

IJCSIS Vol. 11 No. 3, March 2013
ISSN 1947-5500

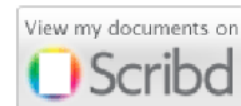
**International Journal of
Computer Science
& Information Security**

© IJCSIS PUBLICATION 2013



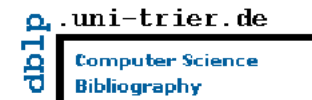
Cogprints

Google scholar



SciRate.com

CiteSeer^x beta



DOAJ DIRECTORY OF OPEN ACCESS JOURNALS



ProQuest

IJCSIS

ISSN (online): 1947-5500

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 2013 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 Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.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, Cornell University Library, ScientificCommons, EBSCO, ProQuest 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

CiteSeer^x beta

dblp.uni-trier.de
Computer Science
Bibliography

DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS



ProQuest

For more information, please visit the journal website (<https://sites.google.com/site/ijcsis/>)

Editorial Message from Managing Editor

International Journal of Computer Science and Information Security (IJCSIS – established in 2009), has been at the forefront of new knowledge dissemination in research areas of computer science and applications, and advances in information security. The journal themes focus on innovative developments, research challenges/solutions in computer science and related technologies. IJCSIS aims to be a high quality publication platform and encourages emerging scholars and academicians globally to share their research output and findings in the fields of computer science.

IJCSIS archives all publications in major academic/scientific databases; abstracting/indexing, editorial board and other important information are available online on homepage. IJCSIS editorial board consisting of international experts solicits your contribution to the journal with your research papers, projects, surveying works and industrial experiences. IJCSIS appreciates all the insights and advice from authors and reviewers. Indexed by the following International agencies and institutions: Google Scholar, Bielefeld Academic Search Engine (BASE), CiteSeerX, SCIRUS, Cornell's University Library EI, Scopus, DBLP, DOI, ProQuest, EBSCO. Google Scholar reported a large amount of cited papers published in IJCSIS.

IJCSIS is currently accepting quality manuscripts for upcoming issues based on original qualitative or quantitative research that explore innovative conceptual framework, or substantial literature review opening new areas of inquiry and investigation in Computer science. Case studies and works of literary analysis are also welcome.

We look forward to your collaboration. For further questions please do not hesitate to contact us at ijcsiseditor@gmail.com.

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

IJCSIS Vol. 11, No. 3, March 2013 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):



IJCSIS EDITORIAL BOARD

Dr. Yong Li

School of Electronic and Information Engineering, Beijing Jiaotong University,
P. R. China

Prof. Hamid Reza Naji

Department of Computer Engineering, Shahid Beheshti University, Tehran, Iran

Dr. Sanjay Jasola

Professor and Dean, School of Information and Communication Technology,
Gautam Buddha University

Dr Riktesh Srivastava

Assistant Professor, Information Systems, Skyline University College, University
City of Sharjah, Sharjah, PO 1797, UAE

Dr. Siddhivinayak Kulkarni

University of Ballarat, Ballarat, Victoria, Australia

Professor (Dr) Mokhtar Beldjehem

Sainte-Anne University, Halifax, NS, Canada

Dr. Alex Pappachen James (Research Fellow)

Queensland Micro-nanotechnology center, Griffith University, Australia

Dr. T. C. Manjunath

HKBK College of Engg., Bangalore, India.

Prof. Elboukhari Mohamed

Department of Computer Science,
University Mohammed First, Oujda, Morocco



TABLE OF CONTENTS

1. Paper 18021308: RGB Color Space Performance Limit for Skin Detection Using Neural Networks (pp. 1-4)

*Saleh Ali Alshehri
Jubail Industrial College, Jubail Industrial City, Saudi Arabia*

Abstract — Although the separation in skin and non-skin pixel classes is high using RGB color space, it is limited due to the high number of pixels that fall in both skin and non-skin classes. To improve the skin detection, some skin texture descriptors have been introduced. The color space components and skin texture descriptors were used in this research study as Neural Network feature vector. It has been demonstrated in this research study that when using a general image database, the skin detection performance rate could not go much beyond 83.3%. The result of this study is consistent with the findings of similar studies.

Keywords- skin detection, neural networks, RGB color space, LBP.

2. Paper 20021310: Location-based Solar Energy Potential Prediction Algorithm for Mountainous Rural Landscapes (pp. 5-12)

*Onabajo Olawale Olusegun & Chong Eng Tan
Faculty of Computer Science & Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Malaysia*

Abstract -- The world is facing critical energy crisis today. As a result the conventional grid energy supplies are not enough to meet the present demand. Many advance researches are in progress to overcome this energy predicament. Power generation and management in disconnected rural villages is challenging. The situation is even more challenging when landscape structure in such environment are irregular. Forces of diffusion, ground reflectance and sky view factor among others, affect the quality of final solar radiation incident on a solar panel. This paper describes the implementation of an algorithm that can be used to predict solar energy potential of irregular landscapes. Location-based Solar Energy Potential Prediction Algorithm (LOSEPPA) takes as input, the geographic latitude and longitude of the location of interest to compute the Solar Irradiance Factor (SIF). Geographic latitude plays an important role in the availability of sufficient solar radiation as well as the state of the atmosphere. Therefore, SIF value serves as a guide to the state of the atmosphere in terms of degree of cloud cover, temperature, humidity and landscape structure; which determines the feasibility of the solar energy implementation. The approach described in this paper can be used for rapidly computing the amount of solar radiation generated on a mountainous landscape surface and in the atmosphere as a function of height parameters. With SIF value known, solar panel can be mounted along specific angle of inclination to the sun. The algorithm design covers one year period and is based on the Digital Elevation Model (DEM) of the location under investigation. The proposed system was simulated using MATLAB1. Result show that the more irregular the landscape is, the lower the solar irradiance factor. SIF value of 400 and above predicts well enough sunshine for solar PV implementation in mountainous landscapes. Sample results show that solar radiation per kernel per day for a given landscape is highest between 12noon and 2.00PM local time; and the radiation per kernel per year for a given landscape have highest sunshine hours in January and December.

Keywords- Geographic latitude, Diffusion, Solar Panel, Landscape, DEM

3. Paper 23021318: Impact of Multipath Routing Performance of Video Traffic (pp. 13-20)

*Adel ECHCHAACHOU, SIME laboratory, E.N.S.I.A.S, Rabat, Morocco
Ali CHOUKRI, SIME laboratory, E.N.S.I.A.S, Rabat, Morocco
Ahmed HABBANI, SIME laboratory, E.N.S.I.A.S, Rabat, Morocco*

Mohammed ELKOUTBI, SIME laboratory, E.N.S.I.A.S, Rabat, Morocco

Abstract — Ad-hoc Network is a wireless environment of transmission which offers a very high mobility with low establishment costs. However, in this mode of communication the throughput and delay are limited, especially if traffic is needs high bandwidth, such as streaming video. In this paper, we study and evaluate the performance of video traffic in Ad-hoc network based on reactive routing protocol. As a first step, we study and compare the behaviour of AODV and AOMDV to carry streaming video traffic.

Keywords—*Performance, Video, AODV and AOMDV*

4. Paper 28021326: Stock Prices Forecast Using Radial Basis Function Neural Network (pp. 21-29)

Julia Fajaryanti, Faculty of Industrial Technology, Gunadarma University, Jalan Margonda Raya 100 Depok - Indonesia

Priyo Sarjono Wibowo, Faculty of Industrial Technology, Gunadarma University, Jalan Margonda Raya 100 Depok - Indonesia

Abstract — Neural Network has been implemented in various applications especially in pattern recognition. This power has attracted several people to use Neural Network for various systems. One of the neural network implementation in the field of finance or investments is forecasting stocks. Assuming that the prediction of the output system is deterministic, than the suitable Neural Network model to predict it is Multilayer Network. To get the solution, Multilayer Neural Network method with supervised algorithm is applied. The supervised algorithm used for stock price prediction is Radial Basis Function. This algorithm can supervise the networks by using previous stock price data, classifying them and putting weight on the networks. This journal illustrates how Radial Basis Function Neural Network method can be used to predict stocks. The result showed that Radial Basis Function Neural Network method is able to forecast and follow the movement of stock data used in the experiment.

Keywords: Stock Prices, Multilayer Neural Network, Radial Basis Function, Supervised

5. Paper 28021327: Model Based Framework for Estimating Mutation Rate of Hepatitis C Virus in Egypt (pp. 30-35)

*Nabila Shikoun, Mohamed ElNahas, Faculty of Engineering, Al Azhar University
Samar Kassim, Faculty of Medicine, Ain Shams University*

Abstract - Hepatitis C virus (HCV) is a widely spread disease all over the world. HCV has very high mutation rate that makes it resistant to antibodies. Modeling HCV to identify the virus mutation process is essential to its detection and predicting its evolution. This paper presents a model based framework for estimating mutation rate of HCV in two steps. Firstly profile hidden Markov model (PHMM) architecture was builder to select the sequences which represents sequence per year. Secondly mutation rate was calculated by using pair-wise distance method between sequences. A pilot study is conducted on NS5B zone of HCV dataset of genotype 4 subtype a (HCV4a) in Egypt.

Keywords: Hepatitis C virus (HCV), Profile Hidden Markov Model (PHMM), Non-structure 5 B(NS5B), Phylogenetic tree, pair-wise distance.

6. Paper 28021329: Exploring Tracer Study Service in Career Center Web Site of Indonesia Higher Education (pp. 36-39)

*Renny, Department of Accounting, Gunadarma University, Depok, Indonesia
Reza Chandra, Department of Information Systems, Gunadarma University, Depok, Indonesia
Syamsi Ruhama, Department of Informatics Management, Gunadarma University, Depok, Indonesia
Mochammad Wisuda Sarjono, Department of Information Systems, Gunadarma University, Depok, Indonesia*

Abstract — Quality competence of worker the present do not meet labor market criteria and the low level of labor productivity, the lack of communication between the labor market with education, changing of socio-economic structure and global political influence labor market, the development of science and technology very rapidly lead to

fundamental changes in terms of qualifications, competencies and requirements for entering the workforce. Tracer Study results can be used by universities to determine the success of the educational process that has been done towards their students. Therefore, universities need a technology services to support the optimization of the use of tracer study. One of that is the use of a website to facilitate the conduct tracer study. Most services tracer study provides information to college, like year graduated, got a job waiting period, the first salary to work, first job, the relevance of the curriculum to the work, and compliance with the major areas of work taken in college. Tracer study feature in Career Center Website affect the popularity website especially in traffic and rich file website.

Keywords — career center, tracer study, traffic, popularity.

