilience and an evaluation study on techniques, existing research works on resilience and the classification of communities. The Challenge of the application of Bayesians Networks in Community base resilience will be also discussed, to finish we will conclude by opening perspectives for future works.

## 1. The review of literature

One of first authors to speak about social resilience in his writings was Pelling in 2013; Obristen in 2006. In their respective works, these authors agreed to support the thesis according to which « Social institutions which organise the distribution, access and use of resources at the level of households are the key elements of resilience », Pelling introduced the concept of potential adaptation in 2003 on page 67, it consists in describing actions which are based on the socioeconomic advantages to improve the resilience of populations. This idea was indeed vague but it was a track towards social resilience. Moreover, other researchers such as Ostrom 2004 also made a study on the actors of resilience. However , in this part, we will introduce the different authors who were interested especially in métá-analysis and in the development of a model based on the actors in other words, a computer based model of the social actors as progress system of autonomous interactive actors so as to test the empirical hypotheses of some researchers. Also, in this article our objective is to offer a contribution in the modelling of social resilience through statistical and computer tools between the community approaches of resilience as a general fact and family resilience in particular. Besides, we shall outline the methods of analysis of data, and even a comparative study with a view to adopt the best one among the ones offered. The approach which will be kept must take into account the dimensions of family resilience.

## 2.1. Different types of modelling of Community resilience

### 2.1.1. The definition of impact strength

Resilience according to the researchers [10] is a set of personal characteristics of the individual (or of a group of individuals), a process and a result. It fits into the approach of training, empowerment and of self-determination across which the individual reinterprets the signification of the situation of adversity and reorientates positively the sense of her life so as to follow her development, while reinforcing her personal or environmental protection factors with, the situation of adversity as new organizer of the individual. So according to another researcher «to resile is to recover, to rebound, to go forward after an illness, a traumatism, a stress [2]. It is to overcome crises of existence that is to say to resist them, then to overcome them and keep on living in the best way possible».

### 1.1.2 The definition of Community or family resilience

The family is the smallest community; it can be defined as a dynamic system which is persistently created and recreated and not a collection of people [3]. As for Community resilience it is defined as the capacity of a community to respond to adversity, by searching a better functioning level. There are such concepts or approaches related to the community resilience such as: «Internet resilience ». This approach was developed by researchers, «*intimacy vs extremity* », and trans-community trans-generational resilience [5]. One of the last approaches of resilience according to the researcher Richardson is: «Resilience is the process of adaptation to *stressors* in adversity, changes and opportunities, which result in identification, reinforcement and enrichment of the protection factors, be they personal or environmental».

### 2.1.2. Analysis of the RIMA community resilience

In the literature, according to the RIMA model that was tested in Eastern Africa precisely in Kenya, it is based on the list of contextualized factors which enable to measure the vulnerability of households [6]. In those factors we can name the ones which make a household vulnerable in front of a specific effect such as : Income (R) access to food (IFA), Access to basic services (ABS), social welfare system (SSN), favorable institutional Environment (EIF), Stability(ies), Capacity of adaptation (AC), Agricultural Assets (AA), non-agricultural Assets (NAA), Practices& agricultural technologies (APT), climate Change (CC). So, according to this model Vulnerability is an (f) function of the exposure of a family to risks and its resilience ( $R_i$ ) opposite those risks [15] :

$$V_i = f(Exposition_{risque}, R_i) \quad (1)$$

According to the UMI resilience, vulnerability is a dimension of resilience. This approach not only studies how disturbances and changes can influence the structure of a system, but also how its functionality can change in order respond these needs.

Resilience is a key factor to determine the vulnerability which would show to what extent: the access of a household to food security, health care, is affected by a traumatic shock. The food security of a household ( $Y_i^1$ ) is a function composed of three main factors before a shock occurs. The probability to be attained by a shock due to geographical situation ( $Pr_g^0$ ); and the likelihood to suffer from a shock due to the characteristics of the means of existence of a household ( $Pr_{se}^0$ ); and the resilience ( $R_i^0$ ) of this household that can be translated by the following equation (2).

$$Y_i^1 = f(Pr_g^0, Pr_{se}^0, R_i^0) \quad (2)$$

In spite of the different approaches we offered in a previous paper which is already published the new approach of the social resilience which takes into account all aforementioned factors to those dimensions of resilience. This new orientation of resilience is meant to be Communitarian. It studies the resilience of the individual first, then his environment and introduces the notions of conditional probability by the use of Bayesians network [16].

## 2. Analysis of the post electoral crisis relating shock

The post electoral crisis in Côte d'Ivoire was a shock, and source of several physical and psychological acts of violence, also it brought about the pauperization of several Households. This section will study the intensity of the shocks suffered by households during ten years of the crisis. It will be also a matter to make a comparative analysis of the prevalence of shocks according to the initial place of residence of households. In term of contribution we offer a model which relies on the model of Richardson and adapts it to the reality of Households worldwide and in Côte d'Ivoire especially (figure 1).

### 3.1.1. Definition of an indicator

An indicator is a tool of Evaluation and aid of decision making, it serves in providing indications, information on the value of a quantity. Its role consists to sum up complex information, to assess the performance of an organization, its politics of progress, its tendencies, to identify sectorial mechanisms. This indicator must indeed reflect the variations of what is supposed to synthesize or to measure. He must highlight links between the different components of the system or of the ecosystem.

<sup>1</sup> Website: <https://dslpitt.org/genie/>

#### 4.1. The modelling of the resilience of households

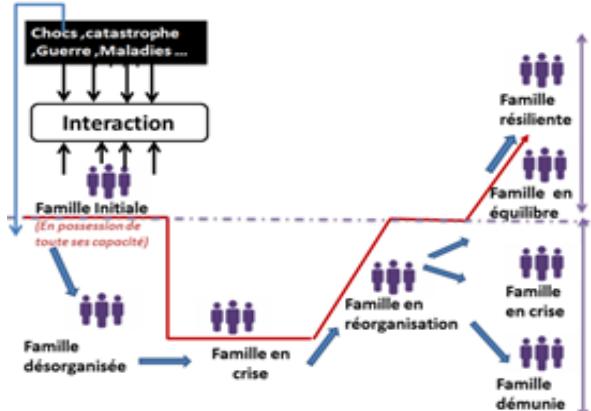


Figure 1: Proposal of model of resilience of family inspired by the card of resilience and by the model of Richardson

#### 4.2. Indicators in résiliométrie household

Within the “résiliométrie<sup>2</sup>” the households use indicators to measure the dimensions of social resilience: capacity of (evaluation of the personality), result (social adaptation, welfare, quality of life, social competence, social participation, depression, aggressiveness...) process (strategies of adaptation).

#### 4.3 Dimensional modelling of the data of household resilience

In Côte d'Ivoire the problems of the vulnerable households or family sheltering orphans and vulnerable children (OVC) care taking led many Organizations to set up systems of monitoring and evaluation of the devices of their care taking [16]. Very often care taking politics emanates from international organizations. The following table introduces the different dimensions included in the device of monitoring and evaluation of vulnerable household targeted by the NGOs Manassé. This NGO is a branch of international charitable organization HOPE with a section in Côte d'Ivoire called (Hope Côte d'Ivoire).

Table 1: CSI<sup>1</sup> monitoring dimension and evaluation of the Households

DIMENSIONS	ATTRIBUTES	MODALITIES
Feeding and nutrition	Food Security	Good, Medium, bad; Very bad
	Growth and Nutrition	Good, Medium, bad; Very bad
Accommodation and care	Accommodation	Good, Medium, bad; Very bad
	Care	Good, Medium, bad; Very bad
Protection	Abuse and working	Good, Medium, bad; Very bad

<sup>1</sup>CSI: Children Record status Index (index of valuation of the status of the child and household)

	Legal protection	Good, Medium, bad; Very bad
Health	Health	Good, Medium, bad; Very bad
	Health services	Good, Medium, bad; Very bad
Psychosocial	Emotion	Good, Medium, bad; Very bad
	Social behaviour	Good, Medium, bad; Very bad
Education and performance	Education	Good, Medium, bad; Very bad
	Performance	Good, Medium, bad; Very bad

The table introduces the structure of monitoring information - evaluation enabling the appreciating the level of vulnerability and the resilience of households. On the basis of this one, it is possible to construct an improved dimensional model, adapted to the installation of a data warehouse of monitoring and evaluation of households. In fact, the dimensional modelling, notably the star diagrams is well known for its effectiveness in the development of solutions for decision making. It is easily exploitable for the development of applications of reporting and performance indicators. The star diagrams based on table is given by the following figure:

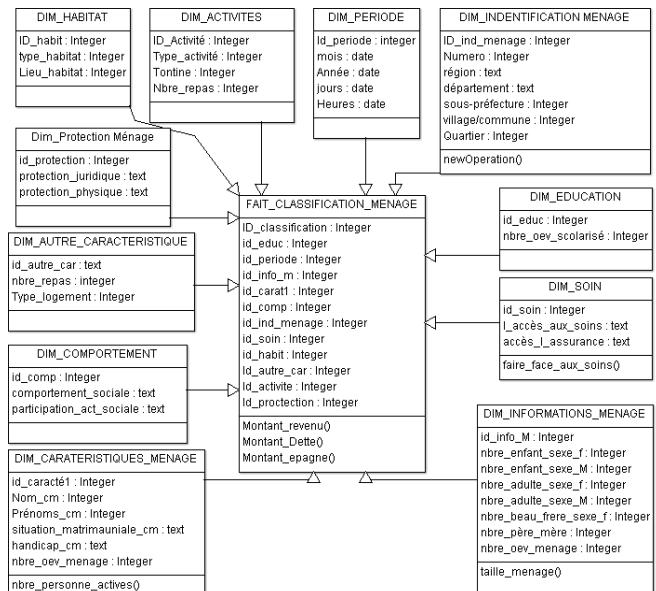


Figure 2: Proposal of star model for household resilience

To access the implementation of the star diagrams, it is necessary that the diagrams of the data base be such as currently used (spread sheet) be converted into the format of the data warehouse by an ELT programs. On top of that, the quality of stored data must be included in the practical implementation data warehouse. To have data of very good quality, it is always possible to extract useful data for a multidimensional search, by accomplishing, by the way, all necessary corrections for their exploitation. The interest of having a data warehouse is to

be able to study regularly the resilience of households to optimize the used of management politics.

According to Richardson: « Resilience is the process of adaptation to stressors in the adversity, changes and opportunities, resulting from identification of the protection factors enrichment and reinforcement be they personal or environmental» [6]. Several non-governmental organizations in the health sector and other structures fighting against the HIV-AIDS have data in abundance, but have not the needed information for good decisions making. The transformation of data coming from known databases in their decision-making process can help those organizations to optimize. It is the reason why, in the field of research on resilience, *Résilométrie*<sup>2</sup> is introduced into so as to make up for the expectation of modelling techniques adapted to the processes of resilience. In the particular case of households, analytical need requires a method easy to update and giving a simple approach for an efficient simulation. In such context, the technology of the Bayesians network is an ideal approach due to its qualitative and quantitative character [9].

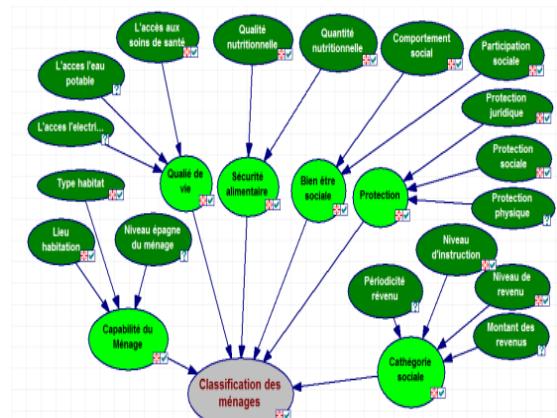
### 4.3. Modelling of the resilience of household by Bayesians networks

Also called probabilistic expert Systems, Bayesian networks are tools of representation of knowledge and of automatic reasoning on this knowledge. They were introduced by Judea Pearl in 1980s and revealed themselves as powerful and very practical tools for the representation of uncertain knowledge and reasoning from incomplete information [4]. Bayesians networks are simulation tools allowing to observe the behavior of a complex system in contexts and conditions which are not necessarily accessible to experimentation. Technically, speaking Bayesians networks are graphic models combining the theory of graphs and the theory of probability. Besides, a Bayesian network is an orientated graph in which nodes represent variables and arches represent the relations of dependency between those different variables. Each node is endowed with a table of conditional probability which is a modelling of beliefs in the occurrence of this or that case when we are in any condition. In the case of the modelling of the process of identification of better actions of resilience, such graph translates identified actions and variables of decision and subsequent actions. This graph depends on envisaged politics and the structure of corresponding interaction which can vary from a study to the other. To model the process of understanding of the resilience of Households, let us note:

- Let  $X = (X_i)_{1 \leq i \leq N}$  be a set of the different attributes taken into account in the system of monitoring and evaluation of the household of CSI, all merged dimensions;
- $K_i$  the number of modalities of attribute  $X_i$  ;
- $\mathcal{P}(X_i)$  The set constituted by variables  $X_j$  and by node parents of  $X_i$ .

<sup>2</sup>The résilométrie: the science which studies social

Attributes  $X_i$  are linked between themselves by causal relations. The following graph gives the structure of relations which maintain different attributes [11] [12] [13]:



Represent 3: Structure of the household of Bayesian network

This graph is an intuitive representation of the process of resilience which governs households on the basis of information (attributes) considered useful by CSI to appreciate the resilience of those one. This graph of dependency constitutes the qualitative part of the corresponding model of Bayesian networks.

On a formal level, a Bayesian network is a couple  $(G, \Theta)$  with:

- $G$  a graph orientated without cycle;
- $\Theta$  a distribution of probability defined on the variables of graph;
- every node of  $G$  is associated with a random variable and a single one;
- Let  $\{X_1, \dots, X_N\}$  be the set of random variables (nodes of graph). The graph without cycle obeys the following property:

$$\Theta = \mathbb{P}(X_1, \dots, X_K) = \prod_{i=1}^K \mathbb{P}(X_i | \mathcal{P}(X_i))$$

A Bayesian network is therefore completely described when they have, in each of its nodes, the conditional probability of this node knowing each of its parents. Any Bayesian network obeys on the condition of Markov, that is, in a Bayesian network, every node is independent conditionally on his descendants, when he knows its parents. In the case of Households, the inference in Bayesians networks consists in calculating the probability

$$\Theta = \mathbb{P}(X_1, \dots, X_K) = \prod_{i=1}^K \mathbb{P}(X_i | \mathcal{P}(X_i))$$

In practice, these calculations are performed thanks to the following rule of the chain of conditional probability:

$$\forall K \in [1, N],$$

$$\mathbb{P}(X_1, \dots, X_K) = \prod_{i=1}^K \mathbb{P}(X_i | X_{i-1}, \dots, X_1)$$

Although conditional probability can be given by experts of monitoring and evaluation of households, the fact for them to

have stored data helps them to get more definite estimates on those probabilities. A data based training of parameters is therefore performed. The data base training of the parameters of Bayesians network consists in estimating the unknown parameter  $\Theta$  on the basis of data  $\mathcal{D}$ . The choice of structure of  $\Theta$  depends upon the law of probability which is supposed to govern the generation of the data  $\mathcal{D}$  collected. For that, once a distribution of probability (Poisson's law, Gaussian's law, etc.) Is postulated, they look for the value of  $\Theta$  which maximizes the probability function defined below:

$$\mathcal{L}(\Theta, \mathcal{D}) = \mathbb{P}(\mathcal{D}|\Theta) = \prod_{i=1}^M \mathbb{P}(x_i | \Theta)$$

In practice, due to the importance of Bayesians networks, several software's have been developed for their implementation and among them a significant number is free. The basics uses of the Bayesians network consist, once the parameter are estimated, to simulate the consequences, a certain number of choices over other actions and the variables taken into account in the developed model. GeNIs software<sup>3</sup>, is a software of Bayesians network complete and especially free as well for research works as for commercial works. In fact, GeNIs, is a free software dedicated to Bayesians network and to its extensions (Bayesians dynamic networks and Diagrams of influence). It includes a very large number of training algorithms for parameters as well as data structure. It also has a very convivial interface and can easily be used by non-specialists in modelling, notably psychologists, environmentalists, economists, sociologists, and epidemiologists etc.

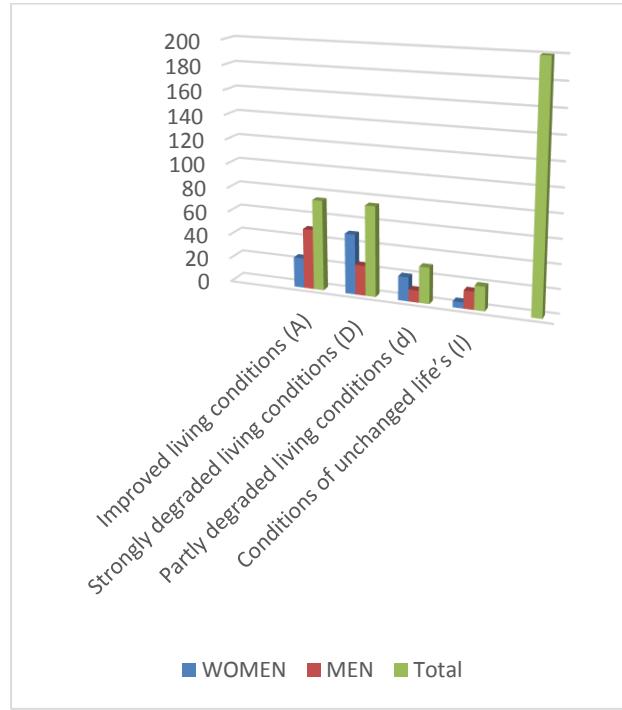
### 3. Experimentation

In order to assess the performances of our contribution we are going to choose a database of an experimental sample of 200 households already identified by the NGO Manassé with a view to classify them. This experiment must be done with two methods. The first is manual technology used by NGOs and the second is the method of automatic classification which we offer. This method comes in support to the first method. It is necessary to note that the sample submitted to the study is based on the head of the households. We are going to simulate our offered method on two hundred (200) households represented by their heads. They may be of the two sexes Male (M) or Female (F) that is a hundred households each. So the results of different simulation will be recorded in a table. This will allow us to test the effectiveness of the model family resilience offered in our approach. Our sample was obtained thanks to the tool of CSI,

**Table2:** summary of CSI evaluation results

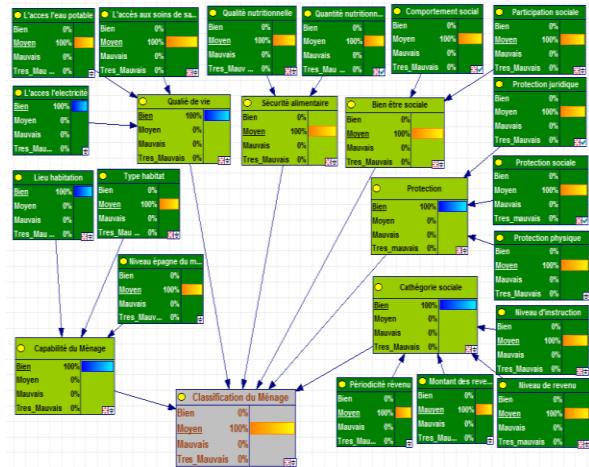
Domestic number having changed status in reference to CSI	WOMEN	MEN	Total
Improved living conditions (A)	25	50	75
Strongly degraded living conditions (D)	50	25	75
Partly degraded living conditions (d)	20	10	30
Conditions of unchanged life's (I)	5	15	20
			200

**Table3:** Bar chart of results of the valuation of CSI



In the graph illustrating the results of CSI inquiries, we see that after the situation of the dramatic post electoral, and political crisis in Côte d'Ivoire, most of the people experienced the deterioration of their living conditions. Also this degradation was more acute for households led by women than those led by men. For a better understanding and a precise interpretation of the results we will rely on the Bayesian model previously mentioned in (figure3). Yet, we can assume that households having a woman as leader suffered much more than others, definitely because women are more vulnerable in the society. Moreover, let us say that though the crisis was dramatic for the large majority of the population, there is a tiny part of the population whose living conditions improved. In addition ,to the comparison made with the Bayesian approach it is not possible to highlight such detailed aspects like the sex of the

victims of the shock, this explains the importance of our approach which uses the Bayesian network in order to try through probability conditional indicators how to appreciate other aspects [14].



**Face 4:** Structure of the Bayesian network of a balanced household with GeNies

Our classification technique gives us interesting details that we can observe through this illustration. So we have the possibility to detect quantitative and qualitative aspects to be reinforced at this household. These different aspects were obtained thanks to different probabilities generated by the values of modalities of attributes of dimensions of social resilience, extracted in the data warehouse in the consolidated databases. All things considered we can state that if the crisis was advantageous for a tiny part of the population, in general, dimensions such as *the accommodation, welfare, social security reveal a great degradation of the living conditions of the majority of households.*

#### 4. CONCLUSION

Within the framework of household care taking politics, the storage of information collected in a data warehouse will improve significantly not only the management of those data but also their exploitation in decision making, and especially in the comprehension processes of household resilience . Within the framework of the study of resilience in general and family social resilience in particular, Bayesian networks are all the more appropriate as they are adapted in situations where one is confronted with incomplete, imprecise and uncertain data .The use of the Bayesian simulation in the management of households will enable organizations who use CSI to better understand the processes of resilience by simulating the impact of modifications of one or several attributes of CSI over the other, owing to available data. Therefore professionals will benefit from both quantitative and qualitative aspects of their study. In the next articles, we intend to develop the approach of resilience and the community vulnerability through online social networks.

#### REFERENCES

- [1] Saha Kouassi Bernard, Achiepo Odilon Yapo, Brou Konan Marcellin, and al. Storage and Bayesian modelling of data one the social resilience: Hut of Orphans and Vulnerable Children (OVCS) in Côte d'Ivoire. International Newspaper of Computer Science Exits (IJCSI), on 2015, flight. 12, N. 4, p. 137.
- [2] Manciaux, M. (2001). Impact strength: resist and be built. Ed. Medicine & Hygiene.
- [3] Anaut M. (2006). Family impact strengths or résilientes families? Reliance19 (1), 14-17.
- [4] Gauvin-Lepage, J., Lefebvre, H., & Malo, D. (2014). Family resilience: defining the concept from has humanist perspective. Interdisciplinary Newspaper of Family Studies XX (2), 22-36.
- [5] Bernard Michelet, PH. D. CRDP Intervalley Girafe-Crir; 2nd annual symposium of CRDP Intervalley Impact strength and readjustment a history to be followed.
- [6] Dumont, Michelle, Leclerc, Danielle, Assembled, Line, and al. Programs of management of stress teenagers as lever of impact strength. Resilience, regulation and quality of life: concepts, assessment and intervention, on 2009, p. 301.
- [7] Békaert, J, Masclet, G. and Caron, highway. The instruments of measure of impact strength at the teenagers having confronted with a traumatism: a magazine of literature. In : Annales Médico-Psychologiques, psychiatric magazine. Elsevier Masson, on 2011. p. 510-516.
- [8] Martin-Breen, P. & Andries M. 2011. Resilience: IN literature review. New York USA, City University of New York and Tucson USA, Arizona State University. Available at following address:<http://www.rockefellerfoundation.org/news/publications/reresilience-literature-review>.
- [9] Anaut, Mary. Family impact strengths or resilience's families? Revival, on 2006, flight. 19, N. 1, p. 14-17.
- [10] Manciaux, Mr (2001). Impact strength, myth or reality? Social Notebooks Medico, 9-10.
- [11] Gauvin-Lepage, J. (2014). Co-building of the elements of an intervention programs in support of the impact strength of families which a teenager is attained of a crano-cerebral traumatism.
- [12] Philippe Leray, Bayesians networks: training and modelling of complex systems, HDR, on 2006.

- [13] P. NaimP. Leray, and al, Bayesians networks, 3rd edition, Eyrolles on 2007.
- [14] Antoine Cornuejols Laurent Miclet Artificial Training, Concept and Algorithm, 2nd edition, Eyrolles on 2010.
- [15] FAO, in 2010, the state of food insecurity in the world 2010: Fight food insecurity during extended crises. Rome, Organisation of the United Nations for feeding and agriculture and worldwide food Programme;
- [16] Kulig J. C., Edge D., & Joyce, B. (2008). Community resiliency ace has measure of collective health status: perspectives from rural communities. *CJNR (Canadian Journal of Nursing Research)*, 40(4) 92-110j.

**First A. Author, SAHA Kouassi Bernard** Is a Computer sciences Engineer (Agitel-Formation Abidjan, Côte d'Ivoire) and a Master degree holder in Computer Science with specialization in Industrial Computer sciences and Business Intelligence (University Nanguui Abrogoua). He is a Ph-D student in Mathematics and Information Technologies (EDP INP-HB Yamoussoukro, Côte d'Ivoire). He is also a Teacher-researcher at the University Felix Houphouet-Boigny, Côte d'Ivoire (ENS), and member of the Research Laboratory in Computer Sciences and Telecommunications of Houphouet-

Boigny National Polytechnic Institute (INP-HB), Abidjan, Côte d'Ivoire. His interests of the research include The Data Mining, and. his works are centered on their research of the database and programming languages.

**Second B. Author, Konan Marcellin BROU** is Doctor in Computer Science and Teacher researcher at the Houphouet-Boigny National Polytechnic Institute (INP-HB) of Yamoussoukro (Côte d'Ivoire). He is the Director of the Department of Mathematics and Computer Science. He is a Member of Laboratory in Computer Sciences and Telecommunications (LARIT/INP-HB) Abidjan Côte d'Ivoire. His interests are information systems, database and programming languages.

**Third C. Author, BABRI Michel** is Doctor Computer Science and Teacher researcher at the Houphouet-Boigny National Polytechnic Institute (INP-HB) de Yamoussoukro (Côte d'Ivoire). He is the second Director of the Laboratory in Computer Sciences and Telecommunications of Houphouet-Boigny National Polytechnic Institute (LARIT/INP-HB), Abidjan, Ivory Coast. His interests to Network security and Telecommunications

**Fourth D. Author, SOULEYMANE Oumtanaga** is a Professor in Computer Science and Teacher researcher at the Houphouet-Boigny National Polytechnic Institute (INP-HB) de Yamoussoukro (Côte d'Ivoire). He is the Director of the Laboratory in Computer Sciences and Telecommunications of Houphouet-Boigny National Polytechnic Institute (LARIT/INP-HB), Abidjan, Ivory Coast. His interests are Futur Network systems, Network security and Telecommunications.

# Secure Approach for Net Banking by Using Fingerprint Authentication in Distributed J2EE Technology

Rachid ALAOUI<sup>1</sup>, Khalid ABBAD<sup>2</sup>, Ahmad EL ALLAOUI<sup>3</sup> and Moulay Abdellah KASSIMI<sup>4</sup>

<sup>1</sup>Laboratory of Systems Engineering and Information Technology (LISTI), ENSA, Ibn Zohr University Agadir, Morocco

<sup>2</sup>SIA Laboratory, FST, FSDM University, FEZ, Morocco

<sup>3</sup>ENSA AL Hoceima and Labo MATSI Mohammed I University OUJDA, Morocco

<sup>4</sup>LGEMS Laboratory, ENSA, Ibn Zohr University Agadir, Morocco

**Abstract:** Today, Net Banking or Internet Banking System is popular technology typically used by individuals to carry out a variety of personal and business financial transactions and banking functions by using mobile technology. Net Banking is used to describe banking transactions through internet application. But there are many security problems like fraudulent websites, fake emails from banks, capturing user IDs and passwords, hacking personal bank accounts and ATM card etc. Security and Authentication of individuals is necessary for our daily lives especially in net Banking. It has been improved by using biometric verification techniques like fingerprints. This research paper gives a security solution mobile through a new model with biometric recognition and SMS service.

**Keywords:** Secure Internet banking, Smartphone, Fingerprint, Banking transaction.

## 1. Introduction

Nowadays, in the self-service banking system has got extensive popularization with the characteristic of offering high-quality 24 hours service for customer. Internet Banking is not only focused on transferring money, but also to conduct many banking transactions with minimum time [Nsouli S et al, 2002]. Every customer can get connected to his bank's website with android smartphone and browser. However, many hacking process is done in internet banking. To avoid these problems, a new model has been developed for secure internet banking with biometric recognition and SMS mobile service. Once user get internet banking access permission, user can access different types of transaction such as balance enquiry, transfer of funds, online payment of bills, accrued interest, fees and taxes, transaction details of each account. The banking services include bill payment, transferring amount, recharging mobile phones, online applications, online purchase, maintaining accounts [Basel Committee Report, 1998]. In existing internet banking, user need to register with bank for accessing internet banking and then bank will provide a user ID and password (PIN) to user. Then, user can login through bank website with user ID and password. If user enters correct user ID and password, user can access to his bank account with internet banking. Some banks provide extra authentication process such as providing another security token code to user mobile phone through SMS message.

### Disadvantages of Existing method

- Internet banking use user ID and password of the user. In this system, There are possibilities of hacking keys or duplicated; signatures could be forged, passwords could be easily stolen or hacked by a specialist people.
- Encryption problems software is used to protect account information. However, there are no perfect systems. Accounts are prone to hacking attacks, phishing, malware and illegal activities.
- Learning – Banks with complicated sites can be cumbersome to navigate and may require one to read through tutorials to navigate them.
- complex transactions– face to face meeting is better in handling transactions problems. Customary banks may call for meetings and seek expert advice to solve issues.

## 2. Literature Review

Automated teller machine (ATM) is a mechanical device that has its roots embedded in the accounts and records of a banking institution [Sri Shimal Das et al , 2011]. Many established banks in developed countries began with ATMs and evolved through Personal Computer-banking, Telephone-banking, Internet-banking. Daniel [1999] explained that the increased competition due to new arrivals, electronic services and increasing security for banking systems considering e-banking. Khorshid and Ghaneh [2009] conducted a research about challenges of e-banking

and identified the problems such as maintaining privacy of customers, security and attaining customer trusts. Main challenges for development of Net banking on customer side arise due to reputation, laws and regulations.

To avoid all these accidental losses; banks and other institutions should enter biometric security and all our fears could be laid to rest. Biometrics security system simply allows identifying yourself by your inherent biological features like eye, finger prints. So fingerprint recognition is widely used due to its reliability [D. Maltoni et al, 2009]. It is widely used in forensic and commercial applications such as criminal investigation, ecommerce, unique ID cards (fig.1) and net banking [Heeseung Choi et al, 2011][ M.Sandeep et al, 2015].



Figure 1. ATM Transaction by ID cards and fingerprint recognition.

Fingerprint recognition is identified from impressions made by unique ridges on fingertips. The finger prints images given through the scanner and enhanced, then converted into a template. Most of the automatic systems use finger print recognition method for minutiae matching (figure 2). The split in the ridges, bifurcation, lake and termination in irregular pattern is called minutiae. In general, ridge ending and ridge bifurcation are used for fingerprint identification [Lin Hong,1998].

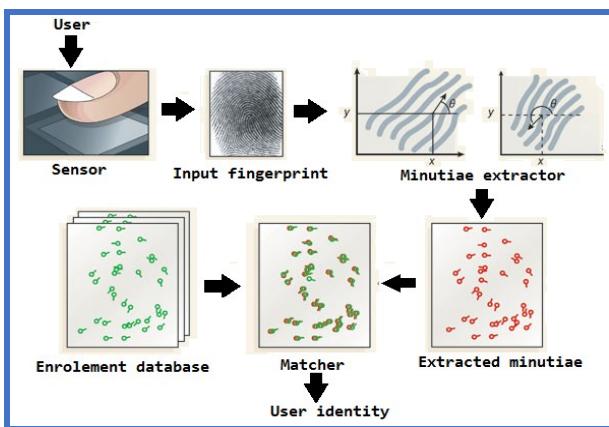


Figure 2. Registering a person in a biometric system.

With the help of sophisticated programming techniques, the websites which resides on a financial institution's network can be hacked by an attacker. Using this, he can access the bank's systems to locate the ATM database and hence collect card information which can be used later to create a clone card. A biometric system recognition provides more accuracy and secrecy than PIN. When a client approaches the branch for opening an account, he is asked to fill in with the questions. Along with the questions the fingerprint images are also collected in the branch.

### 3. The Existing Method for Net Banking

Internet banking identifies a particular set of technological solutions for the development and the distribution of financial services, which rely upon the open architecture of the Internet. With the implementation of an Internet banking system, the banks maintain a direct relationship with the end users via the web and are able to provide a personal characterized to the interface, by offering additional customized service.

Fig 3 explains the Internet Banking Security (IBS); the user should first enter User ID and password which will be verified in the bank website for authorization. If the user ID and password matches the user can login to internet banking system. Otherwise, "Invalid user" is reported to the user. If the user is valid, user can access to internet Banking processing such as balance enquiry, transfer of funds, online payment of bills, accrued interest, fees and taxes, transaction details of each account, accounts, credit card and home loan balances, transfer funds to third party accounts user nominate, open a deposit right from the terminal. The details of the transactions are finally stored in the Database:

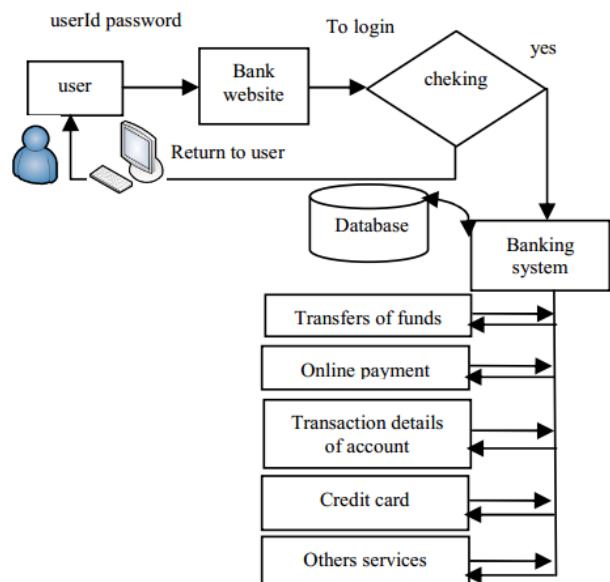


Figure 3 Existing model for Internet Banking system.

Another Approaches for Net banking security combines the usage of pin-number and mobile code [Collin Mulliner et al, 2013]. This validation bank provide extra authentication process such as providing another security token code to user mobile phone through SMS. In online banking web applications for example, the user has to authenticate himself via a valid username and password to initiate a transaction. Directly after this transaction request, the user gets an SMS message containing the One-Time Passwords OTP that must be additionally entered to authorize the transaction. In this application area the OTP is called a mobile Transaction Authorization Number (mobile TAN or mTAN). The Figure 4 presented the online service sends the OTP to the user's mobile phone via

the cellular network, and the user enters the OTP to authenticate or authorize a transaction.

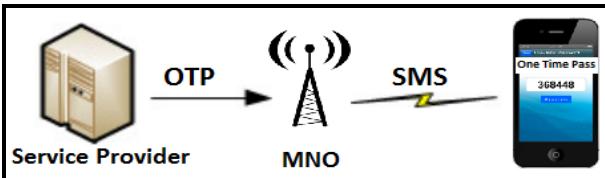


Fig. 4 SMS OTP Principle: The OTP is generated by the service provider and sent to the mobile network operator (MNO) that delivers the OTP via SMS to the user.

## 4. Using a Smartphone for Biometric Authentication

Taking into consideration accuracy and reliability among the various biometric system the most popular are the ones based on fingerprint matching. In Fig 5, the arrangement for sensors can be made in built in the existing smartphone like fingerprint sensor. This makes the mode of identification very attractive and easier. Due to its unique identity and easy accessing, the finger print identification has been increased in civil and law enforcement applications [Zain S. Barham et al, 2011][R.Mourya1 et al, 2015].

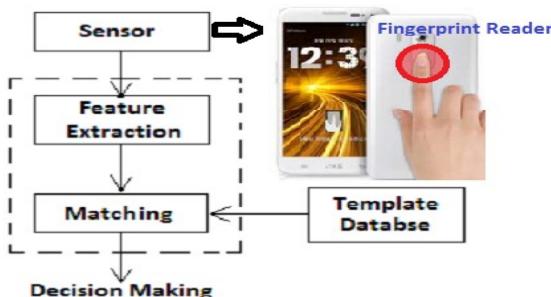


Fig.5 Block diagram of Biometric Process by a smartphone sensoring.

The purpose of this study is to identify security in mobile banking and to provide an authentication method for mobile banking transaction by using a biometric authentication.

## 5. Proposed Method of Internet Banking

Every group bank in order to be able to use the broker will have to subscribe to the services provided. The bank system comprises a module which were developed to demonstrate the full integration of the proposed authentication scheme. This module is an account management system used by the bank's workers (AdminBank) to carry out such management process as creating new accounts, setting up account's details and security levels, adding an additional holder to an existing account, and enrolling user's fingerprints.

In Internet Banking, the user should first enter User ID and password which will be verified in the bank website for authorization. If the user ID and password matches the user can login to internet banking system. Otherwise, "Invalid user" is reported to the user. At the same time user scans his fingerprint through scanner

and checked with fingerprint feature extraction and matching process (Fig.6). The Fingerprint image should match with banking database fingerprint. After that, the customer can access to interface Manager customer bank ATM. When the customer lost the ATM card, he can block the ATM card. After the fingerprint recognition success, a onetime password is generated during registration process. That password is sent to the user's mobile number for authentication. After validity, the user can access to interface Manager customer and start transaction. The details of the transactions are finally stored in the Database.

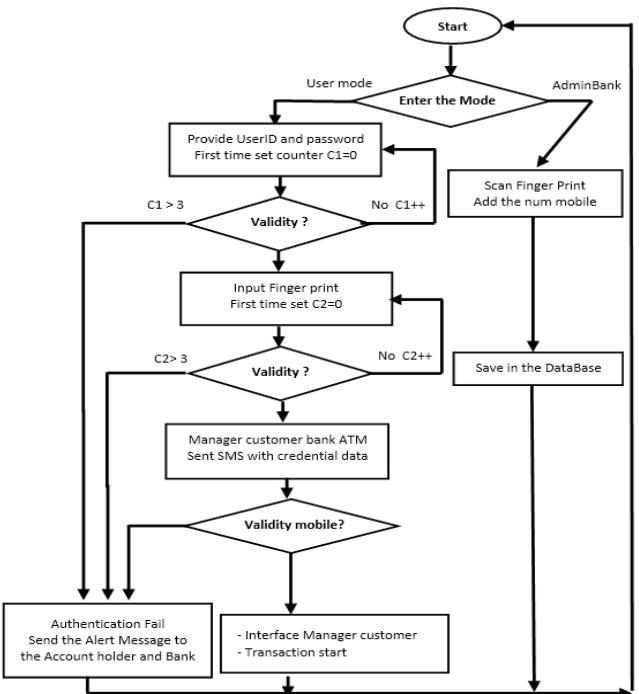


Fig.6 Architectural diagram of the proposed Net Banking system.