7. Paper 28021342: Active Use of ICTs among the Elderly by Positive User Experience (pp. 40-44)

Ayako Hashizume, Faculty of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

Masaaki Kurosu, Center for ICT and Distance Education, The Open University of Japan, 2-11 Wakaba, Mihama, Chiba-shi 261-8586, Japan

Abstract — Recent technological advances have drastically changed our daily life. Information and communication technology (ICT) devices are being used by a wide variety of people to achieve diverse goals in different situations. However, there is an identifiable gap between high-end users and low-end users depending on demographic traits, particularly age. Focusing on ICT usage, we conducted a field survey by using the contextual inquiry method, analyzed the data by applying the modified grounded theory approach, and then summarized the results in a category relationship diagram. We found that motivation, active involvement in communication, and literacy are three principal factors for the use of ICTs.

Keywords - user experience (UX); elderly people; Qualitative approach; information and communication technology (ICT)

8. Paper 28021346: Feasibility Study of Millimeter Wave Transmission (pp. 45-48)

Mamta Agiwal, Associate Professor, CMRIT and Research Scholar at Jain University, Bangalore

Dr. Fathima Jabeen, Professor and P G Coordinator, K S School of Engineering and Management, Bangalore

Kashif Ahmed, Assistant Professor, Dept. of EEE, CMRIT, Bangalore

Abstract — Past few years have witnessed the stupendous growth in wireless communication networks. However much research on wireless communication has focused on Power consumption and the frequency reuse of spectrum in the range of 300 MHz–3 GHz, while the other direction could be in considering mm waves for transmission with a wave band of 30 GHz-300 GHz. Integrating of new technologies of optical fibers, mesh networks, improved CMOS platform and enhanced antenna design can open plethora of opportunities to use mm waves for transmission. The paper discusses advantages, limitations, possibilities and hardware developments to support transmission using mm waves.

Keywords- mm-wave; rain losses; mesh network; narrow beam and frequency reuse; CMOS platform.

9. Paper 28021348: Evaluation of Watermarking Approaches for Arabic Text Documents (pp. 49-54)

Muhammed N. Kabir, Omar Tayan and Yasser M. Alginahi

IT Research Center for the Holy Quran (NOOR), College of Computer Science and Engineering, Taibah University

P.O Box 344, Al-Madinah Al-Munawwarah, Saudi Arabia

Abstract — The embedding of digital watermark data in a cover text media, as in digital text watermarking, is used in numerous applications such as copyright protection, tamper detection, content-authenticity, steganography and other applications. Such issues have been largely studied for the security and protection of digital English-texts with relatively few citations found that address the specifics of other language characteristics. Moreover, the predominance of the text as a communications medium over the Internet suggests that more attention is required to protect online textual data in languages other than English. Hence, the purpose of this paper is to identify the

properties of different watermarking applications for the case of Semitic languages with particular focus on Arabic-text documents by evaluating three invisible text-watermarking approaches; Kashida-based, spacebased and sukun-based watermarking, the latter two of which present newly proposed watermark encoding schemes not found in the Arabic-text literature. This paper investigates the effect of two parameters on the watermarking scheme used, including; the word-group set-size, and, the number of bits embedded per set, before examining their consequent impact on the capacity and imperceptibility properties of the watermarking scheme on the host cover-text for different applications. Experimental results had illustrated the effect of the two encoding parameters on the resultant watermarking properties. It was found that by adjusting those variable watermark parameters, any target Arabic-text application could be optimized to achieve a desired capacity-ratio and level of imperceptibility.

Keywords-component; text watermarking, copyright protection, Kashida; diacritics, sukun.

10. Paper 28021349: Systematic Mapping Study on Security Threats in Cloud Computing (pp. 55-64)

*Carlo Marcelo Revoredo da Silva, José Lutiano Costa da Silva, Ricardo Batista Rodrigues, Leandro Marques do Nascimento, Vinicius Cardoso Garcia
Informatics Center, CIn Federal University of Pernambuco, UFPE, Recife, Brazil*

Abstract — Today, Cloud Computing is rising strongly, presenting itself to the market by its main service models, known as IaaS, PaaS and SaaS, that offer advantages in operational investments by means of on-demand costs, where consumers pay by resources used. In face of this growth, security threats also rise, compromising the Confidentiality, Integrity and Availability of the services provided. Our work is a Systematic Mapping where we hope to present metrics about publications available in literature that deal with some of the seven security threats in Cloud Computing, based in the guide entitled "Top Threats to Cloud Computing" from the Cloud Security Alliance (CSA). In our research we identified the more explored threats, distributed the results between fifteen Security Domains and identified the types of solutions proposed for the threats. In face of those results, we highlight the publications that are concerned to fulfill some standard of compliance.

Keywords: Security Threats, Cloud Computing, Systematic Literature Review, Security Domains, Compliance Issues.

11. Paper 28021354: Survey of Server Virtualization (pp. 65-74)

*Radhwan Y Ameen, Department of Comp. Engineering, College of Engineering, Mosul University, Mosul, Iraq
Asmaa Y. Hamo, Dept. of Software Engineering, College of Computer Sc. and Mathematics, Mosul University, Mosul, Iraq*

Abstract — Virtualization is a term that refers to the abstraction of computer resources. The purpose of virtual computing environment is to improve resource utilization by providing a unified integrated operating platform for users and applications based on aggregation of heterogeneous and autonomous resources. More recently, virtualization at all levels (system, storage, and network) became important again as a way to improve system security, reliability and availability, reduce costs, and provide greater flexibility. Virtualization has rapidly become a go-to technology for increasing efficiency in the data center. With virtualization technologies providing tremendous flexibility, even disparate architectures may be deployed on a single machine without interference This paper explains the basics of server virtualization and addresses pros and cons of virtualization.

Keywords- virtualization, server, hypervisor, Virtual Machine Manager, VMM , para virtualization , full virtualization, OS level server.

12. Paper 28021357: A Proposed DK-PC Algorithm for Code Bloat Control in a Tree-Based Genetic Programming (pp. 75-82)

*Oghorodi, Duke Urhe-otokoh, Department of Computer Science, College of Education, Warri, Delta State, Nigeria,
Asagba, Prince Oghenekaro, Department of Computer Science, University of Port Harcourt, River State, Nigeria*

Abstract — This paper addresses the Genetic Programming (GP) issue of code bloat which is the uncontrolled growth of program codes without a commensurate improvement in the program fitness to solve a given problem. Code Bloat is a serious issue in GP as it consumes computer memory and processing time. Though, several reasons and solutions for code bloat control have been suggested in literature, yet no final solution has been found so far. Against this backdrop, we proposed the Delete lower and Keep higher fitness value Programs after Crossover (DKPC) algorithm which keeps the higher fitness value program and delete the lower value fitness value programs from memory. We tested the Boolean 6-multiplexer and Boolean 11-Multiplexer functions against the preparatory requirements using our proposed algorithm, and we got very impressive results; we observed that the algorithm was able to control bloat to a large degree. However, the algorithm performed better in the Boolean 11-multiplexer function than in the Boolean 6-multiplexer function. Both functions displayed almost the same behaviour; except that the Boolean 11-multiplexer exhibited higher performance result than the Boolean 6-multiplexer in terms of better program size reduction. To this extent, our algorithm performed better in bloat control, based on the benchmark problems used.

Keywords — Genetic Programming, Evolutionary algorithms, Code Bloat

13. Paper 31011325: Attack Resilient & Adaptive Medium Access Control Protocol (pp. 83-91)

Piyush Kumar Shukla, CSE, UIT, RGPV, Bhopal, India

Sarita Singh Bhadauria, ECE, MITS, Gwalior, India

Sanjay Silakari, CSE, UIT, RGPV, Bhopal, India

Abstract — In IEEE 802.11, all nodes contending for the access to a medium needs to perform activities as per the specification of medium access control sub layer. It has been observed that when the number of node increases, it leads to the probability of collisions, which finally causes longer back-off values of the concerned collided nodes. The recent development within the field of computer networking enables everyone to access the Internet in the fastest manner using the Tablet, Mobile, Laptop or traditional Desktop. One common way to achieve Internet connectivity includes the use of Wi-Fi, which also forms the subject matter of present research. In the entire situations' basic requirement at the user's end- is a fair sharing of the available channel bandwidth, adequate quality of service (QoS), for which customers are paying lots of money. Unfairness in the network performance indicates the presence of attackers or some kind of misbehavior by existing users. It has been observed that sometimes to get more share of available bandwidth, several legitimate users show greediness or selfishness, which results in injustice to the other users in the same Wireless Local Area Network. However, it is too difficult to understand about the type and behavior of misbehaving nodes in the common shared environment. Another issue that requires attention is related to QoS that is if a user is availing better service than others within the network, then it will be appreciable. Otherwise, this is the matter of MAC misbehavior and needs to resolve. This research is motivated by selfish node, which manipulates their working (differ from normal MAC protocols) in different ways to increase their share to occupying the access to the channel. This exploitation of the MAC layer protocol may be hidden from the upper layers and in this work a solution has been proposed to embark upon the problem at the MAC Layer itself. A faired, Attacks Resilient and opportunistic Adaptive Medium Access Control protocol has been further modified, and its performance has been compared with existing CSMA/CA base on the Key performance Indicators, i.e. Throughput, Medium Access Delay, Collisions per frame and Fairness Index.

Keywords - Opportunist Mode, Attacks Resiliency, Adaptability, MAC Layer Misbehavior, Selfish Node, IEEE 802.11, DCF, back-off, Attacking Mode, Suspicious Mode.

14. Paper 18021307: A Cloud Computing Architecture for E-Learning Platform, Supporting Multimedia Content (pp. 92-99)

Atlee Gamundani, University of Zimbabwe, Computer Science Department, P.O.Box MP 167, Zimbabwe

Taurayi Rupere, University of Zimbabwe, Computer Science Department, P.O.Box MP 167, Harare, Zimbabwe

Benny M Nyambo, University of Zimbabwe, Computer Science Department, P.O.Box MP 167, Zimbabwe

Abstract — E-learning based platforms that support multimedia content to enhance interactive learning demands large disk space. Despite research ground covered under e-learning circles, less attention has been devoted to solicit the best methods to address the disk space challenges at minimal cost. This research focuses on advancing a best

architecture that meet the need for storage space when developing interactive multimedia e-learning based portals. Simulation was used using the CloudSim toolkit. Findings show that to precisely test the performance of viable architectures, there has to be a robust platform for such experiments. The main conclusions drawn from this research were that, there is room to improve on existing architectures to scale down on development costs so attributed to e-learning portals that are interactive in stature. Storage can be built from exiting personal computers through harnessing the cloud computing functionality designed as most of the personal computers are not fully being used by their owners. This research culminate by recommending the need to explore on best simulator packages that can be used to test the functionality of cloud computing based architecture for e-learning environments.

Keywords-E-learning, Multimedia, Architecture, Cloud Computing, Storage, CloudSim, Simulation, Interactive.

15. Paper 23021319: A Survey of Predicting Close Value in Stock Market (pp. 100-103)

Dharamveer Sisodia (1), Beerendra Kumar (2), Jitendra Kumar Gupta (2), Dr. Saurav Srivastava (3)

(1) M,Tech Scholar, Department of Computer Science & Engineering, SR Group of Institution, CSE Campus Jhansi, India

(2) Assistant Professor, Department of Computer Science & Engineering, SR Group of Institution, CSE Campus Jhansi, India

*(3) Assistant Professor Department of MCA, Bundelkhand University Jhansi, India
College Of Science & Engineering , Jhansi (Affiliated UPTU,Lucknow)*

Abstract — The goal of fundamental analysis is to decide the value of a stock based on the previously mentioned factors and to act on the assumption that the actual stock price will eventually reflect the determined value. Stock price forecasting is an important task for investment/financial decision making challenge. It receives considerable attention from both researches and practitioners. Stock market is highly volatile, complex and dynamic area so stock/price forecasting is a considerable challenging issue. Several approaches have been used for forecasting stock price such as traditional and fundamental methods. In this paper we propose a hybrid combinatorial method of horizontal partition based decision tree and the genetic algorithm for the prediction of close values in the stock market.

Keywords- Stock market prediction, Genetic Algorithm, Decision Tree.

16. Paper 28021330: Lecturer's Homepage Usage in Indonesian Private University (pp. 104-109)

Dessy Wulandari AP, Department of Informatics Management, Gunadarma University, Depok, Indonesia

Abdus Syakur, Department of Informatics Engineering, Gunadarma University, Depok, Indonesia

M. Achsan Isa Al Anshori, Department of Informatics Management, Gunadarma University, Depok, Indonesia

M. Akbar Marwan, Department of Information Systems, Gunadarma University, Depok, Indonesia

Abstract — The use of personal homepage by lecturers is getting popular in universities, especially to support teaching and learning process in the classroom. This study observed a lecturer in Information Systems Department in Indonesian Higher Education. Only 55.53% of the 151 lecturers who actively use the personal homepage. The amount of content and the popularity of the personal homepage on average still lower than the number of web pages viewed, number of documents, referring domain, and total backlinks. Statistical test results showed that there were differences in the number of web pages, the number of documents, and the total backlinks between male and female lecturers. There are significant differences occur only in the total number of web pages and documents, while referring domain and total backlinks shows no significant difference seen from the education lecturer.

Keywords- Lecturer's Homepage; Indonesian Private University; Differences based on gender and educational background

17. Paper 28021335: Comparative Survey of Steganography Techniques (pp. 110-114)

Rhythm walia, Om Pal

Centre for development of advanced computing, Noida, India

Abstract — Steganography in its core is about hiding message in such a manner that it is invisible to any intermediate party. Being undetected is the most important trait for any steganography technique. To be invisible a steganography technique should produce minimum distortion in the cover image. This paper contains a through description of the techniques and also a comparison between different steganography techniques for their security against various attacks.

18. Paper 28021347: Single CCTA-Based Four Input Single Output Voltage-Mode Universal Biquad Filter (pp. 115-119)

S. V. Singh, Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Sect-128, Noida-201304, India

R. S. Tomar, Department of Electronics Engineering, Anand Engineering College, Agra-282007, India

D. S. Chauhan, Department of Electrical Engineering, Institute of Technology, Banaras Hindu University, Varanasi-221005 (India)

Abstract — In this paper a new single current conveyor transconductance amplifier (CCTA)-based four-input single-output (FISO) voltage-mode universal biquad filter is proposed. The proposed filter employs only single CCTA, two capacitors and two resistors. The proposed filter realizes all the standard filter functions i.e. low pass (LP), band pass (BP) and high pass (HP), band reject (BR) and all-pass (AP) filters in the voltage form ,through appropriate selection of the four input voltage signals. The circuit does not require inverting-type input voltage signal(s) and double input voltage signal(s) to realize any response in the design. The filter enjoys attractive features, such as orthogonal electronic tunability of quality factor and pole frequency, low sensitivity performance. The validity of proposed filter is verified through PSPICE simulations.

Keywords-component; CCTA, Biquad, Universal, Filter

19. Paper 28021350: Reliable Multipath Routing Protocol (RMRP) For Mobile Ad Hoc Networks Using Adaptive Video Compression (pp. 120-124)

S. P. Swornambiga, Assistant Professor, Department of Computer Applications & Software Systems, C.M.S. College of Science and Commerce, Coimbatore, India

Antony Selvadoss Dhanamani, Associate Professor and Head, Research Department of Computer Science, NGM College, Pollachi, India

Abstract - This paper presents a reliable multipath routing protocol for mobile ad hoc networks using adaptive video compression. Mobile ad hoc network is the kind of wireless network which consists of mobile nodes and has the characteristic of deploying anywhere anytime. An adaptive video compression mechanism is deployed. Multipath routing mechanism is adapted from the Ad hoc On-demand Distance Vector (AOMDV) routing protocol and the received signal strength is measured for the discovered available paths. The path with the maximum received signal strength is selected and the packets (compressed video packets) are sent through the path. The performance metrics such as packet delivery ratio, throughput, drop, jitter are taken into account for comparison with AOMDV. Simulation results proved that the proposed RMRP outperforms AOMDV in all performance aspects.

20. Paper 28021360: Non-Preemptive Multi-Constrain Scheduling for Multiprocessor with Hopfield Neural Network (pp. 125-130)

Abdellatif H. ALI, ECE Department Modern Academy, Cairo, Egypt

Abstract- In this paper, task scheduling for non-preemptive multi-constrained multi-processor systems presented. The proposed model based on discrete Hopfield Neural network augmented with a methodology for weighting constrains to form overall network energy function. The network augmented with a layer to handle network re-initialization, based on min-max algorithm, case of local minima trapped without an acceptable solution. The proposed neural network solution does not require a predetermined scheduling length. Constrains included in the study are: task time, precedence, resources conflict, task dead time, and favoring tasks of the same setup to run on the same processor to suit reconfigurable hardware.

Keywords: Preemptive, Multi-constrains, Multi-processors, Scheduling, Hopfield, Neural Network.

21. Paper 31011311: Comparative Analysis between Split and HierarchyMap Treemap Algorithms for Visualizing Hierarchical Data (pp. 131-142)

Aborisade D. O., Department of Computer Science, College of Natural Sciences, Federal University of Agriculture, Abeokuta, (FUNAAB) Ogun State, Nigeria.

Oladipupo, O. O., Department of Computer and Information Sciences, Covenant University, Ota, Ogun State, Nigeria.

Oyelade, O. J., Department of Computer and Information Sciences, Covenant University, Ota, Ogun State, Nigeria.

Obembe O. O., Department of Biological Sciences, Covenant University, Ota, Ogun State, Nigeria.

Obagbuwa, I. C., Department of Computer Sciences, Lagos State University, Lagos, Nigeria. (LASU)

Ewejobi, I. T., Department of Computer and Information Sciences, Covenant University, Ota, Ogun State, Nigeria.

Abstract - We carried out comparative analysis between Split treemap algorithm and a more recently introduced treemap algorithm called HierarchyMap. HierarchicalMap and Split are Treemap Visualization methods for representing large volume of hierarchical information on a 2-dimensional space. Split layout algorithm has been developed much earlier as an ordered layout algorithm with capability to preserve order and reduce aspect ratio. HierarchyMap is a newer ordered treemap algorithm developed to overcome certain deficiencies of the Split layout algorithm. The two algorithms were analyzed to compare their rate of complexity. They were also implemented using object-oriented programming tool and compared using a number of standard metrics for measuring treemap algorithms. Their implementation shows that HierarchyMap and Split although maintain the same level of data ordering and usability but HierarchyMap algorithm has better aspect ratio, better readability, low run-time, and less number of thin rectangles compared to Split treemap algorithm. Since aspect ratio is an important metric for determining the efficiency of treemaps on 2-D and small screens, and the result of the analysis shows that HierarchyMap is better efficient than Split treemap algorithm, we conclude that HierarchyMap is more efficient than Split treemap algorithm.

Keywords: Treemap algorithm, Aspect ratio, HierarchyMap, 2-D space, Data Visualization.

22. Paper 29021255: Analysis of Network Security Policy – Based Management (pp. 143-146)

Aliyu Mohammed, Sulaiman Mohd Nor, Muhammad Nadzir Marsono

Universiti Teknologi Malaysia, Faculty of Electrical Engineering

Abstract — Network security and management policy in information communication is the desire to maintain the integrity, validity and consistency of a system or network, its data and its immediate environmental infrastructure. Well established and secured infrastructure would help in no means making the network safe from all kinds of intrusion. Protecting all these resources is another very important concept that is needed of any computer system. Harnessing, accessing and configuring relevant security policies are very important roles to be played in safeguarding the complex network infrastructure. The paper therefore analysis some of the desired policies and assessment guidelines that should be followed by network administrators for effective and strong network management, security facilities and data optimization.

Key words: Network Security; Management Policy; Intrusion; Domain Infrastructure.

RGB Color Space Performance Limit for Skin Detection Using Neural Networks

Saleh Ali Alshehri
Jubail Industrial College
Jubail Industrial City, Saudi Arabia

Abstract—Although the separation in skin and non-skin pixel classes is high using RGB color space, it is limited due to the high number of pixels that fall in both skin and non-skin classes. To improve the skin detection, some skin texture descriptors have been introduced. The color space components and skin texture descriptors were used in this research study as Neural Network feature vector. It has been demonstrated in this research study that when using a general image database, the skin detection performance rate could not go much beyond 83.3%. The result of this study is consistent with the findings of similar studies.

Keywords—skin detection, neural networks, RGB color space, LBP.

I. INTRODUCTION

It has been shown that RGB color space produces the highest skin and non-skin class separation [1,2]. However, it was demonstrated in this study that using RGB color space directly as feature vector for Neural Network approach to detect skin pixels may have a very high error rate. This is because there are many RGB pixel values which are common between skin and non-skin groups [3]. Figure 1 shows this relationship. It shows that the intersection between skin and non-skin can reach 80%. This means that the error in detection can reach 80%. Therefore, using a small image database may give a good result, but without generalization. This is because 19.5% pixels in any given large image samples fall in both skin and non-skin groups [3]. This is true for almost any skin detection technique that uses RGB color space components. NN is one example of such methods. It has been reported that an accuracy rate of up to 99.49% can be achieved using NN applied on RGB color components [4]. This is possible when test samples are chosen from limited resource that does not constitute generalization.

In this research, we contend that the maximum possible performance rate can be achieved when RGB color space and/or some of its transformations are used as NN feature vector. The result is reported when a generalized image database is used for testing.

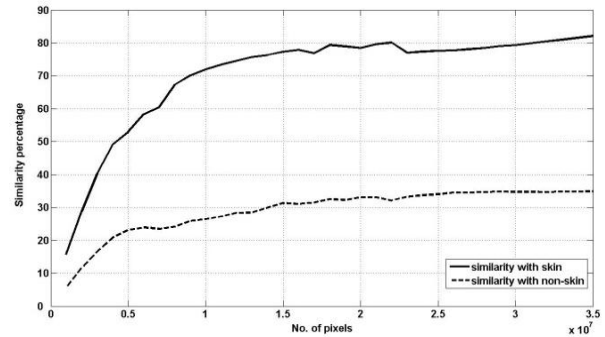


Figure 1. Similarity between skin and non-skin pixels [3].