## 6. Algorithm of Proposed Model for Internet Banking System

Connect personal system bank website by a smartphone

```

1. [Entering into Internet Banking System]
2. [SET banking user id, password]
3. [Validate userID, password]
If bkuserid := userid and bkpasswd := password Then
    Enter into Internet Banking System
Else:
    Write : invalid user;
4. [Finger print recognition]
5. [Scan finger print] Read : fingerprint;
6. [Retrieve finger print]
Set USERfingerprint := fingerprint;
7. [Validate finger print ]
For i:= every valid user in system, do
    If db[i].fingerprint = USERfingerprint Then
        Enter into interface Manager customer bank ATM;
    If card ATM is losed check option blocked card;
    [end if]
        password is sent to the user's mobile number for
        authentication;
    Else
        Write : invalid user
        [end if]
    [end for]
```

## 8. [Validate SMS Mobile]

Enter into Interface Manager customer;  
Start transaction;

### 9. Exit

This algorithm can be used to develop a various number of applications for control access, internet banking or anything else that requires a great level of security

## 7. Architectural and Comparison of Existing Method and Proposed Model

The J2EE platform gives a multitiered distributed application model, the ability to reuse components, a unified security model, and flexible transaction control for a net banking architecture. The Figure 7 shows two multitiered J2EE applications divided into the tiers described in the following list. The J2EE application parts are presented in J2EE Components:

- Client-tier components run on the clients machine.
- Web-tier components run on the J2EE server.
- Business-tier components run on the J2EE server for Net Banking process.
- Enterprise information system (EIS)-tier software runs on the EIS server.

For leveraging the security, our J2EE architecture include more modules integration for secured Net Banking process. The Java Authentication and Authorization Service (JAAS) can be used for authentication and authorization of users to ensure they have the access control rights (permissions).

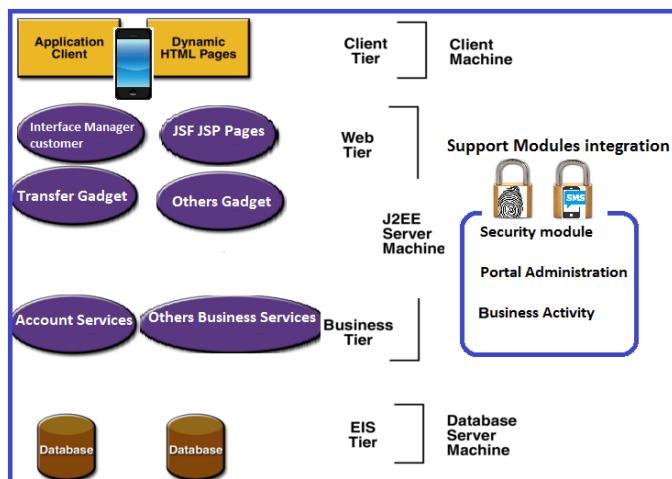


Fig. 7 J2EE Design Patterns for the Net Banking Architecture.

The figure 8 describe how they all work together to process an authentication request. For starters, the following sequence diagram shows the class interaction that occurs during a successful authentication and identifies the key participants and their activities. The Client requests access to a protected J2EE application. The J2EE application verifies the requests using the JAAS authentication Modules and then initiates authentication by forwarding the request to the biometric authentication server and mobile OPT validity.

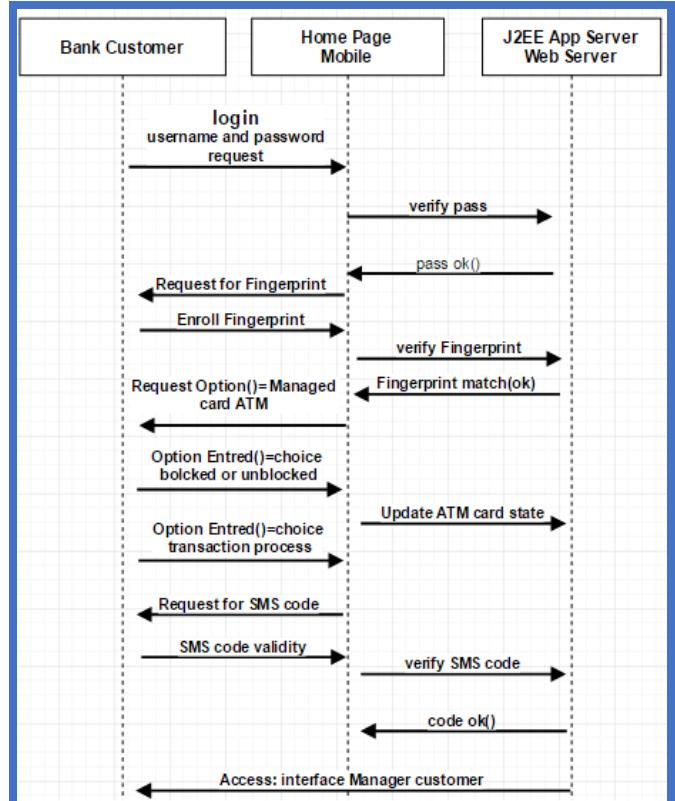


Fig.8 Sequence Diagram authentication process

One of the major problems with the authentication of users via the internet Banking is the inherent lack of security of traditional authentication techniques, passwords PIN numbers and cookies. With the current development of the biometric technology and mobile validity market (TABLE 1), the possibility of identifying someone online has been addressed. Our architecture allows a web page to include a validation check using objects embedded in the web page.

In the proposed solution, even if the mobile phone and card is lost the attacker gets hindered by various levels. This provide enough time for the user to be aware of the issue and he/she can immediately block the ATM card himself or herself. The great advantage of the solution is that it ensures security (TABLE 2) in the worst case where both the card and mobile phone get lost.

Moreover the proposed solution does not demand any change in the infrastructure of the J2EE system. Since this is the era of mobile banking, the proposed solution can be easily integrated into the mobile applications that enable J2EE banking system. All that is needed here is some add-ons to the mobile application and inclusion of some extra functionality to the already existing web service. Hence the solution is cost-effective. Here security is improved by integrating mobile phone into J2EE system.

For a single authentication system, any one can hack user id and password and also they can access the Net banking. So it is not secure authentication method. So a double authentication system is better than single authentication system. Insider is most responsible for the majority of fraud action. Since insider can easily hack username, password as well as user mobile SMS

also. Mostly insider may be family members, colleague or nearby gang.

In our proposed model (TABLE 2), Fingerprint recognition has been used for uniqueness and anybody cannot change finger print of user. Fingerprints became an important identification of complex criminals through finger print recognition. So it is more secure model. Users fingerprint cannot be used anywhere without the knowledge of user. In this architecture, user should scan his fingerprint. But, all systems has not

scanning peripherals by default. So each system or laptop has to be made with scanning facilities inbuilt. For the machines already in use, user can use additional accessories for fingerprint scanning. Already, this fingerprint authentication system is used in ATM. Not only ATM, many departments using this model. But Net bank is most popular and money oriented groups. No one can maintain full secure methods for this process in internet banking.

TABLE I. COMPARISON OF EXISTING METHODS AND PROPOSED MODEL

Existing Method	Proposed model
<p>Single authentication system: User enters User ID and password which will be verified in the bank website for authorization. If the user ID and password matches the user can login to internet banking system. Otherwise, "Invalid user" is reported to the user.</p>	<p>User enters User ID and password which will be verified in the bank website for authorization. At the same time user scan his fingerprint by a smartphone and checked for matches.</p>
<p>Double authentication system:  <ul style="list-style-type: none"> <li>User enters User ID and password which will be verified in the bank website for authorization. If the user ID and password matches the user can login to internet banking system. Otherwise, "Invalid user" is reported to the user.</li> <li>After this validation bank provide extra authentication process such as providing another security token code to user smartphone through SMS</li> </ul> </p>	<p>A biometric authentication and mobile validity market verified in the bank website for authorization. Otherwise, "Invalid user" is reported to the user.</p>

TABLE II. COMPARISON BETWEEN SINGLE AUTHENTICATION SYSTEM, DOUBLE AUTHENTICATION AND PROPOSED MODEL

Method	User id & pass to login	SMS security code	Biometric recognition	Security ATM card	Security level
Model					
<b>Single authentication system</b>	Can hack	.....	.....	.....	Not secured
<b>Double authentication system</b>	Can hack	Insider only can hack	.....	.....	Half secured
<b>Proposed model for IBS with Biometric recognition</b>	Can hack	Insider only can hack	No one can hacking	Secured if ATM loses	Fully secured

## 7. Conclusion

Mobile Net banking has become immensely popular among customers as a suitable method for money transaction. The proposed model has been developed for net banking system with biometric recognition and mobile process. A new technique to access the internet banking process is more secure than existing methods. Because fingerprint recognition method is unique method. If the machines are built with scanning accessories, the user can make the authentication by using user ID, password and finger print recognition, SMS validity. By the interface Manager customer bank ATM, when the ATM card is lost, the customer can block the ATM card with every android smartphone. The transaction would be more secure method. In this model, unauthorized persons cannot surely hack or access the user accounts.

## References

- [1] Basel Committee Report on Banking Supervision. (1998). Risk Management for Banking and electronic money activities. Available From: [www.bis.org/publ/bcbs98.pdf](http://www.bis.org/publ/bcbs98.pdf).
- [2] Daniel, E. (1999), Provision of Electronic Banking in the UK and the Republic of Ireland. International Journal of Bank Marketing, 17(2):72-82.
- [3] Khorshid, S. and Ghane, H. (2009), Ranking the challenges of e-banking with the help of AHP model. Journal of Modiriyate Sanati azad University of Sanandaj. 4(9):89-106.
- [4] Zain S. Barham, "Fingerprint Recognition using MATLAB", 2011
- [5] Lin Hong, "Automatic person identification using fingerprints," Ph. D. Thesis, 1998
- [6] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, "Handbook of Fingerprint Recognition", Second Edition, Springer, 2009, ISBN 978-1-84882-25365
- [7] Collin Mulliner, Ravishankar Borgaonkar, Patrick Stewin, and Jean-Pierre Seifert "SMS-Based One-Time Passwords: Attacks and Defense" Springer-Verlag, DIMVA 2013, LNCS 7967, pp. 150-159, 2013
- [8] Heeseung Choi, Kyoungtaek Choi, and Jaihie Kim, "Fingerprint Matching Incorporating Ridge Features with Minutiae", June 2011

- [9] Salil Prabhakar, Anil K Jain and Sharath Pankanti, "Learning fingerprint minutiae location and type", Pattern recognition 36(2003)- 1847-1857
- [10] Sri Shimai Das, Smt. Jhunu Debbarma"Designing a Biometric Strategy (Fingerprint) Measure for Enhancing ATM Security in Indian E-Banking System"International Journal of Information and Communication Technology Research,ISSN-2223-4985, Volume 1 No.5, September 2011
- [11] M.Sandeep, D.Nagalaxmi "Secure Approach for Net Banking by Using Fingerprint Authentication" International Journal of Engineering Science and Computing IJESC 2015
- [12] Nsouli, S M and A Schaechter (2002): 'Challenges of the E-banking Revolution', Finance and Development, International Monetary Fund, September, Volume 39, Number 3
- [13] Renu Mourya1, Ms.Sarita "FINGERPRINT MATCHING TECHNIQUES: REVIEW"International Journal of Science, Technology & Management Volume No 04, Special Issue No. 01,ISSN (online): 2394-1537, May 2015

# *Towards the design of Fault Tolerant Binary Comparator by Parity Preserving Reversible Logic based Multi Layer Multiplexer*

Biswajit Das

Murshidabad College of Engineering  
& Technology  
Berhampore, India

Shefali Mamataj

Murshidabad College of Engineering  
& Technology  
Berhampore, India

Saravanan Chandran

National Institute of Technology  
Durgapur, India

**Abstract**—Reversible circuits which are Parity-preserving are nowadays getting more weight towards the progress of designing systems having fault-tolerance in the field of nanotechnology. The reversible circuit which preserves parity must have the parity preserving property means the input vector parity must be the same to the output vector parity. It contributes a expansive category of finding faults in the circuit which can be detect at the circuit outputs. Thus in a single word reversible logic circuits which preserves parity will be more beneficial towards the progress of fault free circuit realization. In this paper we have proposed three new fault tolerant reversible gates FTM, FTC and FATOC for optimizing the circuit in terms of the gate number, garbage outputs, hardware complexity and constant inputs. This work targets implementation of reversible Fault Tolerant Comparator (FTCom) by Reversible Logic-based Multi Layer Multiplexer of proposed FTM. Furthermore the design is also presented by the obtainable fault tolerant reversible gates and the proposed gates FTC & FATOC. We have also presented three lemmas to verify the fault tolerance or parity preserving property of these proposed FTM, FTC and FATOC gate respectively.

**Keywords-** *Fault Tolerance, Parity-Preserving Reversible Gate, Reversible Logic, Comparator*

## I. INTRODUCTION

Reversible logic is the most admired conception regarding the energy efficiency in the area of computations. It is promising as a vital area to investigate. It can be used for wide applications in several fields, for instance low power CMOS design, quantum computing and optical information processing. An attractive point of view of the reversible logic is that to construct digital devices which can realize processing unit of computation having almost zero power dissipation. Landauer [1] showed that for the computations of irreversible circuit, for each bit an amount of energy  $k_B T \ln 2$  Joules is lost as a heat. The energy  $E$  bit necessary for one bit of operation is specified by Shannon-Von Neumann-Landauer (SVNL) expression in equation (1).

$$E_{\text{bit}} \geq E_{\text{SVNL}} = k_B T \ln 2 = 0.017 \text{ eV} \dots \dots \dots (1)$$

$k_B$  = Boltzmann constant

$T = 300 \text{ K}$ .

This is the smallest amount of energy necessary for the processing of a bit. Bennett [2] showed that power dissipation may be zero in logical circuit if and only if the circuit is consisting of reversible logic gates. Although reversibility can recover loss of bit but it is unable to identify the bit error in the circuit. Reversible circuits that are fault tolerant must be able of preventing errors at outputs. Any structure consists of fault-tolerant components are capable of the detection and correction of faults easily and simply. In communication field and many other applications, fault tolerance is obtained by means of parity. As a result, parity preserving reversible circuits may be the upcoming designing swing towards the growth of fault tolerant reversible systems. Most gates used in digital design are irreversible. For example the AND, OR and EXOR gates do not carry out reversible operations. Out of the commonly used gates, only the NOT gate is reversible. To design reversible circuits a set of reversible gates is required. A number of such gates have been proposed over the past decades. Among them are the controlled-not (CNOT) which was proposed by Feynman [3], Toffoli, and Fredkin [4, 5] gates.

Comparison between two binary numbers has an extensive variety of application in encryption devices, microprocessors, sorting networks, communication systems etc. Therefore, binary comparator is an imperative circuitry in recent VLSI design and nanotechnology [6]. Therefore in this paper, we have presented a fault tolerant reversible binary comparator circuit which uses a lesser amount of number of gates, a lesser amount of number of garbage output, and a lesser amount of constant input and hardware complexity. Three lemmas are also presented here to prove the parity-preserving property of the proposed three fault tolerant gates to be precise FTM, FTC and FATOC in that order.

This paper is organized as follows. Section II, specifies the ideas about the reversible logic, fault tolerant logic, basic definition of some fault tolerant reversible gates. Section III shows our proposed fault tolerant reversible gates and the proof of their parity preserving property. This section also shows multiplexer design using proposed fault tolerant reversible gate FTM. Section IV describes the design of proposed fault tolerant reversible binary comparator (FTCom)

in different ways. Section V gives the comparison results and summary of the proposed fault tolerant reversible circuits. Finally, the conclusion is given in Section VI.

## II. PRELIMINARIES

### A. Reversible Logic

A logic gate can be defined as reversible if the mapping of its input vectors to output vectors is bijective that means for each specific output is related to the specific input and also the number of inputs is the same as to the number of outputs [7]. The important cost metrics in reversible logic circuits are the gate count, constant input, garbage output, hardware complexity and quantum cost. The cost of every  $2 \times 2$  gate is the unity and the cost of  $1 \times 1$  gate is zero [8]. Any reversible logic can be possible to implement with primitive gates such as  $2 \times 2$  reversible gates and  $1 \times 1$  NOT gates. Reversible logic does not allow the fan outs and feedback paths.

### B. Fault-Tolerant Logic

Fault tolerant system is able to work appropriately even in the occurrence of the failure of some of its elements. Any structure consists of fault-tolerant components are capable of the detection and correction of faults easily and simply. A fault tolerant reversible gate also can be said a conservative gate [9]. The parity of input vectors and output vectors must be equal. Let us consider the input vectors be  $I_v = I_0, I_1, \dots, I_{n-1}$  and output vectors of any fault tolerant gate be  $O_v = O_0, O_1, \dots, O_{n-1}$ , where

- (i)  $I_v$  (Bijective)  $O_v$ ,
- (ii)  $I_1 \oplus I_2 \oplus \dots \oplus I_{n-1} \leftrightarrow O_1 \oplus O_2 \oplus \dots \oplus O_{n-1}$ .

### C. Fault Tolerant Reversible Logic gate

Many parity preserving means fault tolerant reversible gates have been already proposed by many authors. A small number of favourable parity preserving gates are given as follows:

**Feynman Double Gate (F2G):** A  $3 \times 3$  Feynman Double Gate (F2G) is shown in figure 1 [10]. The input vectors are specified as A, B and C and output vectors are given as  $P = A$ ,  $Q = A \oplus B$ , and  $R = A \oplus C$ . Quantum cost is given as equal to 2.



Figure1. Feynman Double Gate

**Fredkin Gate (FRG):** A  $3 \times 3$  Fredkin gate (FRG) is shown in figure 2 [11]. The input vectors are specified as A, B and C and output vectors are given as  $P = A$ ,  $Q = A' B \oplus A C$  and  $R = A' C \oplus A B$ . Quantum cost is equal to 5.

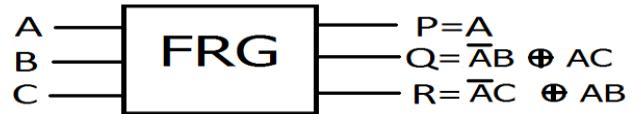


Figure2. Fredkin Gate

**Modified IG Gate (MIG):** A  $4 \times 4$  Modified IG gate is shown in figure 3 [12]. The input vectors are specified as A, B and C and output vectors are given as  $P = A$ ,  $Q = A \oplus B$ ,  $R = A B \oplus C$  and  $S = A B' \oplus D$ .

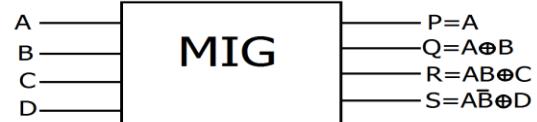


Figure 3.Modified IG gate

**Islam Gate (IG):** A  $4 \times 4$  IG gate is shown in figure 4 [13]. The input vectors are specified as A, B and C and output vectors are given as  $P = A$ ,  $Q = A \oplus B$ ,  $R = A B \oplus C$  and  $S = B D \oplus B(A \oplus D)$ .



Figure 4. Islam Gate

**New Fault Tolerant (NFT):** A  $3 \times 3$  NFT gate is shown in figure 5[14]. The input vectors are specified as A, B and C and output vectors are given as  $P = A \oplus B$ ,  $Q = B C' \oplus A C'$  and  $R = B C \oplus A C$ .

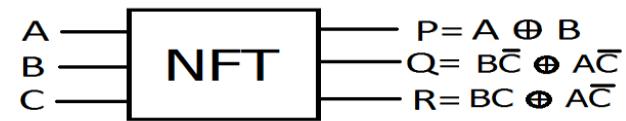


Figure 5. New Fault Tolerant Gate

## III. PROPOSED GATE

In this section, we have proposed three new fault tolerant reversible gates named FTM, FTC and FATOC in subsections III.A, III.B and III.C, respectively. Truth table of these gates is also presented in this section which shows their reversibility as well as their parity preserving property. Three lemmas are also presented in this section to prove the parity preserving property .Also in subsection III.D designing of fault tolerant multiplexers by FTM gate has been shown.

### A. Proposed fault tolerant reversible FTM gate

In this subsection, a new  $3 \times 3$  fault tolerant reversible gate namely FTM gate is proposed. The proposed gate and its truth table is given away in Fig. 6 and Table 1.respectively.It can be

notified from the truth table that the input bit pattern related to a specific output bit pattern can be possible to determine uniquely and find out the input-output bit parity.

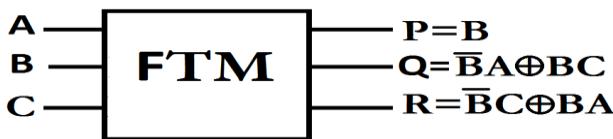


Figure 6. FTM Gate

Therefore it can be said that the proposed FTM gate is a reversible gate.

TABLE I. TRUTH TABLE FOR FTM GATE

Inputs			Outputs		
A	B	C	P	Q	R
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	1	0	0
0	1	1	1	1	0
1	0	0	0	1	0
1	0	1	0	1	1
1	1	0	1	0	1
1	1	1	1	1	1

**Lemma 1:** Proposed FTM Gate is a Fault Tolerant Gate

**Proof:** The input vectors and output vectors of FTM gate are  $I_v = \{A, B, C\}$  and  $O_v = \{B, (B'A \oplus BC), (B'C \oplus AB)\}$  respectively. From Section II.B, we can know that input parity and output parity will be same in fault tolerant or parity preserving gate.

Therefore, the input parity of FTM gate is  $= A \oplus B \oplus C$

$$\begin{aligned}
 \text{Output parity of FTM gate is} &= B \oplus (B'A \oplus BC) \oplus (B'C \oplus AB) \\
 &= B(1 \oplus C) \oplus B'A \oplus (B'C \oplus AB) \\
 &= BC' \oplus B'(A \oplus C) \oplus AB \\
 &= B(C' \oplus A) \oplus B' (A \oplus C) \\
 &= B (C \oplus A)' \oplus B' (A \oplus C) \\
 &= (B \oplus A \oplus C) \\
 &= A \oplus B \oplus C
 \end{aligned}$$

Thus, output parity is equal to input parity. As FTM gate preserves the parity values of input and output, FTM gate is a fault tolerant reversible gate.

#### B. Proposed fault tolerant reversible FTC gate

In this subsection, a new 4\*4 fault tolerant reversible gate namely FTC gate is proposed. The proposed gate and its truth table are shown in Fig. 7 and Table II respectively. It can be

revealed from the truth table that the input pattern related to a particular output pattern can be uniquely determined and find out the input-output bit parity.

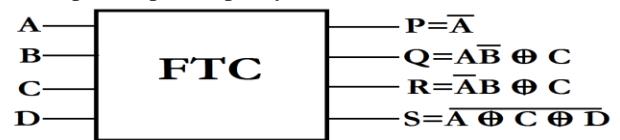


Figure 7. FTC Gate

Therefore it can be said that the proposed FTC gate is a reversible gate.

TABLE II. TRUTH TABLE FOR FTC GATE

Inputs				Outputs			
A	B	C	D	P	Q	R	S
0	0	0	0	1	0	0	1
0	0	0	1	1	0	0	0
0	0	1	0	1	1	1	0
0	0	1	1	1	1	1	1
0	1	0	0	1	0	1	1
0	1	0	1	1	0	1	0
0	1	1	0	1	1	0	0
0	1	1	1	1	1	0	1
1	0	0	0	0	1	0	0
1	0	0	1	0	1	0	1
1	0	1	0	0	0	1	1
1	0	1	1	0	0	1	0
1	1	0	0	0	0	0	0
1	1	0	1	0	0	0	1
1	1	1	0	0	1	1	1
1	1	1	1	0	1	1	0

**Lemma 2:** Proposed FTC Gate is a Fault Tolerant Gate

**Proof:** The input vectors and output vectors of FTC gate be  $I_v = \{A, B, C, D\}$  and  $O_v = \{A', (AB' \oplus C), (A'B \oplus C'), (A \oplus C \oplus D)\}$  respectively. From Section II.B, we can be familiar with that input parity and output parity must have to be same in fault tolerant or parity preserving gate.

Thus, the input parity of FTC gate is  $= A \oplus B \oplus C \oplus D$

$$\begin{aligned}
 \text{Output parity of FTC gate is} &= A' \oplus (AB' \oplus C) \oplus (A'B \oplus C') \oplus (A \oplus C \oplus D)' \\
 &= (C \oplus C) \oplus (A' \oplus AB') \oplus A'B \oplus A \oplus C \oplus D \oplus 1 \\
 &= (A' \oplus 1) \oplus AB' \oplus A'B \oplus (A \oplus C \oplus D) \\
 &= A \oplus AB' \oplus A'B \oplus A \oplus C \oplus D \\
 &= (A \oplus A) \oplus AB' \oplus A'B \oplus C \oplus D \\
 &= AB' \oplus A'B \oplus C \oplus D \\
 &= A \oplus B \oplus C \oplus D
 \end{aligned}$$

#### C. Proposed fault tolerant reversible FATOC gate

In this subsection, a new 4\*4 fault tolerant reversible gate namely FATOC gate is proposed. The proposed gate and its truth table are shown in Fig. 8 and Table III respectively. It

can be said from the truth table that the input pattern related to a particular output pattern can be uniquely determined and can find out the input-output bit parity.

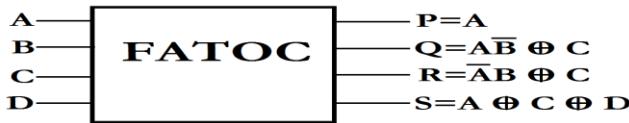


Figure 8. FATOCA Gate

Therefore it can be said that the proposed FATOCA gate is a reversible gate.

TABLE III. TRUTH TABLE FOR FATOCA GATE

Inputs				Outputs			
A	B	C	D	P	Q	R	S
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	1
0	0	1	0	0	1	1	1
0	0	1	1	0	1	1	0
0	1	0	0	0	0	1	0
0	1	0	1	0	0	1	1
0	1	1	0	0	1	0	1
0	1	1	1	0	1	0	0
1	0	0	0	1	1	0	1
1	0	0	1	1	1	0	0
1	0	1	0	1	0	1	0
1	0	1	1	1	0	1	1
1	1	0	0	1	0	0	1
1	1	0	1	1	0	0	0
1	1	1	0	1	1	1	0
1	1	1	1	1	1	1	1

**Lemma 3:** Proposed FATOCA Gate is a Fault Tolerant Gate.

**Proof:** The input vectors and output vectors of FATOCA gate be  $I_v = \{A, B, C, D\}$  and  $O_v = \{A, (AB' \oplus C), (A'B \oplus C), (A \oplus C \oplus D)\}$  respectively. From Section II.B, we can know that input parity and output parity have to be same in fault tolerant or parity preserving gate.

Accordingly, the input parity of FATOCA gate is  $= A \oplus B \oplus C \oplus D$

Output parity of FATOCA gate is

$$\begin{aligned}
 &= A \oplus (AB' \oplus C) \oplus (A'B \oplus C) \oplus (A \oplus C \oplus D) \\
 &= (C \oplus C) \oplus (A \oplus AB') \oplus A'B \oplus A \oplus C \oplus D \\
 &= (A \oplus A) \oplus AB' \oplus A'B \oplus C \oplus D \\
 &= AB' \oplus A'B \oplus C \oplus D \\
 &= A \oplus B \oplus C \oplus D
 \end{aligned}$$

#### D. Proposed fault tolerant reversible FTM gate as Multiplexer

In this subsection fault tolerant multiplexer has been designed by the proposed fault tolerant reversible FTM gate. The designs of 2:1 multiplexer & 4:1 multiplexer are shown in figure 9 and figure 10 respectively. From figure we can see for the realization of 2:1 multiplexer only one FTM gate is

required and garbage is two but for the 4:1 multiplexer designing three FTM gates are required and garbage is five.

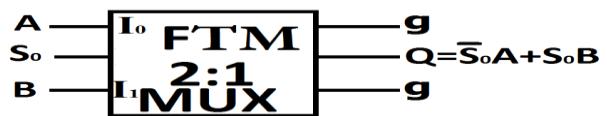


Figure 9. FTM Gate as 2:1 MUX

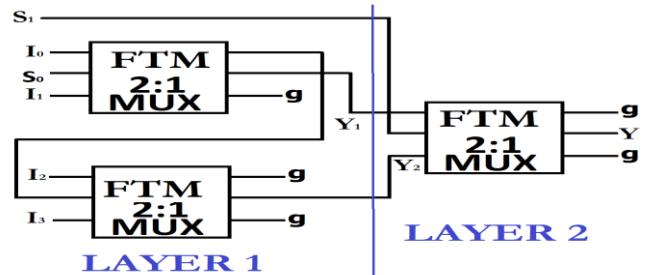


Figure 10. FTM Gate as 4:1 MUX

#### IV. PROPOSED DESIGN FOR FTCom

This section describes the design of proposed fault tolerant reversible binary comparator (FTCom) in different ways. The subsection IV.A describes realization of FTCom by the proposed multiplexer and the subsection IV.B describes realization of FTCom by the existing fault tolerant gates and proposed gates in several ways.

The binary comparator compares two binary numbers ( $A, B$ ) and determines the result among  $X (A=B)$  and  $Y (A < B)$  and  $Z (A > B)$ . The truth table of 1-bit comparator is shown in Table IV.

TABLE IV. TRUTH TABLE FOR 1 BIT COMPARATOR

Inputs		Outputs		
A	B	X(A=B)	Y(A < B)	Z(A > B)
0	0	1	0	0
0	1	0	1	0
1	0	0	0	1
1	1	1	0	0

#### A. Proposed 1 bit FTCom based on fault tolerant reversible multilayer multiplexer of FTM gate

The proposed FTM gate can be used to implement a 1-bit comparator as a multiplexer. It is well known that  $X (A=B) = (A \oplus B)', Y (A < B) = A'B$  and  $Z (A > B) = AB'$ .

#### 1) Proposed FTCom based on fault tolerant reversible multilayer 2:1 multiplexer of FTM gate

To realize 1 bit FTCom by 2:1 fault tolerant multiplexer of FTM gate, three FTM gates and one F2G gate are required

shown in figure 11 which takes two 1-bit binary numbers as input A and B, and four constant inputs. The circuit produces four garbage outputs and three most wanted outputs that are ( $A=B$ ) ( $A < B$ ) and ( $A > B$ ).

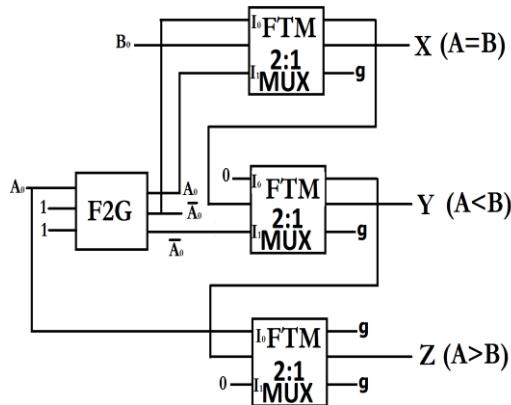


Figure 11. Realization of FTCom by 2:1 MUX of FTM gate

## 2) Proposed FTCom based on fault tolerant reversible multilayer 4:1 multiplexer of FTM gate

To realize 1 bit FTCom by 4:1 fault tolerant multiplexer of FTM gate, nine FTM gates are required shown in figure 12 which takes two 1-bit binary numbers as input A and B, and twelve constant inputs. The circuit produces eleven garbage outputs and three needed outputs that are ( $A=B$ ) ( $A < B$ ) and ( $A > B$ ).

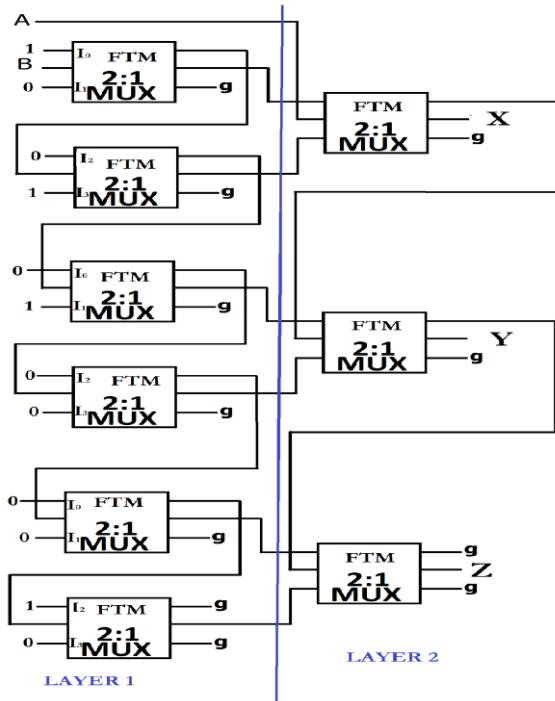


Figure 12. Realization of FTCom by 4:1 MUX of FTM gate

In this way we can construct 2 bit binary comparator by 8:1 multiplexer and 16:1 multiplexer. Moreover comparator for N bit operation can be implemented by  $(2^{2N}:1)$  multiplexer and  $(2^{2N-1}:1)$  multiplexer.

## B. Proposed 1bit FTCom based on fault tolerant reversible logic gate

The 1 bit comparator can also be realized by using fault tolerant reversible gates. This subsection describes several ways of 1 bit FTCom realization.

### 1) Proposed FTCom based on fault tolerant reversible FRG gate

To realize 1 bit FTCom by existing fault tolerant Fredkin gate, two FRG gates and two F2G gates are required shown in figure 13 which takes two 1-bit binary numbers as input A and B, and four constant inputs. The circuit produces four garbage outputs and three required outputs that are ( $A=B$ ) ( $A < B$ ) and ( $A > B$ ).

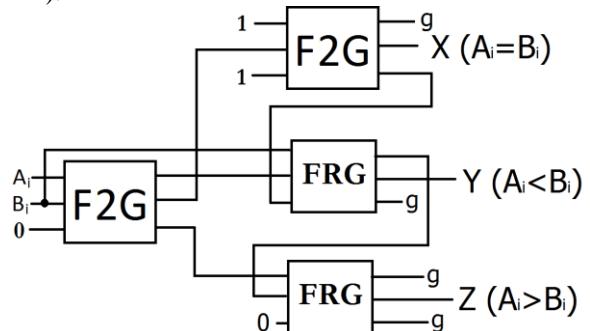


Figure 13. Realization of FTCom by FRG gate

### 2) Proposed FTCom based on fault tolerant reversible MIG gate

To realize 1 bit FTCom by existing fault tolerant Modified Islam gate, two MIG gates and one F2G gates are required shown in figure 14 which takes two 1-bit binary numbers as input A and B, and six constant inputs. The circuit produces six garbage outputs and three wanted outputs that are ( $A=B$ ) ( $A < B$ ) and ( $A > B$ ).

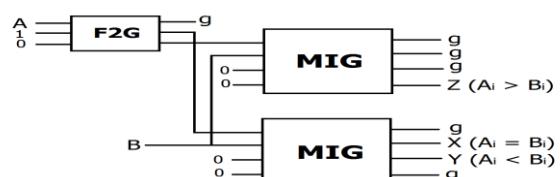


Figure 14. Realization of FTCom by MIG Gate

3) Proposed FTCom based on fault tolerant reversible IGgate

To realize 1 bit FTCom by existing fault tolerant Islam gate, one IG gate and one F2G gate are required shown in figure 15 which takes two 1-bit binary numbers as input A and B, and two constant inputs. The circuit produces two garbage outputs and three wanted outputs that are (A=B) (A<B) and (A>B).

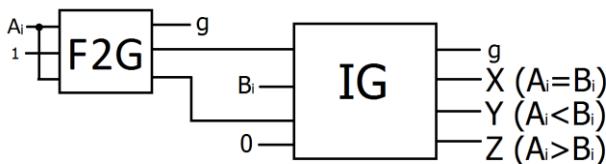


Figure15. Realization of FTCom by IG gate

4) Proposed FTCom based on fault tolerant reversible NFTgate

To realize 1 bit FTCom by existing fault tolerant New Fault Tolerant gate, two NFT gates and two F2G gates are required shown in figure 16 which takes two 1-bit binary numbers as input A and B, and four constant inputs. The circuit produces five garbage outputs and three wanted outputs that are (A=B) (A<B) and (A>B).

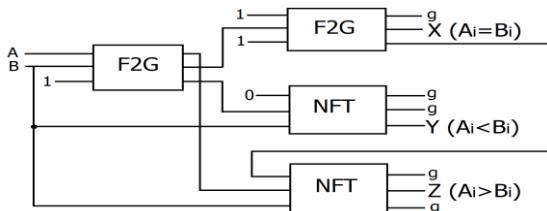


Figure16. Proposed FTCom based on fault tolerant reversible NFT Gate

5) Proposed FTCom based on proposed fault tolerant reversible FATOC gate

To realize 1 bit FTCom by proposed fault tolerant FATOC gate, one FATOC gate and one F2G gate are required shown in figure 17 which takes two 1-bit binary numbers as input A and B, and one constant input. The circuit produces one garbage output and three required desired outputs that are (A=B) (A<B) and (A>B).

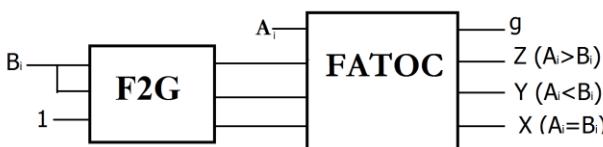


Figure17. Realization of FTCom by FATOC gate

6) Proposed FTCom based on proposed fault tolerant reversible FTC gate

To realize 1 bit FTCom by proposed fault tolerant FTC gate, one FTC gate is needed shown in figure 18 which takes two 1-bit binary numbers as input A and B, and one constant input. The circuit produces one garbage outputs and three wanted outputs that are (A=B) (A<B) and (A>B).

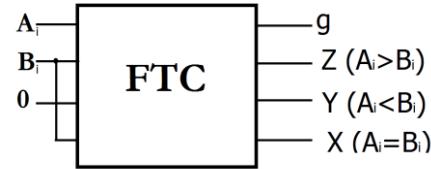


Figure18. Realization of FTCom by FTC gate

C. Proposed 2bit FTCom based on proposed fault tolerant reversible logic gate

The 1-bit comparator compares two 2-bit of two binary numbers ( $A=A_1A_0$ ,  $B=B_1B_0$ ) and determines the result among X (A=B) and Y (A<B) and Z (A>B). The truth table of 2-bit comparator is shown in Table V.