The rest of the paper is organized as follows. Section 2 represents the method which was used in this study. After that, the results of this study are discussed followed by the conclusion.

II. METHD

Compaq image data base was selected to be used in this research [5]. It contains about 4600 skin images and 9000 non-skin images. This database has corresponding masks for each skin image to differentiate between skin and non-skin pixels. The World Wide Web was the source of images. Different angles, brightness and background conditions were followed when collecting the images. Various ethnicities, skin colors and tones are included in the database. Due to the Compaq database's immense size, it was decided to construct a small database by reducing the number of images used in this research to 750 images. We tried to avoid both pornographic images and images noted as very blurry. R, G and B components of the skin and non-skin pixels of all the images were gathered based on the given masks. The result was about 60×10^6 pixels. It is enough to use part of this group of pixels to illustrate the objective of this research.

It has been shown that by simply using the RGB component is not enough for skin detection [3]. So we decided to add more features. Skin texture can be a promising descriptor (descriptor and transformation is used interchangeably throughout this paper). There are many methods which are used for computing skin texture. In addition to the use of R-G and R-B [6], we chose to use three methods. They are:

- Difference with immediate neighbors (DIN).
- Difference between 8-connected neighbors (DBN).
- Local decimal pattern (LDP) based on local binary pattern (LBP) [7].

The pixels distribution shown in Fig. 2 and Fig. 3 has been followed to calculate DIN, DBN and LDP. The use of RG and RB has shown good performance rate [6]. Therefore three descriptor sets were computed from R-G matrix then from R-B matrix separately as follows:

1. LDP :

LBP is calculated first as in Fig. 2 then decimal value is obtained as :

$$LDP = \sum_{i=1}^8 LBP_i X 2^{(9-i)} \quad (1)$$

2. DIN set:

$DIN_i = (CP - P_i)$, where CP is the current pixel, P_i is the i^{th} neighboring pixel to CP and $1 < i < 8$

(2)

3. DBN set:

$DBN_1 = P_8 - P_4$, $DBN_2 = P_7 - P_3$, $DBN_3 = P_6 - P_2$,
 $DBN_4 = P_1 - P_5$, where P_i is the i^{th} neighboring pixel

8	7	6
1	Current pixel (CP)	5
2	3	4

Figure 2. Pixels numbering scheme.

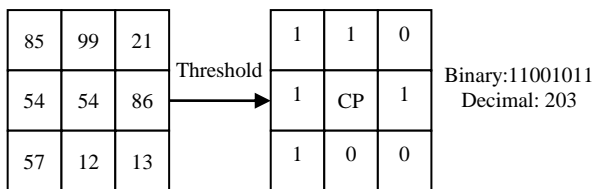


Figure 3. LBP and LDP calculation example [7].

All possible combinations of these transformations have been tried as NN feature vector. The neighboring pixels have been chosen because they show some degree of discrimination between skin and non-skin pixels and hence are good texture descriptors. Table I, Table II and Table III show the mean and

the standard deviation of these transformations which clarify that they can be used as discrimination features.

TABLE I. MEAN AND STD OF R-G AND R-B VALUES OF THE DATASET.

R-G				R-B			
Skin		Non-skin		Skin		Non-skin	
Mean	STD	Mean	STD	Mean	STD	Mean	STD
55.068	25.83	14.02	53.70	77.68	35.22	22.61	60.05

TABLE II. MEAN AND STD OF THE USED TEXTURE DESCRIPTORS.

Descriptor	Skin (R-G only)		Non-skin (R-G only)	
	Mean	STD	Mean	STD
LDP	164.164	87.231	174.992	87.991
DBN_1	0.4735	10.7948	-0.1963	20.2992
DBN_2	0.0674	8.4846	-0.0192	15.1944
DBN_3	-0.3323	10.5443	0.1586	19.6959
DBN_4	0.4405	8.0205	-0.1800	15.2205
DIN_1	-0.1142	5.1946	0.0825	8.7391
DIN_2	0.0658	6.6116	0.0551	11.5983
DIN_3	0.1566	5.4139	-0.0256	8.7281
DIN_4	0.4722	6.8645	-0.1208	11.9586
DIN_5	0.3263	5.3168	-0.0975	8.7530
DIN_6	0.3981	6.6807	-0.1035	11.5921
DIN_7	0.0892	5.4055	-0.0064	8.7090
DIN_8	-0.0013	6.6843	0.0755	11.9450

The NN is designed according to parameters shown in Table III. Since NN with one hidden layer can approximate any function, the number of hidden layers was chosen to be one [8]. Finding how many nodes in the hidden layer can be found by expanding and shrinking technique [9].

TABLE III. NN PARAMETERS.

NN parameters used in MATLAB	Value
Number of nodes in Hidden layer	10
Number of nodes in Output layer	1
Transfer function	tansig
Training function	trainglm
Learning rate	0.4
Momentum constant	0.6
Maximum number of Epochs	10000
Minimum performance gradient	1e-25

The training, validating and testing samples sizes were 500,000, 250,000 and 250,000 pixels respectively. Each set

contains 50% skin pixels and 50% non-skin pixels. For all NN experiments which were conducted with a different feature vector; the training, validation and testing performance were between 88.9% and 91.4%.

III. RESULT

To evaluate the NN performance, three performance indicators were used in this study. They are Correct Detection Rate (CDR), False Acceptance Rate (FAR) and False Rejection Rate (FRR). They can be obtained as follows:

$$CDR = \frac{\text{No. of pixels correctly classified}}{\text{Total no. of pixels}} \quad (3)$$

$$FAR = \frac{\text{No. of non-skin pixels classified as skin pixels}}{\text{Total no. of pixels}} \quad (4)$$

$$FRR = \frac{\text{No. of skin pixels classified as non-skin pixels}}{\text{Total no. of pixels}} \quad (5)$$

Throughout this research study, we attempted many combinations of the different features in order to find which combination produced the best results. Table IV shows the performance results of all possible combinations of the descriptors. The best result was the use of R-G, R-B and LDP combination. It was noticed that there is a very small performance rate difference between the use of LDP and DIN pixels together with R-G and R-B transformations. This is because both are constructed similarly where neighboring pixels are involved in the calculation for both methods.

It can be recognized also that most of the discrimination information are kept in R-G and R-B transformations. One can use only these two values when seeking to detection speed. For the purpose of accuracy, using LDP together with R-G and R-B will be the best choice. The test was done on 590 image files of Compaq image database which contain around 47×10^6 pixels.

TABLE IV. PERFORMANCE RESULT OF THE NN FOR VARIOUS COMBINATIONS OF THE DESCRIPTORS.

Feature vector	CDR	FAR	FRR
R,G & B	76.7	22.4	0.90
R-G & R-B	81.8	17.3	0.96
R-G,R-B& LDP	83.3	15.6	1.0
R-G,R-B & DBN	82.9	16.1	1.0
R-G,R-B & DIN	83.0	16.0	0.97
R,G,B & LDP	77.5	21.5	0.90
R,G,B & DBN	80.7	18.5	0.86
R,G,B & DIN	82.1	17.1	0.80
R,G,B,R-G,R-B & LDP	76.8	22.4	0.90
R,G,B,R-G,R-B & DBN	68.9	20.3	0.81
R,G,B,R-G,R-B & DIN	81.4	17.8	0.81

Some notes on the results presented in Table IV:

1. R-G, R-B & LDP produced the best performance rate which is equal to 83.3%. Figure 4 shows the ROC curve for this trail which indicates how appropriate the result is.
2. Using R-G, R-B alone gives better result than using R,G,B components. This is because as it was mentioned before that there is intersection between skin and non-skin pixels which produces decision confusion.
3. Even though LDP with R-G and R-B give the best result, using DIN with any other combination always produces result close to the best possible one. It can be inferred that DIN represent the best texture descriptors among other descriptor in this research.

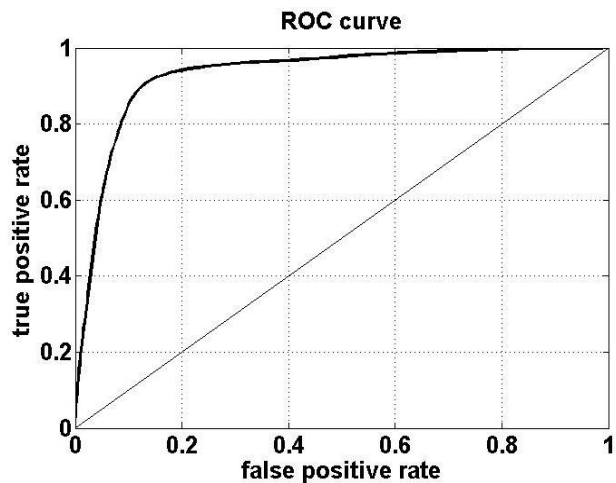


Figure 4. ROC curve for R-G, R-B & LDP combination.

The result reported in this research is consistent with the findings of other similar research studies. For example, Chelsia *et al.* reported a CDR of 83.98%. However, different color space was used [8]. Hamideh *et al.* showed CDR result of 84.5% when merging NN and genetic algorithm [10].

IV. CONCLUSIONS

It has been shown throughout this paper that the maximum performance detection rate of skin pixels cannot go much beyond 83.3%. This is the case when RGB color space is used. We proved the result by using R-G, R-B and some skin texture measure as NN feature vector. We believe that using other methods will not improve the result much. This is due to pixel values distribution of skin and non-skin RGB components.

REFERENCES

- [1] V. Vezhnevets, V. Sazonov and A. Andreeva, "A Survey on Pixel-Based Color Detection Techniques", Proceedings of the GraphicCon, 2003, pp. 85-92.
- [2] M. C. Shin, K. I. Chang and L.V. Tsap, "Does Color Space Transformation Make Any Difference on Skin Detection," Proceedings of the 6th IEEE Workshop on Application of Computer Vision, Dec. 3-4 2002, IEEE Computer Society, Orlando, FL, pp. 275-279.

- [3] S. Alshehri, "Neural Networks Performance for Skin Detection," *Journal of Emerging Trends in computing and Information Sciences*, vol. 3, no. 12, 2012, pp.1582-1585.
- [4] K. K. Bhoyar and O. G. Kakde, "Skin Color Detection Model Using Neural Networks and its Performance Evaluation", *Journal of Computer Science*, vol. 6, 2010, pp. 963-968.
- [5] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection", *International Journal of Computer Vision*, vol. 46, 2002, pp. 81-96.
- [6] S. Alshehri, "Skin Detection Using Simple Arithmetic Operations," *Journal of Computer Technology and Application*, vol. 3, 2012, pp. 797-800.
- [7] T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, 2006, pp. 2037 – 2041.
- [8] C. Doukim, J. Darpham, A. Checkima and S. Omatu, "Combining Neural Networks For skin Detection", *Signal & Image Processing: An International Journal*, vol.1, no.2, 2010, pp. 1-11.
- [9] C. M. Bishop, *Neural Networks for Pattern Recognition: Oxford University Press*, 1997.
- [10] H. Zolfaghari, A. Nekonam and J. Haddadnia, "Color-Base Skin Detection using Hybrid Neural Network & Genetic Algorithm for Real Times," *International Journal of Computer Science and Information Security*, vol. 9, No. 10, 2011, pp.67-71.

Location-based Solar Energy Potential Prediction Algorithm for Mountainous Rural Landscapes

Onabajó Olawale Olusegun,

Faculty of Computer Science & Information Technology,
Universiti Malaysia Sarawak, 94300 Kota Samarahan, Malaysia

Chong Eng Tan,

Faculty of Computer Science & Information Technology,
Universiti Malaysia Sarawak, 94300 Kota Samarahan,

Abstract. -- *The world is facing critical energy crisis today. As a result the conventional grid energy supplies are not enough to meet the present demand. Many advance researches are in progress to overcome this energy predicament. Power generation and management in disconnected rural villages is challenging. The situation is even more challenging when landscape structure in such environment are irregular. Forces of diffusion, ground reflectance and sky view factor among others, affect the quality of final solar radiation incident on a solar panel. This paper describes the implementation of an algorithm that can be used to predict solar energy potential of irregular landscapes. Location-based Solar Energy Potential Prediction Algorithm (LOSEPPA) takes as input, the geographic latitude and longitude of the location of interest to compute the Solar Irradiance Factor (SIF). Geographic latitude plays an important role in the availability of sufficient solar radiation as well as the state of the atmosphere. Therefore, SIF value serves as a guide to the state of the atmosphere in terms of degree of cloud cover, temperature, humidity and landscape structure; which determines the feasibility of the solar energy implementation. The approach described in this paper can be used for rapidly computing the amount of solar radiation generated on a mountainous landscape surface and in the atmosphere as a function of height parameters. With SIF value known, solar panel can be mounted along specific angle of inclination to the sun. The algorithm design covers one year period and is based on the Digital Elevation Model (DEM) of the location under investigation. The proposed system was simulated using MATLAB¹.*

Result show that the more irregular the landscape is, the lower the solar irradiance factor. SIF value of 400 and above predicts well enough sunshine for solar PV implementation in mountainous landscapes. Sample results show that solar radiation per kernel per day for a given landscape is highest between 12noon and 2.00PM local time; and the radiation per kernel per year for a given landscape have highest sunshine hours in January and December.

Keywords-Geographic latitude, Diffusion, Solar Panel, Landscape, DEM

SYMBOLS AND ABBREVIATIONS

α - Geographic Latitude

\in - Angle of rotation of the XYZ coordinates

δ - Solar declination angle

β - Angle between the earth's axis and the XYZ coordinate Z axis

This work is supported partly by Centre for Graduate Studies Universiti Malaysia Sarawak (UNIMAS).

t_0, t_1 – Time of rotation of the XYZ coordinate from t_0 to t_1

\mathbf{n} – Vector normal to the grid cell surface

S_0 – Solar vector at noon local time

S_v – Solar vector at time t

S_C – Solar Constant (1367 Wm^{-2})

Kernel – 3X3 Grid Window

SKV – Sky View Factor

DEM – Digital Elevation Model

SIF – Solar Irradiance Factor

FCN - Four Closest Neighbors

MAPE – Mean Absolute Percentage Error

ANN – Artificial Neural Network

MBE – Mean Bias Error

RMSE – Root Mean Square Error

I. INTRODUCTION

This paper is the first of two papers describing the implementation of intelligent algorithms useful in solar energy potential prediction, generation and management in topographically challenging rural areas. In response to the growing concern over the use of fossil fuels, renewable energy industries are becoming significant economic drivers in different parts of the world. Disconnected rural communities are cut off from government economic transformation agenda as a result of not being connected to the national grid. Many remote residences, businesses and communities located in the sparsely populated and rugged terrains; faces serious challenge in accessing uninterrupted wireless broadband as a result of intermittent electricity supply. An alternative energy supply system in the form of solar electricity, supported by indigenous communities has been widely accepted as a provisional escape route for the rural folks from abject poverty caused by digital divide. The stand-alone photo-voltaic energy system is a well tested energy alternative in an environment where grid electricity is completely absent. However, in mountainous areas, amount of solar radiation obtained on a landscape surface is impacted by numerous environmental factors such as cloud cover, humidity, zenith angle of the sun, diffusion, ground reflectance, air-mass ratio, sky view factor (SVF) and the general albedo of the land surface among others; that must be taken into consideration in order to get the net solar radiation incidence on the solar panel. Terrain parameters derived from DEMs, such as slope, aspect and

cell surface area can be represented as vectors normal to the surface and used in conjunction with the minimum unit area of the DEM that is enclosed between data points to get the net solar radiation.

II. BACKGROUND

The main concern in extending grid electricity to rural areas is installation and costs. The term rural areas connote underserved region. Underserved communities in most part of the world are communities that live in the remote areas of countries with no electricity, water, essential public services, or a shortage of such services. Governments of all nations around the world endeavor to reduce the effect of digital divide on their citizens living in rural areas so as to bring administrative activities closer to the people. Recent government-driven initiatives include commissioning of stand-alone solar photovoltaic sites for use of the rural people. Implementation of solar energy project is actually not a new idea. But most installations do not last due to a number of issues; namely viability and initial investment cost of such projects. Lack of technical specialists, cost of maintenance and access to spare parts are additional concerns that cause the system to fail in generating power consistently. Furthermore, existing solar project during the low peak period, part of the power produced is wasted. Effective and intelligent management of the power generated remains a challenge, especially in location where environmental landscape greatly affect what is produced. Numerous past approaches used in the estimation of solar radiation, considered horizontal surfaces where the effects of irregular landscapes are not taken into account; rather a general equation is set up based on relative duration of sunshine for a number of locations. The equation obtained is then used with some site-specific variables to estimate solar radiation [1]. While most of these approaches produces practical results, the objective of setting them up are specific. However, current research initiative is moving towards general-purpose algorithms that could find useful application in different environment.

Focus of this paper is on the development of a predictive algorithm that could serve as a guide to the solar radiation potential of any given mountainous landscape, taking into consideration environmental parameters such as geographic latitude and longitude of the location, zenith angle of the sun, diffusion, sky view factor and the general albedo of the land surface. The motivation for this research is to develop a functional procedure that could be used to perform routine check on a given mountainous landscape and estimate the viability of solar projects before investment decisions are taken. For instance, the number of solar panels required and the angle of placement can be obtained based on the value of SIF. Location-based Solar Energy Potential Prediction Algorithm (LOSEPPA) takes as input the geographic latitude and longitude of the location, and computes the solar irradiance factor of the landscape through gradient and aspect estimation on per kernel basis. Several methods were used in the literature for calculating slope and aspect from gridded Digital Elevation Models (DEMs) [2]. One of such techniques is the Four Closest Neighbor (FCN) which is used in the design of the proposed algorithm. The solar irradiance factor is a reflection of the solar potential of the landscape.

III. RELATED WORK

Inhabitants of rural areas where public utilities for most of the time are not available, often make use of generators as source of electricity to power appliances. Lately, solar project implementation is taking over the conventional diesel generators. The solar PV systems have advantages as sources of small amounts of electrical power in remote areas that could last longer compared to the diesel generator. The work of [3] estimated the monthly average daily global solar radiation, H of different empirical models based on Angstrom-PreScott model using only the sunshine duration hours. The hourly solar radiation data measured at Kuala Terengganu station during the period (2004–2007) were used to calculate the monthly mean values of H using selected models. The selected models were compared on the basis of the statistical error tests such as the mean bias error, the mean percentage error, the root mean square error, Nash–Sutcliffe equation, correlation coefficient and the t -test. From the statistical results obtained, a new linear model $H / H_0 = 0.2207 + 0.5249(n / N)$, based on modified Angstrom model were recommended for the estimation of monthly average daily global solar radiation for Terengganu, a state in Malaysia; and other places with similar climatic conditions where radiation data is missing or unavailable. This work evaluated various models for the estimation of monthly average daily global radiation on a horizontal surface from bright sunshine hours and to select the most appropriate model for Terengganu state. Statistical error tests such as MBE, RMSE, MPE and coefficient of correlation, r were used to test the linear relationship between predicted and measured values. Results show that models 1 to 9 of the 10 selected, responded well to the statistical error tests and can be used for daily solar radiation measurements for Kuala Terengganu.

A neural network based algorithm for estimating solar radiation on a flat surface was presented by [4]. Least squares support vector machines (LS-SVM) was used to develop the global solar radiation model using the conventional meteorological data, which is then mapped to the global solar radiation resources in China. LS-SVM is a variant of SVM, which employs least squares error method in the training error function [5]. Input to LS-SVM includes latitude of the observed stations and other environmental parameters. To provide better convergence and accuracy of the learning process, all input and output were normalized within the range [0, 1] according to equation (1):

$$V_n = \frac{V - V_{\min}}{V_{\max} - V_{\min}} \quad (1)$$

where V_{\min} and V_{\max} are the minimum and maximum domain values of the input or output value V , and V_n is the normalized equivalent. According to the author, artificial neural networks (ANN) have been applied in solar radiation prediction in earlier researches; however, ANNs was found to be unstable predictors due to the local minima errors, and overfitting problems. Because of these shortcomings, support vector machine (SVM) have actually replace ANN. Comparison between the two shows that SVM has more advantages on forecast because it is based on the statistical learning theory and structural risk minimization, which can get the best solution of entire data set and better ability of generalization. The current work even though is a good contribution, is limited to location where required data for investigation is accessible. In most rural areas, meteorological data are not readily available.

Another solar radiation prediction algorithm was designed by [6]. This paper presents a solar energy prediction method using artificial neural networks (ANNs). An ANN was used to predict clearness index which is applied in the calculation of global and diffuse solar irradiation. The ANN model is based on the feed forward multi-layer perception model with four inputs and one output. The inputs are latitude, longitude, day number, and sunshine ratio; the output is the clearness index. Data from 28 weather stations were used in this research, and 23 stations were used to train the network, while 5 stations were used to test the network. In addition, the measured solar irradiations from the sites were used to derive an equation to calculate the diffused solar irradiation, a function of the global solar irradiation and the clearness index. The proposed equation reduces the mean absolute percentage error (MAPE) in estimating the diffused solar irradiation compared with the conventional equation. Twenty eight (28) weather stations data were used overall; 23 stations data were used to train the network, and 5 sites were used to test it. Based on the results, the average MAPE, mean bias error (MBE) and root mean square error (RMSE) for the predicted global solar irradiation were discovered to be 5.92%, 1.46%, and 7.96%. MAPE estimate of the diffused solar irradiation was found to be 9.8%. A comparison with previous work was done, and the proposed approach was found to be more efficient and accurate than previous methods.