TABLE V. TRUTH TABLE FOR 2 BIT COMPARATOR

Inputs		Outputs		
A(A <sub>1</sub> A <sub>0</sub> )	B(B <sub>1</sub> B <sub>0</sub> )	X(A=B)	Y(A<B)	Z(A>B)
00	00	1	0	0
00	01	0	1	0
00	10	0	1	0
00	11	0	1	0
01	00	0	0	1
01	01	1	0	0
01	10	0	1	0
01	11	0	1	0
10	00	0	0	1
10	01	0	0	1
10	10	1	0	0
10	11	0	1	0
11	00	0	0	1
11	01	0	0	1
11	10	0	0	1
11	11	1	0	0

To realize 2 bit FTCom by proposed fault tolerant gate, four FTC gate, one FTM and one F2G gate are required shown in figure 19 which takes two 2-bit binary numbers as input  $A=A_1A_0$  and  $B=B_1B_0$  and six constant inputs. The circuit produces ten garbage outputs and three wanted outputs that are (A=B) (A<B) and (A>B).

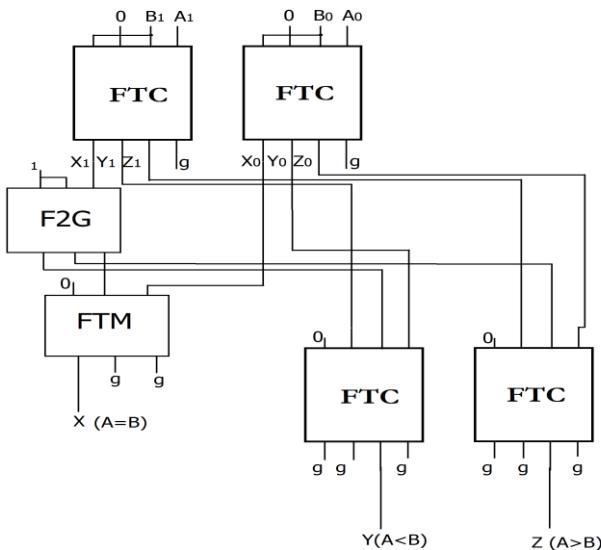


Figure 19. Realization of 2-bit FTCom by FTC & FTM gate

Furthermore N-bit FTCom can also be possible to design by these proposed fault tolerant gates.

## V. COMPARISON RESULTS

In this section comparison results are shown. Table VI shows that we can construct any bit of fault tolerant comparator (FTCom) by the fault tolerant multiplexer which is based on the proposed fault tolerant reversible FTM gate. The required multiplexer, number of gates, garbage output and hardware complexity is given to design 1bit, 2 bit and N bit FTCom. Also Table VII shows the comparison among the various proposed 1-bit FTCom and the existing 1-bit FTCom.

TABLE VI. COMPARISON OF FTCom BY MULTIPLEXER OF FTM GATE

Sl no	Required Multiplexer	Gate Count	Garbage	Hardware Complexity
1-bit F T C o m	2:1 Multiplexer	4	4	$3(2\alpha+4\beta+\delta)+2\alpha$
	4:1 Multiplexer	9	11	$9(2\alpha+4\beta+\delta)$
2-bit F T C o m	8:1 Multiplexer	23	28	$21(2\alpha+4\beta+\delta)+4\alpha$
	16:1 Multiplexer	45	53	$53(2\alpha+4\beta+\delta)$
N-bit F T C o m	$2^{2N-1}:1$ Multiplexer	$3(2^{2N-1}+N)$	$3(2^{2N}+2N-2)+1$	$3(2^{2N-1})(2\alpha+4\beta+\delta)+N2\alpha$
	$2^{2N}:1$ Multiplexer	$3(2^{2N-1}+N-1)$	$3(2^{2N}+2N-3)+2$	$\{3(2^{2N-1}+N-1)\}(2\alpha+4\beta+\delta)$

TABLE VII. COMPARISON OF 1 BIT FTCom BY FAULT TOLERANT REVERSIBLE GATE

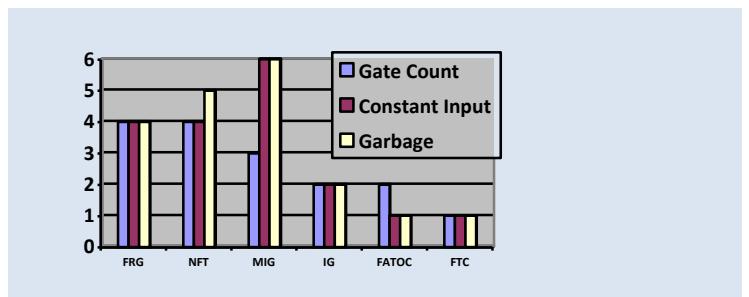
Sl no	Gate Count	constant input	Garbage	Hardware Complexity
proposed 1-bit FTCom (using FRG gate)	4	4	4	$8\alpha+8\beta+2\delta$
proposed 1-bit FTCom (using MIG gate)	3	6	6	$8\alpha+4\beta+2\delta$
proposed 1-bit FTCom (using NFT gate)	4	4	5	$8\alpha+6\beta+2\delta$
proposed 1-bit FTCom (using IG gate)	2	2	2	$6\alpha+3\beta+\delta$
Proposed 1-bit FTCom (using proposed FATOC gate)	2	1	1	$6\alpha+2\beta+2\delta$
Proposed 1-bit FTCom (using proposed FTC gate)	1	1	1	$4\alpha+2\beta+3\delta$
Existing 1-bit FTCom (using AG gate) [15]	1	3	2	$7\alpha+4\beta+4\delta$

In this paper various designing of 1 bit FTCom are represented by the existing fault tolerant gates FRG, MIG, NFT, IG and also by our proposed fault tolerant reversible gates FTC, FATOC and FTM. In terms of Gate count, constant input, garbage output and hardware complexity all these designs are given in Table VII. Furthermore it can be seen that our proposed 1 bit FTCom using proposed fault tolerant gate FTC and FATOC are most efficient with other designs with respect to gate count, constant input, garbage output and hardware complexity. Also a comparison of 2-bit FTCom is shown in Table VIII. It can be seen that our proposed 2-bit FTCom is more efficient with respect to the existing one.

TABLE VIII. COMPARISON OF 2 BIT FTCom BY FAULT TOLERANT REVERSIBLE GATE

Sl No	Gate Count	Const ant Input	Garb age	Hardware Complexity
Existing 2-bit FTCom [15]	6	11	10	$20\alpha+14\beta+7\delta$
Proposed 2-bit FTCom (using proposed Fault tolerant gate)	6	6	10	$22\alpha+16\beta+14\delta$

A chart is given below to describe the required number of gates, constant input, garbage output for 1 bit FTCom by FRG, MIG, NFT, IG, FTC and FATOC gates.



## VI. CONCLUSION

This work represents towards the design of Fault Tolerant Binary Comparator by Parity Preserving Reversible Logic based Multi Layer Multiplexer. The design is very useful for the future computation in the field of very low power application zone such as digital circuits along with quantum computers. It is shown that the proposed circuit is highly optimized with respect to number of reversible logic gates, constant input, number of garbage and hardware complexity. The design methods are definitely useful for the construction of future computer and other computational structures in nanometric fashion. Synthesis of these designs and QCA implementation of these designs may be the future scope of work.

## ACKNOWLEDGMENT

The authors wish to thank the CSE and ECE Department of Murshidabad College of Engineering and Technology, Berhampore, for supporting this work

## REFERENCES

- [1] R. Landauer, “Irreversibility and heat generation in the computing process,” IBM Journal of Research and Development, vol. 5,
- [2] C. H. Bennett, “Logical reversibility of computation,” *IBM Journal of Research and Development*, vol. 17, no. 6, pp. 525–532, 1973
- [3] R. Feynman, —Quantum Mechanical Computers!, Optics News, 11, 1985, pp. 11-20.
- [4] E. Fredkin and T. Toffoli, —Conservative logic, *Intl. Journal of Theoretical Physics*, pp. 219-253, 1982.
- [5] T. Toffoli, —Reversible Computing!, Tech. Memo MIT/LCS/TM-151, MIT Lab for CS, \_80.
- [6] M. Nielsen and I. Chuang, “Quantum computation and quantum information” in Cambridge University Press, 2000.
- [7] T. Toffoli, “Reversible computing,” in Proceedings of the 7th Colloquium on Automata, Languages and Programming, pp. 632–644, Springer, London, UK, 1980.
- [8] M. Perkowski, M. Lukac, P. Kerntopf et al., “A hierarchical approach to computer-aided design of quantum circuits,” in Proceedings of the 6th International Symposium on Representations and Methodology of Future Computing Technology, pp. 201–209, 2003.
- [9] J. W. Bruce, M. A. Thornton, L. Shivakumaraiyah, P. S. Kokate, and X. Li, “Efficient adder circuits based on a conservative reversible logic gate,” in Proceedings of IEEE Symposium on VLSI, pp. 83–88, Washington, DC, USA, 2002.
- [10] B. Parhami, “Fault tolerant reversible circuits”, in Proceedings of 40th Asimolar Conf. Signals, Systems, and Computers, *Pacific Grove*, CA, pp. 1726-1729, October 2006.
- [11] E. Fredkin and T. Toffoli, “Conservative logic”, *Intl. Journal of Theoretical Physics*, pp. 219-253, 1982.
- [12] Islam S. and M. Mahbubur Rahman, 2009b. Efficient Approaches for Designing Fault Tolerant Reversible Carry Look-Ahead and Carry-Skip Adders, *MASAUM Journal of Basic and Applied Sciences*, 1(3): 354-360.
- [13] M. S. Islam, M. M. Rahman, Z. Begum, M. Z. Hafiz and A. A. Mahmud, “Synthesis of fault tolerant reversible logic circuits”, In Proc. IEEE International Conference on Testing and Diagnosis, Chengdu, China, 28-29 April, 2009.
- [14] M. Haghparast and K. Navi, “A novel fault tolerant reversible gate for nanotechnology based systems”, Am. J. of App. Sci., vol. 5, no.5, pp.5 19-523, 2008.
- [15] Avishek Bose and Ankur Sarker, “A Novel Approach for Constructing Reversible Fault Tolerant  $n$ -Bit Binary Comparator”, 3rd International Conference on Informatics, Electronics & Vision (ICIEV), 23-24 May, 2014, Dhaka, Bangladesh

# Sentiment Analysis of Twitter data using Hybrid Method of Support Vector Machine and Ant Colony Optimization

Jasleen Kaur <sup>#1</sup>, Sukhjit Singh Sehra <sup>\*2</sup>, Sumeet Kaur Sehra <sup>\*3</sup>

<sup>#</sup> Research Scholar, Guru Nanak Dev Engineering College  
Ludhiana, India, 141006

<sup>1</sup> jasleenbhullar66@gmail.com

<sup>\*</sup> Assistant Professor, Guru Nanak Dev Engineering College  
Ludhiana, India, 141006

<sup>2</sup> sukhjitsehra@gmail.com

<sup>3</sup> sumeetksehra@gmail.com

**Abstract**—Sentiment analysis is the process of elicitation, comprehension, classification and illustration of opinions or sentiments expressed by various users concerning a topic or object. It has become prominent due to the increase in crowd-sourced information on social media. Social media has bestowed users with much more power than they possessed before its advent. Presently, Twitter is a prominent micro-blogging platform which empowers its users to post their opinions in form of “tweets”. These can be utilised to gain insights into opinions and sentiments of people for better decision making and marketing. This research aims to use Twitter data to inspect sentiments of the crowd regarding a particular subject. Retrieved tweets are classified into two opinion classes: Positive or Negative. This classification is performed by using a hybrid strategy of Machine Learning algorithm Support Vector Machine (SVM) and Ant Colony Optimization (ACO). Unigrams are employed for feature extraction with term frequency-inverse document frequency as feature weighting criteria. The average accuracy of classification enhances from 75.54% (using SVM) to 86.74% (using SVM-ACO).

**Index Terms**—Crowdsourced data, Machine Learning Techniques, Sentiment Analysis, Twitter

## I. INTRODUCTION

Enterprises require acquiring data points promptly for solving complex decision problems. Enterprises can now store and analyse the huge variety of data acquired through different platforms using predictive analysis and big data. The recent advances in technologies have accelerated the pace of computation process to seconds from hours. However, the main bottleneck of this process is the collection of data. Enterprises have now switched to crowdsourcing for overcoming this bottleneck. Crowdsourcing is the process of engaging a large group of people for generating and providing the required data [1]. It increases the speed of data collection. Nowadays, a large proportion of people are interlinked through networking with Web 2.0 [2] and they share their data or information on social media.

Social media has bestowed users with much more power than they possessed before its advent. Twitter is one of the social networking services which enable its users to post and

read tweets. Tweets are short messages of 140 characters. The users who have registered accounts on Twitter can post as well as read all tweets but the unregistered ones cannot post them. Hu et al. [3] developed a statistical model for prediction of real life events from tweets by considering the effect of their engagement on Twitter on events of the real world. They found that social network structure of tweets of a particular user along with his prior activities proved to be the best indicators of happening of a real-world event as well as their degree of engagement in it.

Sentiment Analysis is the process of elicitation, comprehension, classification and representation of opinions or sentiments expressed by various users regarding a topic or object. It is a model that requires input in form of text or document and after analysing it, summarizes the opinions or sentiments present in that text or document. Sentiment analysis finds application in systems that present summarization of reviews, dialogue systems, analysis of media applications. It imparts companies that offer any service or product a method for scrutinizing reviews published by different users and hence estimation of product acceptance can be done. Generating automatic labels for review documents like products or movies regarding their polarities is useful in business intelligent and recommending systems [4]. For consumers, this plenty of data and opinionated content from different sources assists them to utilise the wisdom and views of crowds, in order to improve their decisions [5].

Many research works have been carried out in the field of sentiment analysis. Bermingham and Smeaton [6] compared performance of SVM with Naïve Bayes (NB) for analysis of microblogs and microreviews using different features and found sentiment classification easier in microblogs as compared to blogs. SVM outperformed NB in blogs but the opposite happened in the microblogs. They concluded that syntactic patterns represented by the POS n-gram features contain more authentic knowledge than unigrams. Davidov et al. [7] utilised framework of Twitter tags and smileys as sentiment labels. The basic feature types utilised by them included pattern, n-gram, single word and punctuation features. Results were obtained

for binary as well as multi class classification.

Go et al. [8] classified tweets with help of emoticons and the resulting labelled data was used for training. Using only unigrams, SVM performed the best followed by Naïve Bayes and Maximum Entropy. Bigrams diminished the accuracy of classification using SVM and Maximum Entropy. Pak and Paroubek [9] tagged the input data or text using Part of Speech tags. Then after tagging the input, the tags were matched using pre-defined rules to the already specified tags which determined if the document was subjective or objective. N-grams and POS tags were utilised to extract features and Multinomial Naïve Bayes classifier was employed. They gained the best accuracy using bigrams.

Amolik et al. [10] employed various machine learning algorithms for opinion mining of reviews of movies. They extracted twitter specific features to form feature vector. SVM was found to have higher accuracy (75%) than Naïve Bayes (65%) using this feature vector. Esuli and Sebastiani [11] developed a resource containing lexicons to be used for classification- “SentiWordNet”. This is available for Opinion Mining in public domain. Singh et al. [12] performed opinion mining of reviews of movies and blog posts and evaluated classifier’s performance by assigning different weights to adverb and adjective SentiWordNet scores. This helped to assign a weight to the adjective with help of a score that modified or altered its importance in accordance to the adverb which preceded it. Any of the SentiWordNet approaches did not perform as well as NB and SVM.

Basari et al. [13] implemented a new approach for opinion mining in movie domain by hybridization of Support Vector Machine with PSO (Particle Swarm Optimization). N grams were employed for feature extraction and theses were further filtered by using weighting techniques like tf and tf-idf. The selection of best subsets from all the possible subsets was performed by PSO. This optimal selection of features helped SVM to achieve better accuracy. This hybrid method achieved better accuracy (77%) and precision than using only SVM. However, it led to a decrease in recall and increase in time complexity.

## II. RESEARCH DESIGN

The flow diagram in Figure 1 shows the research design employed in present work.

### A. Data collection

Twitter API (Application Programming Interface) provides a programmatic method for retrieving tweets containing the required keyword. A secured access to this API is provided by a protocol OAuth. The Twitter API contains an option for specifying the language. This parameter has been set to English in this work because our training data also consists of only English tweets. Specific queries are provided and Tweets containing that keyword are fetched using the twitter APIs. A query can belong to any domain, it can be a brand, product or a service. The number of tweets to be downloaded for each query has been set to 1000.

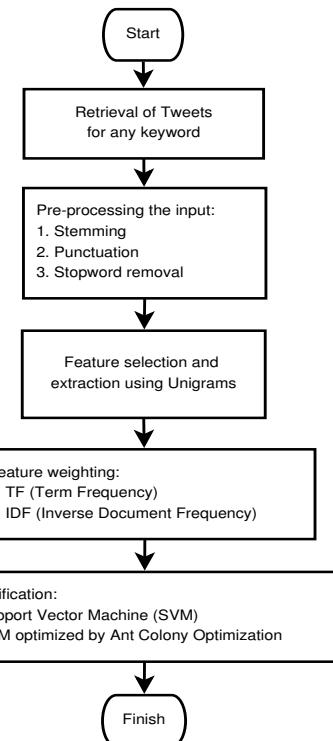


Fig. 1. Methodology for Sentiment Analysis of Tweets

### B. Pre-processing

Pre-processing is the process of identifying and eliminating noise or irrelevant data from the input. Hence it cleans the data for further processing. Pre-processing has a huge impact on the accuracy of classification. As crowdsourced data contains content that may not have any impact on the polarity of the text, inclusion of such content will only lead to increasing the dimensionality of feature vector thus increasing the complexity of the classification process.

The pre-processing steps that have been incorporated in our work in order to improve accuracy include Punctuation erasure for erasing punctuation characters like period, exclamation point, comma, apostrophe, question mark, quotation mark etc., Numbers filter for filtering all those terms that consist only of numbers, Case converter for converting all the terms present in the text to lower case. Snowball stemmer and stopword filter have also been utilised.

### C. Feature creation

Features are those terms or phrases which impart a positive or negative polarity to the text from which they are generated. The feature selection can be based on unigrams, bigrams, part of speech tags etc. Widely used feature weighting techniques are Term frequency(tf) and Term frequency-inverse document frequency (tf-idf). The present work employs unigrams as features. Tf-idf is utilised for creating feature vector. Tf-idf weight is combination of two terms: TF and IDF which are computed using equations 1 and 2.

$$TF = \frac{\text{Number of times term } t \text{ appears in a document}}{\text{Total number of terms in the document}} \quad (1)$$

$$IDF = \log_e \left( \frac{\text{Total number of documents}}{\text{Number of documents with term } t \text{ in it}} \right). \quad (2)$$

#### D. Classification using SVM

SVM builds a hyperplane in high-dimensional space for classification. Best performance is attained by a hyperplane that can maximize the distance of nearest training instance of both classes i.e. maximize the functional margin. The main objective is to decrease the generalization error and make it resistant to over fitting.

Consider a classification task :  $\{x_i, y_i\}, i \in 1, 2, \dots, l$ ,  $y_i \in \{-1, 1\}$  and  $x_i \in R$  where  $x_i$  and  $y_i$  denote data point and its corresponding label respectively. The hyperplane for separation is given by equation 3.

$$w^T x + b = 0 \quad (3)$$

where  $w$  and  $b$  denote the coefficient vector and offset from origin respectively. The ideal margin for separation is obtained by solving the optimization objective stated in equation 4 subjected to conditions in stated in equation 5.

$$\text{Minimize } g(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (4)$$

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (5)$$

Here,  $C$  is the generalization parameter and  $\xi$  is positive slack variable. This optimization problem is reduced to Lagrangian dual problem by introducing Lagrangian multipliers  $\alpha_i (i = 1, 2, \dots, n)$  and by Karush Kuhn Tucker (KKT) condition, we can obtain  $w$  and  $b$ . Hence the linear discriminant function is represented by equation 6.

$$g(x) = \text{sgn} \left( \left[ \sum_{i=1}^n \alpha_i y_i x_i^T x \right] + b \right) \quad (6)$$

When the classes are not separable by a linear line, the original feature space is mapped to feature space of some high dimensionality. The new decision function is represented by equation 7.

$$g(x) = \text{sgn} \left( \left[ \sum_{i=1}^n \alpha_i y_i \phi(x_i)^T \phi(x) \right] + b \right) \quad (7)$$

where  $x_i^T x$  of the input space is depicted as  $\phi(x_i)^T \phi(x)$  in feature space.  $\phi(x_i)$  can be computed by using kernel function. There exist many kernel functions which can be utilised by Support Vector Machines. Present work uses Gaussian kernel (or RBF) which is defined by equation 8 where  $\gamma$  denotes the predefined parameter that controls width of Gaussian kernel.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (8)$$

#### E. Classification using SVM-ACO

Ant Colony Optimization is one of the approaches that utilize swarm intelligence in solving real world problems. It is a model that solves optimization problems by designing meta heuristic algorithms [14], [15]. Set of computational asynchronous and concurrent ants move across to find the solutions

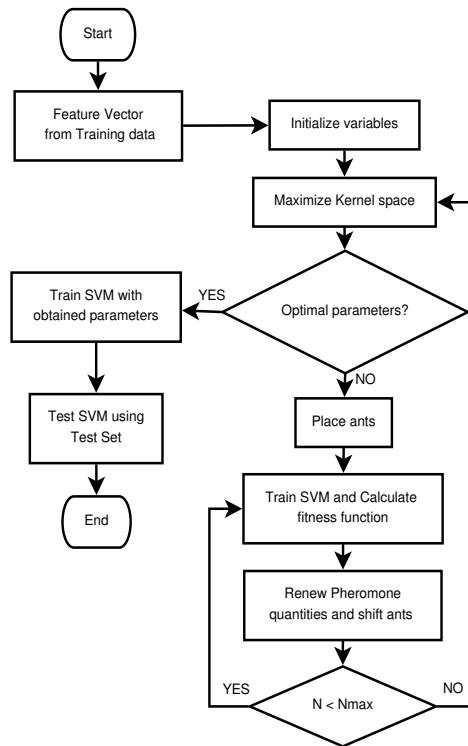


Fig. 2. Optimization of SVM using ACO

of sub problems which correspond to partial solutions to the main problem to be solved. Each ant builds its own solution for the problem incrementally after each movement. An ant updates its trail value according to the components used in its solution after evaluating and completing its solution or at its construction phase. The search of the future ants will be directed and affected to a large extent by this pheromone value.

For any ant, the probability  $p_{xy}^k$  of changing its state from  $x$  to  $y$  is dependent on prior attraction of that change/move denoted by  $\tau_{xy}$  and the posterior information of trail level about the proficiency in the past to make that move denoted by  $\eta_{xy}$ . It is computed using equation 9.

$$p_{xy}^k(t) = \frac{[\tau_{xy}(t)]^\alpha \cdot [\eta_{xy}]^\beta}{\sum_{z \in \text{allowed}_z} [\tau_{xz}]^\alpha \cdot [\eta_{xz}]^\beta} \quad (9)$$

where  $\alpha \geq 0$  and  $\beta \geq 1$  parameters are used to control the influence of  $\tau_{xy}$  and  $\eta_{xy}$  respectively.  $\tau_{xz}$  and  $\eta_{xz}$  represent attractiveness and trail level for all other transitions that are possible. The trail level is altered using equation 10:

$$\tau_{xy}(t) \leftarrow (1 - \rho) \cdot \tau_{xy}(t) + \sum_k \Delta \tau_{xy}^k(t) \quad (10)$$

where  $\rho$  indicates the pheromone evaporation constant,  $\tau_{xy}$  indicates amount of pheromone that is deposited for transition on that state and  $\tau_{xy}^k$  is the pheromone deposited by the  $k^{th}$  ant.

Performance of SVM is subjected to value of two parameters to a great extent. These are: The generalization parameter  $C$  and the parameter for kernel function  $\gamma$ . The most suitable values of  $C$  and  $\gamma$  have to be selected for obtaining maximum accuracy and best classification performance using this model.

TABLE I  
COMPARISON OF SVM AND SVM-ACO

Parameter	SVM	SVM-ACO
Accuracy	75.54%	86.74%
Precision	72.77%	84.85%
Recall	70.98%	85.05%

The major problem lies in setting these two parameters such that the generalization error can be minimized. In the proposed model, the values of these parameters used by SVM are optimized by implementation of meta-heuristic process of dummy ants. Then classification is performed by Support Vector Machine model which takes these optimized values of parameters for construction of hyperplane. The process used by the proposed method is shown in Figure 2.

### III. RESULTS AND DISCUSSION

For comparison of performance of classification using SVM and SVM-ACO, three parameters: average accuracy, precision and recall have been used and Table I presents the values of these parameters for both SVM and SVM-ACO. This work uses Unigrams for feature extraction and Term frequency-inverse document frequency for feature weighting. These help to create an efficient feature vector.

As ascertained from Table I, average accuracy, precision and recall obtained with developed SVM-ACO system fare much better than using only SVM for the same dataset. The comparison is presented in Figure 3.

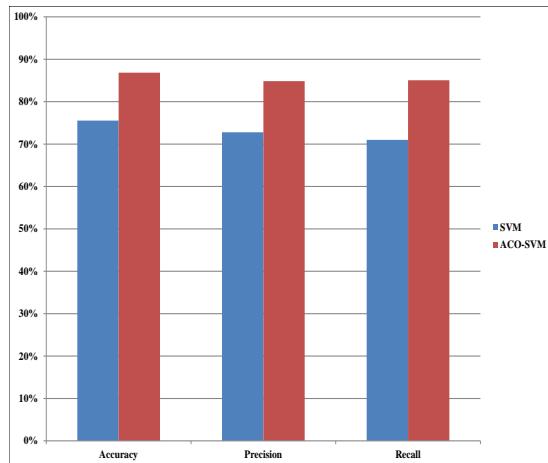


Fig. 3. Comparison of SVM and SVM-ACO

### IV. CONCLUSION

Crowdsourcing is the process of engaging a large group of people for generating and providing the data required by Enterprises for decision making processes. It increases the speed of data collection. This data is facilitated through Web 2.0 sites. These allow interaction of one user with others

and enable them to actively alter contents of any website, which distinguish it from traditional non-interactive websites which limit their users to only view the provided information passively. Twitter is one of the social networking services which enables its users to post and read “Tweets”. The present work is concerned with Sentiment analysis of tweets. Tweets are retrieved using Twitter APIs for a specific keyword. These are further pre-processed for noise removal and Unigrams have been used as feature extraction technique. TF-IDF is used for assigning weights to these features. The present work shows that Ant Colony Optimization effects the accuracy of Support Vector Machine. SVM, being a parametrised classifier, is highly dependent on values of these parameters. These parameters are optimized with the use of Ant Colony Optimization. The best average accuracy achieved is 86.74% by SVM-ACO model.

### REFERENCES

- [1] M. K. Poetz and M. Schreier, “The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?: The Value of Crowdsourcing,” *Journal of Product Innovation Management*, vol. 29, no. 2, pp. 245–256, Mar. 2012.
- [2] P. Andersen, *What is Web 2.0?: ideas, technologies and implications for education*. JISC Bristol, UK, 2007, vol. 1, no. 1.
- [3] Y. Hu, S. Farnham, and K. Talamadupula, “Predicting user engagement on twitter with real-world events,” in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*. AAAI, 2015.
- [4] Y. Hu and W. Li, “Document sentiment classification by exploring description model of topical terms,” *Computer Speech & Language*, vol. 25, no. 2, pp. 386–403, Apr. 2011.
- [5] V. Sindhwani and P. Melville, “Document-word co-regularization for semi-supervised sentiment analysis,” in *Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on*. IEEE, 2008, pp. 1025–1030.
- [6] A. Bermingham and A. F. Smeaton, “Classifying sentiment in microblogs: is brevity an advantage?” in *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010, pp. 1833–1836.
- [7] D. Davidov, O. Tsur, and A. Rappoport, “Enhanced sentiment learning using twitter hashtags and smileys,” in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 2010, pp. 241–249.
- [8] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” *CS224N Project Report, Stanford*, vol. 1, p. 12, 2009.
- [9] A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining.” in *LREC*, vol. 10, 2010, pp. 1320–1326.
- [10] A. Amolik, N. Jivane, M. Bhandari, and M. Venkatesan, “Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques.” *International Journal of Engineering and Technology*, vol. 7, no. 6, pp. 2038 – 2044, 2015.
- [11] A. Esuli and F. Sebastiani, “Sentiwordnet: A publicly available lexical resource for opinion mining,” in *Proceedings of LREC*, vol. 6. Citeseer, 2006, pp. 417–422.
- [12] V. K. Singh, R. Piryani, A. Uddin, and P. Waila, “Sentiment analysis of Movie reviews and Blog posts,” in *Advance Computing Conference (IACC), 2013 IEEE 3rd International*. IEEE, 2013, pp. 893–898.
- [13] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, “Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization,” *Procedia Engineering*, vol. 53, pp. 453–462, 2013.
- [14] M. Dorigo, M. Birattari, and T. Stutzle, “Ant colony optimization,” *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28–39, Nov. 2006.
- [15] M. Dorigo and T. Stützle, “Ant colony optimization: overview and recent advances,” *Techreport, IRIDIA, Universite Libre de Bruxelles*, 2009.

# Defending Against Attacks from the Dark Web

## Using Neural Networks and Automated Malware Analysis

eng. Mihai-Gabriel IONITA\*

Computer Sciences and Information Technology Doctoral  
School  
Military Technical Academy  
Bucharest, Romania

Prof. Victor-Valeriu PATRICIU

Computer Sciences and Information Technology Doctoral  
School  
Military Technical Academy  
Bucharest, Romania

**Abstract**— In an Internet connected world, cyber security assurance is critical for protecting an organization's critical infrastructures. For this task, we propose a connected infrastructure that offers various types of malware analysis capabilities. This infrastructure's architecture is based on customized open-source projects. This proposed implementation has been integrated into an already built platform that aims to protect an organization's geographically distributed network. Our proposed implementation is based on software defined network components, and it uses artificial neural networks for protecting these critical infrastructures. The malware analysis component is based upon three sub-components that perform static and behavioural analysis against suspected pieces of code, documents or traffic. In addition, when attacks that involve zombie computers come from the Dark Web, the proposed platform tries to uncover their true source, so it can inform the unsuspecting users or defer them to justice. As detecting Tor traffic is not a trivial task, the platform includes a dedicated module for scanning and making a risk assessment of inbound and outbound connections. An intelligent firewall separates the protected infrastructure from malicious internet traffic by telling apart malevolent Tor traffic from other benign traffic flows. The platform also offers added protection against 0-day vulnerabilities and APT attacks by using its behavioural analysis techniques.

**Keywords-** cyber security, artificial neural networks, automated malware analysis, Tor, dark-web

### I. INTRODUCTION

When discussing about the Dark Web we refer to the services and the content exchanged over darknets such as Tor, through overlay networks or local friend-to-friend networks. As cyber criminals tend to move their operations to the Dark Web for added anonymity and protection, the face of cyber security will change completely. Privacy seeking users, which are scared of the government spying on them, also try to use the Dark Web, in some cases being infected with malware, or being targeted by government investigations, because of their anonymity seeking actions.

In today's cyber security context, everything is changing. From the 2013 Distributed Denial of Service (DDoS) attack on Spamhaus [1], which generated an unearthly 300Gbps traffic,

the DDoS attack against BBC [2] servers generated a hardly measurable 602Gbps figure, which makes the former attack seem like child's play. Other problems appear from the malware campaigns used to take down public utilities or factories, which bring financial harm or even physical injury to human beings by the so-called kinetic cyber-attacks. Such an example involves the malware campaign [3], where the Black Energy framework was used to target the Ukrainian power plant Prykarpattyabolenergo. A similar incident took place when the now famous Stuxnet malware was used to sabotage Iran's nuclear program. These kinds of targeted malware campaigns are impossible to detect with traditional antimalware applications. Classical antimalware applications, such as antivirus products, rely on comparing pieces of analysed code against malicious code definitions that are provided when an analyst detects a piece of code as being malicious, or infections are reported. The same situation appears for web security applications where they also block only known malicious websites or domains. To make matters worse, attackers started using the Dark Web and Tor nodes for Denial of Service attacks and for spreading malware. This predicament slows down any effort of identifying the aggressors by forensic investigators and law enforcement teams.

#### A. Concerns regarding the future of SCADA systems in utility monitoring and control which could lead to kinetic cyber-attacks

Unfortunately, a recent study [4], produced by the security company Tripwire, states that over 80% of the respondents have seen an increase in the successful cyberattacks over the past year. Alarmingly, the study has highlighted the fact that for the majority 53 percent of the surveyed companies in this business sector, the number of attacks has even doubled over the past year. On the bright side of things, at least 69% of the questioned employees have said that their company does not detect all cyber-attacks and for over 72% of the involved companies a single executive has security responsibility for both IT and operational technologies (OT). As this information is not a pleasant one, it is important that high-level executives from these companies understand their

current standing regarding the problem and are willing to invest in the safety of their business.

In the wake of the Dark Web's expansion and that of Tor traffic, more and more cyber criminals get interested in attacking important public utilities. From state sponsored groups that look after industrial secrets and employ in industrial sabotage actions to the more benign script-kiddies which are just looking to try out their newly acquired "powers" and get some attention, everybody is rejoicing for the added anonymity the two latter communication options offer.

This added anonymity has to be used to better protect well-intended citizen's privacy and not aid criminals in planning secret attacks, which could lead to the loss of human lives. From recent reports [5], [6], it appears that the intelligence agencies of countries that invest highly in cyber security are closely monitoring these forms of communication and are actively trying to breach their security for intercepting the messages that are exchanged there. Even though for the society the fact that the government minimizes an internet user's right to be anonymous is upsetting, it also has to come to mind that state agencies' actions stop terrorist attacks. These attacks could lead to worse situations than an individual that does not benefit from maximum privacy.

Unfortunately, a large number of small and medium-sized organizations own networked SCADA systems. These systems have to be protected, irrespective of the funds such a company allocates for its IT budget.

For a civilian organization the only way to protect against these kind of attacks is to deploy defence in depth techniques, as presented in our other paper [7], for keeping operations under control. Nonetheless, when proposing such an architecture, an organization's size has to be taken into account. Any equipment acquisition has to be tailored accordingly to the organization's budget. As described thoroughly in Zhao's paper [8] for this kind of situation, a cost-benefit analysis is mandatory, following a thorough risk analysis. Only after completing these two important steps, an acquisition can be entitled for public posting.

In our opinion, one of the hardest sectors to regulate when it comes to SCADA security is the financial one. Not only the public utilities use SCADA systems for their internal mechanisms, financial institutions also hold such systems and use them. Such systems are installed especially in the USA and the UK. Unfortunately, these are also vulnerable to cyber-attacks and are more desirable for the motivated hackers. Regulating the use of SCADA systems inside financial institutions is hard because these structures are financially motivated. They take investment decisions only based on revenue and financial gain after a thorough risk analysis. There are usually two kinds of motivations for orchestrated cyber-attacks: financial gain, by manipulating data [9] or

disruption of services, by the means of sabotage [10]. Manipulation of data can lead to the manipulation of a stock's exchange value concerning a specific share at a specific moment in time. Alternatively, tampering with data regarding the sum of money available in a specific bank account can make somebody rich overnight just by pushing a button. Of course, there are different security solutions in place to stop these kind of actions, but critical vulnerabilities tend to appear all the time. Firewalls, Web Application Firewalls (WAF) and Intrusion Detection Systems (IDS) or Intrusion Prevention Systems (IPS) usually protect such platforms. These systems are usually cascaded and setup in high-availability mode, especially for repelling any kind of known attack types. However, serious vulnerabilities can be discovered in core components of such appliances, as was the case with the Heartbleed [11] Transport Layer Security (TLS) vulnerability which over exposed sensitive data in a large proportion, about 55% of HTTPS supporting sites [12], servers, firewalls, security appliances and software applications. Similar large-scale vulnerabilities have been discovered in applications that are largely deployed like OpenSSH [13]. In OpenSSH's versions spanning from 5.4 to 7.1, a vulnerable client, connecting to a malicious server, would expose its private SSH keys as well as its virtual memory. Another recent vulnerability affected the widely used 7-Zip archive manager, which is also imbedded in several security appliances, like FireEye and software applications, including antivirus products like Malwarebytes and Comodo Antivirus [14]. The vulnerability discovered by Cisco's Talos Security [15] may lead to arbitrary code execution and corruption of files. These events happened in the past and will surely happen in the future. In addition, you can never know when an employee is willing to tip the scales in favour of the competitors, or just make a mistake, which will have the same result. Failsafe systems that are in place will not be taken into discussion, but inarguably human error remains a possibility that has to be taken into account when making a risk analysis.

For this specific sector, an interesting proposal comes from this article [16], where the authors study the effectiveness of an Early Warning System tailored to NFIs (Non-Financial Institutions) this is similar to our own idea of such a system, presented in [17], but fundamentally different in the respect of applicability. Our proposed architecture is a multi-purpose cyber security system designed for event detection and investigation, in some way more familiar to a SIEM (Security Information and Event Management) in functionality. On the other hand, their system is a condition detection system particular to the financial domain. In any respect, more initiatives of this kind will help shape the financial domain's security.

### B. Assessing the Tor-Web's size and the number of posted hidden services

When it comes to defending against cyber-attacks that come from the dark-web of Tor's network, an organization has to know what it is facing.

According to the following figures, captured from the Tor project entitled TorMETRICS [18], there are some interesting conclusions that can be seen.

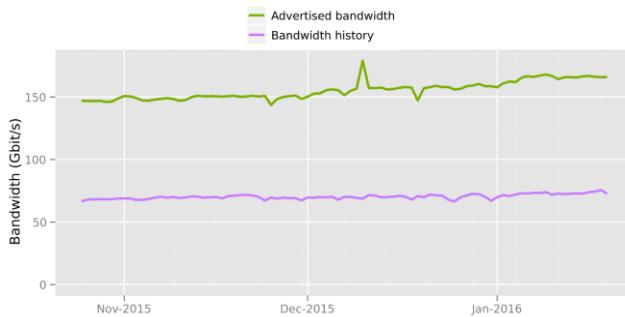


Figure 1 - Tor Total Relay Bandwidth [18]

According to the above figure, fig.1, the total relay bandwidth of the Tor network increased steadily after December 2015, over the 150 Gbit/s limit it was at the end of the year 2015. As stated above, Tor traffic is rising slowly with spikes of around 175 Gbit/s, usually in correlation with events that necessitate more anonymity. As a fact, Tor traffic is expected to rise in the near future. This can be confirmed only by looking at the number of unique onion hidden services created in the first part of 2016 and correlating it with censorship related events from around the world [19]. This brings a very serious issue into discussion. How do you tell apart criminals from security minded users, which try to go by unnoticed on the web and don't want to be victims of large ad selling companies, or which don't want to be followed on the web for identity theft? All of these are legitimate fears. Moreover, they will appear even more often if products and protocols are not regulated accordingly.