These previous work have good results. However, current research has varied differences from all of them based on the following contributions: (a) use of algebraic approach instead of the usual artificial neural networks which was found to be unstable predictors due to local minima errors and overfitting problems as

mentioned in [4]; (b) all previous work considered flat and regular landscapes whereas LOSEPPA is designed for irregular landscapes; (c) Another unique difference from previous work is that DEM of the location is part of the input parameters to the simulation; (d) In most remote rural areas, meteorological data are not readily available; (e) Most algorithms read in the literature were designed for specific location; but today a general-purpose algorithm that could take local input parameters from any mountainous landscape to predict solar radiation behavior, is much desired.

IV. THE PROPOSED LOSEPPA ALGORITHM

Rural landscapes with spontaneous mountains and hills require additional input parameters such as DEM, nature of the landscape (direct effect on radiation scattering), sky view factor and prevailing cloud conditions when implementing solar energy projects. Irregular landscapes have slopes and aspects that provide additional information on the nature of the surface. There are several methods for calculating slope and aspect from gridded Digital Elevation Models (DEMs). Generally, their determination is based on neighborhood estimation where calculations are made for a cell based on the values of the cells that are spatially adjacent in the grid [7]. One of such methods is the Four Closest Neighbor (FCN) technique which is used in this paper. FCN uses the four cardinal neighbors; those to the north, south, east and west, to estimate slope and aspect for an irregular surface. The elevations at these four closest neighbors are used to define two orthogonal components of slope, the slope in x and y, which define the steepness and downhill direction at the point of interest [8].

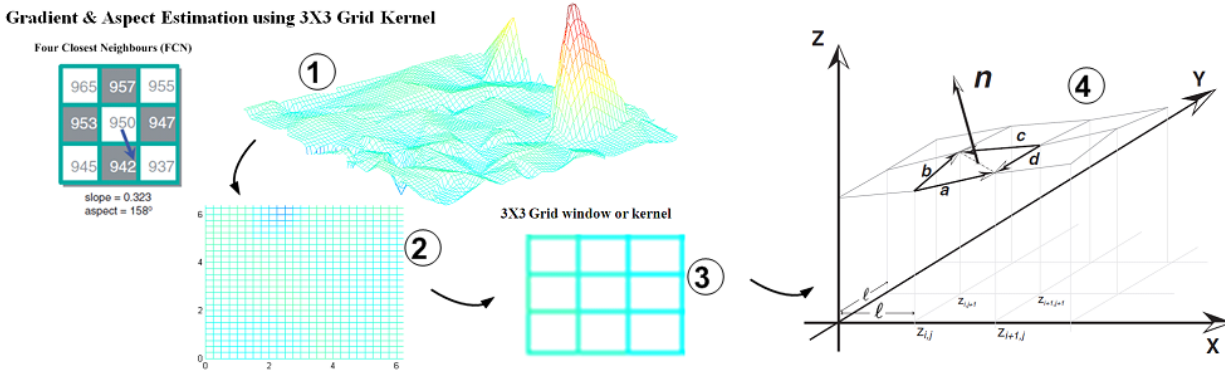


Figure 1: LOSEPPA Algorithm Implementation

A. Landscape Modeling

DEM is a digital representation of a portion of the earth surface, or any planet's surface, derived from elevation measurements at regularly spaced horizontal sampling intervals (Figure 1(1)). For effective LOSEPPA algorithm execution, DEM of the landscape is required in the simulation. The landscape model was produced from a meshed three-dimensional surface using mathematical function of the form:

$$f(z) = f\sin((f(y^2 + x^2)^{0.5}) + g\cos(y)) + h\tan(2x) + 2\cos(h\tan(2x)) \quad (2)$$

where f , g , and h are appropriate constants. Depending on the pattern required for the landscape, mountains and valley representation can be modified by appropriately changing the constant values in the function. Landscape area of size 15.4 square

kilometers were used in the simulation by generating 200 to 500 random points that are locations representing mountains, valleys or level ground. One cell is approximately 200 sq meters. Therefore one kernel is 600 sq meters. A plain ground has a height value (z) of zero.

B. Inclined Surfaces Vector Representation

Digital elevation models can be used to calculate local environmental parameters, such as terrain gradient magnitude (slope) and direction (aspect), which affect mainly the direct radiation component. Gradient and aspect estimation is done based on 3 by 3 grid kernel (Figure 1(3)). The effect of surrounding topography on solar radiation can be modeled by calculating the

local horizon angle for a given solar azimuth [9]. We consider the smallest surface unit in a regularly gridded DEM (Figure 1 (4)) as the plane enclosed by four data points: $z_{i,j}$, $z_{i+1,j}$, $z_{i,j+1}$, $z_{i+1,j+1}$, where $z_{i,j}$ is the elevation of a point at row i , column j . We may say it is unreliable interpolating values from the interior of a cell on a mountainous surface characterized by non-uniform change. However, as a plane is defined by three points in space (x, y, z) , then the four points at the corners of the grid cell may define more than one single plane. A good approximation is to find an average between the two triangles formed at both sides of the cell diagonal. It can be shown that the result is the same whichever diagonal we use. A vector normal to these surfaces is defined by half the sum of cross products of vectors along the sides of the grid cell, defined by equation (3).

$$\begin{cases} a = (l, 0, \Delta z_a), \text{ with } \Delta z_a = z_{i+1, j} - z_{i, j} \\ b = (0, l, \Delta z_b), \text{ with } \Delta z_b = z_{i, j+1} - z_{i, j} \\ c = (-l, 0, \Delta z_c), \text{ with } \Delta z_c = z_{i, j+1} + z_{i+1, j+1} \\ d = (0, -l, \Delta z_d), \text{ with } \Delta z_d = z_{i+1, j} - z_{i+1, j+1} \end{cases} \quad (3)$$

The vector normal to the grid cell surface \mathbf{n} will therefore be:

$$\mathbf{n} = \frac{(a \times b) + (c \times d)}{2} = \frac{1}{2} \begin{vmatrix} i & j & k \\ l & 0 & \Delta z_a \\ 0 & l & \Delta z_b \end{vmatrix} + \frac{1}{2} \begin{vmatrix} i & j & k \\ -l & 0 & \Delta z_c \\ 0 & -l & \Delta z_d \end{vmatrix} \quad (4)$$

The surface area of the cell under consideration will be $|\mathbf{n}|$, the length of vector \mathbf{n} . If equation (4) is simplified further, and expressed in terms of the natural coordinates (x, y, z) it gives the components of the vector normal to the surface in relation to grid elevation points and cell spacing, from which the slope and aspect can be calculated using Four Closest Neighbor technique. FCN is a method of slope and aspect estimation that uses the four cardinal neighbors, those to the north, south, east and west, representing a second order finite difference relationship [10]. The slope and aspect are calculated from equations 5 and 6 as follows:

$$\text{Slope at each grid kernel} = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} \quad (5)$$

$$\text{Aspect} = \tan^{-1}\left(\frac{\partial z / \partial x}{\partial z / \partial y}\right) \quad (6)$$

C. Sun Position Vector Representation

To understand the modeling of the position of the sun, and be able to follow the conventions used in this paper, as obtained in other solar radiation studies such as the work of [2, 4, 7, 9 and 11], the coordinate axes have to be defined sufficiently in relation to the environment under consideration. The coordinate axes are defined as follows:

- The X axis is tangential to the earth surface in the direction East-West, and positive eastward.
- The Y axis is tangential in the direction of North-South and positive southward; and
- The Z axis lies along the earth's radius and it's positive upward.

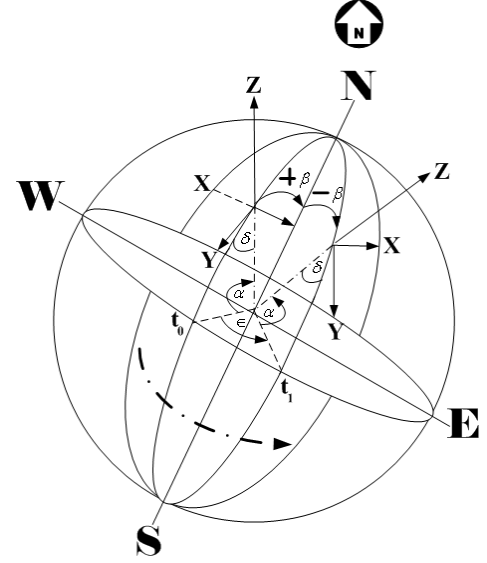


Figure 2: Rotation of the topocentric coordinate system xyz at an angle ϵ from time t_0 to t_1

By definition, the sun lies on the ZY plane (vertical plane) at noon local time. At this time, the x-coordinate of a unit vector pointing to the sun (solar vector) will be null. The solar declination (δ) is the angle between the solar rays and the plane of the earth's equator. The geographical latitude α is the angle between the radius of the earth at the observer position (that is the z-axis) and the equatorial plane. Thus, the solar vector, at noon local time will be:

$$S_o = (0, \sin(\alpha - \delta), \cos(\alpha - \delta)) \quad (7)$$

At any given time t , the earth will have rotated away from noon an hour angle ϵ and angular speed of 2π radians or 360° per day. The hour angle is the angle between the observer meridian and the solar meridian, the convention follows that at noon it is zero and positive before noon [11]. At this time the topocentric coordinate system will have changed position in relationship to the sun at noon. This movement can be decomposed in to three rotations (Figure 2): one around the X-axis, to place the Z-axis parallel to the axis of rotation of the earth; a second rotation around the Z-axis at an angle ϵ and a third rotation back around the X-axis to the observer position. To find the coordinates of the solar vector in the new, rotated reference system we multiply the original coordinates by 3 rotational matrices describing these movements. Therefore, at any time, and assuming no atmospheric refraction occurs, the solar vector S_v will be:

$$S_v = r_x(\beta)r_z(\epsilon)r_x(-\beta)S_o \quad (8)$$

where \mathbf{r} is a rotation matrix around axis in subscript and angle in brackets, β is the angle between earth's axis and the topocentric coordinate system Z-axis, and the hour angle ϵ is zero at noon and has the following value in radians at any time t given in hours and decimal fraction:

$$\epsilon = \pi \left(\frac{t}{12} - 1 \right) \quad (9)$$

D. Sky View Factor

The sky view factor (SVF) is a dimensionless parameter denoting the quantity of visible sky at certain location. It is a measure of the openness of a surface: SVF of 1 means an unobstructed view of the sky and SVF of 0 means a completely obstructed view of the sky. The sky view factor at ground level has been shown to be related to environmental phenomena such as solar radiation characteristics, heat dissipation effect, air pollution and surface energy budget. [9, 10] The average sky view factor at ground level has been computed using DEM modeling.