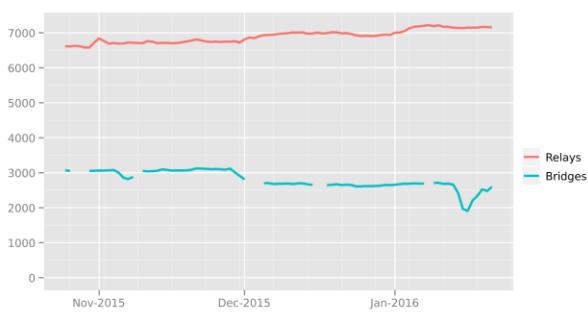


Figure 2 - number of Relays and Bridges in the Tor network [18]

Another interesting aspect that should be noted is the growth of relay numbers, as indicated in the above figure, fig. 2. Relays are publicly listed nodes, in a directory authority, which forward traffic for clients. On the other hand, the second tendency depicted in fig. 2 is the reduction of the

number of bridges. These are the most important to our research, because these are hidden. They are not publically listed and are most of the time used in combination with the pluggable-transport option. This alternative transport protocol alongside hidden nodes and bridges help Tor users circumvent ISP (Internet Service Providers) or governmental blocking of Tor traffic at the ISO/OSI transport layer. Traffic coming in this combination towards our protected resource is of utmost importance to the research described in this paper. Because this is as anonymous as you can get using the Tor network. In addition, this is the method most used by attackers. In essence, when using the pluggable-transport option, Tor traffic is obscured to seem like a regular HTTP connection to an unblocked server [20]. All the obfuscation techniques are employed against ISP's which use Deep Packet Inspection (DPI) techniques.

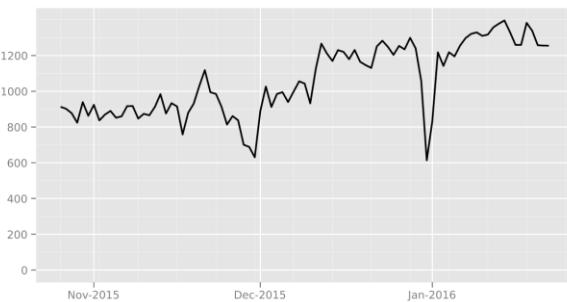


Figure 3 - Hidden-service traffic in Mbit/s [18]

There are also some interesting articles written about hidden service discovery and traffic monitoring. One particularly relevant for hidden-service determination and statistical data is the one written by Kadianakis et.al. [21] from the Tor Project Organization. They explain the mechanism behind the figures similar to that of fig. 3. They explain how they are able to determine the traffic of hidden services when only 1% of traffic from specific reporting relays is analysed. Everything is done by extrapolating from the small amount of data they receive. Such extrapolation, alongside a good statistical analysis, is a good starting point, if not all that is needed to determine servers that host hidden-services. The U.S. Naval Research Laboratory, along with some other researchers, has done similar interesting research from the Tor Project Organization in the article written about hidden-service statistics [22]. In this article, they present the evaluation criteria for gathering statistics as well as the obfuscation methodology used for this analysis. Another initiative is underlined by the article written by Loesing et.al. [23] , in which another fundamental characteristic of the Tor network is analysed, the QoS (Quality of Service) parameters. In short, QoS refers to the time it takes a user to get his requested resource, after making a request. In strong connection with our research is the article written by Zander et.al. [24]. This paper provides us with a method in combination with the article written by Loesing et.al. [23] that can permit an analyst to determine if a particular server is hosting a Tor hidden-service. This is all done by estimating

clock-skew on a legitimate server that is supposed to run a hidden-service, in response to a series of queries, while monitoring its QoS parameters. Another interesting approach is that of Biryukov et.al. in their paper [25], which presents flaws in the way Tor hidden-services function that allow an attacker to de-anonymize or even disable a hidden service.

There are also more aggressive techniques for determining a hidden user's identity. Such as fingerprinting a browser's audio stack by using the HTML5 AudioContext API as described in [26] or by using the HTML5 Canvas to dynamically generate an image used for identification and tracking as presented in [27]. However, these are going to be presented in the following chapters of this paper.

### C. Malware spread using Tor and its analysis

There were different campaigns [28], [29] that demonstrated Tor could be used against its users. There are usually three motives for which Tor services and relays can be tainted to distribute malware:

- Espionage – different organizations try to find out why other entities use Tor and what kind of data is transported through it. In addition, user profiling is recommended because using Tor means that users possess valuable information.
- Law enforcement surveillance – With terrorism in mind and other threats to the citizen's lives, governmental agencies will most likely keep under close observation covert communication channels, such as Tor.
- Criminal activities that involve financial interest – There are people which try to make payments across the Tor network and getting free bitcoins is enough for some criminals. Alternatively, a hacker could try to connect an innocent user's computer to a botnet, transforming it into a

zombie. These zombies can do anything from mining bitcoin for the bot-herders to conducting large scale DDoS attacks, like presented in our other article [30].

Confirming fears expressed in one of our previous papers regarding the implication of malware distribution attacks in cyber security [17], a recent threat report [31] from F-Secure Labs has demonstrated that malicious Tor relays are spreading malware. Moreover, they presented their analysis of the involved samples, in which a malicious Tor relay in Russia was tainting all the downloaded executables that passed through it with a Trojan dropper. The now modified executable contained also the original binary, the one the user was looking for and a Trojan dropper that would install the OnionDuke backdoor, alongside the legitimate application, as depicted in the flow chart from their report, fig. 4. After the drop has taken place, this bot, part of the MiniDuke malware family, connects to several C&C (Command and Control) servers which are hardcoded in its configuration file to receive new orders like: download new components, self-destruct, attack host etc. Another important thing to note from their research is that the hard-coded C&C servers were not dedicated servers, but compromised innocent web sites. This variant of the MiniDuke family would be able to use specific twitter accounts as a backup C&C server for receiving commands. Other papers delve into the analysis of malware present on the Tor network as well as the Dark Web, like this one from Trend Micro [32]. Which talks about not only the Tor network and its hidden services but also other dark nets like the following:

- I2P – Invisible Internet Project

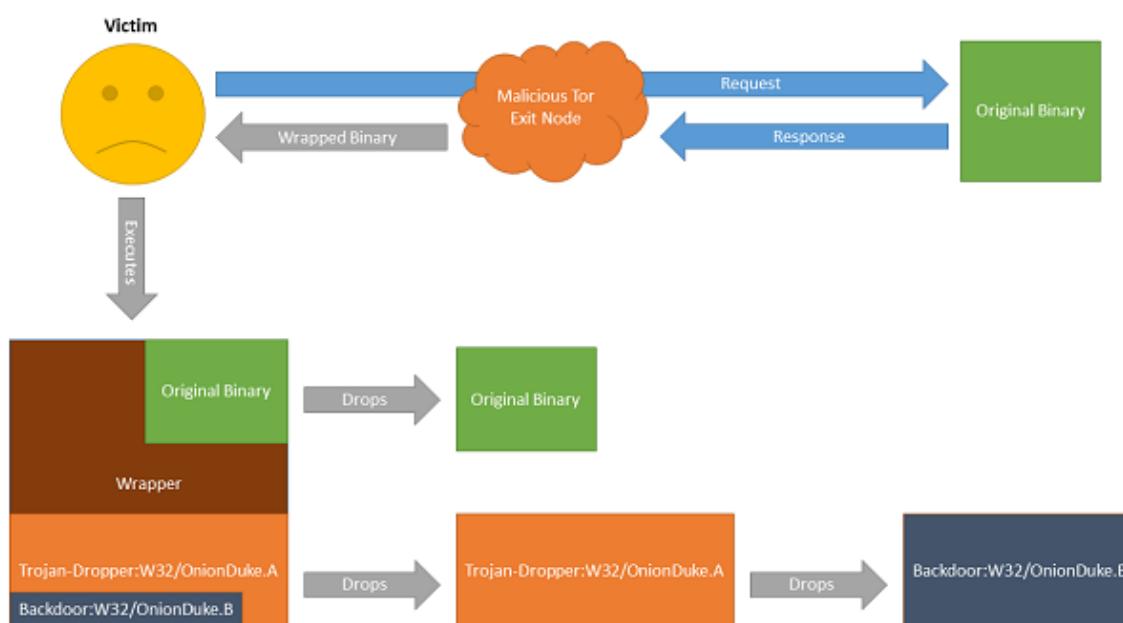


Figure 4 - TOR OnionDuke infection flow chart [31]

- Freenet
- Alternative Domain Root (Rogue TLDs)
  - Namecoin
  - Cesidian root
  - Namespace.us
  - OpenNIC
- Tor Marketplaces
  - Silk Road
  - Atlantis
  - Black Market Reloaded
  - Sheep
  - Underground Market Boards
  - Underground Marketplaces

During Tor's lifetime, different malware spreading campaigns were made public, like the following: The Skynet Botnet, the Mevade Botnet, the Atrax Crimekit, 64bit Zeus, ChewBacca Malware, Bifrose Malware, Cryptowall version 2.0 (Tor Cryptowall), the Citroni Ransomware(CTB-Locker), the TorRAT, ZBOT, etc.

In addition, I2P is another popular dark net for malware, with a few discovered samples: Criptowall 3.0 and the Dyre Trojan.

Moreover, a very recent article [29] reveals that the FBI used malware on TorMail to capture paedophiles. During this operation, the FBI deployed an NIT (network investigative technique) which bombarded users of the TorMail service with malware. This happened also to legitimate users, not only to criminals.

## II. ANALYSING TRAFFIC FOR DETECTING ANOMALIES AND ATTACKS

With Tor analysis, the process of de-anonymizing users is straightforward. Whoever controls the exit nodes controls Tor user's traffic. This is what law enforcement agencies around the world are trying to do and have been exposed by security researchers [33]. With around 7500 active Tor relays, about 1000 are exit nodes, from which 65 were „malicious” exit nodes, according to the results of the research paper published by Winter et.al. [34].

They are marked as malicious because they are actively trying to intercept traffic or to find the originator of the transmission. As indicated in the same paper, 6.5% out of all the exit nodes is not a huge number, but it can be considered important when it comes to governmental agencies' surveillance techniques. The problem with this kind of attacks is that malicious nodes are flagged quickly and are abandoned by users.

These types of attacks do not represent a novel approach. Web anonymity has been a long desiderate of cryptographers and users alike. Crowd or MIX systems had been proposed as an alternative to the way people exchange messages as far as the 1980's through Chaum's "Untraceable electronic mail, return addresses, and digital pseudonyms [35]." However, unfortunately, this method of exchanging messages would not be of use for the other protocols that make up the internet which are of interest and used more. Another proposition on which the Tor system could be based is the Crowds or MIX systems, as described in this paper [36]. A newer approach is presented and analysed in [37], where their proposed model lets every network user choose his own level of protection, regardless of the other user's choice.

The information that the FBI (Federal Bureau of Investigation) used a Metasploit module for apprehending criminals involved in child pornography shook the whole cyber security world. It used the "Metasploit Decloack Engine" which employs a combination of client-side technologies alongside locally installed applications for uncovering the IPs of the visitors of a certain site. These all happened during the FBI's "Operation Torpedo". The techniques required the user that was accessing the monitored site to have at least one of the following applications installed on their system:

- Microsoft Word
- Oracle Java Runtime
- Adobe Flash Player
- Apple QuickTime
- Apple iTunes

Even when accessing the monitored site from the Tor network, these locally installed applications, in combination with a script written in JavaScript or a Java applet would make the user connect directly to the monitored site, thus revealing his true identity. All of these would happen instantly without the user suspecting anything. This is how the FBI apprehended 25 US citizens and many more abroad.

A similar technique is used for discovering the true IP address of an attacker if he uses the Tor network through cyber-retaliation techniques, which are described in one of our other papers [38].

For such reactive measures, the defence system has to detect that it is dealing with Tor traffic. For this to happen Tor traffic detection has been setup as follows.

Using the exact "Metasploit Decloack Engine" attack model is not feasible any more. It is known and Tor Browser developers are actively trying to counter it. In this regard, the only feasible approach and the least intrusive is to sniff traffic coming toward and out of the protected organization. However, here appears another problem: Tor is designed to tunnel everything through the TLS protocol on standard ports, while respecting the RFC 2246. If you look at Tor traffic in a traffic analyser as Wireshark, you will see only legitimate SSL

connections to different hosts. Therefore, there are two viable options: first, you could determine Tor traffic based on its origin, correlating it with the updated list of Tor exit nodes from the Torproject. Secondly, you could look for anomalies or particularities in the inbound/outbound traffic of the protected organization, in hope of telling apart Tor tunneled traffic from regular, benign TLS traffic.

Using the first approach on the exterior border firewall, a list with all the Tor exit nodes addresses is kept up to date by using the page provided by the TorDNSEL, a branch of the Torproject. This list can be further refined by using the same project and providing an IP address and a port as destination. Nevertheless, our wish is to detect traffic from all Tor exit nodes so the platform will not refine the list. It will just trim the provided list for the column “ExitAddress” field value, which is an IP address. In our network setup, Tor traffic will be considered suspicious regardless of any other parameters, its source is sufficient to make it suspicious.

In the second approach, and in the hope of catching traffic between updates, or in case of malfunction of the first mechanism, all traffic is scanned with the open source project “Bro Network Security Monitor” which is running a special script tailored to our needs, called “detect-tor [39]” contributed by Seth Hall to the framework’s repository.

This module monitors traffic from a network tap in our infrastructure and outputs alerts like this one when analysing PCAP files of recorded network traffic from the Skynet Tor botnet (Trojan.Tbot):

```
#open 2016-01-31-21-39-41
#fields note msg src dst actions suppress_for dropped
#types enum string addr addr set[enum] interval bool
DetectTor:Found 172.16.253.131 was found using Tor by connecting to servers with at least 10 unique weird
certs 172.16.253.131 - Notice::ACTION_LOG 3600.000000 F
```

From the Wireshark forum, it appears that all Tor nodes

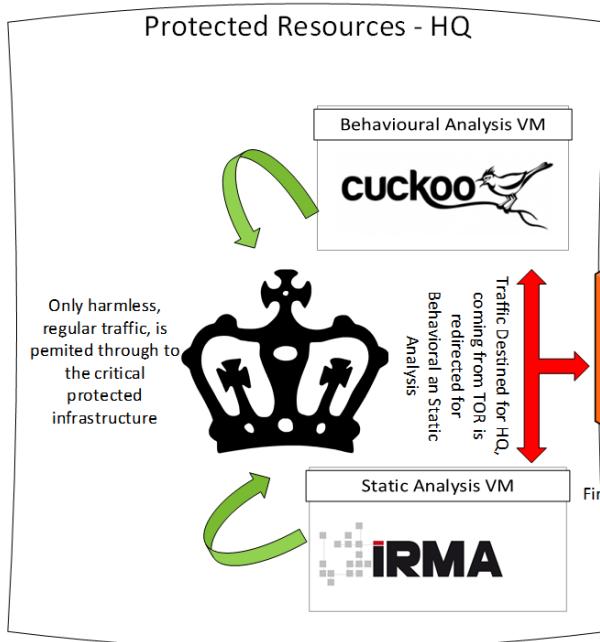


Figure 5 - The proposed setup for protecting critical infrastructures against Deep Web Attacks

present a certificate with a random name like these, from the PCAP analysed above:

- CN=www.ferqncujta3wvl.net
- CN=www.tl6ou6ap7fjroh2o.net
- CN=www.4iru7bub5avg.net

In combination with checking the IP’s in the provided Tor exit nodes list and looking for tell-tale ports like 9001 and 9030, detection of Tor traffic becomes a certainty.

The diagram from fig. 5 depicts the proposed implementation for protecting a critical infrastructure using automated malware analysis. As said earlier, fig. 5 illustrates the traffic flow of suspicious packets that arrive to the border firewall and are addressed to the protected critical infrastructure. The border firewall uses a neural network to classify traffic as suspicious regarding different patterns that appear in usual communications. If traffic originates from Tor exit nodes, it will be processed by the neural network and forwarded, if needed, based on its type to one, or both, of the analysis machines installed on premises.

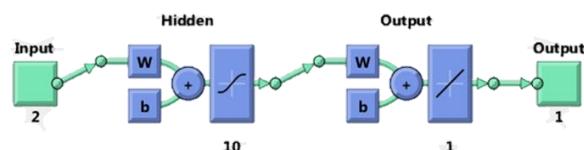
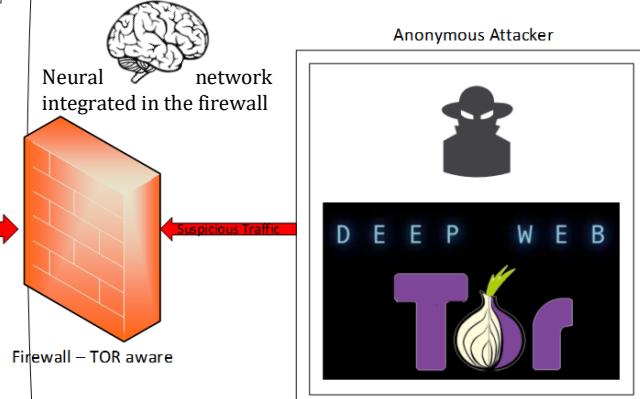


Figure 6 - The Implemented Artificial Neural Network

Figure 6 depicts the proposed architecture that implies a Feed-Forward Backward-Propagating neural network based on two input layers, ten hidden layers and one output layer. The



training was done using 1000 input values and 1000 output values captured from a network of sensors formed by local agents, based on the processed security events. These agents are locally installed and are based on the OSSEC Host-IDS platform. They monitor system file's integrity and processes running on the system, they also transfer all the important system logs as well as any anomalous behaviour to the SIEM central component as fully described in [40]. The training method used was Levenberg-Marquardt. Performance was calculated using the Mean Square Error approach. The architecture is fully described in our other paper [40].

After the suspicious traffic is analysed for malware it is also scored against our safety score matrix based on intelligent risk assessment [40]. A simple formula is used for risk calculation:

$$\text{Risk} = (\text{Probability} \times \text{Harm})^{(\text{Distress\_signal} + 1)} \quad (1)$$

The “Distress\_signal” used in the above equation is inspired from the theory of biological danger and is further described in our paper [40].

Moreover, it is forwarded to the protected infrastructure if it is considered safe, or it is dropped and the attacker is served a special page, which embeds the resources available in our custom version of the Metasploit „Decloack Framework” module. The framework that was described earlier as being used by the FBI for getting a user's real address is used only if the user's real address cannot be determined in another way. Other methods used represent browser fingerprinting techniques like scroll-wheel monitoring and HTML5 canvas observation as in [41]. These work only when the attacker has accessed the protected site without spoofing his IP, or for confirmation, after deploying the Decloack Framework. It is only used in extreme cases when repeated attacks appear from the same IP address. After determining the real address, retaliatory techniques are used for stopping the attacks on our network.

Other techniques, which do not perform as well, as the presented implementation, can be seen in [42].

The screenshot shows the Cuckoo Sandbox web interface. At the top, there are two input fields: 'File' (with a browse icon) and 'URL'. Below them is a large text input field with a 'Select' button. A note above the text field says 'Network routing through *dirty line* or VPN'. A dropdown menu next to it is set to 'None (no internet access)'. The main configuration area is titled 'Advanced Options' and contains several sections:

- Analysis Package:** Set to 'Detect Automatically'.
- Timeout:** An empty text input field.
- Options:** An empty text input field.
- Priority:** Set to 'Low'.
- Machine:** Set to 'cuckoo1'.
- Custom:** An empty text input field.
- Checkboxes at the bottom:**
  - No Injection (disable behavioral analysis)
  - Process Memory Dump
  - Full Memory Dump (if the "memory" processing module is enabled, will launch a Volatility analysis)
  - Enforce Timeout

**Figure 7 - The web GUI (Graphical User Interface) for submitting files in our custom version of cuckoo sandbox**

#### A. Using custom automated behavioural analysis for detecting threats

Contrary to older beliefs, an automated sandbox is not the holy grail of cyber security. It definitely helps a malware analyst, but it cannot finish the job alone. Moreover, the new types of malware have debugger, virtual machine and sandbox evading techniques built in. New malware creation toolkits have so many sandbox and debugger evasion techniques that the user is presented only with a checkbox named “Enable evasion techniques”. Which enables all the aforementioned options. In this way, running cuckoo sandbox and thinking you are safe is a grave mistake.

The custom version of the Cuckoo Sandbox which is used is enhanced so that it passes the Pafish framework checks for analysis environments. Out of 58 controls involving:

- Debugger detection
- VM detection based on CPU information
- Generic sandbox detection techniques
- Hook detection
- Wine detection
- VirtualBox, VMware, Qemu, Boschs and Cuckoo detection

Only 4 controls had warnings:

- Using mouse activity
- disk size <= 60GB via GetDiskFreeSpaceExA()
- MAC address starting with 00:05:69, 00:0C:29, 00:1C:14 or 00:50:56

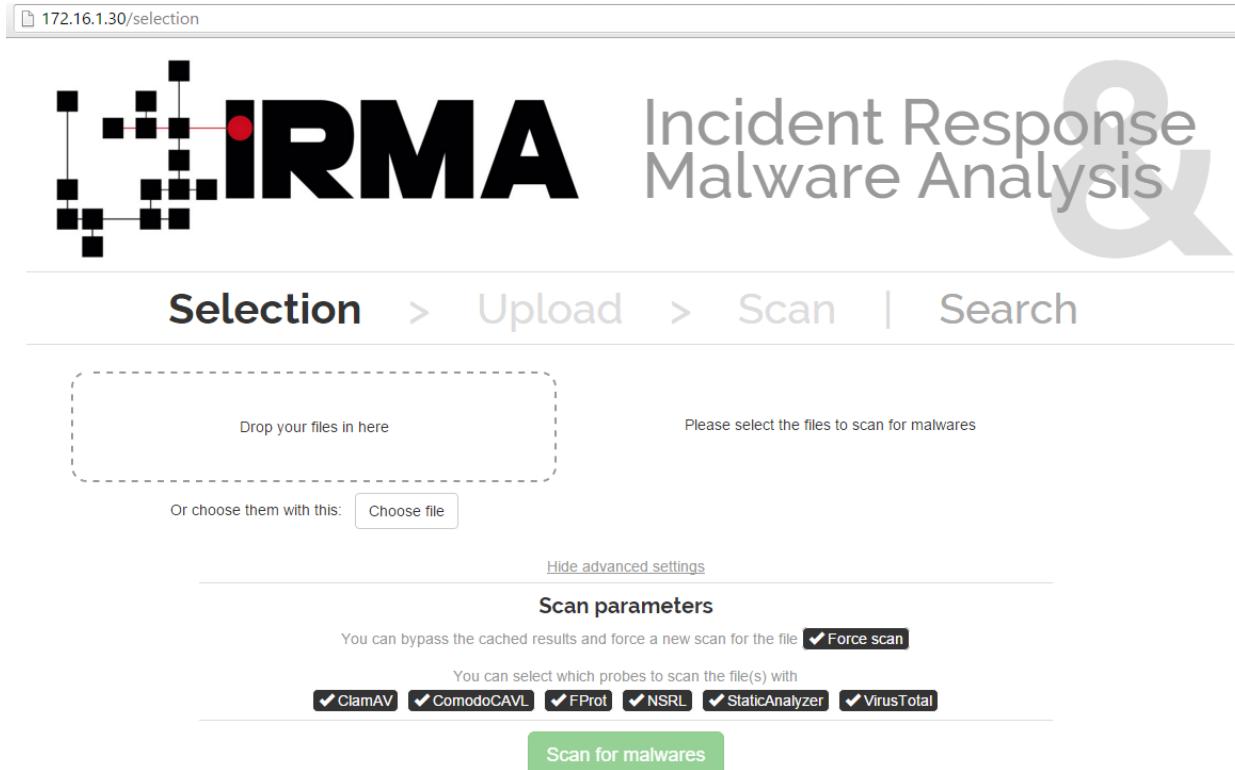
- Looking for pseudo devices

The virtual machines (VM) used for analysis are deployed using Jurriaan Bremer’s VMCloak tool for leaving less of a VM footprint for analysis. In addition, it has installed CheckPoint’s „CuckooSploit” module that allows the analysis of web-based exploits. For enabling PCAP emulation, Omri Herscovici’s “CapTipper” module is also installed. Malware samples are submitted automatically using the mechanism described in fig. 5, or an analyst or a user, using a web interface like the one in fig. 7, below, can also submit them manually.

This component gives the platform its ability to dynamically assess if its supplied file or sample of traffic is malicious. It not only analyzes its static properties but also opens the file in the intended application, similar to what a user would do on a workstation. If the analysis involves a traffic capture sample, then this is processed similar to a usual workstation on the network. The outcome of any of the aforementioned actions is closely analysed and documented correspondingly. The results can refer to dropped files, changed registries, added mutex-es, contacting C&C servers, etc. All these details are included in the Risk Analysis which is made to decide if a specific sample is malicious or not and if it can be forwarded to the protected infrastructure.

#### B. Using a custom version of the IRMA platform for statically analysing threats

Our static analysis machine is based on a custom version of QuarksLab’s version 1.2.1 of the Incident Response and



**Figure 8 – Web GUI for manually submitting samples to the static analysis VM, a custom version of the IRMA platform**

Malware Analysis (IRMA) platform. Compared to the stock version, it has enabled all the supported antivirus engines that do not require a subscription and all the online checks in the VirusTotal database are made using SHA1 hashes of the files, without uploading them to Google's servers. The activated components, which are used for document and file scanning, are as follows:

- ClamAV
- ComodoAV
- F-Prot antivirus
- VirusTotal is set for offline use. Only hashes of the analysed files will be uploaded to check if the files have already been scanned by Google's servers.
- NSRL – The NSRL (National Software Reference Library) is a large database supported by NIST (National Institute of Standards and Technologies) and the US DHS (Department of Homeland Security). It is made up of different RDS's (Reference Data Sets) which contain digital signatures of known files and applications. This is very helpful when deciding if an item is worth investigating or not. There are two ways of querying this database, an online fashion, where you upload hashes, or the offline one in which you rebuild the whole database on the analysis machine, about 20GB of digital signatures, which are kept offline. The latter one was chosen for protecting the organization's privacy. This great asset is used for creating whitelists of known, benign, files.

As described in fig. 8, below, files can be submitted through the CLI for automated analysis based on detections of the Neural Network or through the pictured web GUI, which also presents the different option for analysis, out of the six available options for scanning. The latter option is useful for analysts or users that want to check the security of their files.

- Similar research has been carried out in [43] where they use an approach inspired by the human immune system through the leverage of self – non-self-detection.

### III. EXPERIMENTAL RESULTS

For testing and proving the functionality of the proposed platform, 13 malware families, which use Tor as their C&C servers or documents that are dropped during attacks related to the Dark Net were chosen for analysis. These include malware and traffic samples involved in the following attacks:

- OnionDuke
  - Traffic and PDF files

- The 64bit variant of Zeus, which communicates using Tor
  - DOC, RTF and PPT files
- Documents that exploit the vulnerability described in CVE-2010-0188, CVE-2010-3333 and CVE-2012-0158. Which are still actively exploited in the wild, even if they are a few years old.
  - RTF, PDF and PPS documents

The experiments that were run include also some benign known PDF and DOC files for acting as witness files. A few important aspects have to be taken into consideration. While running the experiment, from the provided output it can be seen that all of the malicious documents have been marked as malware, either using the categorical scenario or using the majoritarian scenario. The equation used for determining the result of an analysis that is run on the platform looks like this:  

$$2*cuckoo\_result + irma\_result = final\_result \quad (2)$$

Equation 2 is used to determine the final result of the analysis as follows:

- cuckoo\_result can have values 0 or 1:
  - 0 benign
  - 1 malware
- irma\_result can have values 0 or 1:
  - 0 benign
  - 1 malware
- final\_result can have values from 0 to 3:
  - 0 harmless
  - 1 suspicious
  - 2 malicious
  - 3 malware

These values integrate the partial results of the components used. Moreover, they are used to simplify and match the results obtained from each of the two components in question. For example, cuckoo sandbox has analysis ratings that range from 0 to 10, with the possibility to go over 10 if specific conditions are met, while IRMA gives out a binary rating of green or red based on results from six subcomponents.

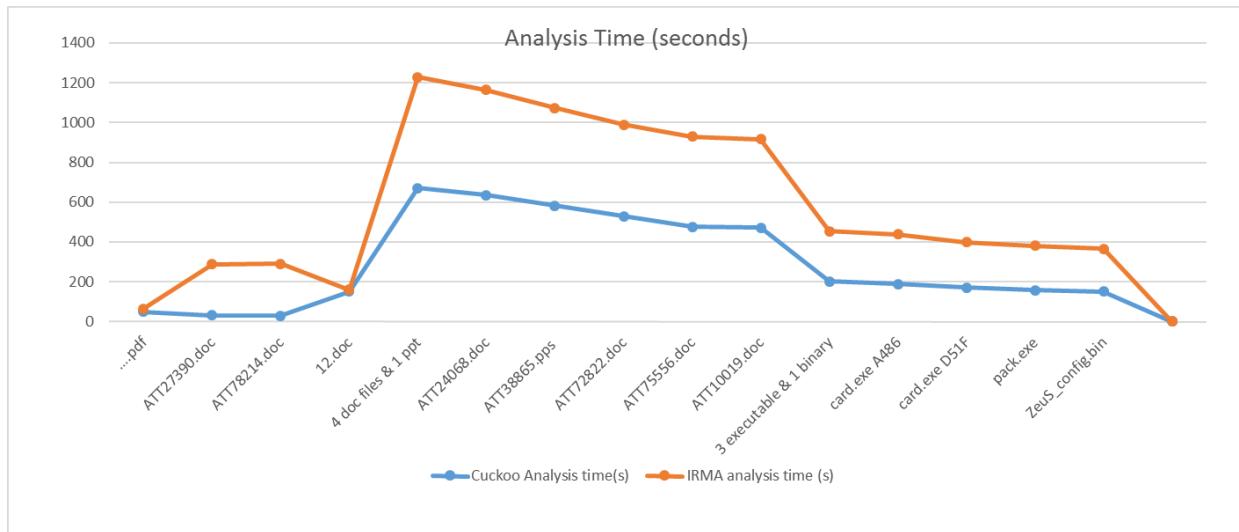
As described above, the categorical scenario appears when both of the platform's subcomponents flag a document or traffic sample as malware. If only one component flags the document or flow of traffic, then the second equation is applied. As can be seen from it, decisions made by the Cuckoo Sandbox component have a greater weight, because it includes more techniques of analysing files and traffic. It does a deeper and faster analysis of files and after experimenting with about a thousand samples, it is more accurate in analysing them. Moreover, it has a deeper perspective when unpacking files in a live environment and seeing exactly what system calls the files or what files it drops. A static analysis would never see this kind of actions. This is why the IRMA component is used in triaging the samples and the sandbox analyzes them afterwards, having more weight in the decision taking process.

As depicted in fig. 9 the total time for the two subcomponents is different. Moreover, as can be observed, with the orange line, IRMA's analysis lasts even eight or nine times as more as that of Cuckoo Sandbox. In addition, in some cases it can take less than that of the sandbox to complete, especially when using the full memory dump and analysis available in the Cuckoo component. This is caused by the fact that the

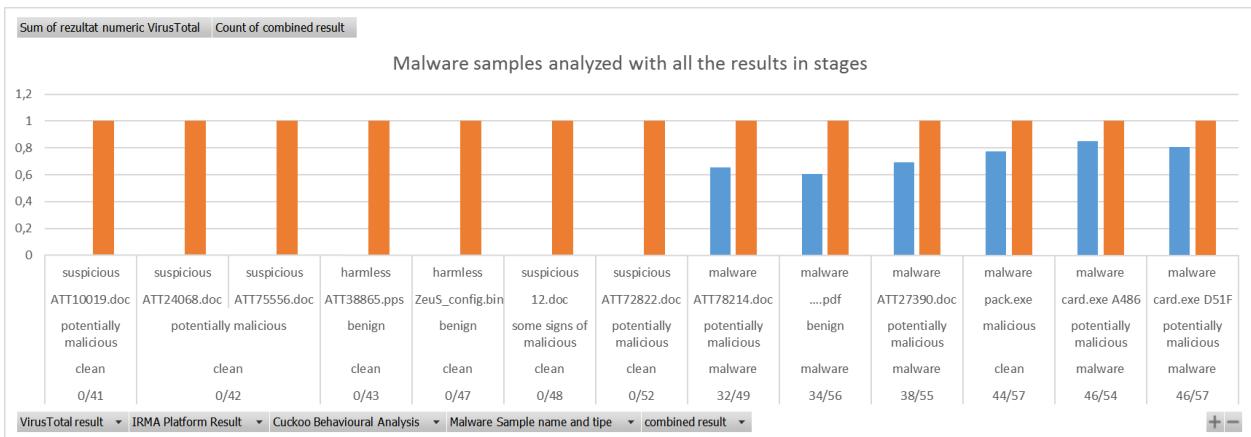
16GB of RAM, where each virtual machine uses 6 GB of RAM and 4 CPU cores.

In the below fig. 10, the results of the analysis during every stage are presented as follows, in order, from top to bottom, the labels displayed are as follows:

- The final result, as a combination of all the results bellow it



**Figure 9 – Comparison for the time it takes for the subcomponents to analyze documents from the 13 families of malware.**



**Figure 10 – Diagram of the most important malware samples, from the analyzed families, alongside the results obtained in every stage of the analysis and a witness value.**

IRMA subcomponent queries, for every item it examines, the offline image of the NSLR database that has around 20GB of digital signatures. Even by indexing it, performance does not vary greatly, due to its size. For a production environment, the machine that hosts the IRMA subcomponent could benefit from the added speed of storing its databases on an SSD (Solid State Disk). Overall, fully analysing a family of supposed malware samples in a maximum period of 20 minutes, statically and dynamically is a satisfying performance. The machine on which the tests were run is an Intel Core i7 with

- The name of the analysed sample
- The result obtained from the Cuckoo Sandbox subcomponent
- The result obtained from the IRMA subcomponent
- The result obtained from the VirusTotal platform, online, without using any of the platform's tools.

The independent result from VirusTotal is posted as a witness test. It is there to show that the results are comparable

to Virus Total. Even if in some cases neither of the antivirus engines used by VirusTotal detected something malicious, the proposed platform flagged the samples as suspicious and even potentially malicious. Which demonstrates the benefit the proposed platform brings to an organization's security. VirusTotal only uses classic antivirus engines to detect threats. The proposed platform, in addition to these engines, opens also the documents or processes traffic samples inside VMs that simulate normal workstations of users inside an organization. This allows an analyst go deeper inside an analysis, faster and safer in many circumstances.

#### IV. CONCLUSIONS AND FUTURE WORK

This novel implementation, which combines static and dynamic analysis for discovering a piece of code's intentions, helps reduce by more than a half the number of samples and security events that a security analyst has to parse.

With all the precautions implemented in the proposed architecture, the authors assume that this gives reasonable protection against attacks from the Dark Web and Tor connections. Undoubtedly more protection than using an offline scanning station that has some antivirus engines. These "disinfection" workstations are usually offline. Even supposing that all the virus definitions and program components are up to date for every scan, which is a daunting task even when you have a small number of antivirus products from different vendors, there would still be malicious files that would pass unnoticed. Moreover, 0-day vulnerabilities are not taken into consideration. Those would certainly pass unnoticed. In contrast, the proposed platform has a good chance of detecting them. These pose the largest threat for large organizations, because the latest vulnerabilities are usually bought from the Dark Web or other underground sources. Some 0-day vulnerability markets like TheRealDeal or WabiSabiLabi were present on the Dark Web, at some time and have sold, using hidden services, the latest vulnerabilities to unknown actors for large sums of bitcoin, for added anonymity. There are also legitimate firms that sell 0-day vulnerabilities to government agencies, like the French security team Vupen [44].

In combination with retaliatory techniques described in our other articles, the proposed setup would give strong protection against attacks, even in a WAN (Wide Area Network) environment with geographically distributed critical infrastructures. The platform's design includes increased protection for APT (Advanced Persistent Threat) attacks.

As a future evolution of this project the authors have already began work on setting up a similar setup in a SDN (Software Defined Network) for further attack evasion capabilities, including moving servers and services in real time.

The proposed SDN architecture would benefit from an increase in flow detection when using the multi-layered IDS concept described in [45].