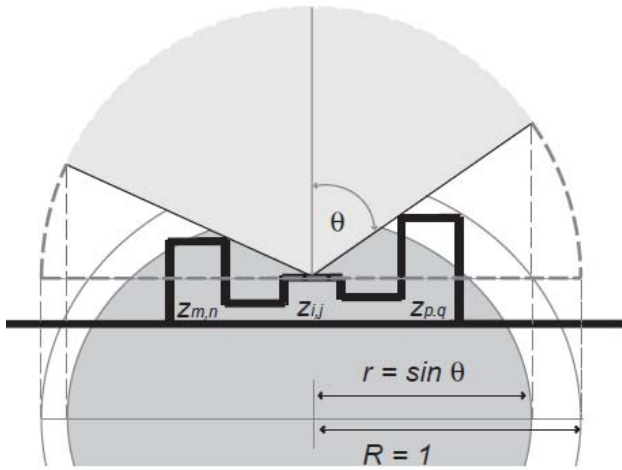


Figure 3: Effect of sky view factor on solar radiation

DEM data about a given landscape serves as database for the quantification of relevant realistic climatic conditions in mountainous and complex areas. The estimation of the sky view factor is based on the knowledge of each of the angle elements of the landscape environment and of the associated elevation angle θ that produces shadow (Figure 3) and azimuth angle α . \mathbf{r} is the length of the adjacent side of the right angle triangle formed with angle θ using height $Z_{i,j}$ as angle in focus. Each cell height $Z_{m,n}$; $Z_{i,j}$ and $Z_{p,q}$ are taking into consideration in the estimation of the sky view factor. Accordingly, the sky view factor Ψ_s can be assumed to be the sum of all this angle information over the entire landscape (eqn 10).

$$\psi_s = \sum_{i=1}^n \sin^2 \theta_i \left(\frac{r_i}{360} \right) \quad (10)$$

Sample estimation for a terrain with slope up to 45° and at 15° azimuthal intervals will require $45 \times 24 \times N$ operations if N is the number of cells in the DEM.

E. Hill Shade Effect

The sun is considered as a point light source at some infinite distance away, and therefore, all illumination rays arriving at a grid cell surface can be considered parallel. For estimation convenience we consider an illumination plane perpendicular to the solar rays (Figure 4). All solar rays pass through this plane at a right angle. By checking the projection of a grid cell over this plane, following the direction of the sun, we can determine whether a point is in the sun or in the shade of another cell. In Figure 4 this is illustrated with a two dimensional example: the projection of P_1 , that is, P'_1 , has a value higher than any previous point (since it is the first point to be scanned), so it is in the sun. This is the same for P'_2 and P'_3 , however, P'_4 has a lower value than P'_3 and therefore is in its shadow.

Therefore, the projection of a point P_i on the solar plane SP is the dot product of the vector $\overrightarrow{OP_i}$ and the unit vector S_p , which is a vector in the direction of the plane SP and perpendicular to the solar vector \overrightarrow{S} . A cell will shade itself if the angle between the sun and the vector normal to the cell's surface is higher

than $\frac{\pi}{2}$. The vector from the origin to any point $P_{i,j}$ will then be:

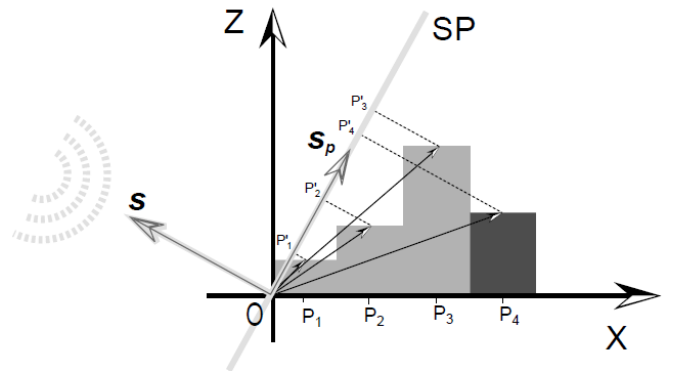
$$\overrightarrow{OP_{i,j}} = (l_i, l_j, Z_{i,j}) \quad (11)$$


Figure 4: Effect of Hill shade on irradiance

V. RESULTS AND DISCUSSION

In this work SIF for various landscapes were determined in order to identify their insolation potential. Result show that the more irregular a landscape is, the lower the SIF value. SIF value of 400 and above predicts excellent sunshine hours for solar energy project implementation. Scattering at the surface of the landscape are possibly responsible for this observation. SIF is defined according to equation (12).

$$SIF = \frac{\text{irradiance_value}}{S_c} \quad (12)$$

where *Irradiance_value* is a measure of the rate of energy received per unit area, and has units of watts per square metre (W/m^2). *Sc* is the solar constant, defined by F.S Johnson [12] as the amount of incoming solar electromagnetic radiation per unit area that would be incident on a plane perpendicular to the rays, at a distance of one astronomical unit (AU). Its unit is also in watts per square metre (W/m^2). Thus SIF is a dimensionless quantity which characterizes a given landscape. Sample landscape structures used in the determination of SIF are shown in Figures 5 and 6. An intense hill landscape such as Figure 5 has more height parameters which affect the insolation capacity of the landscape due to environmental forces of scattering, diffusion, ground reflectance, air-mass ratio, sky view factor etc. From results obtained, Figure 6, a moderate hill landscape is more open to the insolation from the Sun, and therefore SIF is higher.

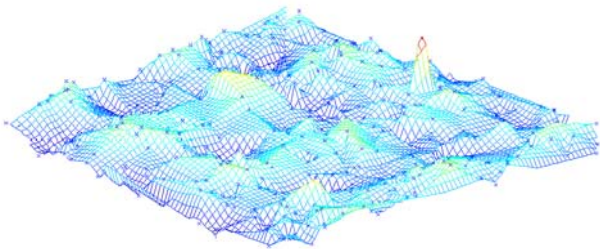


Figure 5: Intense hill landscape

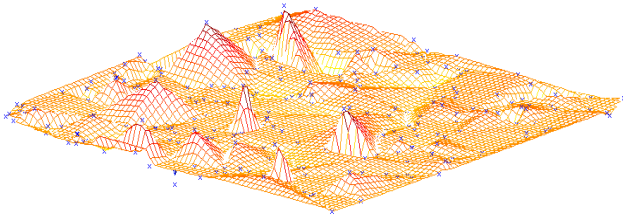


Figure 6: Moderate hill landscape

Simulations carried out on thirty different landscape structures follow the graph pattern in Figure 7. Hourly solar radiation per kernel per day (*hsrpkpd*) \rightarrow Total hours of daily sunshine (*hds*) X Hour angle at sunrise (*hs*) X day length (*dl*) X clear sky solar radiation (*cssr*) X size of kernel (*ks*) X hourly sun declination angle (*hsda*). Estimation of *hsrpkpd* is given by equation (13):

$$hsrpkpd = hds * hs * dl * cssr * ks * hsda \quad (13)$$

Likewise, monthly solar radiation per kernel per year (*msrpkpy*) is given by equation (14):

$$\text{Therefore, } msrpkpy = \frac{hsrpkpd * 365}{12} \quad (14)$$

Hourly solar radiation per kernel per day for a given landscape is shown in Figure 8, and monthly solar radiation per kernel per year for a given landscape is also shown in Figure 9. It is difficult locating previous effort that this work could be compared. All previous work consulted were either based on flat landscape or treated only segments of the radiation parameters. While current work engages the use of DEM in the input parameters, this parameter was not explicitly considered by others because flat landscapes were the basis for the implementation. In this case input parameters to the simulation differ from that of the current work.

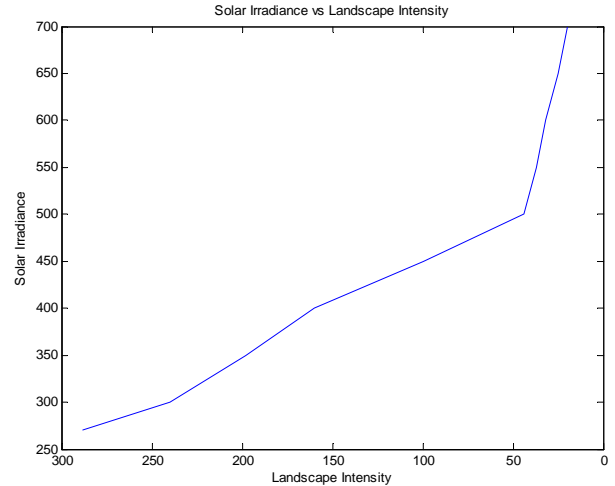


Figure 7: Plot of Solar Irradiance Factor (SIF) against Landscape Intensity

The algorithms described in this work are particularly suitable for surfaces of high and variable relief such as mountainous terrain (Figures 5 and 6). They minimize the area under evaluation through keeping extreme values within minimum smoothing. One of the component algorithms, the algorithm for the computation of the solar position permits straightforward operation with the other terrain parameters, expressed as vectors. The main advantages of LOSEPPA is the facility to operate on vectors, the consistency of vector representation for all the procedures involved, easier conceptual visualization in three dimensions and optimization of array handling capabilities in computer languages, which results in a fast and efficient implementation.

The estimation of SIF as shown in this paper relies principally on the digital elevation model (DEM) of the location under investigation. A gridded representation allows for easy access and manipulation since the elevations can be stored as a simple matrix. While this approach has not been considered in any previous work to the best of our knowledge, various degrees of DEM data are available at United States Geological Survey (USGS) website for some fees, making estimation of SIF less cumbersome. Gridded DEMs as opposed to other terrain model types are normally easy to obtain in digital form for a given area without the need of further manipulation. It is a well known fact that solar radiation/meteorological data are deficient or completely missing for most remote rural regions, which make proper estimation of solar radiation difficult if not impossible. This is not required in the current work. Development of LOSEPPA which uses algebraic approach is a major departure from the usual norm of using artificial neural networks which was found to be an unstable predictor due to local minima errors and overfitting problems as mentioned in [4]. Most algorithms in the literature were designed for flat landscape and specific location which make their application limited. LOSEPPA is a general-purpose algorithm that takes local input parameters from any mountainous landscape to predict solar radiation behavior. A general purpose algorithm like LOSEPPA have a wider application domain compared to those designed for specific location, which means it can be applied irrespective of the location of the mountainous landscape.