#### REFERENCES

- [1] L. Constantin, "DDoS attack against Spamhaus was reportedly the largest in history," *Techworld*, 28-Mar-2013. [Online]. Available: <http://www.techworld.com/security/ddos-attack-against-spamhaus-was-reportedly-largest-in-history-3437607/>. [Accessed: 22-Jun-2016].
- [2] S. Khandelwal, "602 Gbps! This May Have Been the Largest DDoS Attack in History," *The Hacker News*, 08-Jan-2016. [Online]. Available: <http://thehackernews.com/2016/01/biggest-ddos-attack.html>. [Accessed: 22-Jun-2016].
- [3] J. Titcomb, "Ukrainian blackout blamed on cyber-attack," 05-Jan-2016.
- [4] Tripwire, "Tripwire 2016 Energy Survey: Oil and Gas," *Tripwire*, 2016. [Online]. Available: <http://www.tripwire.com/company/research/tripwire-2016-energy-survey-oil-and-gas/>. [Accessed: 22-Jun-2016].
- [5] Die Zeit, "NSA helps German domestic intelligence agency: XKeyscore - the document," *Die Zeit*, Hamburg, 26-Aug-2015.
- [6] Leo Kelion, "NSA and GCHQ agents 'leak Tor bugs', alleges developer," *BBC News*, 22-Aug-2014. [Online]. Available: <http://www.bbc.com/news/technology-28886462>. [Accessed: 22-May-2016].
- [7] M. G. Ionita and V. V. Patriciu, "Autoimmune Cyber Retaliation Supported by Visual Analytics," *J. Mob. Embed. Distrib. Syst.*, vol. 6, no. 3, pp. 112–121, Sep. 2014.
- [8] L.-R. ZHAO, S.-E. MEI, and W.-J. ZHONG, "AN ECONOMIC ANALYSIS OF THE INTERACTION BETWEEN FIREWALL, IDS AND VULNERABILITY SCAN," *Econ. Comput. Econ. Cybern. Stud. Res.*, no. 4, 2015.
- [9] M. Overfelt, "Prepare for hackers to change your credit score," *CNBC*, 09-Mar-2016. [Online]. Available: <http://www.cnbc.com/2016/03/09/the-next-big-threat-in-hacking--data-sabotage.html>. [Accessed: 22-May-2016].
- [10] S. J. Shackelford, *Managing Cyber Attacks in International Law, Business, and Relations: In Search of Cyber Peace*. Cambridge University Press, 2014.
- [11] Codenomicon, "Heartbleed Bug," 07-Apr-2014. [Online]. Available: <http://heartbleed.com/>. [Accessed: 22-May-2016].
- [12] Z. Durumeric, M. Payer, V. Paxson, J. Kasten, D. Adrian, J. A. Halderman, M. Bailey, F. Li, N. Weaver, J. Amann, and J. Beekman, "The Matter of Heartbleed," 2014, pp. 475–488.
- [13] E. Kovacs, "OpenSSH Vulnerability Exposes Servers to Brute Force Attacks," *SecurityWeek.Com*, 23-Jul-2015. [Online]. Available: <http://www.securityweek.com/openssh-vulnerability-exposes-servers-brute-force-attacks>. [Accessed: 22-May-2016].
- [14] M. Smith, "Researchers reveal flaws in 7-Zip, users and security vendors affected," *Network World*, 12-May-2016. [Online]. Available: <http://www.networkworld.com/article/3069937/security/researchers-reveal-flaws-in-7-zip-users-and-security-vendors-affected.html>. [Accessed: 22-May-2016].
- [15] J. Schultz, "Cisco Talos Blog: Multiple 7-Zip Vulnerabilities Discovered by Talos," 11-May-2016. .
- [16] B. Moinescu and A. Costea, "Towards an early-warning system of distressed non-banking financial institutions," *Econ. Comput. Econ. Cybern. Stud. Res.*, vol. 48, no. 2, pp. 75–90, 2014.
- [17] M. G. Ionita and V. V. Patriciu, "Cyber Incident Response Aided by Neural Networks and Visual Analytics," in *2015 20th International Conference on Control Systems and Computer Science*, 2015, pp. 229–233.
- [18] N. Hopper, E. Y. Vasserman, and E. Chan-TIN, "How Much Anonymity Does Network Latency Leak?," *ACM Trans Inf Syst Secur*, vol. 13, no. 2, p. 13:1–13:28, Mar. 2010.
- [19] S. Gallagher, "Whole lotta onions: Number of Tor hidden sites spikes—along with paranoia," *Ars Technica*, 04-Mar-2016. [Online]. Available: <http://arstechnica.com/information-technology/2016/03/whole-lotta-onions-number-of-tor-hidden-sites-spikes-along-with-paranoia/>. [Accessed: 22-May-2016].

- [20] The Tor Project, “doc/AChildsGardenOfPluggableTransports – Tor Bug Tracker & Wiki,” 2015. [Online]. Available: <https://trac.torproject.org/projects/tor/wiki/doc/AChildsGardenOfPluggableTransports>. [Accessed: 22-May-2016].
- [21] G. Kadianakis and K. Loesing, “Extrapolating network totals from hidden-service statistics,” Tor Tech Report 2015-01–001, Jan. 2015.
- [22] D. Goulet, A. Johnson, G. Kadianakis, and K. Loesing, “Hidden-service statistics reported by relays,” DTIC Document, 2015.
- [23] K. Loesing, W. Sandmann, C. Wilms, and G. Wirtz, “Performance Measurements and Statistics of Tor Hidden Services,” in *International Symposium on Applications and the Internet, 2008. SAINT 2008*, 2008, pp. 1–7.
- [24] S. Zander and S. J. Murdoch, “An Improved Clock-skew Measurement Technique for Revealing Hidden Services.,” in *USENIX Security Symposium*, 2008, pp. 211–226.
- [25] A. Biryukov, I. Pustogarov, and R. Weinmann, “Trawling for Tor Hidden Services: Detection, Measurement, Deanonymization,” 2013, pp. 80–94.
- [26] S. Englehardt, C. Eubank, P. Zimmerman, D. Reisman, and A. Narayanan, “OpenWPM: An automated platform for web privacy measurement,” Mar. 2015.
- [27] K. Mowery and H. Shacham, “Pixel perfect: Fingerprinting canvas in HTML5,” *Proc. W2SP*, 2012.
- [28] J. Cox, “The FBI’s ‘Unprecedented’ Hacking Campaign Targeted Over a Thousand Computers,” *Motherboard*, 05-Jan-2016. [Online]. Available: <http://motherboard.vice.com/read/the-fbis-unprecedented-hacking-campaign-targeted-over-a-thousand-computers>. [Accessed: 24-May-2016].
- [29] R. De Souza, “FBI Randomly Used Malware on TORMail Users While Busting Pedophiles,” *HackRead*, 24-Jan-2016. .
- [30] P. V.-V. Ionita Mihai-Gabriel, “Achieving DDoS resiliency in a software defined network by intelligent risk assessment based on neural networks and danger theory,” presented at the IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI), 2014, 2014, vol. CINTI.
- [31] M. Hyppönen, “Threat report 2015\_v1,” 21-Mar-2016.
- [32] V. Ciancaglini, M. Balduzzi, M. Goncharov, and R. McArdle, “Deepweb and Cybercrime It’s Not All About TOR,” *Trend Micro*, 2014. [Online]. Available: <https://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-deepweb-and-cybercrime.pdf>. [Accessed: 08-Jul-2016].
- [33] S. Gallagher, “Law enforcement seized Tor nodes and may have run some of its own,” *Ars Technica*, 10-Nov-2014. [Online]. Available: <http://arstechnica.com/security/2014/11/law-enforcement-seized-tor-nodes-and-may-have-run-some-of-its-own/>. [Accessed: 24-May-2016].
- [34] P. Winter, R. Köwer, M. Mulazzani, M. Huber, S. Schrittwieser, S. Lindskog, and E. Weippl, “Spoiled onions: Exposing malicious Tor exit relays,” in *International Symposium on Privacy Enhancing Technologies Symposium*, 2014, pp. 304–331.
- [35] D. L. Chaum, “Untraceable electronic mail, return addresses, and digital pseudonyms,” *Commun. ACM*, vol. 24, no. 2, pp. 84–90, 1981.
- [36] B. Muir, “TOR Packet Analysis - Locating Identifying Markers,” 25-Feb-2014.
- [37] K.-S. Wong and M. H. Kim, “Towards a respondent-preferred k i - anonymity model,” *Front. Inf. Technol. Electron. Eng.*, vol. 16, no. 9, pp. 720–731, Sep. 2015.
- [38] M.-G. Ionita, “Autoimmune Cyber Retaliation Based on Collaborative Defense Techniques,” presented at the The 7th International Conference on Security for Information Technology and Communications (SECITC 14), 2014, vol. SECITC 14, pp. 193–202.
- [39] S. Hall, “sethhall/bro-junk-drawer,” *GitHub*, 24-Apr-2015. [Online]. Available: <https://github.com/sethhall/bro-junk-drawer>. [Accessed: 22-Jun-2016].
- [40] M. G. Ionita and V. V. Patriciu, “Biologically inspired risk assessment in cyber security using neural networks,” in *2014 10th International Conference on Communications (COMM)*, 2014, pp. 1–4.
- [41] J. C. Norte, “Advanced Tor Browser Fingerprinting,” Mar-2016. [Online]. Available: <http://jcarlosnorte.com/security/2016/03/06/advanced-tor-browser-fingerprinting.html>. [Accessed: 25-May-2016].
- [42] A. Karim, R. B. Salleh, M. Shiraz, S. A. A. Shah, I. Awan, and N. B. Anuar, “Botnet detection techniques: review, future trends, and issues,” *J. Zhejiang Univ. Sci. C*, vol. 15, no. 11, pp. 943–983, Nov. 2014.
- [43] W. Wang, P. Zhang, Y. Tan, and X. He, “Animmune local concentration based virus detection approach,” *J. Zhejiang Univ. Sci. C*, vol. 12, no. 6, pp. 443–454, Jun. 2011.
- [44] A. Greenberg, “Meet The Hackers Who Sell Spies The Tools To Crack Your PC (And Get Paid Six-Figure Fees),” *Forbes*, 21-Mar-2012. [Online]. Available: <http://www.forbes.com/sites/andygreenberg/2012/03/21/meet-the-hackers-who-sell-spies-the-tools-to-crack-your-pc-and-get-paid-six-figure-fees/>. [Accessed: 22-Jun-2016].
- [45] M. M. K. Alanezi and N. B. Aldabagh, “An Immune Inspired Multilayer IDS,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 9, no. 10, p. 30, 2011.

# Reduce collisions and increase the efficiency of the RFID network system by using Manchester encoding

Fahimeh Afkhamnia

Department of Computer Eng., Faculty of Eng., **Isfahan (khorasan)** branch, Islamic Azad University, Isfahan, Iran.

Mohammad Reza Soltan Aghaei

Department of Computer Eng., Faculty of Eng., **Isfahan (khorasan)** branch, Islamic Azad University, Isfahan, Iran.

**Abstract**— RFID networks represent a system that uses radio waves to transmit information. This network plays a key role in a wide range of applications such as traffic control, transportation, military and medical use. In such networks, data collision is inevitable. The thing that made it difficult and seriously affected the desire to progress in the field of practical applications of radio networks is the problem of collision. Collision as a key problem in the RFID system, can waste energy consumption and bandwidth and leading to an increasing the time requirement for the process of tags identification.

In this article, we review some adversaries to consider anti-collision algorithm first of all, and then present a method that use Manchester encoding to reduce collision, which aims to increase the system efficiency, reducing the amount of energy consumption and collision. Finally, evaluate of the proposed algorithm in system efficiency parameters such as the number of collision. The result of the comparison shows that the performance of the proposed algorithm will reduce energy consumption and increase the system efficiency.

**Keywords-** Data collision, Radio networks, System efficiency, Slot, Manchester encoding Commas

## I. INTRODUCTION

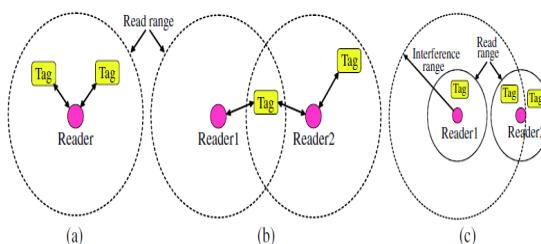
The speed of growing technology rise every day. Sometimes it is so fast that we have to stay behind and yet, we can't use technology properly that we meet new one. In the age of speed, the human needs products and technologies to help him, on doing his affairs with more efficiency and accuracy and speed. Nowadays, we deal with identifications system, data collection, smart cards, barcode, and so on, that they all are tools to help human to identify and collecting data. Identification by the radio waves that call the internet of things too, is another integration achievements for computer and telecommunications industry that can play a vital role in daily life [1].

Using this technology brings hopeful future for the power industry, But there is an objection refers to this electrical system, among the problems the system faces a collision. Collision occurs in telecommunication systems including RFID systems, while it happens in a specified region, transmitters or receivers on a channel sending and receiving operations at a specified time. This action causes the transmission is not successful [2].

RFID system has two basic hardware component with the name of the sender and the review. These two components are

called in order of tag and tag reader. The primary task of RFID tag is saving data and sending it to a tag reader (review). Tag is including an electronic microchip and antenna that this antenna transmit information task between tag and tag reader by using radio waves. Tag readers are elements of communication between tags and systems management. Tag reader send the radio waves at a specified frequency that these waves received by tag and tag reader identify them [7].

RFID systems use radio transmissions to send energy to a RFID tag, while the tag emits a unique identification code (ID) back to the RFID reader. If multiple tags are to be identified simultaneously, collision would be occurred. Hence, anti-collision protocols need to be devised between the tags and the reader to minimize collisions. The RFID collision problems could be summarized and classified into: tag-to-tag collision (or tag collision), reader-to-tag collision, and reader-to reader collision, as shown in Fig.1. Both reader-to-tag collision and reader-to reader collision are called reader collision in general. The tag-to-tag collision, as shown in Fig.1 (a), occurs when a plurality of the tags responds to one reader's inquiry simultaneously and therefore the reader cannot identify any tag. Reader-to-tag collision, as shown in Fig.1(b), occurs when one tag is simultaneously located in the fields of two or more readers and more than one reader attempts to communicate with that tag at the same time. Reader-to-reader collision, as shown in Fig.1(c), occurs when a reader transmits a signal that interferes with the operation of another reader, thus preventing the second reader from communicating with tags in its interrogation zone. In other words, the reader collision indicates that a plurality of readers' requests inquiries to one tag concurrently, so it is confusing for the tag to identify the inquiries [8]. In this article, we focus on tags collision.



**Figure 1. Types of collision on the network**

It is clear that in order to prevent the occurrence of data collisions, we have to use specific algorithms. The purpose of the implementation of anti-collision algorithm will ensure the ability of tag reader due to communicating with several tag. In this article we will introduce an algorithm that uses Manchester encoding to reduce collisions. Manchester encoding is one of the most effective techniques for collision detection bits in the network. Diagnosis the real location of collision bit lead to reduce the collisions rate and unnecessary idle slots [4]. But this coding alone cannot lead to increase the system efficiency. As regards, the number of tags is unknown before the identification process. Adding a movable situation, for nodes due to highlighting initial frame of length is necessary.

## II. RELATED WORKS

In conventional protocols for passive RFID tag anti-collision, ALOHA-based protocols are very popular. The idea of ALOHA-based protocols is to divide access time of tags into a number of slots, and each tag responds at a random slot. If tags collide in a slot, which means that at least two tags responses in the slot. We hereby introduce the algorithm related to signal collision in a RFID system. There are many scholars who offer many solutions for RFID systems, the three that have been most widely discussed are the ATSA Algorithm, the Tree Slotted Aloha Algorithm (TSA), and the Dynamic Assigned Tree Slotted Aloha Algorithm (DyATSA). Each of these features their own unique properties and although there are many improved algorithms successively that have been offered, we will now focus on these three as a starting point.

### A. The ATSA Algorithm

ATSA algorithm in RFID network is one of the anti-collision algorithm .The essential idea behind our protocols is to assign a unique ID prefix to each frame and slot. In a frame, tags first match their ID with the frame prefix. When matched, tags then match their tpre with the slot prefix and reply in the matching slot. During the frame, the reader records the prefixes of every collision slots. At the end of the frame, the reader uses the Vogt estimation algorithm to estimate the number of colliding tags in each collision slot. According to the prefixes of collision slots and the estimated tag number, the reader assigns the frame prefixes to the subsequent frames and splits colliding tags into disjoint subgroups. Then, the reader recursively identifies colliding tags. During the identification process of ATSA, the reader keeps a query queue Q to record pre and F of each frame. At the beginning of a frame, the reader obtains pre and F from queue Q, and then broadcasts a Query (pre, F) command. In a frame, the reader uses a QueryRep (spre) command to start each slot except the first slot. On the tag side, a tag first turns its state into inactive, and then executes the following two matching processes:

Frame prefix matching: After receiving the Query (pre, F) command, a tag first checks whether its ID contains the same prefix with pre. If matched, it turns into the active state, and

then obtains the tag's slot prefix tpre by recording its first P-bit ID next to pre. If tpre is an all-zero string, the tag replies to the reader immediately. Otherwise, the tag waits for reader's slot starting command QueryRep (spre) and goes to the following slot prefix matching process.

Slot prefix matching: After receiving the Queryrep (spre) command, an active tag compares its tpre with spre. If tpre=spre, the tag replies in the current slot [3, 4].

The performance of the ATSA protocol relies on the accuracy of the tag estimation function. Specifically, we employ the well-known Vogt algorithm to estimate the number of colliding tags in a frame as it can achieve accurate estimation with easy implementation. Vogt utilized Chebyshev's inequality to obtain the estimation function and determines the optimal  $n$  by minimizing the distance  $\varepsilon_{vd}$  between the real values and the expected numbers of three kinds of slots. In mathematical terms,

$$\varepsilon_{vd(N,c_0,c_1,c_k)} = \min \left( \begin{pmatrix} a_0^{N,n} \\ a_1^{N,n} \\ a_{\geq 2}^{N,n} \end{pmatrix} - \begin{pmatrix} c_0 \\ c_1 \\ c_k \end{pmatrix} \right) \quad (1)$$

The estimated number of tags with the number of slots (N), used in each phase and the results of the previous step states as ( $c_0$ ,  $c_1$ ,  $c_k$ ) that represent the number of slots-empty (idle) the number of slots with a label (successful slot), and slots with more than one label (slot collision).

The top relation  $(a_0^{N,n}, a_1^{N,n}, a_{\geq 2}^{N,n})$  define the empty slots, slots filled with a label and the collision slots where  $N$  is the number of slots and  $n$  specifies the number of tags. With a value of  $n$ ,  $N$ , the number of 0,1,r for labels in a slot distributed as polynomial and the expected value for them is expressed in the following equation:

$$a_r^{n,N} = (X=r) = N \binom{n}{r} \left( \frac{1}{N} \right)^r \left( 1 - \frac{1}{N} \right)^{n-r} \quad (2)$$

Given that the tag number is not specified And ATSA randomly assigned the length of the frame, the process of identifying face a problem and lead to increase unnecessary idle slots.

### B. The TSA Algorithm

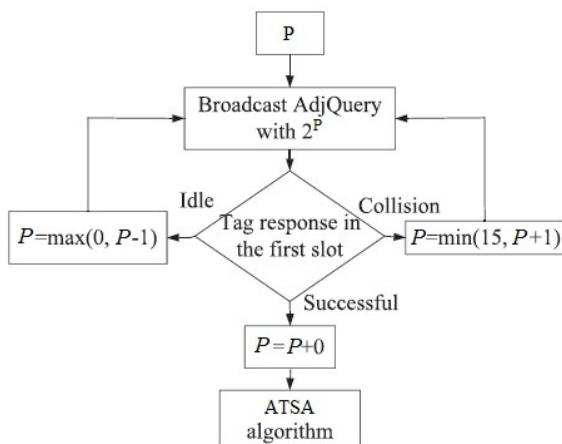
The basic idea of our TSA identification protocol, is to solve a collision as soon as it happens. In Framed Slotted Aloha protocols, two tags not colliding in a frame, can collide in the next frame. In our approach, the above situation is avoided, since when a collision occurs in a slot, only the tags generating such collision are queried in the next read cycle. The protocol is performed in several tag reading cycles. A reading cycle consists of two steps:

In the first step, the reader broadcasts a request for data by specifying the frame size [ $L_i$ ], in the second step each tag in the communication range of the reader, selects its response slot by generating a random number in the range  $[1, \dots, L_i]$  and

transmits its ID in such a slot. The reader identifies a tag when it receives the tag ID without collisions the behavior of the protocol follows a tree structure. The root node is the frame in the first reading cycle at the end of each reading cycle, if the reader realizes that collisions occurred, it starts a new reading cycle for each slot where there was a collision. The problem with this algorithm is that the labels need to know the size and frame synchronization circuit is [10].

### C. The DyATSA Algorithm

DyATSA algorithm define the length of the frame dynamically for the process of recognition was introduced. With this algorithm, the tag reader, select reasonable value and account for the frame. Fig. (2) Shows a flowchart of dynamic ATSA protocol.



**Figure 2** Flowchart defines the frame length dynamically

Initially, a frame length  $F=2^P$ , in fact,  $P$  as data logging, the value of 2, and if desired can be determined and be sent to the label. Then, a reader broadcasts an AdjQuery command with  $L$ . After a tag receives the command with  $L$ , the tag's Counter selects a random integer from 0 to  $L-1$ .

Tags whose Counters select 0 can transmit their IDs, and the reader will detect the tags responses in the first slot. If the first slot is collisional,  $Q=Q+1$ , and the reader will broadcast a command with a new  $L$  and then detect tags responses in the first slot of next frame. If the first slot is idle,  $Q=Q-1$ , and the reader will also broadcast a new  $L$  and then detect tags responses. If the first slot is successful,  $Q$  will not be changed and the reader's operation will transit to ATSA algorithm [5]. One disadvantage of this method is, when the number of tags increase more than 2,500 the dynamic will lose performance and a great number of slot collision with a large slope will increase.

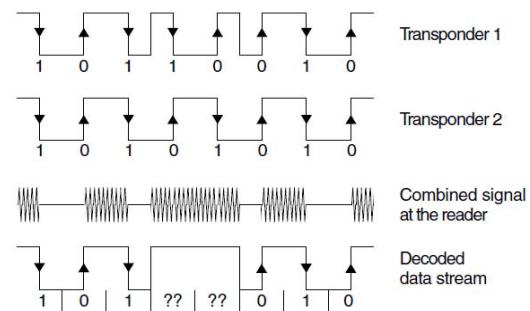
### III. METHOD

This paper presents an anti-collision algorithm to enhance the system efficiency. This algorithm is an acronym for Dynamic Improved Assigned Tree\_Slotted Aloha (DyImATSA). In all algorithms except DATA frame length is always constant and does not change much during the identification process As a

result, when the number of tags is low slot in each frame is wasted and when the number of tags will have much impact. So performance is highly dependent algorithms are defined and influenced by the length of the frame. DyATSA algorithm Dependence on performance to the frame length using a dynamically defined time frame using an estimate of the number of tags improve. The definition frame must be considered dynamically alone can't reduce collision of tags in a lot of cases, so need to use the method to accelerate the process of identifying and reducing our collision, the Manchester encoding used for this purpose.

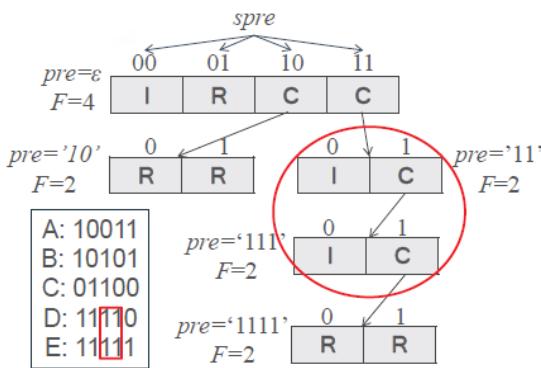
The purpose of the proposed algorithm DyImATSA is that, in addition to reducing the number of collision accidents occurring in detection process and reduce the number of unnecessary idle Slots, improve system efficiency. Manchester coding is used to determine the actual location collision bits. The use of this code is that the recipient must be able to properly start, end or middle of each bit without the need for an external clock signal is detected.

In Manchester encoding, the value of a bit is defined by the change in level within a bit window. A logic 0 or 1 is coded by a positive and negative transition, respectively. If two (or more) transponders simultaneously transmit bits of different values then the positive and negative transitions of the received bits cancel each other out, resulting in an error (see Fig.3) .By means of this encoding scheme, the reader can effectively find where the colliding bit is. When the tag density is very low, the reader cannot find the actual colliding bit in ATSA. In this case, the reader needs to experience many collision and idle slots to identify the location of the actual colliding bit. Manchester encoding is used to locate the actual colliding bit. Thus, the reader can reduce the unnecessary collision and idle slots especially in sparse tag environments.



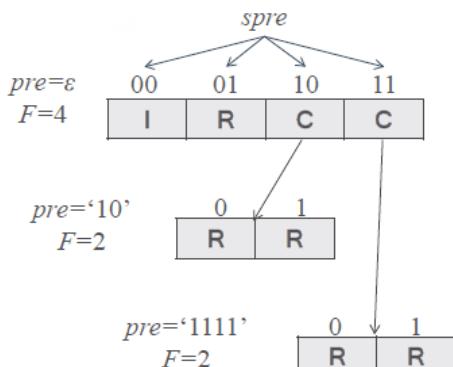
**Figure 3** Example of Manchester encoding

Herein, we give an example to demonstrate the effectiveness demonstrate the effectiveness of the ImATSA protocol. Suppose that there are 5 tags to be identified, which are A (10011), B (10101), C (01100), D (11110) and E (11111). The identification process using ATSA can be illustrated by a tree structure in Fig (4). In the third frame, tags D and E check the third bit of their ID to match the slot prefix. But their third bits are the same, causing an idle slot and a collision slot. Since the case of the fourth frame is the same as that of the third frame, tags D and E are not identified until the fifth frame.



**Figure 4. The process of identifying label without using Manchester encoding**

If Manchester encoding is employed in DyImATSA, the first colliding bit of tags D and E can be detected in the first frame so that we can set the frame prefix of the third frame to be 1111\_. As a result, tag D and E will be identified in the third frame, as shown in Fig. (5). In other words, the performance is improved by saving the idle and collision slots in the third and fourth frames. This situation happens more frequently, especially when  $n$  is very small, i.e., in sparse environments.



**Figure 5. The process of identification labels from Manchester**

#### IV. SIMULATION

##### A. Simulation Parameters

The simulation software which is used in this research is the NS2 software. The wireless network is the type of network used. Of wave propagation two ray ground (a sort of propagation delay in the network) for passing the consumer at 802.11 mac layer protocols and routing algorithms for DSDV (a routing protocol on NS2) is used. Table 1 shows the simulation parameters.

**Table 1. Simulation parameters**

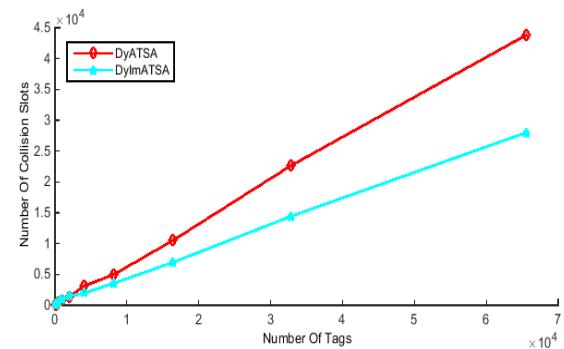
Row	Parameter	Amount
1	Area Simulation	782 x 545 m <sup>2</sup>
2	Domain Transfer	100 m <sup>2</sup>
3	Primary energy	100 J

4	Mac protocol	Mac 802.11
5	Bandwidth	250 kbps
6	Receipt energy	0.76
7	Connection Energy	0.28
8	The number of Reader	1
9	The number of nodes	1...150

##### B. Simulation Results

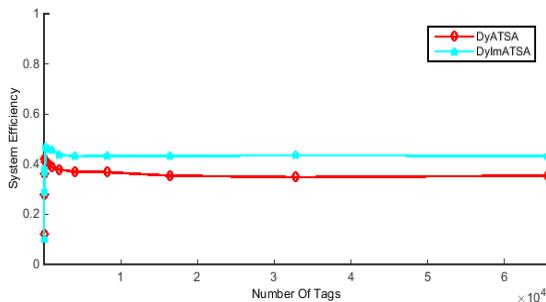
In this paper, the proposed algorithm with ATSA, TSA and DyATSA anti-collision algorithms and system efficiency in terms of the number of collision occurred, were compared. Although the total length of the tag ID in EPC Global C1G2 is 96 bits, in many identification applications the header, domain manager, sometimes even the object class of tags, are the same. Therefore, for simplicity we only focus on specific tag ID length of 96 bits. But according to the explanations given before the start of the evaluation algorithms in a moving address, the proposed algorithm with DyATSA in fixed network nodes examined, we put the goal of this study is to show the use of dynamic mode improve system efficiency and reduce alone cannot lead to a collision.

In Fig. 6, the proposed method and DyImATSA and DyATSA collision algorithm in terms of the number of slots in the fixed network nodes are compared with each other. As you can see, when the number of tags to be more than  $2^{14}$  DyATSA method loses its performance and the collision with the steep rise. In fact, growth in the number of collision slots DyATSA method proposed by DyImATSA faster.



**Figure 6.The number of collision slots in the fixed network nodes**

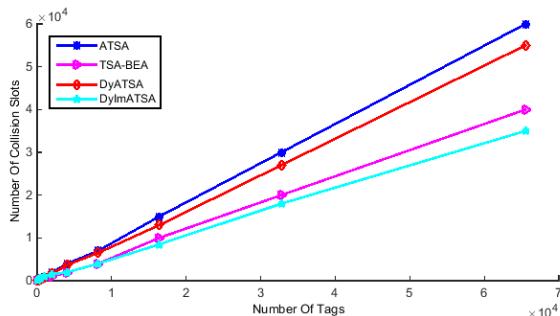
The main objective of the various protocols are provided, achieve higher efficiency in the identification tag is an RFID network. In Fig. 7 you can see the system's efficiency in our fixed network node. System efficiency is inversely proportional to the total number of slots collision and idle. As you can see in the figure proposed by the Manchester encoding method and delete unnecessary slots, more efficient than the method DyATSA.



**Figure 7. System efficiency in fixed network nodes**

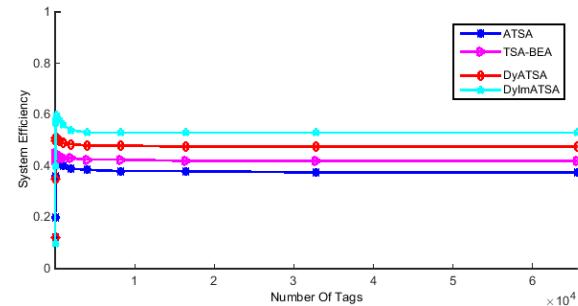
So far in Fig. 6 and Fig. 7 in the fixed network nodes examined. We're going to the number of slots collision and system efficiency for network system with mobile nodes have the tags movement examine.

Fig.8 number of slot collision on the network with mobile nodes and the 96-bit ID length shows. As is evident from the graph, the proposed algorithm has better performance and substantially more than the other three algorithms. Note that DyATSA algorithm that uses dynamic mode only than ATSA algorithm is less than the number of slots collision detection label, and the number of tags increases, it becomes more visible.



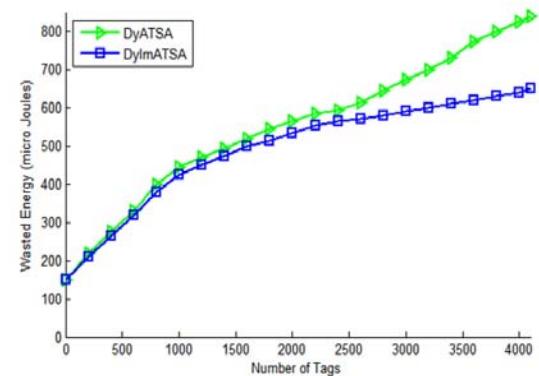
**Figure 8. The number of slots collision in the proposed method with mobile nodes in the network**

The system efficiency of an RFID network in the ratio of the total number of slots has slots (slots successful+ slots collision + slots idle). Fig.9 is system efficiency graph indicates on the network with mobile nodes. The graph shows that dynamic and have a greater impact on the use of Manchester encoding system efficiency in the detection label's RFID network and efficiency the proposed method is about 60%.

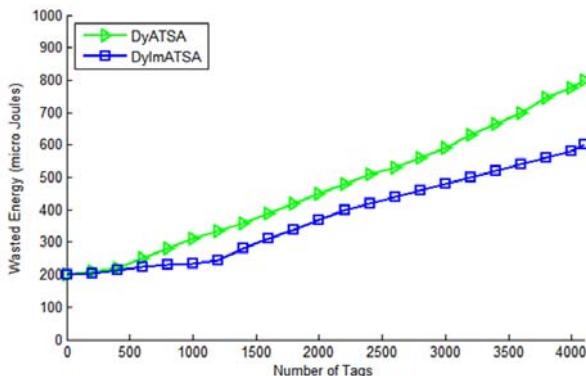


**Figure 9. System efficiency in the proposed method with mobile nodes in the network**

The purpose of this research is to reduce energy consumption in the process of detection tags. As proposed that the minimum number of slots collision system more efficient and consequently have the lowest energy consumption. We're going to implement the method and the proposed method DyImATSA and DyATSA in both stationary and mobile network nodes, the energy figures compare with each other. The algorithm results in both graph and this difference was DyImATSA algorithm is better DyATSA algorithm applied in determining the use of Manchester encoding is the label that less energy is wasted in the process. The difference in the number of tags higher than the number of tags the lower stems. This means that the number of tags RFID network growth rate of the proposed method is more energy DyATSA method. Fig.10 and Fig.11 show energy consumption diagrams.



**Figure 10. The chart in energy consumption in the proposed method with mobile nodes in the network**



**Figure 11.**The chart in energy consumption in the proposed method with fixed nodes in the network

#### IV. CONCLUSION

The main objective of this study was to increase system efficiency and reduce energy consumption in the process of detection of tags in the RFID network. The system is more effective means labels and the lowest number of collision has identified more successfully. So in this article we use the dynamic mode and Manchester encoding, Idle essential slots and reduce the number of collision which have led to increase system efficiency and reduce the energy consumption nodes. The results show that the proposed DyImATSA algorithm to increase system efficiency and reduces energy consumption has been more successful than other methods. The only parameter that has a negative impact, the overhead parameter is the value of this parameter in the proposed method than other methods increased 3.2%.

Something that could be used for future efforts to use energy harvesting methods that can be applied to the heat of the human body. But this method requires an increase in the cost of producing tags and reprogramming. The method can be used for filtering data by selecting the appropriate cluster collision can be eliminated, leading to savings in energy consumption.

#### REFERENCES

- [1] C. K.-W. R. Klair Dheeraj K, “A survey and tutorial of RFID anti-collision protocols”, *Communications Surveys and Tutorials, IEEE*, vol. 12, no. 3,2012, pp. 400-421.
- [2] H. He, Q. Li, and Z. H. Zhang, “RFID Security Authentication Protocol Based on Hash for the Lightweight RFID Systems”, *Applied Mechanics and Materials*, vol. 380,2011, pp. 2831-2836.
- [3] Chiu-Kuo Liang; Hsin-Mo Lin, “Using Dynamic Slots Collision Tracking Tree Technique Towards an Efficient Tag Anti-collision Algorithm in RFID Systems”, *Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing (UIC/ATC), 9th International Conference on 2012*, vol., no., Sept 4-7, 2012, pp.272,277.
- [4] Zhang, Lijuan; Zhang, Jin; Tang, Xiao , “Assigned Tree Slotted Aloha RFID Tag Anti-Collision Protocols ”,Wireless Communications, IEEE Transactions on , vol.12, no.11, November 2013, pp.5493,5505.
- [5] Hai Feng Wu; Yu Zeng; Jihua Feng; Yu GU, “Binary Tree Slotted ALOHA for Passive RFID Tag Anti-collision”,*Parallel and Distributed Systems, IEEE Transactions on* , vol.24,no.1,June2013,pp.19,31.
- [6] Ullah, S., Alsalihi, W., Alsehaim, A., and Alsadhan, N. “A review of tags anti-collision and localization protocols in RFID networks”, *Journal of medical systems*, vol 20, 2012, pp. 4037-4050.
- [7] QIAN, Z., and WANG, X. “An Overview of Anti-Collision Protocols for Radio Frequency IdentificationDevices”,*ChinaCommunications*,vol.12, 2014, pp.256,262.
- [8] Shih, T.-F., and Teng, C.-H. “The Design of a Stride Query Tree Algorithm for RFID Systems”.Paper presented at the ComputationalIntelligence, Communication Systems and Networks (CICSyN), 2011 Third International Conference on.
- [9] Jia, X., Feng, Q., and Yu, L. “Stability analysis of an efficient anti-collision protocol for RFID tag identification”, *Communications, IEEE Transactions on*, 60(8),2013,pp. 2285-2294.
- [10] Rennane, A., Saadi, H., Touhami, R., and Yagoub, M. C. “A comparative performance evaluation study of the basic binary tree and aloha based anti-collision protocols for passive RFID system”, Paper presented at the Microelectronics(ICM),2012 24thInternational Conference on.

# Numerical Solution of Nonlinear Optimality Problem by PSO&GA

H. Hossein zadeh and E.Salehpour\*

**Abstract**— The present research aims to introduce a combined method for solving optimization problems namely PSO-GA. In this algorithm, particle swarm optimization (PSO) is operated in order to improve vector while genetic algorithm is used in order to improve decision vectors through genetic algorithms. A balance between exploration capabilities and exploitation is improved in PSO algorithm through genetic operators namely cut and swap. Defined limitations are used in the problem through penalty function without parameter. Empirical results of optimization problems are compared to different kinds of methods in the published paper. The obtained solution compared with better suggested method of the solution existing in this paper and published texts. Moreover, empirical results show that the proposed method is the best solution for engineering problems.

**Keywords**—Particle swarm optimization, Genetic algorithm, Constraint optimization, PSO -GA.

## I. INTRODUCTION

USEFUL optimization is an important part of every problem in engineering and industry. Most of the optimization of real world has different kinds of limitations improving shape of search space. During the past decades, a broad range of meta-heuristic algorithm is designed and used in order to solve useful optimization problems. In a way that because of different kinds of constraint (equality or inequality) and their mutual relation between target functions, useful optimization problems are harder than unconstrained optimization problems. These kinds of problems are generally a mixture of continues and non-continues, non linear target functions and non linear constraints. On the purpose of optimizing these kinds of problems, two come on methods namely mathematic programming and Meta heuristic methods are used different mathematic programming methods are: linear programming, harmonious linear programming, integer programming, dynamic programming, non linear programming that are used for these kinds of problems. These methods used gradient information in order to search a solution space with the primary start point, generally, gradient based methods are become harmonious rapidly and can obtain solution more carful than random methods during its duty of local search. While on the purpose of performing effectively these methods, variables and cost function of these generators

should be continues. Moreover, choosing successfully a good start point for these methods is essential. In most of the optimization problems, forbidden regions, border constraints, rough cost function should be studied. As a result, these rough optimization problems cannot be solved by old method of mathematic programming. Although dynamic programming or mixture of non linear and integer programming choices proposed in solving rough problems but they have high computational cost.

As a replacement for common mathematic methods, innovative optimization techniques are used for general optimization solutions or semi general solution. Regarding their exploration capabilities and finding optimistic regions in search area in a correct time, these methods are really appropriate for general search. Furthermore, in these methods, there is not any need for the continuous cost functions and variables are used for mathematic optimization. Although these methods are approximate, for example, their solution is good but essentially it is not optimized, they don't have derivation of target function or limitations and used changeably probable transition laws as real one. So, researchers focused on exploration capabilities as the best solutions in reasonable time. Although, traditional algorithms are derived for all non linear limitation functions in order to evaluate systematic usability. But in real problems, because of high computational complexity of system, conclusion is really difficult. So the purposes of overcoming this problem, algorithms which areinspired through nature, are the last feature of aesthetic algorithms. These algorithms work well with optimization problems so other problems of classic methods like classic methods in the nature are not flexible. Through most of the researchers, it is determined that genetic algorithms, particle swarm optimization, harmonic search (HS) and evolutionary strategies (ES) are attractive because they don't influence mathematic assumption regarding optimization problems and have better general search capabilities than common optimization algorithm. these meta innovative techniques are used in different kinds of useful optimization problems, For example, engineering problems, structural design problems, validity optimization, decision. For example, GA is mixed with other developmental algorithms in order to reach high productivity from the computational point.

Coel, Dimopoulos used genetic algorithm for solving optimization problem of mixed integer engineering design. Coello and Montespresented a dominance based program in order to integrate limitations to genetic algorithm fitness function using for general optimizations. Tsai presented a new method for solving nonlinear fractional programming in management and engineering design. Montes and et.al

\*Corresponding Author: ElhamSalehpour, Department of Mathematics, University of Mazandaran, Babolsar, Iran. (E-mail: e.salehpour@stu.umz.ac.ir)

presented an improved version of differential evolutionary algorithm for solving design problems in which a criteria is used on the basis of feasibility and mechanism difference for maintaining impracticable solution Omran and Salman introduced a meta innovative algorithm without new parameter namely CODEQ that is a mixture of concepts of chaotic search, contrast based learning, differential evaluation and quantum mechanics. He and et al, He and Wang, Shi and Eberhart used particle swarm optimization for solving mixed integer design optimization. Moreover, Coelho presented quantum behavior method through swap operator with Gaussian distribution while Cagnina and et al introduced simple constraint particle swarm optimization for solving engineering optimization. Geem, Lee, Fesanghary and et al used Harmonic search algorithm for this kind of problems. Kaveh and Talatahari extended a mixed algorithm which is based on particle swarm optimization with inactive population, Ant colony algorithm (ACO), harmonic search approach in order to solve optimization problems. Gandomi and et al used firefly algorithm in order to solve continues/non-continues structural problem. Mehta and Dasgupta presented a method for bad optimization problems in which limitation used with simple search method of Mead and Nelder, Kaveh and Talatahari presented an improved ant colony optimization in order to solve engineering constraint problems including continues and non continuous range. Hedar and Fukushima presented simulated annealing method in order to solve optimization constraint problems. In their approach, filter-set-based approach presented in their suggested approach and finally an intensity plan is performed as a final step in order to overcome continuity of simulated annealing methods. Garh solve structural engineering design through penalty strategy of bee colony algorithm.

As it was shown in the published papers, the above mentioned optimization techniques are used in order to solve constraint optimization problems. While method of obtaining optimization solution or near optimization has significant difference. So, last developed algorithms such as random optimization techniques, innovative algorithm have weakness like low coherence, premature convergence, not using old knowledge, not exploiting local search information and problem in dealing with large scale optimization problems. On the purpose of overcoming this problem, meta-innovative/developmental algorithms are choose and used successfully for optimization problems. Most of the innovative and meta-innovative algorithms are obtained by the behavior of biologic systems and physical systems in the nature. Two principle roles for each of the meta-innovative algorithm are searching the best solution and then choosing the best candidate then making sure that algorithm can explore effectively search space than producing random number. On the basis of these important roles, meta innovative algorithms are superior than others in order to solve special optimization problems. Modern meta-innovative algorithms are developed with the aim of general search with three principal goals: rapid solving of problems, solving large problems and obtaining algorithms. GA and PSO are different examples of these algorithms and used successfully in order to solve optimization problems of different engineering design while each of the algorithms has weak and strong point. In GA, if thing dose not choose, then

the information of that thing will be omitted but PSO has memory. Although without choose operator, PSO may loosed resources. While GA can find a right solution and are good in reaching a good region, it will increase the interaction of PSO for optimized solution. So keeping advantages of each algorithms, the present research aims to develop a useful combined method namely PSO-GA in order to find optimized solution of non linearconstraint optimization problems.

So in this study, usage of developmental algorithms for multimodal and mixed-variable optimization problems is studied. So a mixed technique namely PSO-GA is presented by considering advantages of each 2 GA and PSO algorithms in order to solve non linear optimization problems. In this method, PSO operators are used in order to improve vector while GA is used in order to improve decision vectors through genetic algorithms. Suggested algorithm is tested with engineering and mathematic optimized problems. Numerical results show that a suggested algorithm is the powerful search algorithm for different optimization problems. The remained contents of this paper areorganized like the following. Part 2 presented the indexes using in this paper. Part 3 is the whole formula of optimization method and problem in order to manage limitations and explained PSO-GA combined method and also a summary of GA and PSO algorithms. Computational results of structural design problem is considered and compared with the existed methods in part 4.

## II. Non Linear Constrained Optimization Problem

The general optimization problem has the form:

$$\begin{aligned} \min f(x), \quad & x \in R^n \\ \text{s.t.} \quad & g_i(x) = 0 \quad i = 1, \dots, m_e \\ & g_i(x) \leq 0 \quad i = m_e + 1, \dots, m \\ & x_l \leq x \leq x_u \end{aligned}$$

In particular, if  $m=0$ , the problem is called an unconstrained optimization problem. In this course we intend to introduce and investigate algorithms for solving this problem. In constrained optimization problem, finding a solution regarding the presence of each kind of limitations in the framework of equality and inequality is not simple. On the purpose of managing these constraints, different methods are presented. The most common method in developmental algorithm population is using fine functions. While using a penalty function, a value violating limitations for penalty is not applicable in a way that applicable solution is used through choosing process. Regarding the familiarity of penalty function, they have some weakness that its principle one is that it has many parameters for arranging and finding a correct mixture may not be simple. Also during it, search is very slow and there is not any assurance that optimization is obtained. On the purpose of overcoming these constraints, Deb improved these algorithms through using free parameter concept of penalty function. For example, an attempt in order to solve a non constraint problem in search space through improved target function such as:

$$F(x) = \begin{cases} f(x) & x \in S, \\ f_w + \sum_{j=1}^M g_j & x \notin S. \end{cases}$$

$X$  is the set of obtained solutions and  $f_w$  is the worst solution in the population.

### III. COMBINED METHOD PSO-GA

In this part, genetic algorithm and new particle swarm optimization are introduced in order to solve constraint optimization problems.

#### A. Particle swarm optimization algorithm

Particle swarm optimization algorithm or PSO, also known as the swarm, one of the most powerful and popular algorithms for optimization. Mostly because of the relatively high rate of convergence is to be used. These algorithms are a little old, but has managed in many application areas, older algorithms, such as genetic algorithms, surpass and to be considered as a first choice.

PSO is one of the optimization methods inspired with nature, it has been developed for solving numerical optimization with very large search space without having to inform the gradient of the objective function. This was invented the first time in 1995 by two people named Kennedy and Eberhart, at the beginning was used to simulate the mass flight of birds, but was observed after the initial simple algorithm that it is doing actually a type of optimization algorithm and for this reasons, it can also be used to solve other optimization problems.

The algorithm is inspired by the lives of a group of animals, including insects (such as ants, bees, etc.), birds and fish has been invented. To solve an optimization problem, a population of candidate solutions move using a simple formula accidentally the problem of domain, and it explores with aims for the global optimal answers. PSO algorithm referred each candidate's answers as a particle, and every bit in flight as corresponding to example of the birds in a flock. PSO algorithm is similar to genetic algorithms.

From this as in point of view a population of solutions randomly generate with the algorithm are working within the domain the problem to answer. However, unlike the PSO algorithm genetic algorithm to optimize the problem potential of each answer is assigned a random speed, so that is displaced at each iteration according to the velocity of the particles in the problem atmosphere. Also, unlike genetic algorithms PSO algorithm have obtained the best solution for the optimization problem (from the beginning of the program until the last repetition) by each particle is also storage. Like the genetic algorithm is inherently PSO algorithm for solving unconstrained maximization is in continuous mode. However, you can do changes in the way defined objective function for solving optimization problems (such as minimizing or maximizing) the state under the constraint (always) be used. PSO not need algorithm to optimize functions any combination of practical information and only uses basic mathematical operators, in order to adjust it to the minimum parameters required. The algorithm performance with growth

of research space will also not be lost. PSO method is one of the new species evolutionary methods the potential has been proved for use in optimization problems with continuous functions. In this way, move toward the optimal point, based on two data sets is done. One of the best-point of information obtained from each of the initial population.

PSO algorithm can be explained:

First, in the search space, a number of points are selected as the initial population. Points in different categories are based on Euclidean distance. For example category i includes three factors is search, function value for each of the factors included in the search space is calculated and in each category is determined whether the value of the minimum or maximum, depending on the intended purpose has.

In this way, the point is determined in each category. On the other hand the availability of information in the past, each factor may be the best spot that has been discovered by which to identify. The optimal point information of each category and each agent specified. The first knowledge is corresponding to the global optimal point in each group and local knowledge second is corresponding to the optimal point.

With this information, the motion vector is given to each factor. This method is not affected much by the size and nonlinear problem and good results in static environments, noise and environment is continuously changing. Simplicity of implementation, lack of commitment to the continuity of the objective function and the ability to adapt to a dynamic environment makes the algorithm used in many different areas. Accordingly, it can be concluded that the targeted nature of the behavior of particles in PSO method based on two principles.

These two principles are:

- i. Individual knowledge: The individual moves to the best of their knowledge gained new knowledge.
- ii. Social science: The person in terms of his relationship with the community uses the best information for continuing the movement.

In Article ii, Individual relationship with the community is important to the topology structure of the community, this feeds defined for different topologies by society. This algorithm has advantages and disadvantages which are mentioned below.

Advantages algorithm include the following:

1. This algorithm is compared to other less regulated parameters optimized algorithms.
2. Implementation is easy and simple concepts.
3. For various issues, effective and enforceable.
4. Also used for discrete states and the continuum concepts.
5. The algorithm's performance will not disappear with the growth of research space.
6. The above algorithm optimization function application does not require any combination of information and only uses basic math operators.
7. Does not require heavy mathematical operations such as gradients.
8. Population-based approach.
9. The partnership uses particles.

According to mathematic law, particle swarm optimization is valued randomly in search space and movement took place through D space in order to search for new solutions. Assume that  $x_k^i$  and  $v_k^i$  are respectively situation and particle speed  $i$  in search space in k repetition then speed and situation of these particles will be updated in repetition (k+1) through the following equations:

$$v_{k+1}^i = \underbrace{w \cdot v_k^i}_{\text{inertia}} + \underbrace{c_1 \cdot r_1 \cdot (p_k^i - x_k^i)}_{\text{local best}} + \underbrace{c_2 \cdot r_2 \cdot (p_k^g - x_k^i)}_{\text{global best}}$$

$$x_{k+1}^i = x_k^i + v_{k+1}^i$$

In which  $r_1$  and  $r_2$  are the random number between 0 and 1 and  $c_1$  and  $c_2$  are fixed,  $p_k^i$  shows the best situation of i particle and  $p_k^g$  relates to the best situation in the swap to k repetition. One of the principle steps of particle swap optimization can be summarized as a pseudo code in algorithm 1.

**Algorithm 1** Pseudo code of Particle swarm optimization (PSO).

```

1: Objective function:  $f(x)$ ,  $x = (x_1, x_2, \dots, x_D)$ ;
2: Initialize particle position and velocity for each particle and set k=1.
3: Initialize the particle's best known position to its initial position i.e.
    $P_k^i = X_k^i$ .
4: do
5: Update the best known position ( $P_k^i$ ) of each particle and swarm's best
   known position ( $P_k^g$ ).
6: Calculate particle velocity according to the velocity equation.
7: Update particle position according to the position equation.
8: While maximum iterations or minimum error criteria is not attained

```

### B. Genetic Algorithms

Genetic Algorithms (abbreviated symbol GA) search techniques in computer science to find approximate solutions to optimization and search problems. GA is a special kind of evolutionary algorithms evolutionary biology such as used inheritance and mutation techniques. The algorithm was first introduced by John Holland. In fact, Darwin's natural of selection principles of genetic. Genetic algorithms are often a good option for forecasting techniques based on regression. Algorithms to find the optimal formula for predicting or use pattern matching. In artificial intelligence, genetic algorithm (or GA) is a programming technique that makes use of genetic evolution as a problem-solving model. The problem must be solved, which has inputs modeled by a process of genetic evolution becomes a solution then solutions assessed as candidates by the evaluation function (Fitness Function) and if the exit condition is provided , the algorithm ends. Overall, iterative genetic algorithm is an algorithm that much of it will be selected for random processes. These algorithms consist of the following sections: Fitness function, Display, Select, change.

Genetic algorithm (GA) is a programming technique that makes use of genetic evolution as a problem-solving model. Entry is a problem that must be solved and the solutions are based on an encoding scheme that is the fitness function and each solution evaluates candidates, most of who were randomly selected. Genetic Algorithm is a search technique in computer science to find the optimal solution and search issues. Genetic algorithms are evolutionary algorithms that one of the types of biology such inspired as inheritance, mutation, selection of a sudden (Biology), natural selection

and composition. Generally solutions for 0 and 1 are shown, evolution of a totally random collection of entities begins and is repeated in the next generation. In each generation, the best choice is not the best. One solution to the problem, with a list of parameters is shown that they say chromosome or genome. Chromosomes are displayed generally in the form of a simple string of data of course a variety of other data structures can also be used. First, several features randomly generated to create the first generation. During each generation, each character is evaluated and the value of fitness is measured by a fitness function. The next step is to create the second generation of community that based selection processes, production from the properties selected with genetic operators: chromosomes connect to each other. For each person, a parent pair is selected. There is several patterns choice: roulette wheel, tournament and etc.

Genetic algorithms are termination condition:

A fixed number of generations come to;

In all budgets (computation time / money);

A person (child produced) found that the minimum (lowest) to meet the criteria;

Most of fit offspring or the other does not yield better results.

Manual inspection;

High compounds.

A fitness function is used in order to evaluate people and success of reproduction is different from fitness. Pseudo code GA algorithm is introduced in algorithm 2.

**Algorithm 2** Pseudo code of Genetic algorithm (GA).

```

1: Objective function:  $f(x)$ 
2: Define Fitness F (eg.  $F \propto f(x)$  for maximization)
3: Initialize population
4: Initial probabilities of crossover ( $p_c$ ) and mutation( $p_m$ )
5: do
6: Generate new solution by crossover and mutation
7: if  $p_c > \text{rand}$ , Crossover: end if
8: if  $p_m > \text{rand}$ , Mutate: end if
9: Accept the new solution if its fitness increases.
10: Select the current best for the next generation.
11: While maximum iterations or minimum error criteria is not attained

```

### C. Combined method PSO-GA

Motivation of developing PSO-GA method is mixing genetic algorithm and optimization of particle swap optimization. Through using genetic operators in standard PSO, a balance between exploration and exploitation capability will be improved, while each of them has weak and strong point. In GA, if a thing is not choose, then the information will be loosed but PSO has memory. Although without choosing operator, PSO will destroy the resources. So the principle idea in PSO-GA is the combination of collective thinking in PSO with the capability of GA local search. As an example, GA and PSO are on the basis of population, PSO-GA method is the population based algorithm so finding solutions is the key parts. This method is started from the first valuing step in which particle swap and their speed will be produced randomly in search space, it means that primary situation  $x_0^i$  of i particles are obtained randomly  $x_0^i \sim U(x_{\min}, x_{\max})$  from uniform distribution of  $[x_{\min}, x_{\max}]$ . In which  $x_{\min}$  and  $x_{\max}$  show variables of high and low border. While, speed vector is used for updating the current situation of each particle in swap and it is a whole on the basis of the memory obtaining through

a particle and also a knowledge which is obtained by swap. So situation of each particle in swap will be arranged with respects to its own and its neighbors' experience. The first one is the best individual situation and the second one as a best situation.

After repetition, situation of each particle will be updated according to equation. After producing new generation in PSO repetitions, some of the particles will choose new population and then GA will be applied for each of them separately, because size of the particles is very big so on the purpose of time saving, GA will not be applied to the whole population. So this question is presented that how many sets will be developed in PSO generation so regarding this, from the sum of population size, number is considered by GA Num that each developed one in each generation is called PSO.

$$GA_{Num} = GA_{NumMax} - \left( \frac{PSO_i}{psomaxitrer} \right)^{10} \times (GA_{NumMax} - GA_{NumMin})$$

$PSO_i$  is the repetition of current PSO and  $PSO_{MaxIter}$  is the maximum number of generations in PSO. After choosing the best people of population, the algorithm aims to produce new population through changing the points in the current population with better points through genetic principles for example through choose operators, swap and cut. One point cutting is used for new mixture of two families through Roulette wheels.

After choose, cut and mutation operations, a shape of meritocracy is done for keeping best solutions in the population through the following equitation:

$$x_{i+1} = \begin{cases} y_i & f(y_i) < f(x_i) \\ x_i & o.w \end{cases} \quad i = 1, 2, \dots, GA_{PS}$$

After evaluating new population, population size and maximum number of repetition is defined for GA changes regarding PSO repetition and their relation as the following:

$$GA_{PS} = GA_{MinPS} + \left( \frac{PSO_i}{psomaxitrer} \right)^{10} \times (GA_{MaxPS} - GA_{MinPS})$$

And

$$GA_{MaxIter} = GA_{MinIter} + \left( \frac{PSO_i}{psomaxitrer} \right)^{15} \times (GA_{MaxIter} - GA_{MinIter})$$

Through a repetitive process, population reproduction is directed to the whole optimization and it is portrayed in figure 1.

- 1: Objective function:  $f(x)$ ;
- 2: Define Fitness F (eg.  $F \propto f(x)$  for maximization)
- 3: for  $PSO_i=1$  to  $PSO_{MaxIter}$
- 4: Initialize the position and velocity for each particle and evaluate them.
- 5: Initialize the particle's best known position to its initial position ( $p_{best}$ ) and update swarm's best known position ( $g_{best}$ ).
- 6: Update Velocity and position of each particle in the swarm.
- 7: Update GA parameters:  $GA_{ps}$ ,  $GA_{MaxIter}$ ,  $GA_{Num}$
- 8: for  $i=1$  to  $GA_{Num}$

```

9: Choose one of the individuals(X)
10: Initialize vector with  $GA_{ps}$  particles randomly
11: put  $Chrome(1)=X$ 
12: Select best global particle of GA
13: Initial probabilities of crossover (pc) and mutation (pm)
14: for  $GAi=1$  to  $GA_{MaxIter}$ 
15: Keep elite
16: Apply crossover and mutation operators
17: Evaluation of each particle
18: Select the current best for the next generation.
19: Update best global particle of GA
20: end;
21: X= best global particle of GA
22: end;
23: end;

```

#### IV. NUMERICAL SOLUTION

It is described by Arora [2] and Belegundu [3]. This includes minimizing the weight of the compression spring (shown in Figure 3). Restrictions on the minimum deviation, pressure, shear (shear) wave frequencies, restrictions on outside diameter and design variables.

Design variables are, average reactance ( $x_1$ ), wire diameter ( $x_2$ ) and the number of active reactance ( $x_3$ ). It turns formulating mathematical formula is as follows:

$$\begin{aligned} \min \quad & f(x) = (x_3 + 2)x_2 x_1^2 \\ \text{s.t.} \quad & g_1(x) = 1 - \frac{x_2^3 x_3}{71785 x_1^4} \leq 0 \\ & g_2(x) = \frac{4x_2^2 - x_1 x_2}{12566(x_2 x_1^3 - x_1^4)} + \frac{1}{5108 x_1^2} - 1 \leq 0 \\ & g_3(x) = 1 - \frac{140.45 x_1}{x_2^2 x_3} \leq 0 \\ & g_4(x) = \frac{x_1 + x_2}{1.5} - 1 \leq 0 \\ & 0.05 \leq x_1 \leq 2 ; 0.25 \leq x_2 \leq 1.3 ; 2 \leq x_3 \leq 15 \end{aligned}$$

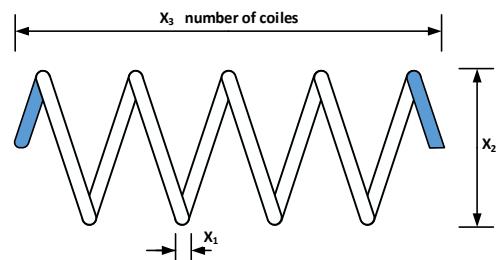


Fig.1. Design of the Pressure / Compression springs

This is by Belegundu [3] using mathematical optimization techniques have solved eight. (Only the best results are shown).

Arora solution this problem by using numerical optimization method called modified restrictions on a fixed cost. Coello [6] and Coello and Montes [7] the problem was solved with GA-based methods. Moreover, He, Wang [8] particle swarm optimization company products (CPSO) was used. Montes, Coello [14] used to solve this problem of the development of the strategies.

Table 1 shows that the best answer to this problem is caused by the PSO-GA algorithm and answer it with that of other earlier research by other authors in different ways, and

compares and in Table 8 by corresponding simulation results are shown.

The best result obtained by the PSO-GA is,  
 $f(\mathbf{X}) = 0.0126652327883$

In accordance with  $\mathbf{X} = [x_1, x_2, x_3] = [0.051689156131, 0.356720026419, 11.288831695483]$

And constraints

other methods.

Although these observations computing the answer by Hu et al [11], Kaveh and Tatahari [12] and He and colleagues [9], one of the restrictions violate collection and the non-feasible. Also shown in Table 2 is made, on average, PSO-GA search quality is better than other methods. In addition, the standard deviation of PSO-GA results in 30 independent run for it is the smallest.

TABLE I  
COMPARISON OF THE BEST ANSWER TO THE PROBLEM OF PRESSURE / COMPRESSION SPRING WITH DIFFERENT METHODS

Method	Design variables			$f(\mathbf{X})$
	$x_1$	$x_2$	$x_3$	
Belegundu [3]	0.05000000000000	0.31590000000000	14.250000000000	0.01283340000000
Arora [2]	0.05339600000000	0.39918000000000	9.18540000000000	0.01273030000000
Coello [6]	0.05148000000000	0.35166100000000	11.63220100000000	0.01270478000000
Ray and Saini [19]	0.05041700000000	0.32153200000000	13.97991500000000	0.01306000000000
Coello and Montes [7]	0.05198900000000	0.36396500000000	10.89052200000000	0.01268100000000
Ray and Liew [18]	0.05216021700000	0.36815869500000	10.64844225900000	0.01266924934000
Hu et al. [11]	0.05146636900000	0.35138394900000	11.60865920000000	0.0126661409 <sup>a</sup>
He et al. [9]	0.05169040000000	0.35674999000000	11.28712599000000	0.012665285 <sup>a</sup>
Hedar and Fukushima [10]	0.05174250340926	0.35800478345599	11.21390736278739	0.01266528500000
Raj et al. [17]	0.05386200000000	0.41128365000000	8.68437980000000	0.01274840000000
Tsai [20]	0.05168906000000	0.35671780000000	11.28896000000000	0.01266523000000
Mahdavi et al. [13]	0.05115438000000	0.34987116000000	12.07643210000000	0.01267060000000
Montes et al. [15]	0.05168800000000	0.35669200000000	11.29048300000000	0.01266500000000
He and Wang [8]	0.05172800000000	0.35764400000000	11.24454300000000	0.01266523300000
Cagnina et al. [4]	0.05158300000000	0.35419000000000	11.43867500000000	0.01269800000000
Zhang et al. [21]	0.05168906140000	0.35671774690000	11.28896533820000	0.01266523300000
Montes and Coello [14]	0.05164300000000	0.35536000000000	11.39792600000000	0.01269800000000
Omran and Salman [16]	0.05168374580000	0.35658983520000	11.29647171070000	0.01266523750000
Kaveh and Talatahari [12]	0.05186500000000	0.36150000000000	11.00000000000000	0.0126432 <sup>a</sup>
Coelho [5]	0.05151500000000	0.35252900000000	11.53886200000000	0.01266500000000
Akay and Karaboga[1]	0.05174900000000	0.35817900000000	11.20376300000000	0.01266500000000
Present study	0.05165095294742	0.35580165025062	11.34287600897656	0.01266525929387

<sup>a</sup> infeasible solution as they violate one of the constraint set

$$[g_1(\mathbf{X}), \dots, g_4(\mathbf{X})] =$$

$$\begin{aligned} & [-2.5313084961 \times 10 \\ & - 13, -5.7553961596 \times 10 \\ & - 13, -4.0537846722, -0.7277291363] \end{aligned}$$

## V. CONCLUSION

It can be seen from Table 1 that the best practical solution obtained by PSO-GA better than the best solution found by

In this paper, a combined method of new penalty direction namely PSO-GA algorithm is presented for constraint

TABLE II  
STATISTICAL RESULTS OF DIFFERENT METHODS FOR PRESSURE / COMPRESSION SPRING

Method	Best	Mean	Worst	Std Dev	Median
Belegundu [3]	0.0128334000000	NA	NA	NA	NA
Arora [2]	0.0127303000000	NA	NA	NA	NA
Coello [6]	0.0127047800000	0.0127692000000	0.0128220800000	$3.9390 \times 10^{-5}$	0.0127557600000
Ray and Saini [19]	0.0130600000000	0.0155260000000	0.0189920000000	NA	NA
Coello and Montes [7]	0.0126810000000	0.0127420000000	0.0129730000000	$5.9000 \times 10^{-5}$	NA
Ray and Liew [18]	0.0126692493400	0.0129226690000	0.0167172720000	$5.92 \times 10^{-4}$	0.0129226690000
Hu et al. [11]	0.0126661409000	0.0127189750000	NA	$6.446 \times 10^{-5}$	NA
He et al. [9]	0.0126652812000	0.0127023300000	NA	$4.12439 \times 10^{-5}$	NA
He and Wang [8]	0.0126747000000	0.0127300000000	0.0129240000000	$5.1985 \times 10^{-5}$	NA
Zhang et al. [21]	0.0126652330000	0.0126693660000	0.0127382620000	$1.25 \times 10^{-5}$	NA
Hedar and Fukushima [10]	0.0126652850000	0.0126652990000	0.0126653380000	$2.2 \times 10^{-8}$	NA
Montes et al. [15]	0.0126650000000	0.0266600000000	NA	$2.0 \times 10^{-6}$	NA
Montes and Coello [14]	0.0126980000000	0.0134610000000	0.1648500000000	$9.6600 \times 10^{-4}$	NA
Cagnina et al. [4]	0.0126650000000	0.0131000000000	NA	$4.1 \times 10^{-4}$	NA
Kaveh and Talatahari [12]	0.0126432000000	0.0127200000000	0.1288400000000	$3.4888 \times 10^{-5}$	NA
Omran and Salman [16]	0.0126652375000	0.0126652642000	NA	NA	NA
Coelho [5]	0.0126650000000	0.0135240000000	0.0177590000000	0.001268	0.0129570000000
Akay and Karaboga [1]	0.0126650000000	0.0127090000000	NA	0.012813	NA
Present study	0.01266540913117	0.01276203779028	0.01302966583665	$1.0234 \times 10^{-4}$	0.01271758256821

optimization problem. The ability of the algorithm to explore and exploit simultaneously, a growing amount of theoretical justification, and successful application to real-world problems strengthens the conclusion that PSO-GA is a powerful, robust optimization technique. On the purpose of showing the effect and power of algorithm comparing with optimization methods, constraint optimization problems of engineering design, including designing containers under the pressure, boiling bars are studied. In this optimization problem, the goal is minimizing design cost relating to different non linear constraints. Comparison of results with other developmental algorithm shows that PSO-GA algorithm is effective. Determining region, a general solution should be tested. Simulation result is done regarding the mean, the worst, the best and standard deviation. Moreover, standard deviation of design cost is lower and it is shown that this method is valuable in order to solve up-timing problems of engineering design.

#### REFERENCES

- [1] B. Akay and D. Karaboga, "Artificial bee colony algorithm for large-scale problems and engineering design optimization", *Journal of Intelligent Manufacturing*, vol. 23, pp. 1001–1014(2012).
- [2] J. S. Arora, "Introduction to Optimum Design", McGraw-Hill", New York, (1989).
- [3] A. D. Belegundu, "A Study of Mathematical Programming Methods for Structural Optimization", *PhD thesis, Department of Civil and Environmental Engineering, University of Iowa,Iowa, USA*, (1982).
- [4] L. C. Cagnina, S. C. Esquivel and C. A. C. Coello, "Solving engineering optimization problems with the simple constrained particle swarm optimizer, *Informatica*", vol.32, pp. 319–326,(2008).
- [5] L. S. Coelho, "Gaussian quantum-behaved particle swarm optimization approaches for constrained engineering design problems, *Expert Systems with Applications*", vol.37, pp.1676– 1683,(2010).
- [6] C. A. C. Coello, "Use of a selfadaptive penalty approach for engineering optimization problems", *Computers in Industry*, vol.41, pp.113–127,(2000).
- [7] C. A. C. Coello and E. M. Montes, "Constraint- handling in genetic algorithms through the use of dominance-based tournament selection", *Advanced Engineering Informatics*, vol.16,pp. 193–203, (2002).
- [8] Q. He and L. Wang, "An effective co - evolutionary particle swarm optimization for constrained engineering design problems", *Engineering Applications of Artificial Intelligence*, vol. 20,pp. 89–99, (2007).
- [9] S. He, E. Prempain and Q. H. Wu, "An improved particle swarm optimizer for mechanical design optimization problems", *Engineering Optimization*, vol. 36, pp. 585–605,(2004).
- [10] A. R. Hedar and M. Fukushima, "Derivative - free filter simulated annealing method for constrained continuous global optimization", *Journal of Global Optimization*, vol. 35, pp. 521–549,(2006).
- [11] X. H. Hu, R. C. Eberhart and Y. H. Shi, "Engineering optimization with particle swarm", *Proceedings of the 2003 IEEE Swarm Intelligence Symposium*, pp. 53–57,(2003).
- [12] A. Kaveh and S. Talatahari, "An improved ant colony optimization for constrained engineering design problems", *Engineering Computations*, vol. 27, pp.155–182,(2010).
- [13] M. Mahdavi, M. Fesanghary and E. Damangir, "An improved harmony search algorithm for solving optimization problems", *Applied Mathematics and Computation*, vol. 188, pp. 1567–1579, (2007).
- [14] E. M. Montes and C. A. C. Coello, "An empirical study about the usefulness of evolution strategies to solve constrained optimization problems", *International Journal of General Systems*, vol. 37, pp. 443–473, (2008).
- [15] E. M. Montes, C. A. C. Coello, J. V. Reyes and L. M. Davila, "Multiple trial vectors in differential evolution for engineering design", *Engineering Optimization*, vol. 39, pp. 567–589,(2007).
- [16] M. G. H. Omran and A. Salman, "Constrained optimization using CODEQ", *Chaos, Solitons & Fractals*, vol. 42, pp. 662–668,(2009).
- [17] K. H. Raj, R. S. Sharma, G. S. Mishra, A. Dua and C. Patvardhan, "An evolutionary computational technique for constrained optimization in engineering design", *Journal of the Institution of Engineers India Part Me Mechanical Engineering Division*, vol.86, pp. 121–128 (2005).
- [18] T. Ray and K. M. Liew, Society and civilization: "An optimization algorithm based on the simulation of social behavior", *IEEE Transactions on Evolutionary Computation*, vol. 7,pp. 386– 396,(2003).
- [19] T. Ray and P. Saini, "Engineering design optimization using a swarm with an intelligent information sharing among individuals", *Engineering Optimization*, vol.33, pp.735–748,(2001).
- [20] J. Tsai, "Global optimization of nonlinear fractional programming problems in engineering design", *Engineering Optimization*, vol. 37, pp. 399–409,(2005).
- [21] M. Zhang, W. Luo and X. Wang, "Differential evolution with dynamic stochastic selection for constrained optimization", *Information Sciences*, vol. 178, pp.3043–3074, (2008).

# GPASS: A Graphical Password Scheme using alphanumeric characters and pictures

Shah Zaman Nizamani  
Department of Information Technology  
Quaid-e-Awam University of Engineering, Science & Technology  
Nawabshah, Pakistan  
shahzaman@quest.edu.pk

Syed Raheel Hassan  
Department of Computer Systems Engineering  
Quaid-e-Awam University of Engineering, Science & Technology  
Nawabshah, Pakistan  
raheel.hassan@quest.edu.pk

Muhammad Mubashir Khan  
Department of Computer Science & IT  
NED University of Engineering & Technology  
Karachi, Pakistan  
mmkhan@neduet.edu.pk

**Abstract**—Authentication is very important for secure use of any computerized system. Textual password is serving to authentication since long time, but it is vulnerable to different kinds of attacks. To make authentication process more secure and easy to memorize, graphical password authentication has been introduced. This approach solved most of the problems present in textual passwords. However shoulder surfing attack is common in graphical password schemes. Anyone monitoring the process of login, through camera or some kind of recording software can recognize the password easily. To overcome this issue researchers developed different graphical password schemes but most of them suffer from usability and memorability issues. Therefore a graphical password scheme is required, which is resistant to shoulder surfing and similar attacks along with better usability and memorability. In this paper a combined textual and graphical password scheme (GPASS) is proposed with its implementation and usability results. In the GPASS scheme users select password by clicking on a group of four password elements which help to improve the authentication process. Security analysis of GPASS scheme is also presented along with comparison of other recognition based graphical password schemes.

**Index Terms**—Graphical Authentication, Security, Usability, Alphanumeric password

## I. INTRODUCTION

Authentication is the first step towards securing a system. Textual authentication scheme is most widely used, but it has some security issues. In order to overcome the shortcomings of textual authentication scheme, different types of schemes have been proposed. Each type of authentication technique has its own advantages and disadvantages. In the below section hierarchy of authentication techniques as shown in Figure 1, is briefly introduced.

Authentication techniques can be divided into three major categories. Token based and Biometric based techniques require special hardware, while knowledge based techniques depend upon the information provided by the user. Knowledge based techniques are easy to implement and use, that is why they are most widely used. Authentication techniques are briefly described below.

**Token based authentication:** A user possesses some hardware device, on the basis of that device, system identifies

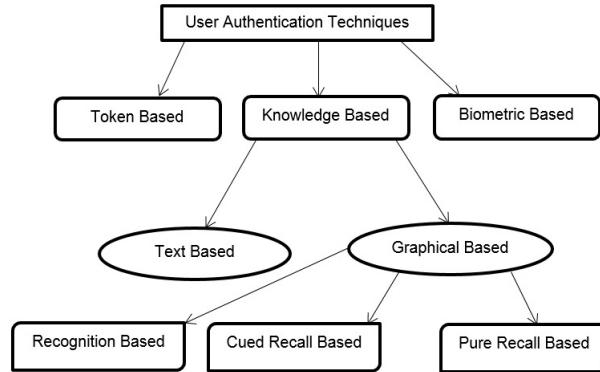


Fig. 1: User Authentication Techniques

the user. Bank ATM card for user authentication is an example of token based authentication.

**Biometric based authentication:** System identifies users, by recognizing their physical features. Iris scan and thumb impression are the examples of this technique.

**Knowledge based authentication:** System identifies users, through the information provided by the user. User name and password or the answer of few questions may be required for this kind of authentication. Textual and graphical authentication are two sub categories of knowledge based technique.

**Textual or text based authentication:** In this category, alphanumeric input is required from users. For example, through provided username and password the system recognizes the users.

**Graphical authentication:** In this technique a graphical window is used for input. For example users have to identify some pictures or draw some lines for successful authentication. Following are three sub categories of graphical authentication.

- 1) **Pure recall based authentication:** Users have to insert password without any clues. For example DAS (Draw-A-Secret) [1] scheme, where users have to draw lines which were inserted during registration.
- 2) **Cued recall based authentication:** In this type of authentication some hints are given to the user before

inserting password, therefore schemes under this category are easier to memorize than pure recall based schemes, PassFaces [2] scheme is the best example of this scheme.

- 3) **Recognition based authentication:** In this category a user has to identify the set of images which were selected during registration. Deja vu [3] scheme belongs to this category.

The remaining paper is divided into five sections. In section 2 literature review of different graphical authentication schemes is presented. General issues in graphical password schemes are discussed in section 3. In section 4 GPASS scheme is explained in detail. Analysis of GPASS scheme with respect to security, memorability and usability is given in section 5. Finally conclusion and future work is given in section 6.

## II. RELATED WORK

The concept of graphical passwords was first proposed by Blonder [4] in 1996. In his scheme users have to identify one or more points into the image as their password. This scheme does not allow users to select any point within the image, but users can select points from predefined areas of the image. Memorization of points is the key issue in this proposed scheme. After Blonder [4] many researchers proposed their schemes, such as Passpoints, proposed by Wiedenbeck et al. [5]. In this scheme a user is allowed to click on any point in the displayed image and set their password. Although this scheme is more secure than Blonder's [4] suggested scheme but users have difficulty to memorize and select the password.

Jermyn et al. [1] proposed Draw-A-Secret (DAS) scheme in which users draw lines as a password into  $N \times N$  grid. A study on this scheme is done by Nali and thrope [6]. In that study they found that users mostly draw lines in the center of the screen. Secondly most users draw 4 or less strokes and use symmetric shapes (e.g. rectangles and crosses), therefore dictionary attacks can be applied on DAS scheme.

Hai Tao [7] suggested a scheme named as Pass-Go in which grid size for images were increased from 5 to 9. Therefore dictionary attacks are difficult to apply. Besides security advantages this scheme suffers from usability issues, the users usually take extra time to draw the lines accurately.

Dhamija et al. [3] proposed Deja vu scheme, which is a recognition based graphical password scheme. In Deja vu scheme abstract images are used as password pictures. Guessability attacks are difficult to apply in abstract images but such images are difficult to remember. Therefore passwords in Deja vu scheme are difficult to remember.

Chiasson et al. [8] proposed a click based scheme known as Persuasive Cued Click Points (PCCP), in which security issue of hotspot is tried to solve by using persuasive technology. In this scheme password is entered by selecting a rectangular area known as viewport, into the password picture and then clicking of password points within the viewpoint. Memorization of password points are difficult in PCCP scheme.

Brostoff et al. [2] proposed PassFaces scheme in which users select certain human faces as a password. This scheme creates memorization issue when selected human faces are

unfamiliar, whereas familiar images create security breach issue.

Wiedenbeck et al. [9] suggested a shoulder surfing resistant scheme. In this scheme, users have to identify their pass objects or pictures among many objects present in the login screen. Pass objects form a convex-hull in the login screen and the users have to click inside the convex-hull for successful authentication. They suggested 1000 objects on the login screen. Although 1000 objects create security advantage but users find difficulty to quickly find their pass objects.

Aljahdali et al. [10] studied the effects of culturally familiar objects on security in graphical password schemes. Their study suggests that culturally familiar objects have memorability advantage at the cost of security.

Mokgadi et al. [11] explored the problems in graphical password schemes and they suggested that password images should be difficult to observe, record and guess by attackers. Secondly authentication schemes should be easy to use.

Akpulat et al. [12] suggested a mixed text based and graphical password scheme. In their proposed scheme a user has to assign alphanumeric code to some pass objects at the time of registration. In order to login, a user has to select and insert alphanumeric code on each of the selected pass object.

Khodadadi et al. [13] developed a metric for judging the security level of recognition-based graphical password schemes. Based on their developed metric they analyzed different recognition-based graphical password schemes and suggested security level of each of the scheme.

Nisha et al. [14] developed a new kind of password scheme based upon virtual reality, named as "3D passwords". In this scheme users have to perform some activities or actions on 3D password screen, to successfully authenticate.

Gao et al. [15] discussed the effects of colors on usability and security in the login screen of graphical password. Zhu et al. [16] utilized Captcha in development of graphical password scheme and described the effect of captcha, in security and usability of developed scheme.

Zhao et al. [17] proposed a shoulder surfing attack resistant scheme, in which user first set textual passwords for registration. Login process starts by clicking on three password characters which creates an invisible triangle, then user has to write the textual password character inside the triangle. This process will continue till last character of password.

A hybrid scheme known as "Click a Secret" was proposed by Eluard et al. [18]. This scheme is the combination of recognition and cued recall based schemes. At the time of registration user clicks on a specific region from the original image, and selects a variation for that region which is provided by the system. At login time user has to repeat same actions for authentication.

### A. Problems in Modern Graphical Password Schemes

For any authentication scheme, usability and memorability are also important beside security. Usability is an issue in most graphical password schemes [18], [14], [12]. Secondly password memorization is also an issue with graphical password schemes as users have to remember different kind of



Fig. 2: Registration Screen

passwords [2], [3]. Therefore a perfect balance among security, usability and memorability is required, which is missing in all the graphical password schemes.

Another problem with graphical password schemes is that, very few schemes provide resistance from shoulder surfing attack [9], [17]. While the schemes which are shoulder surfing resistant, take too much time for inserting password.

At present there are many types of computing devices, such as cursor based desktop systems, tablets and smart mobiles. Authentication scheme should be easy to use on every type of device. However efficiency of current graphical password schemes varies too much in different type of devices.

### III. PROPOSED SCHEME

Most graphical password schemes suffer from shoulder surfing attacks, while remaining schemes suffer from usability and memorability. GPASS scheme is an attempt to resist shoulder surfing, dictionary, phishing, spyware, and guessing attacks along with usability and memorability advantages. This scheme can be divided into two parts; one part is related with frontend graphical user interface while other part belongs to backend processing.

#### A. Frontend graphical user interface

Usability of an authentication scheme depends upon how efficiently user interface is designed. Users interact with GPASS scheme by login and registration screens. Users' interaction with the screens are discussed below.

*1) Registration screen:* Registration screen as shown in Figure 2 is used for collecting profile and authentication information from new users. All information except password can be entered through keyboard in the registration screen.

Users can insert password by clicking on different password elements available in the right side of registration screen. The alphanumeric characters and pictures may be called as password elements.

Registration process consists of the following steps:

- (i) User inserts personal information to create the profile.
- (ii) User inserts the username, if it is already present in the system then the user will be asked to insert another username.

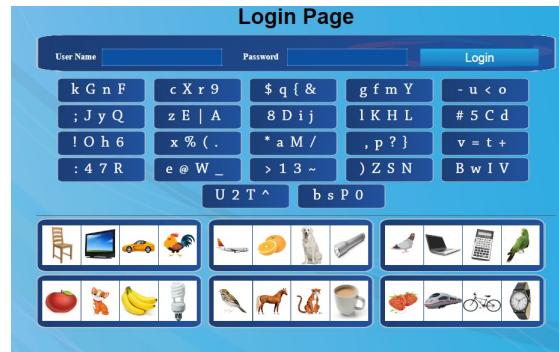


Fig. 3: Login Screen

- (iii) Finally, user enters the password, which can be either alphanumeric, graphical or the combination of both. User can insert password by clicking on any box representing alphanumeric character or password picture. Suggested minimum length of password is five characters, pictures or the combination of both.
- (iv) After password entry, user is required to click on register button for completing registration process.

*2) Login screen:* Users' authentication information is collected and sent to the server through login screen as shown in Figure 3.

In the login screen username can be entered through keyboard, while password can be entered by clicking on different boxes where password elements are present. Alphanumeric and picture password elements are presented separately in the login screen, so that user can easily locate their password elements.

Login process consists of the following steps:

- (i) First user needs to insert username by using login screen.
- (ii) Next step is to insert the password. Users can insert password by clicking on the groups where their chosen password elements are present. If the user's password length is five and it consists of two characters "De", and three pictures "horse, apple and car", then the user has to click on the following groups if the password elements arrangement is such as Figure 3.
  - (a) Third group of second row, for character "D".
  - (b) Second group of fourth row, for character "e".
  - (c) In the picture section, second group of second row, for image "horse".
  - (d) In the picture section, first group of second row, for image "apple".
  - (e) In the picture section, first group of first row, for image "car".

- (f) Finally user have to click on “Login” button for completing login process.
- (iii) When a user inserts correct entries for username and password, then the system allows to login.
- (iv) If a user inserts wrong username or password then system will refresh the login screen, i.e. all password elements will be rearranged. A user will have to insert username and password again.
- (v) Maximum three wrong login attempts are recommended.

Note that in the login screen, arrangement of password elements changes with every login session. This random arrangement of password elements is done through Algorithm III.1.

Following algorithm rearranges password elements for every login session.

#### **Algorithm III.1: RANDOM ARRANGEMENT(*APool*, *FPool*)**

```

APool[] ← List of alphanumeric characters
FPool[] ← List of Figure names
comment: APool & FPool holds all password elements
ALength ← 88
comment: 88 alphanumeric characters are used
FLength ← 24
comment: In GPASS scheme 24 figures are used

comment: Randomly arrange alphanumeric characters
for i ← 0 to 87
  do {ALength ← ALength – 1
    tempElement ← NULL
    ind ← Random(0, ALength)
    tempElement ← APool[ind]
    APool[ind] ← APool[ALength]
    APool[ALength] ← tempElement
  }

comment: Randomly arrange figure names
for i ← 0 to 23
  do {FLength ← FLength – 1
    tempElement ← NULL
    ind ← Random(0, FLength)
    tempElement ← FPool[ind]
    FPool[ind] ← FPool[FLength]
    FPool[FLength] ← tempElement
  }

```

#### **B. Backend Processing**

In this section two processes of GPASS scheme are discussed, the processes are login screen generation and password matching.

1) *Login screen generation:* Every time a user opens the login screen all password elements are randomly arranged. This random arrangement is important for securing shoulder

TABLE I: Lookup table

SNo.	Image	Password	GroupId
1	-	k	1
2	-	G	1
3	-	n	1
4	-	F	1
...	...	...	...
89	chair.jpeg	chair	25
90	television.jpeg	television	25
91	car.jpeg	car	25
92	chicken.jpeg	chicken	25
...	...	...	...

surfing attack. Furthermore login screen is generated through the following steps:

- (i) Login screen generation starts from rearranging of all password elements through Algorithm III.1. Initially alphanumeric characters and figure names are sequentially arranged into APool and FPool arrays respectively. Algorithm III.1 then randomly rearranges alphanumeric characters and figure names into APool and FPool arrays, as shown in Figure 4. Index number in both arrays shows the order in which characters and figures will be shown by login screen.
- (ii) After setting the order of password elements, GroupId is assigned to each password element. Same GroupId is assigned to consecutive four password elements and saved temporarily as shown in Table-I. This table is called as temporary lookup table, which has four columns. Order of password elements in the login screen depends upon the sequence created in the “Password” column and rendering of password elements depends upon “Image” column of the lookup table.
- In the “Image” column, dashes are given against all character based password elements as shown from serial number 1 - 4, because characters can be printed in the login screen by taking value from “Password” column. While picture names are given in the same “Image” column against all picture based password elements as shown in serial number 89 and onwards, because images can be shown in the login screen, through the physical path of the servers.
- (iii) Lookup table will be saved into session variable or in database for current authentication session. Sequence of password elements in the lookup table changes for every login session.
- (iv) Finally HTML code of login screen will be sent to the client machine.

2) *Password matching:* In this stage, username and password submitted through login screen are compared against the authentication data stored in the database. Password matching process consist of two parts, authentication data submission and authentication data comparison.

a) *Authentication data submission:* It this stage password elements clicked by the user are saved into client machine, then this information is sent to server. Authentication data submission is done through following steps.

- (i) User inserts the username.

<b>Before random ordering</b>							
APool Array							
Character	a	b	c	....	#	\$	....
Index	0	1	2	....	54	55	....
FPool Array							
Figures	Cat	Dog	Horse	....			
Index	0	1	2	....			

<b>After random ordering</b>							
APool Array							
Character	k	G	n	F	c	X	....
Index	0	1	2	3	4	5	....
FPool Array							
Figures	chair	television	car	....			
Index	0	1	2	....			

Fig. 4: Visual representation of Algorithm III.1

- (ii) User clicks on a group of four password elements. Every time user clicks on a group, respective GroupId will be stored into client browser or other client application. For example when user clicks on third group of second row, which consists of “8 D i j” password elements as shown in Figure 3, GroupId “8” will be saved in client machine.
- (iii) Step two will continue until all password elements are clicked by a user.
- (iv) Finally, user clicks on the login button to send, username and GroupIds to the server. Each GroupId represents four password elements, therefore every group must contain user corresponding password element. In current scenario GroupIds “8,17,27,26,23” are required for successful authentication.

b) *Authentication data Comparison:* In this stage, user's provided authentication information is compared with stored authentication information. Password comparison is done through the following steps.

- (i) Username and GroupIds are received by the server.
- (ii) Based upon provided username, password is fetched from the database.
- (iii) Password stored in the database is decrypted. In the current scenario the password is “D e horse apple car”. Each password element is separated by space, which can be changed with any other character or string.
- (iv) Based upon received GroupIds, system fetches all the elements of groups from temporary lookup table as shown in Table-I.
- (v) First element of the decrypted password is matched against all password elements of first clicked group,

second password element with second group and so on. In current scenario first password element is “D” and corresponding password elements of 8th group are “8, D, i , j”. Here the system will successfully matches with 8th group, as “D” is present in the group “8, D, i , j”.

- (vi) Step four will continue until all password elements are matched.
- (vii) If all password elements are successfully matched, then the system will allow login otherwise user will be redirected to the login page.

#### IV. ANALYSIS OF GPASS SCHEME

In this section, GPASS and some well known recognition based authentication schemes are compared under different attacks. Table-II contains six columns where each column indicates unique attack. “Y” and “N” options are given in Dictionary, Shoulder Surfing and Guessing attacks mentioned in Column one, two and six. For different well known recognition based authentication schemes mentioned in the first column, “Y” represents that the scheme is resistant to the attack while “N” represents not resistant to the attacks. Phishing attack mentioned in the fourth column show two values (medium and hard) according to the effort required for successful attack. Spyware attack mentioned in column five has high possibility that screen scraper attacks are successful in majority of the schemes while only few are resistant to these types of attack.

##### A. Password Space (Brute Force) Attack

In brute force attack an attacker tries all possible password combinations until, finds the correct password. To resist brute force attack the theoretical password space must have to be huge. Based upon standard American keyboard which has 95 alphanumeric keys, theoretical password space of textual password scheme is

$$\sum_{i=1}^{95} 95^i \quad (1)$$

In the GPASS scheme password space is

$$\sum_{i=1}^{112} 112^i \quad (2)$$

When 5 is the minimum length of password, then total number of combinations becomes

$$\sum_{i=5}^{112} 112^i \quad (3)$$

In GPASS scheme users have option to select password from 112 elements. This huge number of elements helps in creating different combinations of passwords. Therefore brute force attack is difficult to get success.

##### B. Shoulder Surfing

In Shoulder surfing attack an attacker closely observes the process of password entry. It can be done by closely sitting beside the user or by recording the password entry process through some recording device. The GPASS scheme resists

TABLE II: Security comparison of recognition based schemes [19]

Scheme	Dictionary Attack	Shoulder Surfing	Phishing Attack	Spyware Attack	Guessing attack
Deja vu [3]	Y	N	Medium	Screen-scrappers	Y
PassFaces [2]	N	N	Medium	Screen-scrappers	N
Story [20]	Y	N	Medium	Screen-scrappers	Y
Cognitive Authentication [21]	Y	Y	Hard	Y	Y
Convex Hull Click [9]	Y	Y	Hard	Y	Y
Use Your Illusion [22]	Y	N	Medium	Screen-scrappers	Y
Colorlogin [15]	Y	Y	Hard	Y	N
Picture Password [23]	Y	N	Medium	Screen-scrappers	Y
GPI/GPIS [24]	Y	N	Medium	Screen-scrappers	N
Proposed scheme (GPASS)	Y	Y	Hard	Screen-scrappers	Y

the shoulder surfing attack by not allowing user to insert exact password, as user clicks on the box of four password elements. Probability of hacker success in shoulder surfing attack in the GPASS scheme is given by the following formula.

$$P(S) = \left(\frac{1}{4}\right)^n \quad (4)$$

Here "n" represents the length of user password.

#### C. Guessing Attack

Guessing attack is used when an attacker knows about the user, such as user profile, likes and dislikes. In the GPASS scheme well known pictures and alphanumeric characters are utilized for password selection. Therefore, even if an attacker knows the user, cannot successfully guess the password. Well known objects have equal chance of password selection.

#### D. Dictionary Attack

Dictionary attack is a subset of brute force attack, in which an attacker tries only the password combinations which have high percentage of success. Dictionary creation is very difficult task for GPASS scheme because user can set password with the combination of text and graphical pictures.

#### E. Phishing or Forming Attack

In this attack an attacker deceives the user and gets the confidential information such as password or credit card information. Attacker deceives user by redirecting to fake website which looks like authentic website. Once user inserts confidential information then it can be easily collected by the attacker. Phishing attacks are difficult to apply in the GPASS scheme because user clicks on the set of four password elements. The specific password element cannot be identified by the attacker.

#### F. Spyware attack

Spyware is a program, which collects users information without their knowledge. Users do not deliberately install this program, but it is automatically installed by some infected website, compact disk or other installation devices. Key loggers, keyboard loggers and screen scrapers are some examples of spyware programs. GPASS scheme provides resistance to key logger Spywares, as user clicks on different locations on each login session. Keyboard loggers do not work in

GPASS scheme because user does not give input through keyboard. For Screen scrapers GPASS scheme is less resistant. Multiple recordings of login activity through screen scrapers can facilitate attacker to break the password.

#### G. Usability

In GPASS scheme, all password elements are grouped into set of four elements. This grouping creates an overhead for a user to find out the boxes where password elements are present. This extra step increases the time to insert password but on the other side this approach resists many attacks, especially shoulder surfing attack. Alphanumeric elements and graphical password elements are grouped separately, in order to reduce the password entry time.

While considering usability, users feel difficulty in adopting the sudden change of graphical user interface. People move easily from one system to another system when changes are made gradually. In GPASS scheme users have freedom to set their old text based password or with the combination text based and graphical elements. This freedom helps users to easily move towards GPASS scheme.

1) *Testing for Registration and Login Time:* In order to test the timing for registration and login processes of GPASS scheme a web based application was developed, based upon scheme's prototype as shown in Figure 2 and Figure 3. Application was developed by using "PhP" programming language and MySQL database.

For testing purpose, 40 volunteer users were selected from Quaid-e-Awam University. Out of 40 participants 10 were female while remaining 30 were male. Professionally 25 students participated from different departments of the university, 10 users were faculty members and remaining 5 users were administrative staff.

Before asking the participants for registration and login through GPASS scheme, purpose of the testing was described. Secondly demonstration was given to participants about how to register and login through GPASS scheme. When every participant fully understood how to use the testing application, then testing process started.

Participants were free to choose any type of password such as alphanumeric only, pictures only or by mixing both. Only one restriction imposed that all participants must set the password size of five elements. This restriction was imposed in order to clearly judge the timing of GPASS scheme on the specific threshold.

*2) Testing Results:* Registration and login time for GPASS scheme was calculated and saved into the database by the demonstration application. After completing the registration and login activity of 40 users, results were collected from database. GPASS scheme require 11.42 seconds for registration, 9.26 for password confirmation and 24.47 seconds for login, when password length is five. In the Table-III a comparison is shown among GPASS scheme and three other schemes. Timing for Passface and Passpoint schemes were taken from the research done by Jali [25] and timings of PCCP scheme were taken from the research done by Chaturvedi and Sharma [26].

TABLE III: Mean and SD Time in seconds for Different Schemes

Scheme	M/S	Register	Confirm	Login
PassFace	Mean	34	17	17
...	SD	21	6	6
PassPoint	Mean	12	8	8
...	SD	7	4	5
PCCP	Mean	50.7	15.7	16.2
...	SD	-	-	-
GPASS	Mean	11.42	9.26	24.47
...	SD	3.16	4.63	8.22

The login time of the GPASS scheme is higher than registration time, because in login screen all the password elements are presented in the set of four random elements. Users have to scan and find out their password element first and then they can click on their password element. While this process is easy in the case of registration as each password element is shown individually see Figure 2.

#### H. Memorability

GPASS scheme allows user to set textual, graphical password or even combination of both. Due to this feature, users can set their old textual password within GPASS scheme.

In recognition based schemes, users set pictures as password. Therefore the kind of pictures available for password selection, plays very important role of memorability of the scheme. There are many objects present in the world which can be used as password pictures. Hundreds of pictures cannot be used for an authentication scheme, because huge number of pictures will negatively affect the usability and memorability areas of the scheme. Culturally familiar pictures are easy to memorize [27], but guessability attacks on such password pictures are easy to apply. Therefore in the GPASS scheme, pictures are used which are familiar to almost all kind of users. It reduces the chance of guessability attacks and improves the memorability of the scheme.

## V. DISCUSSION

GPASS scheme is flexible enough to make some minor changes in the algorithms and graphical user interface of the scheme. Graphics of the authentication screens may be changed according to the actual system where the proposed scheme will be implemented. Mechanism of lookup table generation, password storage and matching can also be changed.

In the lookup table both textual and picture based password elements are represented by their original names as shown in the “Password” column of Table-I. These Password elements can be represented by some random alphanumeric characters. For example picture of “Apple” can be represented with some random characters such as “ax33b” or some unicode character of different language. This conversion of password element names improves the security of GPASS scheme.

GPASS scheme gives more importance to security than usability, this is a reason why a user requires on average 24 seconds to enter password. The time for password entry can be reduced if users are allowed to enter textual portion of their password through keyboard. Although this will harm the security of the scheme as key-loggers can send the password to the attackers. Users can use this facility if they are sure enough that no spyware is installed into the machine.

Password interference is a problem in all knowledge based authentication schemes, where users have to remember different password for different applications. This problem can arise in GPASS scheme as well. The intensity of password interference may be reduced in some level if same password pictures are used among different applications while implementing the proposed model.

GPASS scheme improved the security of recognition based graphical password authentication, especially against shoulder surfing attack. However this scheme needs to be improved for Screen scraping attacks. In this attack a computer program send, display information of the login screen to the attacker.

#### A. Secure Password Transmission and Storage

To reduce the risk of being intercepted the stolen passwords and restricted data must be encrypted when they are sent over the network. The latest version of TLS (version 1.2 or later) may be used for sending passwords or any secret content from web-browsers or web-based applications over the Internet [28]. Slower hashes are considered as a suitable mechanism for passwords storage. Instead of using faster hash functions like MD5, SHA-1, and SHA-256 using a slower hash like the bcrypt algorithm provides better protection against brute force attacks. Since each password takes more time to compute [29]. It has been observed that many users use same password for their multiple accounts on the internet, the use of salt is highly recommended for the security of web applications. Another benefit of salt is in the scenario when two users choose the identical passwords, or the same user might choose same passwords on different machines. The system stores such passwords with the same hash value. This would reveal the fact that two accounts have the same password, which may be exploited by the person who knows the password of one of the accounts to access the other account. By applying salt on the passwords with two random characters this problem may be avoided even if the same password is used for two different accounts. Salt scheme can also be used at the time of lookup table creation for adding more security to the scheme. GroupIds and image names of lookup table can be encrypted with salt scheme before sending information to the client.

## VI. CONCLUSION

GPASS scheme provides a secure mechanism for authentication. This scheme has bigger password space compare to textual password scheme. GPASS scheme resists many security attacks specially shoulder surfing attack, which is a threat to majority of graphical password schemes. Password memorization is easy in GPASS scheme. Users can set their old textual password into this scheme. User interface of GPASS scheme is very simple and easy. Users do not require lot of time for training before using this scheme. Password entry time of GPASS scheme is larger than textual scheme because users need to search and click password elements rather to type them. This usability disadvantage is minimal against the security advantages provided by GPASS scheme.

## ACKNOWLEDGEMENT

The authors acknowledge the students and faculty members of Quaid-e-Awam University of Engineering, Science and Technology, Pakistan, for spending their time to perform the usability tests of the GPASS scheme.

## REFERENCES

- [1] I. Jermyn, A. J. Mayer, F. Monrose, M. K. Reiter, A. D. Rubin, et al., "The design and analysis of graphical passwords," in *Usenix Security*, 1999.
- [2] S. Brostoff and M. A. Sasse, "Are passfaces more usable than passwords? a field trial investigation," in *People and Computers XIV Usability or Else!*, pp. 405–424, Springer, 2000.
- [3] A. Perrig and R. Dhamija, "Déjà vu: A user study using images for authentication," in *USENIX Security Symposium*, 2000.
- [4] G. E. Blonder, "Graphical password," Sept. 24 1996. US Patent 5,559,961.
- [5] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon, "Passpoints: Design and longitudinal evaluation of a graphical password system," *International Journal of Human-Computer Studies*, vol. 63, no. 1, pp. 102–127, 2005.
- [6] D. Nali and J. Thorpe, "Analyzing user choice in graphical passwords," *School of Computer Science, Carleton University, Tech. Rep. TR-04-01*, 2004.
- [7] H. Tao, *Pass-Go, a new graphical password scheme*. PhD thesis, University of Ottawa, 2006.
- [8] S. Chiasson, E. Stobert, A. Forget, R. Biddle, and P. C. Van Oorschot, "Persuasive cued click-points: Design, implementation, and evaluation of a knowledge-based authentication mechanism," *Dependable and Secure Computing, IEEE Transactions on*, vol. 9, no. 2, pp. 222–235, 2012.
- [9] S. Wiedenbeck, J. Waters, L. Sobrado, and J.-C. Birget, "Design and evaluation of a shoulder-surfing resistant graphical password scheme," in *Proceedings of the working conference on Advanced visual interfaces*, pp. 177–184, ACM, 2006.
- [10] H. M. Aljahdali and R. Poet, "Educated guessing attacks on culturally familiar graphical passwords using personal information on social networks," in *Proceedings of the 7th International Conference on Security of Information and Networks*, p. 272, ACM, 2014.
- [11] M. Rasekgala, S. Ewert, I. Sanders, and T. Fogwill, "Requirements for secure graphical password schemes," in *IST-Africa Conference Proceedings, 2014*, pp. 1–10, IEEE, 2014.
- [12] M. Akpulat, K. Bicakci, and U. Cil, "Revisiting graphical passwords for augmenting, not replacing, text passwords," in *Proceedings of the 29th Annual Computer Security Applications Conference*, pp. 119–128, ACM, 2013.
- [13] T. Khodadadi, M. Alizadeh, S. Gholizadeh, M. Zamani, and M. Darvishi, "Security analysis method of recognition-based graphical password," *Jurnal Teknologi*, vol. 72, no. 5, 2015.
- [14] N. Salian, S. Godbole, and S. Wagh, "Advanced authentication using 3d passwords in virtual world," *International Journal of Engineering and Technical Research*, vol. 3, no. 2, 2015.
- [15] H. Gao, X. Liu, R. Dai, S. Wang, and X. Chang, "Analysis and evaluation of the colorlogin graphical password scheme," in *Image and Graphics, 2009. ICIG'09. Fifth International Conference on*, pp. 722–727, IEEE, 2009.
- [16] B. B. Zhu, J. D. YAN, G. Bao, M. Yang, and N. Xu, "Captcha as graphical passwordsa new security primitive based on hard ai problems," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 6, pp. 891–904, 2014.
- [17] H. Zhao and X. Li, "S3pas: A scalable shoulder-surfing resistant textual-graphical password authentication scheme," in *Advanced Information Networking and Applications Workshops, 2007. AINAW'07. 21st International Conference on*, vol. 2, pp. 467–472, IEEE, 2007.
- [18] M. Eluard, Y. Maetz, and D. Alessio, "Action-based graphical password:"click-a-secret"," in *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, pp. 265–266, IEEE, 2011.
- [19] H. Gao, X. Liu, S. Wang, H. Liu, and R. Dai, "Design and analysis of a graphical password scheme," in *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on*, pp. 675–678, IEEE, 2009.
- [20] D. Davis, F. Monrose, and M. K. Reiter, "On user choice in graphical password schemes.," in *USENIX Security Symposium*, vol. 13, pp. 11–11, 2004.
- [21] D. Weinshall, "Cognitive authentication schemes safe against spyware," in *Security and Privacy, 2006 IEEE Symposium on*, pp. 6–pp, IEEE, 2006.
- [22] E. Hayashi, R. Dhamija, N. Christin, and A. Perrig, "Use your illusion: secure authentication usable anywhere," in *Proceedings of the 4th symposium on Usable privacy and security*, pp. 35–45, ACM, 2008.
- [23] W. Jansen, "Authenticating mobile device users through image selection," *The Internet Society: Advances in Learning, Commerce and Security*, vol. 1, pp. 183–194, 2004.
- [24] K. Bicakci, N. B. Atalay, M. Yuceel, H. Gurbaslar, and B. Erdeniz, "Towards usable solutions to graphical password hotspot problem," in *Computer Software and Applications Conference, 2009. COMPSAC'09. 33rd Annual IEEE International*, vol. 2, pp. 318–323, IEEE, 2009.
- [25] M. Z. Jali, *A study of graphical alternatives for user authentication*. PhD thesis, The university of Plymouth, 2011. Available at <https://pearl.plymouth.ac.uk/handle/10026.1/881>.
- [26] S. Chaturvedi and R. Sharma, "Securing text & image password using the combinations of persuasive cued click points with improved advanced encryption standard," *Procedia Computer Science*, vol. 45, pp. 418–427, 2015.
- [27] H. M. Aljahdali and R. Poet, "The affect of familiarity on the usability of recognition-based graphical passwords: Cross cultural study between saudi arabia and the united kingdom," in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on*, pp. 1528–1534, IEEE, 2013.
- [28] C. Heinrich, "Transport layer security (tls)," in *Encyclopedia of Cryptography and Security*, pp. 1316–1317, Springer, 2011.
- [29] N. Provos and D. Mazieres, "A future-adaptable password scheme.," in *USENIX Annual Technical Conference, FREENIX Track*, pp. 81–91, 1999.

# Concept Based Text Document Clustering with Vector Suffix Tree Document Model

Dr.N.Sandhya  
Professor  
CSE Department  
VNRVJIET, Hyderabad

Dr.A.Govardhan  
Professor & Principal  
CSE Department  
JNTUH,Hyderabad

Dr.G.Rameshchandra  
Professor  
CSE Department  
VNRVJIET, Hyderabad

**Abstract—** The most popular way for representing documents is the vector space model, because of its speed and versatility. The vector space model has some drawbacks. To overcome the bag of words problems, text documents are treated as a sequence of words and documents are retrieved based on sharing of frequent word sequences from text databases. The sequential relationship between the words and documents is preserved using a suffix tree data structure. Syntax based disambiguation is attempted by enriching the text document representations by background knowledge provided in a core ontology. Word Net is used for this purpose in our model.

This work aims to extend a document representation model which is elegant by combining the versatility of the vector space model, the increased relevance of the suffix tree document model and also retains the relationship between words like synonyms. The effectiveness and the relevance of this concept based model compared to the existing models is evaluated by a partitioning clustering technique and then a systematic comparative study of the impact of similarity measures in conjunction with different types of vector space representation on cluster quality is performed. This document model will be called the Concept Based Vector Suffix Tree Document Model (CBVSTDM).

**Keywords-** Text Clustering, Similarity Measures, Suffix tree WordNet, Cluster Accuracy

## I. INTRODUCTION

Text clustering is the process of grouping similar documents into clusters. Clustering is the task of grouping similar objects together [1]. Initially clustering was a technique used for improving the precision or recall in an Information Retrieval System [2] [3]. Clustering has evolved as the best technique for browsing a collection of documents [4] or organizing the results returned by a search engine in response to user's query [5] or to help users quickly identify and focus on the relevant set of results. Document clustering is an unsupervised learning from unstructured textual data and aids in improving the efficiency for various information retrieval (IR) tasks.

Most text mining approaches are based on the idea that a text document can be represented by a set of words, that is, a bag-of-words representation. However, in order to be able to define at least the importance of a word within a given document, usually a vector representation is used, where for

each word a numerical "importance" value is stored. The vector space model [6], is currently the predominant approach based on this idea.

The vector space model is versatile because vector representation can be used as a feature vector for a large number of clustering algorithms. The vector-based document models do not have the information about the order by which the words occur in a document. A document model that is more sophisticated and that preserves the complete word order information is the STD model [13]. Here the similarity between two documents is defined in terms of string overlaps in their common suffix tree. The concept of the words can be preserved by preprocessing the documents by enriching their representations with the external background knowledge provided in the core ontology [18].

## II. CONCEPT BASED VECTOR SUFFIX TREE DOCUMENT MODEL

An approach that combines suffix trees with the vector space model was proposed earlier [7] [8]. This approach uses the same Suffix Tree Document Model (STD) proposed by Zamir and Etzioni but they map the nodes from the common suffix tree to a M dimensional space in the VSM model. Thus, a feature vector containing the weights of each node can be used to represent the documents. Once the vector of weights is obtained, any similarity measure and clustering algorithm used with the Vector Space Model can be applied. This research extends the VSTDM [9] model by achieving the word sense disambiguation. The original Suffix Tree Clustering Model suffers from problem of the lack of an effective measure for the quality of clusters which is overcome in this model. An improved model has been developed in our proposed work.

### A. Constructing a Suffix Tree Document Model

Suffix tree document model considers a document  $d = w_1w_2:::w_m$  as a string consisting of words  $w_i$  and not characters ( $i = 1; 2; :::m$ ). A suffix tree of document  $d$  is a compact trie containing all suffixes of document  $d$ .

### B. Building the Document-Term Matrix

Once a generalized suffix tree of the document collection is built, the document-term matrix, or the TF-IDF matrix can be obtained. This document-term matrix can be built with a single traversal that is DFS traversal of the suffix tree as these values are already stored for each node.

### C. Steps to Build CBVSTDM

The following steps are carried out in building CBVSTDM:

- Collecting the document dataset
- Perform POS tagging
- Remove stop words
- Apply stemming using the WordNet stemmer
- The stemmed words are then looked up in the WordNet, the lexical database to replace the words by their synset IDs which corresponds to a set of word forms which are synonyms
- The words with the same synonyms are merged and are assigned a unique ID
- Unique suffixes are generated
- Build a concept based generalized suffix tree
- Constructing a document-term matrix from the generalized suffix tree
- Perform DFS traversal and obtain all word sequences
- Retain those frequent word sequences which satisfy the minimum support. The minimum support of the frequent word sequences is usually in the range of 4-15%. When the minimum support is too large, the total number of frequent words would be very small, so that the resulting compact documents would not have enough information about the original data set
- Thus the feature selection used is DF-thresholding
- Attach weights to the obtained word sequences using either TF or TF-IDF method
- The model is now evaluated using K-means clustering. The four similarity measures used in the analysis are Cosine, Jaccard, Euclidean and Pearson Correlation Coefficient

### III. EXPERIMENTAL SETUP

The aim of this work is to explore the benefits of the new model that contains the elegance of the “*best of three worlds*” that is preserving word sequences order, partial disambiguation of words by their POS, the inclusion of WordNet [16] concepts and benefits of vector space representation. It is very difficult to conduct a systematic study of comparing the impact of similarity measures on cluster quality with suffix tree clustering. In practice, manually assigned category labels are usually used as baseline criteria for evaluating clusters. As a result, the clusters, which are generated in an unsupervised way, are compared to the pre-defined category structure, which is normally created by human experts.

This work experiments with two data sets. One is the benchmark dataset Classic dataset collected from uci.kdd repositories and another dataset containing abstracts. Classic dataset consists of four different collections, namely CACM, CISI, CRAN and MED.

The second dataset consists of abstracts from four different fields which are downloaded from the web. Here the aim is to conduct WordNet based clustering of the downloaded abstracts from different research related topics. The abstracts of the following four research topics are collected: Network Security, Image Processing, Natural Language Processing and Data Mining. 100 documents of each research topic are selected totaling to 400 abstract documents for conducting the experiments.

### IV. RESULTS AND ANALYSIS

Many algorithms are available for implementing text document clustering [15]. A partitional k-means approach is used in this work for document clustering [12][14][17]. Here the seed points are statically chosen. However, efficiency can further be improved if seeds selected are random or run the code more than once to check the efficiency. From the previous study [10] it is proven that boolean representation with these similarity measures did not perform better. It is also seen that Euclidean measure performs worst. So, here analysis of these clusters using frequency and TF-IDF representation with Cosine, Jaccard and Pearson Correlation Coefficient measures is attempted [11]. Again Jaccard and Pearson measures emerge as the better techniques for comparing similarity between documents.

#### A. Evaluation of CBVSTDM using Classic Dataset

Here analysis of these clusters is performed using term frequency and TF-IDF representation with Cosine, Jaccard and Pearson Correlation Coefficient measures. As shown in Tables 1.1 and 1.2 Pearson measure with frequency and Jaccard measure with TF-IDF representations performs better. It is also observed from Table 1.3, the overall entropy representation table that for frequency count and TF-IDF representations with Cosine shows NaN values as some of the clusters are empty. Tables 1.4 and 1.5 show partitions as generated by the frequency count representation. Tables 1.6 and 1.7 show partitions as generated by the TF-IDF using Classic dataset. Figures 1 and 2 show the analysis of Jaccard and Pearson similarity measures with term frequency and TF-IDF representation. It is also observed that unlike the previous results TF-IDF with Pearson measure performs the worst with this CBVSTDM.

The clustering accuracy is used as a measure of a clustering result. Clustering accuracy  $r$  is defined as

$$r = \frac{\sum_{i=1}^4 a_i}{n}$$

where  $a_i$  is the number of instances occurring in both cluster  $i$  and its corresponding class and  $n$  is the number of instances in the dataset. Using the cluster accuracy formula for Classic dataset CBVSTDM has shown an accuracy of above 87%.

Table 1.1 Entropy results for TF-IDF representation using Classic dataset with CBVSTDM

	Cosine	Jaccard	Pearson
Cluster[0]	NaN	0.2166	0.3477
Cluster[1]	NaN	0.1955	0.1084
Cluster[2]	NaN	0.1628	0.2474
Cluster[3]	NaN	0.1409	0.0427

Table 1.2 Entropy results for frequency count representation using Classic dataset with CBVSTDM

	Cosine	Jaccard	Pearson
Cluster[0]	NaN	0.3369	0.2315
Cluster[1]	NaN	0.3340	0.0953
Cluster[2]	NaN	0.2727	0.2507
Cluster[3]	NaN	0.2393	0.1902

Table 1.3 Total Entropy results using Classic dataset with CBVSTDM

	Cosine	Jaccard	Pearson
Term Frequency	NaN	0.3053	0.1962
TF-IDF	NaN	0.1785	0.1774

Table 1.4 CBVSTDM Clustering results for frequency count representation with Jaccard Measure using Classic dataset

	CACM	CISI	CRAN	MED	Label
Cluster[0]	131	53	23	18	CACM
Cluster[1]	24	25	161	59	CRAN
Cluster[2]	38	107	7	8	CISI
Cluster[3]	7	15	9	115	MED

Table 1.5 CBVSTDM Clustering results for frequency count representation with PCC measure using Classic dataset

	CACM	CISI	CRAN	MED	Label
Cluster[0]	178	36	9	4	CACM
Cluster[1]	3	1	165	4	CRAN
Cluster[2]	15	155	8	10	CISI
Cluster[3]	4	8	18	182	MED

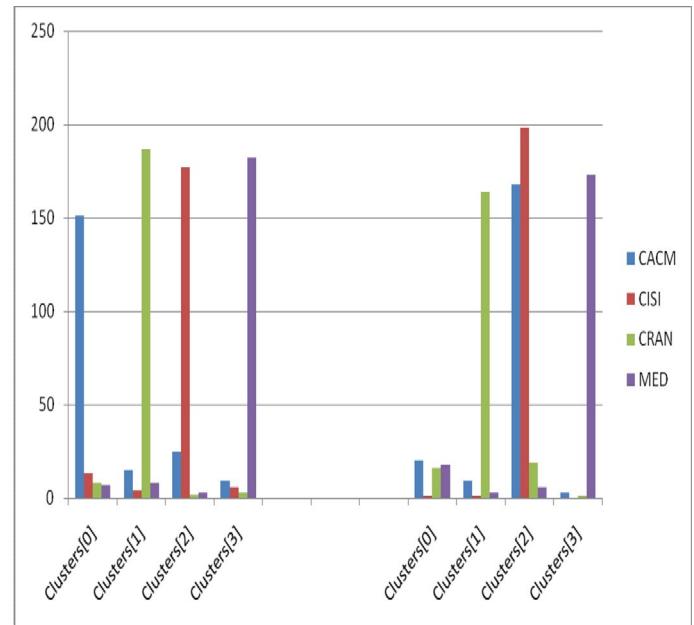


Figure 1: Analysis of Jaccard and PCC with frequency count representation for CBVSTDM

Table 1.6 CBVSTDM Clustering results for TF-IDF representation with Jaccard measure using Classic dataset

	<b>CACM</b>	<b>CISI</b>	<b>CRAN</b>	<b>MED</b>	<b>Label</b>
<b>Cluster[0]</b>	151	13	8	7	<b>CACM</b>
<b>Cluster[1]</b>	15	4	187	8	<b>CRAN</b>
<b>Cluster[2]</b>	25	177	2	3	<b>CISI</b>
<b>Cluster[3]</b>	9	6	3	182	<b>MED</b>

Table 1.7 CBVSTDM Clustering Results for TF-IDF representation with PCC measure using Classic dataset

	<b>CACM</b>	<b>CISI</b>	<b>CRAN</b>	<b>MED</b>	<b>Label</b>
<b>Cluster[0]</b>	20	1	16	18	<b>CACM</b>
<b>Cluster[1]</b>	9	1	164	3	<b>CRA N</b>
<b>Cluster[2]</b>	168	198	19	6	<b>CISI</b>
<b>Cluster[3]</b>	3	0	1	173	<b>MED</b>

#### A. Evaluation of CBVSTDM using Abstracts Dataset

As shown in Tables 1.8 and 1.9 Jaccard measure performs better with both frequency and TF-IDF representations. It is also observed from Table 1.10 that frequency count and TF-IDF representation with Cosine do not have non-zero clusters. But the cluster accuracy with Cosine value is not that significant. On an average, the Jaccard measure is slightly better in generating more coherent clusters, which means the clusters have lower entropy scores. Tables 1.11 and 1.12 show one partition as generated by the frequency count representation using abstracts dataset. Tables 1.13 and 1.14 show one partition as generated by the TF-IDF. It is also observed here that Pearson measure with TF-IDF performs very poor. Abstracts dataset has shown an accuracy of above 79% with Jaccard measure in TF-IDF representation.

Table 1.8 Entropy results for TF-IDF representation using Abstracts dataset with CBVSTDM

	<b>Cosine</b>	<b>Jaccard</b>	<b>Pearson</b>
<b>Cluster[0]</b>	0.3478	0.2268	0.3607
<b>Cluster[1]</b>	0.1766	0.1733	0.1142
<b>Cluster[2]</b>	0.1400	0.2934	0.3632
<b>Cluster[3]</b>	0.2757	0.1854	0.1622

Table 1.9 Entropy results for term frequency representation using Abstracts dataset with CBVSTDM

	<b>Cosine</b>	<b>Jaccard</b>	<b>Pearson</b>
<b>Cluster[0]</b>	0.2320	0.2244	0.3377
<b>Cluster[1]</b>	0.3312	0.1283	0.06048
<b>Cluster[2]</b>	0.2936	0.2767	0.2067
<b>Cluster[3]</b>	0.3032	0.1873	0.36146

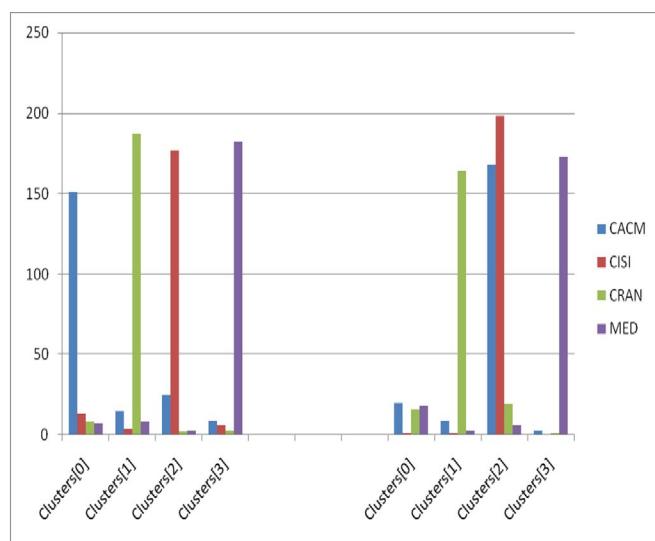


Figure 2: Analysis of Jaccard and PCC with TF-IDF representation for CBVSTDM

Table 1.10 Total Entropy results using Abstracts dataset with  
CBVSTDM

	Cosine	Jaccard	Pearson
Term Frequency	0.2929	0.2131	0.25739
TF-IDF	0.2897	0.2298	0.2745

Table 1.11 CBVSTDM Clustering results for frequency count representation with Jaccard measure using Abstracts dataset

	DM	IP	NLP	NS	Label
Cluster[0]	4	0	49	26	<b>NLP</b>
Cluster[1]	2	99	3	4	<b>IP</b>
Cluster[2]	94	0	30	30	<b>DM</b>
Cluster[3]	0	1	18	40	<b>NS</b>

Table 1.12 CBVSTDM Clustering results for frequency count representation with PCC measure using Abstracts dataset

	DM	IP	NLP	NS	Label
Cluster[0]	30	3	51	58	<b>NS</b>
Cluster[1]	1	83	2	0	<b>IP</b>
Cluster[2]	63	0	8	9	<b>DM</b>
Cluster[3]	6	14	39	33	<b>NLP</b>

Table 1.13 CBVSTDM Clustering results for TF-IDF representation with Jaccard measure using Abstracts dataset

	DM	IP	NLP	NS	Label
Cluster[0]	3	1	9	66	<b>NS</b>
Cluster[1]	1	97	6	9	<b>IP</b>
Cluster[2]	92	2	24	18	<b>DM</b>
Cluster[3]	4	0	61	7	<b>NLP</b>

Table 1.14 CBVSTDM Clustering results for TF-IDF representation with PCC measure using Abstracts dataset

	DM	IP	NLP	NS	Label
Cluster[0]	18	21	81	86	<b>NS</b>
Cluster[1]	1	77	2	3	<b>IP</b>
Cluster[2]	10	2	12	6	<b>NLP</b>
Cluster[3]	61	0	5	5	<b>DM</b>

The entire code for carrying the above steps is implemented in Java.

## V. CONCLUSION

The vector space model representation cannot handle lexical variation, semantic variation, syntactic variation and morphological variation. The suffix tree document model preserves the order of words in a document. Keeping word ordering increases the relevance of the document representation and also reflects better quality on various algorithms. But, with the suffix tree document model the similarity measures that are widely applied with the vector space model cannot be applied, thus making it inapplicable on a large scale, on a large number of scenarios.

A new approach called Concept Based Vector Suffix Tree Document Model (CBVSTDM) is proposed to overcome the VSM and the STD model problems and the efficiency of clustering algorithm is improved by including the background knowledge using WordNet in the document representation. This model is then evaluated using a partitioning clustering technique and then a systematic comparative study is carried out for the impact of similarity measures in conjunction with

different types of vector space representation on cluster quality.

The conclusion is that this model can be used in practice for text mining tasks. The future work will focus on the scalability of this algorithm with larger volumes of data.

## REFERENCES

- 1) Anil K. Jain and Richard C. Dubes. Algorithms for clustering data. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.
- 2) C.J.Van Rijsbergen,(1989), Information Retrieval, Butterworth, London, Second Edition.
- 3) G. Kowalski, Information Retrieval Systems – Theory and Implementation, Kluwer Academic Publishers, 1997.
- 4) D.R. Cutting, D.R. Karger, J.O. Pedersen, and J.W. Tukey, Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections, SIGIR '92, Pages 318 – 329, 1992.
- 5) O. Zamir, O. Etzioni, O. Madani, R.M. Karp, *Fast and Intuitive Clustering of Web Documents*, KDD '97, Pages 287-290, 1997.
- 6) Salton, G., Wong, A., Yang, C.S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11):613-620.
- 7) Hung Chim and Xiaotie Deng. *A new suffix tree similarity measure for document clustering*. In WWW '07: Proceedings of the 16<sup>th</sup> international conference on World Wide Web, pages 121–130, NewYork, NY, USA, 2007. ACM.
- 8) Hung Chim and Xiaotie Deng. *Efficient phrase-based document similarity for clustering*. Knowledge and Data Engineering, IEEE Transactions on, 20(9):1217–1229, Sept. 2008.
- 9) Horatiu Mocian. *Text mining with suffix trees*, 2009.
- 10) N.Sandhya, Y.Srilalitha, K.Anuradha, Dr.A.Govardhan. “Analysis of Stemming Algorithm” International Journal of Computer Science and Issues.
- 11) Anna Huang *Similarity Measures for Text Document Clustering* published in the proceedings of New Zealand Computer Science Research Student Conference 2008.
- 12) D. Arthur and S. Vassilvitskii. *k-means++ the advantages of careful seeding*. In Symposium on Discrete Algorithms, 2007.
- 13) Yanjun Li, Soon M. Chung, John D. Holt *Text document clustering based on frequent word meaning sequences*.
- 14) P. Pantel and D. Lin. *Document clustering with committees*. In Proc. Of SIGIR'02, Tampere, Finland, 2002.
- 15) M. Steinbach, G. Karypis, and V. Kumar. *A Comparison of Document Clustering Techniques*. In KDD Workshop on Text Mining, 2000.
- 16) Miller, G.: WordNet: a lexical database for English. Communications of the ACM, Volume 38, Issue 11 , pp.39- 41 (1995)
- 17) Technische Universit`at Dresden, *An Empirical Study of K-Means Initialization Methods for Document Clustering*.
- 18) A. Hotho, S. Staab, and G. Stumme. *Ontologies improve text document clustering*. In Proc. IEEE Int. Conf. on Data Mining (ICDM 03), pages 541–544, 2003.

## IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA  
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia  
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA  
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway  
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India  
Dr. Amogh Kavimandan, The Mathworks Inc., USA  
Dr. Ramasamy Mariappan, Vinayaka Missions University, India  
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China  
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA  
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico  
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India  
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania  
Dr. Junjie Peng, Shanghai University, P. R. China  
Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia  
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India  
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain  
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India  
Dr Li Fang, Nanyang Technological University, Singapore  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia  
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India  
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand  
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India  
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia  
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India  
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India  
Dr. P. Vasant, University Technology Petronas, Malaysia  
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea  
Dr. Praveen Ranjan Srivastava, BITS PILANI, India  
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong  
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia  
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan  
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria  
Dr. Riktesh Srivastava, Skyline University, UAE  
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia  
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt  
and Department of Computer science, Taif University, Saudi Arabia  
Dr. Tirthankar Gayen, IIT Kharagpur, India  
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan  
Prof. Ning Xu, Wuhan University of Technology, China  
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen  
& Universiti Teknologi Malaysia, Malaysia.  
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India  
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan

Prof. Syed S. Rizvi, University of Bridgeport, USA  
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan  
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghata (HP), India  
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal  
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P  
Dr. Poonam Garg, Institute of Management Technology, India  
Dr. S. Mehta, Inha University, Korea  
Dr. Dilip Kumar S.M, Bangalore University, Bangalore  
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan  
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University  
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia  
Dr. Saqib Saeed, University of Siegen, Germany  
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]  
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt  
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India  
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India  
Dr. Muhammad Sohail, KUST, Pakistan  
Dr. Manjaiah D.H, Mangalore University, India  
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India  
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada  
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia  
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India  
Dr. M. Azath, Anna University, India  
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh  
Dr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia  
Dr. Suresh Jain, Devi Ahilya University, Indore (MP) India,  
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia  
Dr. Hanumanthappa. J. University of Mysore, India  
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)  
Dr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria  
Dr. Santosh K. Pandey, The Institute of Chartered Accountants of India  
Dr. P. Vasant, Power Control Optimization, Malaysia  
Dr. Petr Ivankov, Automatika - S, Russian Federation  
Dr. Utkarsh Seetha, Data Infosys Limited, India  
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal  
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore  
Assist. Prof. A. Neela madheswari, Anna university, India  
Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India  
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh  
Dr. Atul Gonsai, Saurashtra University, Gujarat, India  
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand  
Mrs. G. Nalini Priya, Anna University, Chennai  
Dr. P. Subashini, Avinashilingam University for Women, India  
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat  
Mr Jitendra Agrawal, : Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal  
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India  
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof, Kanwalvir Singh Dhindsa, B.B.S.B.Engg. College, Fatehgarh Sahib (Punjab), India  
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah  
Mr. Nitin Bhatia, DAV College, India  
Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India  
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia  
Assist. Prof. Sonal Chawla, Panjab University, India  
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India  
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia  
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia  
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India  
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France  
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India  
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology, Durban, South Africa  
Prof. Mydhili K Nair, Visweswaraiah Technological University, Bangalore, India  
M. Prabu, Adhiyamaan College of Engineering/Anna University, India  
Mr. Swakkhar Shatabda, United International University, Bangladesh  
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan  
Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India  
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India  
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India  
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran  
Mr. Zeashan Hameed Khan, Université de Grenoble, France  
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow  
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria  
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India  
Dr. Maslin Masrom, University Technology Malaysia, Malaysia  
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India  
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City  
Dr. Mary Lourde R., BITS-PILANI Dubai, UAE  
Dr. Abdul Aziz, University of Central Punjab, Pakistan  
Mr. Karan Singh, Gautam Budtha University, India  
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India  
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia  
Assistant Prof. Yasser M. Alginahi, Taibah University, Madinah Munawwarah, KSA  
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India  
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India  
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India  
Dr. M. Munir Ahmed Rabbani, B. S. Abdur Rahman University, India  
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India  
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia  
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India  
Asst Prof. Jasmine. K. S, R.V. College of Engineering, India  
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius  
Prof. Mahesh Goyani, G H Patel College of Engg. & Tech, V.V.N. Anand, Gujarat, India  
Dr. Mana Mohammed, University of Tlemcen, Algeria  
Prof. Jatinder Singh, Universal Institution of Engg. & Tech. CHD, India

Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim  
Dr. Bin Guo, Institute Telecom SudParis, France  
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia  
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia  
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius  
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore  
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India  
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India  
Dr. C. Arun, Anna University, India  
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India  
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran  
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology  
Subhabrata Barman, Haldia Institute of Technology, West Bengal  
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan  
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India  
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India  
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand  
Dr. P. Chakrabarti, Sir Padampat Singhania University, Udaipur, India  
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.  
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran  
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India  
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA  
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India  
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India  
Mr. Serguei A. Mokhov, Concordia University, Canada  
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia  
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India  
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA  
Dr. S. Karthik, SNS Collegeof Technology, India  
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain  
Mr. A.D.Potgantwar, Pune University, India  
Dr. Himanshu Aggarwal, Punjabi University, India  
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India  
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai  
Dr. Prasant Kumar Pattnaik, KIST, India.  
Dr. Ch. Aswani Kumar, VIT University, India  
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA  
Mr. Arun Kumar, Sir Padam Pat Singhania University, Udaipur, Rajasthan  
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia  
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA  
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia  
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India  
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India  
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia  
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan

Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA  
Mr. R. Jagadeesh Kannan, RMK Engineering College, India  
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India  
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh  
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India  
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia  
Mr. R. Mohammad Shafi, Madanapalle Institute of Technology & Science, India  
Dr. F.Sagayaraj Francis, Pondicherry Engineering College, India  
Dr. Ajay Goel, HIET, Kaithal, India  
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India  
Mr. Suhas J Manangi, Microsoft India  
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded, India  
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India  
Dr. Amjad Rehman, University Technology Malaysia, Malaysia  
Mr. Rachit Garg, L K College, Jalandhar, Punjab  
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India  
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan  
Dr. Thorat S.B., Institute of Technology and Management, India  
Mr. Ajay Prasad, Sir Padampat Singhania University, Udaipur, India  
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India  
Mr. Syed Raifiul Hussain, Ahsanullah University of Science and Technology, Bangladesh  
Mrs Fazeela Tunnis, Najran University, Kingdom of Saudi Arabia  
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India  
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA  
Mr. Anand Kumar, AMC Engineering College, Bangalore  
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India  
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India  
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India  
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India  
Dr. V V S S Balaram, Sreenidhi Institute of Science and Technology, India  
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India  
Prof. Niranjana Reddy. P, KITS, Warangal, India  
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India  
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India  
Dr. A. Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai  
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India  
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan  
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India  
Dr. Tossapon Boongoen, Aberystwyth University, UK  
Dr. Bilal Alatas, Firat University, Turkey  
Assist. Prof. Jyoti Praaksh Singh, Academy of Technology, India  
Dr. Ritu Soni, GNG College, India  
Dr. Mahendra Kumar, Sagar Institute of Research & Technology, Bhopal, India.  
Dr. Binod Kumar, Lakshmi Narayan College of Tech.(LNCT)Bhopal India  
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan  
Dr. T.C. Manjunath, ATRIA Institute of Tech, India  
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan

Assist. Prof. Harmunish Taneja, M. M. University, India  
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India  
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India  
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad  
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India  
Mr. G. Appasami, Dr. Pauls Engineering College, India  
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan  
Mr. Yaser Miaji, University Utara Malaysia, Malaysia  
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh  
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India  
Dr. S. Sasikumar, Roever Engineering College  
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India  
Mr. Nwaocha Vivian O, National Open University of Nigeria  
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India  
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India  
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore  
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia  
Dr. Dhuha Basheer abdullah, Mosul university, Iraq  
Mr. S. Audithan, Annamalai University, India  
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India  
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India  
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam  
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India  
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad  
Mr. Deepak Gour, Sir Padampat Singhania University, India  
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India  
Mr. Ali Balador, Islamic Azad University, Iran  
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India  
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India  
Dr. Debojyoti Mitra, Sir padampat Singhania University, India  
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia  
Mr. Zhao Zhang, City University of Hong Kong, China  
Prof. S.P. Setty, A.U. College of Engineering, India  
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India  
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India  
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India  
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India  
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India  
Dr. Hanan Elazhary, Electronics Research Institute, Egypt  
Dr. Hosam I. Faiq, USM, Malaysia  
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India  
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India  
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India  
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan  
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India  
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia  
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India

Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India  
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India  
Prof Anupam Choudhary, Bhilai School Of Engg.,Bhilai (C.G.),India  
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya  
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.  
Dr. Kasarapu Ramani, JNT University, Anantapur, India  
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India  
Dr. C G Ravichandran, R V S College of Engineering and Technology, India  
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia  
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia  
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India  
Dr. Nikolai Stoianov, Defense Institute, Bulgaria  
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode  
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India  
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh  
Mr. Hemanta Kumar Kalita , TATA Consultancy Services (TCS), India  
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria  
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela  
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India  
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia  
Dr. Nighat Mir, Effat University, Saudi Arabia  
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India  
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore  
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore  
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US  
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India  
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India  
Mr. P. Sivakumar, Anna university, Chennai, India  
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia  
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India  
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia  
Mr. Nikhil Patrick Lobo, CADES, India  
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India  
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India  
Assist. Prof. Vishal Bharti, DCE, Gurgaon  
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India  
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India  
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India  
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India  
Mr. Hamed Taherdoost, Tehran, Iran  
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran  
Mr. Shantanu Pal, University of Calcutta, India  
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom  
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria  
Mr. P. Mahalingam, Caledonian College of Engineering, Oman  
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt

Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India  
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India  
Mr. Muhammad Asad, Technical University of Munich, Germany  
Mr. AliReza Shams Shafiqh, Azad Islamic university, Iran  
Prof. S. V. Nagaraj, RMK Engineering College, India  
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India  
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia  
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India  
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India  
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco  
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India  
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India  
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India  
Mr. Sunil Taneja, Kurukshetra University, India  
Mr. Fauzi Adi Rafrastra, Dian Nuswantoro University, Indonesia  
Dr. Yaduvir Singh, Thapar University, India  
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece  
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore  
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia  
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia  
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran  
Assoc. Prof. Dhirendra Mishra, SVKM's NMIMS University, India  
Prof. Shapor Zarei, UAE Inventors Association, UAE  
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India  
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India  
Prof. Anant J Umbarkar, Walchand College of Engg., India  
Assist. Prof. B. Bharathi, Sathyabama University, India  
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia  
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India  
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India  
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore  
Prof. Walid Moudani, Lebanese University, Lebanon  
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India  
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India  
Associate Prof. Dr. Manuj Darbari, BBD University, India  
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India  
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India  
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India  
Dr. Abhay Bansal, Amity School of Engineering & Technology, India  
Ms. Sumita Mishra, Amity School of Engineering and Technology, India  
Professor S. Viswanadha Raju, JNT University Hyderabad, India  
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India  
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India  
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia  
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia  
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India  
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia

Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India  
Mr. Shervan Fekri Ershad, Shiraz International University, Iran  
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh  
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh  
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India  
Ms. Sarla More, UIT, RGTU, Bhopal, India  
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India  
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India  
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India  
Dr. M. N. Giri Prasad, JNTUCE,Pulivendula, A.P., India  
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India  
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India  
Assist. Prof. Navnish Goel, S. D. College Of Enginnering & Technology, India  
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya  
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh  
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India  
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh  
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan  
Mr. Mohammad Asadul Hoque, University of Alabama, USA  
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India  
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan  
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA  
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India  
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina  
Dr S. Rajalakshmi, Botho College, South Africa  
Dr. Mohamed Sarrab, De Montfort University, UK  
Mr. Basappa B. Kodada, Canara Engineering College, India  
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India  
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India  
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India  
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India  
Dr . G. Singaravel, K.S.R. College of Engineering, India  
Dr B. G. Geetha, K.S.R. College of Engineering, India  
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon  
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran  
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India  
Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)  
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India  
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India  
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)  
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India  
Assist. Prof. Maram Balajee, GMRIT, India  
Assist. Prof. Monika Bhatnagar, TIT, India  
Prof. Gaurang Panchal, Charotar University of Science & Technology, India  
Prof. Anand K. Tripathi, Computer Society of India  
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India  
Assist. Prof. Supriya Raheja, ITM University, India

Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.  
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India  
Prof. Mohan H.S, SJB Institute Of Technology, India  
Mr. Hossein Malekinezhad, Islamic Azad University, Iran  
Mr. Zatin Gupta, Universti Malaysia, Malaysia  
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India  
Assist. Prof. Ajal A. J., METS School Of Engineering, India  
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria  
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India  
Md. Nazrul Islam, University of Western Ontario, Canada  
Tushar Kanti, L.N.C.T, Bhopal, India  
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India  
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh  
Dr. Kashif Nisar, University Utara Malaysia, Malaysia  
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA  
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan  
Assist. Prof. Apoorvi Sood, I.T.M. University, India  
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia  
Mr. Swapnil Soner, Truba Institute College of Engineering & Technology, Indore, India  
Ms. Yogita Gigras, I.T.M. University, India  
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College  
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad  
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India  
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad  
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India  
Mr. Mehdi Golsorkhtabar Amiri, Islamic Azad University, Iran  
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India  
Assist. Prof. P.Oliver Jayaprakash, Anna University,Chennai  
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India  
Dr. Asoke Nath, St. Xavier's College, India  
Mr. Masoud Rafighi, Islamic Azad University, Iran  
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India  
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India  
Mr. Sandeep Maan, Government Post Graduate College, India  
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India  
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India  
Mr. R. Balu, Bharathiar University, Coimbatore, India  
Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India  
Prof. P. Senthilkumar, Vivekanandha Institue of Engineering and Techology for Woman, India  
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India  
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India  
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India  
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran  
Mr. Laxmi chand, SCTL, Noida, India  
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad  
Prof. Mahesh Panchal, KITRC, Gujarat  
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode

Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India  
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhania University, India  
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India  
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India  
Mr. Srikanta Kumar Mohapatra, NMIFT, Orissa, India  
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan  
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India  
Prof. Elboukhari Mohamed, University Mohammed First, Oujda, Morocco  
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia  
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.  
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India  
Mr. G. Premsankar, Ericsson, India  
Assist. Prof. T. Hemalatha, VELS University, India  
Prof. Tejaswini Apte, University of Pune, India  
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia  
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran  
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India  
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India  
Mr. Vorugunti Chandra Sekhar, DA-IICT, India  
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia  
Dr. Aderemi A. Atayero, Covenant University, Nigeria  
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan  
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India  
Mr. Hassen Mohammed Abdulla Alsaifi, International Islamic University Malaysia (IIUM) Malaysia  
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan  
Mr. R. Balu, Bharathiar University, Coimbatore, India  
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar  
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India  
Prof. K. Saravanan, Anna university Coimbatore, India  
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India  
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN  
Assoc. Prof. S. Asif Hussain, AITS, India  
Assist. Prof. C. Venkatesh, AITS, India  
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan  
Dr. B. Justus Rabi, Institute of Science & Technology, India  
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India  
Mr. Alejandro Mosquera, University of Alicante, Spain  
Assist. Prof. Arjun Singh, Sir Padampat Singhania University (SPSU), Udaipur, India  
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad  
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India  
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India  
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia  
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India  
Mr. Hassen Mohammed Abdulla Alsaifi, International Islamic University Malaysia (IIUM)  
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA  
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu  
Dr. K. Reji Kumar, , N S S College, Pandalam, India

Assoc. Prof. K. Seshadri Sastry, EIILM University, India  
Mr. Kai Pan, UNC Charlotte, USA  
Mr. Ruikar Sachin, SGGSIET, India  
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India  
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India  
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt  
Assist. Prof. Amanpreet Kaur, ITM University, India  
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore  
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia  
Dr. Abhay Bansal, Amity University, India  
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA  
Assist. Prof. Nidhi Arora, M.C.A. Institute, India  
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India  
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India  
Dr. S. Sankara Gomathi, Panimalar Engineering college, India  
Prof. Anil Kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India  
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India  
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology  
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia  
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh  
Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India  
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India  
Dr. Lamri LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept. Computer Science, UBO, Brest, France  
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India  
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India  
Mr. Ram Kumar Singh, S.V Subharti University, India  
Assistant Prof. Sunish Kumar O S, Amaljyothi College of Engineering, India  
Dr Sanjay Bhargava, Banasthali University, India  
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India  
Mr. Roohollah Etemadi, Islamic Azad University, Iran  
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria  
Mr. Sumit Goyal, National Dairy Research Institute, India  
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India  
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur  
Dr. S.K. Mahendran, Anna University, Chennai, India  
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab  
Dr. Ashu Gupta, Apeejay Institute of Management, India  
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India  
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus  
Mr. Maram Balajee, GMR Institute of Technology, India  
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan  
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria  
Mr. Jasvir Singh, University College Of Engg., India  
Mr. Vivek Tiwari, MANIT, Bhopal, India  
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India  
Mr. Somdip Dey, St. Xavier's College, Kolkata, India

Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China  
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh  
Mr. Sathyaprakash P., S.K.P Engineering College, India  
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India  
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India  
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India  
Mr. Md. Abdul Ahad, K L University, India  
Mr. Vikas Bajpai, The LNM IIT, India  
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA  
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India  
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai  
Mr. A. Siles Balasingh, St.Joseph University in Tanzania, Tanzania  
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India  
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India  
Mr. Kumar Dayanand, Cambridge Institute of Technology, India  
Dr. Syed Asif Ali, SMI University Karachi, Pakistan  
Prof. Pallvi Pandit, Himachal Pradesh University, India  
Mr. Ricardo Verschueren, University of Gloucestershire, UK  
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India  
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India  
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India  
Dr. S. Sumathi, Anna University, India  
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India  
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India  
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India  
Assist. Prof M. Anand Kumar, Karpagam University, Coimbatore, India  
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex  
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India  
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India  
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat  
Mr. Sivakumar, Codework solutions, India  
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran  
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA  
Mr. Varadala Sridhar, Varadhaman College Engineering College, Affiliated To JNTU, Hyderabad  
Assist. Prof. Manoj Dhawan, SVITS, Indore  
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India  
Dr. S. Santhi, SCSVMV University, India  
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran  
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh  
Mr. Sandeep Reddivari, Mississippi State University, USA  
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal  
Dr. Hazra Imran, Athabasca University, Canada  
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India  
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India  
Ms. Jaspreet Kaur, Distance Education LPU, India  
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman  
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India

- Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India  
Mr. Khaldi Amine, Badji Mokhtar University, Algeria  
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran  
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India  
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India  
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia  
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India  
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India  
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India  
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany  
Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India  
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India  
Dr. Nadir Bouchama, CERIST Research Center, Algeria  
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India  
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco  
Dr. S. Malathi, Panimalar Engineering College, Chennai, India  
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India  
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India  
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan  
Dr. G. Rasitha Banu, Vel's University, Chennai  
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai  
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India  
Ms. U. Sirthuja, PSG college of arts &science, India  
Dr. Ehsan Saradar Torshizi, Urmia University, Iran  
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India  
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India  
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim  
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt  
Dr. Nishant Gupta, University of Jammu, India  
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India  
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India  
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus  
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Dr. Rahul Malik, Cisco Systems, USA  
Dr. S. C. Lingareddy, ALPHA College of Engineering, India  
Assistant Prof. Mohammed Shuaib, Interl University, Lucknow, India  
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India  
Dr. T. Thambidurai, Sun Univercell, Singapore  
Prof. Anandkumar Telang, BKIT, India  
Assistant Prof. R. Poorvadevi, SCSVMV University, India  
Dr Uttam Mande, Gitam University, India  
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India  
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India  
Dr. Mohammed Zuber, AISECT University, India  
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia  
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India

Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India  
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India  
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq  
Dr. Urmila Shrawankar, G H Raisoni College of Engineering, Nagpur (MS), India  
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India  
Dr. Mukesh Negi, Tech Mahindra, India  
Dr. Anuj Kumar Singh, Amity University Gurgaon, India  
Dr. Babar Shah, Gyeongsang National University, South Korea  
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India  
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India  
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India  
Assistant Prof. Ankit Garg, Amity University, Haryana, India  
Assistant Prof. Rajashe Karappa, SDMCET, Karnataka, India  
Assistant Prof. Varun Jasuja, GNIT, India  
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India  
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India  
Dr. Faouzi Hidoussi, UHL Batna, Algeria  
Dr. Naseer Ali Husieen, Wasit University, Iraq  
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai  
Dr. Ahmed Farouk Metwaly, K L University  
Mr. Mohammed Noaman Murad, Cihan University, Iraq  
Dr. Suxing Liu, Arkansas State University, USA  
Dr. M. Gomathi, Velalar College of Engineering and Technology, India  
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia  
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India  
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India  
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran  
Dr. Thiyyagu Nagaraj, University-INOU, India  
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe  
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India  
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India  
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India  
Dr. Shenshen Liang, University of California, Santa Cruz, US  
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia  
Mr. Snehasis Banerjee, Tata Consultancy Services, India  
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania  
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia  
Dr. Ying Yang, Computer Science Department, Yale University, USA  
Dr. Vinay Shukla, Institute Of Technology & Management, India  
Dr. Liviu Octavian Mafteiu-Scai, West University of Timisoara, Romania  
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq  
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India  
Dr. Timothy Powers, University of Hertfordshire, UK  
Dr. S. Prasath, Bharathiar University, Erode, India  
Dr. Ritu Shrivastava, SIRTS Bhopal, India  
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India  
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania

Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India  
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India  
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India  
Assistant Prof. Mallikarjun C Sarsamba Bheemnna Khandre Institute Technology, Bhalki, India  
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India  
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India  
Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq  
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco  
Dr. Parul Verma, Amity University, India  
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco  
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India  
Assistant Prof.. G. Selvavinayagam, SNS College of Technology, Coimbatore, India  
Assistant Prof. Madhavi Dhingra, Amity University, MP, India  
Professor Kartheesan Log, Anna University, Chennai  
Professor Vasudeva Acharya, Shri Madhwa vadira Institute of Technology, India  
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia  
Assistant Prof., Mahendra Singh Meena, Amity University Haryana  
Assistant Professor Manjeet Kaur, Amity University Haryana  
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt  
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia  
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India  
Assistant Prof. Dharmendra Choudhary, Tripura University, India  
Assistant Prof. Deepika Vodnala, SR Engineering College, India  
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA  
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India  
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan  
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India  
Assistant Prof. Chirag Modi, NIT Goa  
Dr. R. Ramkumar, Nandha Arts And Science College, India  
Dr. Priyadarshini Vydhalingam, Harathiar University, India  
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka  
Dr. Vikas Thada, AMITY University, Pachgaon  
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore  
Dr. Shaheera Rashwan, Informatics Research Institute  
Dr. S. Preetha Gunasekar, Bharathiyar University, India  
Asst Professor Sameer Dev Sharma, Uttarakhand University, Dehradun  
Dr. Zhihan Lv, Chinese Academy of Science, China  
Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, Amritsar  
Dr. Umar Ruhi, University of Ottawa, Canada  
Dr. Jasmin Cosic, University of BiHac, Bosnia and Herzegovina  
Dr. Homam Reda El-Taj, University of Tabuk, Kingdom of Saudi Arabia  
Dr. Mostafa Ghobaei Arani, Islamic Azad University, Iran  
Dr. Ayyasamy Ayyanar, Annamalai University, India  
Dr. Selvakumar Manickam, Universiti Sains Malaysia, Malaysia  
Dr. Murali Krishna Namana, GITAM University, India  
Dr. Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India  
Professor Vimalathithan Rathinasabapathy, Karpagam College Of Engineering, India

Dr. Sushil Chandra Dimri, Graphic Era University, India  
Dr. Dinh-Sinh Mai, Le Quy Don Technical University, Vietnam  
Dr. S. Rama Sree, Aditya Engg. College, India  
Dr. Ehab T. Alnfrawy, Sadat Academy, Egypt  
Dr. Patrick D. Cerna, Haramaya University, Ethiopia  
Dr. Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), India  
Associate Prof. Dr. Jiliang Zhang, North Eastern University, China  
Dr. Sharefa Murad, Middle East University, Jordan  
Dr. Ajeet Singh Poonia, Govt. College of Engineering & technology, Rajasthan, India  
Dr. Vahid Esmaeelzadeh, University of Science and Technology, Iran  
Dr. Jacek M. Czerniak, Casimir the Great University in Bydgoszcz, Institute of Technology, Poland  
Associate Prof. Anisur Rehman Nasir, Jamia Millia Islamia University  
Assistant Prof. Imran Ahmad, COMSATS Institute of Information Technology, Pakistan  
Professor Ghulam Qasim, Preston University, Islamabad, Pakistan  
Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women  
Dr. Wencan Luo, University of Pittsburgh, US  
Dr. Musa PEKER, Faculty of Technology, Mugla Sitki Kocman University, Turkey  
Dr. Gunasekaran Shanmugam, Anna University, India  
Dr. Binh P. Nguyen, National University of Singapore, Singapore  
Dr. Rajkumar Jain, Indian Institute of Technology Indore, India  
Dr. Imtiaz Ali Halepoto, QUEST Nawabshah, Pakistan  
Dr. Shaligram Prajapat, Devi Ahilya University Indore India  
Dr. Sunita Singhal, Birla Institute of Technologyand Science, Pilani, India  
Dr. Ijaz Ali Shoukat, King Saud University, Saudi Arabia  
Dr. Anuj Gupta, IKG Punjab Technical University, India  
Dr. Sonali Saini, IES-IPS Academy, India  
Dr. Krishan Kumar, MotiLal Nehru National Institute of Technology, Allahabad, India  
Dr. Z. Faizal Khan, College of Engineering, Shaqra University, Kingdom of Saudi Arabia  
Prof. M. Padmavathamma, S.V. University Tirupati, India  
Prof. A. Velayudham, Cape Institute of Technology, India  
Prof. Seifeide Kadry, American University of the Middle East  
Dr. J. Durga Prasad Rao, Pt. Ravishankar Shukla University, Raipur  
Assistant Prof. Najam Hasan, Dhofar University  
Dr. G. Suseendran, Vels University, Pallavaram, Chennai  
Prof. Ankit Faldu, Gujarat Technological Universiry- Atmiya Institute of Technology and Science  
Dr. Ali Habiboghi, Islamic Azad University  
Dr. Deepak Dembla, JECRC University, Jaipur, India  
Dr. Pankaj Rajan, Walmart Labs, USA  
Assistant Prof. Radoslava Kraleva, South-West University "Neofit Rilski", Bulgaria  
Assistant Prof. Medhavi Shriwas, Shri vaishnav institute of Technology, India  
Associate Prof. Sedat Akleylek, Ondokuz Mayis University, Turkey  
Dr. U.V. Arivazhagu, Kingston Engineering College Affiliated To Anna University, India  
Dr. Touseef Ali, University of Engineering and Technology, Taxila, Pakistan  
Assistant Prof. Naren Jeeva, SASTRA University, India  
Dr. Riccardo Colella, University of Salento, Italy  
Dr. Enache Maria Cristina, University of Galati, Romania  
Dr. Senthil P, Kurinji College of Arts & Science, India

Dr. Hasan Ashrafi-rizi, Isfahan University of Medical Sciences, Isfahan, Iran  
Dr. Mazhar Malik, Institute of Southern Punjab, Pakistan  
Dr. Yajie Miao, Carnegie Mellon University, USA  
Dr. Kamran Shaukat, University of the Punjab, Pakistan  
Dr. Sasikaladevi N., SASTRA University, India  
Dr. Ali Asghar Rahmani Hosseiniabadi, Islamic Azad University Ayatollah Amoli Branch, Amol, Iran  
Dr. Velin Kralev, South-West University "Neofit Rilski", Blagoevgrad, Bulgaria  
Dr. Marius Iulian Mihailescu, LUMINA - The University of South-East Europe  
Dr. Sriramula Nagaprasad, S.R.R.Govt.Arts & Science College, Karimnagar, India  
Prof (Dr.) Namrata Dhanda, Dr. APJ Abdul Kalam Technical University, Lucknow, India  
Dr. Javed Ahmed Maher, Shah Abdul Latif University, Khairpur Mir's, Pakistan  
Dr. B. Narendra Kumar Rao, Sree Vidyanikethan Engineering College, India  
Dr. Shahzad Anwar, University of Engineering & Technology Peshawar, Pakistan  
Dr. Basit Shahzad, King Saud University, Riyadh - Saudi Arabia  
Dr. Nilamadhab Mishra, Chang Gung University  
Dr. Sachin Kumar, Indian Institute of Technology Roorkee  
Dr. Santosh Nanda, Biju-Pattnaik University of Technology  
Dr. Sherzod Turaev, International Islamic University Malaysia  
Dr. Yilun Shang, Tongji University, Department of Mathematics, Shanghai, China  
Dr. Nuzhat Shaikh, Modern Education society's College of Engineering, Pune, India  
Dr. Parul Verma, Amity University, Lucknow campus, India  
Dr. Rachid Alaoui, Agadir Ibn Zohr University, Agadir, Morocco  
Dr. Dharmendra Patel, Charotar University of Science and Technology, India  
Dr. Dong Zhang, University of Central Florida, USA  
Dr. Kennedy Chinedu Okafor, Federal University of Technology Owerri, Nigeria  
Prof. C Ram Kumar, Dr NGP Institute of Technology, India  
Dr. Sandeep Gupta, GGS IP University, New Delhi, India  
Dr. Shahanawaj Ahamad, University of Ha'il, Ha'il City, Ministry of Higher Education, Kingdom of Saudi Arabia  
Dr. Najeed Ahmed Khan, NED University of Engineering & Technology, India  
Dr. Sajid Ullah Khan, Universiti Malaysia Sarawak, Malaysia  
Dr. Muhammad Asif, National Textile University Faisalabad, Pakistan  
Dr. Yu Bi, University of Central Florida, Orlando, FL, USA  
Dr. Brijendra Kumar Joshi, Research Center, Military College of Telecommunication Engineering, India

# **CALL FOR PAPERS**

## **International Journal of Computer Science and Information Security**

**IJCSIS 2016**  
**ISSN: 1947-5500**  
**<http://sites.google.com/site/ijcsis/>**

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

### ***Track A: Security***

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity  
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.,) Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on

its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

#### **Track B: Computer Science**

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com). Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes>.

**© IJCSIS PUBLICATION 2016  
ISSN 1947 5500  
<http://sites.google.com/site/ijcsis/>**