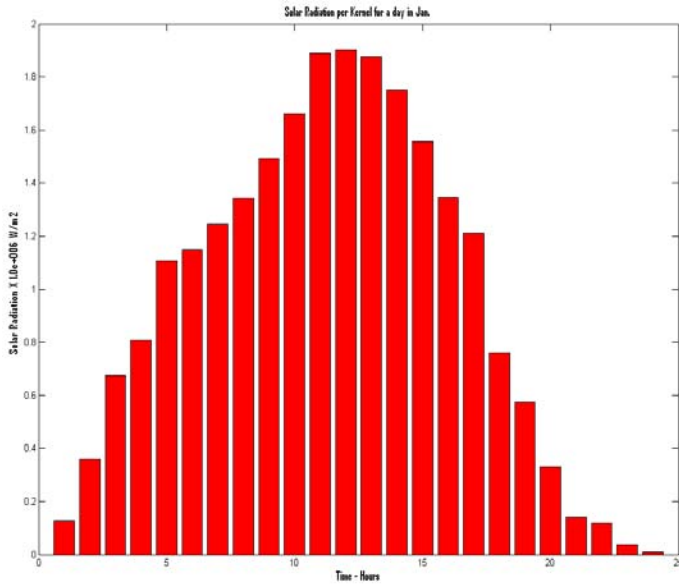


Figure 8: Hourly solar radiation per kernel per day for a given landscape

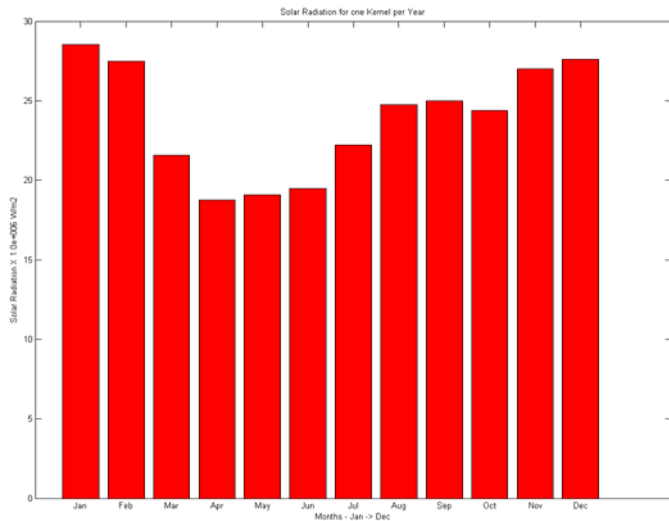


Figure 9: Monthly solar radiation per kernel per year for a given landscape

VI. CONCLUSION

This paper describes the implementation of an algorithm that is useful in the prediction of solar energy potential of irregular landscapes. Significant contribution of this work include: (a) use of algebraic approach instead of the usual artificial neural networks which was found to be unstable predictors due to local minima errors and overfitting problems as mentioned in [4]; (b) use of DEM (height parameters) in the estimation of solar radiation; (c) all previous work considered flat and regular landscapes whereas LOSEPPA is designed for irregular landscapes; (d) In most remote rural areas, topographic/meteorological data are not readily available, therefore use of algebraic approach can be used; (e) Most algorithms read in the literature were designed for specific location; but today a general-purpose algorithm that could take local input

parameters from any mountainous landscape to predict solar radiation behavior, is much desired.

Result show that the more irregular the landscape is, the lower the solar irradiance factor. SIF value of 400 and above predicts well enough sunshine for solar PV implementation in mountainous landscapes. Sample results show that solar radiation per kernel per day for a given landscape is highest between 12Noon and 2.00PM local time; and the radiation per kernel per year for a given landscape have highest sunshine hours in January and December.

VII. REFERENCES

- [1] Gopinathan K.K, "A simple method for predicting global radiation on a horizontal surface", *Solar and Wind Technology, Volume 5, Issue 5, 1988, pages 581-583*, National University of Lesotho, Roma, Lesotho, South Africa,
- [2] Lynn D R, Michael J Collins, "The effect of errors in gridded digital elevation models on the estimation of topographic parameters", Department of Geography, University of Calgary, Canada", April, 2005.
- [3] Norhazwani B.A, Hasril B.H, Adlansyah B.A and Mohd Nasharuddin B.M, "An Improved Solar PV System for Malaysian Rural Electrification – Part-II: Implementation and Economic Analysis", Proceedings of 201 0 IEEE Student Conference on Research and Development (SCORED 2010), 13 - 14 Dec 2010, Putrajaya, Malaysia
- [4] Fangpin D, Gaoli S, Chuang L and Zhengxing W, "Prediction of Solar Radiation Resources in China using the LS-SVM Algorithms ", Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing China, Vol. 5, 2008
- [5] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, "Applying support vector machine to predict hourly cooling load in the building," *Appl. Energy*, vol.86, pp.2249-2256, 2009.
- [6] Tamer Khatib, AzahMohamed, K. Sopian, and M.Mahmoud, "Solar Energy Prediction for Malaysia Using Artificial Neural Networks", *International Journal of Photoenergy*, Hindawi Publishing Corporation, Volume 2012, Article ID 419504, January 2012
- [7] Lynn D. Raaflaub, Michael J. Collins, "The effect of error in gridded digital elevation models on the estimation of topographic parameters", Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada T2N 1N4, 2005
- [8] Guth, P.L., "Slope and aspect calculations on gridded digital elevation models: examples from a geomorphometric toolbox for personal computers", 1995
- [9] Dozier, J., and Frew, J., 1990, Rapid calculation of terrain parameters for radiation modeling from digital elevation data. *IEEE Transactions on Geoscience and Remote Sensing*, 28, 963–969.
- [10] Hoffer, R.M., Fleming, M.D., Bartolucci, L.A., Davis, R.F., Nelson, S.M., 1979. *Digital Processing of Landsat MSS and Topographic data to Improve Capabilities for Computerized mapping of forest cover types*. LARS Technical Report 011579. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana
- [11] Iqbal M, "An Introduction to Solar Radiation", First Edition, 1983, Academic Press, New York, Page 15
- [12] F.S Johnson, "The Solar Constant", *Journal of Meteorology*,

VIII. AUTHORS BIOGRAPHIES

ONABAJO OLAWALE OLUSEGUN was born in Lagos, Nigeria in 1977, and obtained his Bachelor of Computer Science degree from University of Wollongong, Wollongong, Australia in 2005. In 2011, he obtained his Master of Science degree in Computer Science from the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak (UNIMAS). He is currently



with the same faculty for his PhD degree. Mr. Onabajo is a student member of IEEE and fellow of the Institute of Data Processing and Management, London U.K. He was the vice chair for the 2010 edition of UNIMAS Faculty of Computer Science and Information Technology colloquium. Onabajo has authored two published papers and has many scholarship awards to his credit.



TAN CHONG ENG (cetan@ieee.org) received his B.Eng. and M. Eng degrees in Electronic and Telecommunications Engineering from Universiti Malaysia Sarawak (UNIMAS), and Ph.D. degree in

Information Engineering from the University of Cambridge. He is currently an associate professor at the Faculty of Computer Science and Information Technology, UNIMAS. His current areas of research include broadband wireless and affordable rural connectivity initiatives. He is a member of the IEEE and IEM.

Impact of Multipath Routing Performance of Video Traffic

Adel ECHCHAACHOU
SIME laboratory
E.N.S.I.A.S
Rabat, Morocco

Ali CHOUKRI
SIME laboratory
E.N.S.I.A.S
Rabat, Morocco

Ahmed HABBANI
SIME laboratory
E.N.S.I.A.S
Rabat, Morocco

Mohammed ELKOUTBI
SIME laboratory
E.N.S.I.A.S
Rabat, Morocco

Abstract—*Ad-hoc Network is a wireless environment of transmission which offers a very high mobility with low establishment costs. However, in this mode of communication the throughput and delay are limited, especially if traffic is needs high bandwidth, such as streaming video. In this paper, we study and evaluate the performance of video traffic in Ad-hoc network based on reactive routing protocol.*

As a first step, we study and compare the behaviour of AODV and AOMDV to carry streaming video traffic.

Keywords—*Performance, Video, AODV and AOMDV*

I. INTRODUCTION

Ad-hoc network is a wireless infrastructure which is based on mobile nodes communicating via radio in a dynamic and decentralized topology [1].

In Ad-hoc network, data packets are routed from one node to another by a routing system that changes frequently according to the mobility of these nodes. This is ensured by protocols whose operation depends on the way to discover the route and build the routing table.

A proactive routing protocol builds its routing table before demand and tries to define the topology of the network at any time. The best known of these protocols are OLSR and DSDV.

A reactive routing protocol waits that a request is made to start the route discovery and update its routing table. The best known of these protocols are AODV and DSR [2].

The proactive protocols have the advantage of producing less routing load while reactive protocols are characterized by the fact that they offer more reliability and more speed [3, 4].

Two types of reactive protocols are considered:

- A single-path protocol that uses a single way between a source and destination for data transmission, such as AODV.
- A multi path protocol who can use most ways for reception and transmission of data packets

- A multi-path protocol that can be deployed to maintain or load balance traffic, taking several paths between the source and the destination, such as AOMDV.

Numerous studies have been done on the performance comparing between AODV and AOMDV. However, they have not addressed the traffic that requires high throughput and delays greatly reduced like live video. Consumption of bandwidth by video services increases exponentially, which involves the problem of infrastructure that must support this type of communication more greedy in terms of flows.

In this paper, we study and evaluate the impact of multipath routing on video traffic performance. We chose MPEG-4 which provides high-quality of video and supports a large set of multimedia application. As simulator, we selected the NS2 tool that will allow us to see and compare the behavior of AODV and AOMDV to transport an MPEG-4 video traffic.

II. REACTIVE ROUTING PROTOCOL AND VIDEO TRAFIC

A. Reactive routing Protocol

The reactive routing protocols can be classified into two categories according to the number of ways that can be deployed for routing of traffic between a source and a destination. The first category uses a single path while the second category can process a multiple transmission routes. In this project, we study and evaluate the performance of two protocols AODV and AOMDV representing, respectively, the first and second category.

Ad-hoc On-Demand Distance Vector (AODV) is an algorithm that allows dynamic routing between mobile nodes, allowing them to quickly get the roads to new destinations without having to maintain routes to destinations that are not in active communication[5]. Operation of AODV is loop-free, since it uses a destination sequence number for

each entry in the road, thus avoiding the problem of Bellman-Ford "infinite loop" and providing rapid convergence when topology is mobilized.

AODV is a routing protocol designed for mobile ad-hoc networks of small and large population of nodes. It can support high rates of mobility, reduce overhead and the spread of traffic control. AODV defines and uses four messages to manage its routing system: Route Request (RREQ), Route Reply (RREP), Route Reply Acknowledgment (RREP-ACK) and Route Error (RERR). In AODV environment, when a node needs a route to a new destination, it broadcasts a RREQ. When the destination or a node with an active route to the destination receives the RREQ, a RREP message is generated and sent to the source (Fig.1, Fig.2). At the reception, when it is requested, the node sends the message back (RREP-ACK) in acknowledgment the reception of the RREP. When a link is broken making one or more destinations inaccessible to neighboring nodes, a RERR message is broadcast to the nodes for updating their routing tables and repairing broken links.

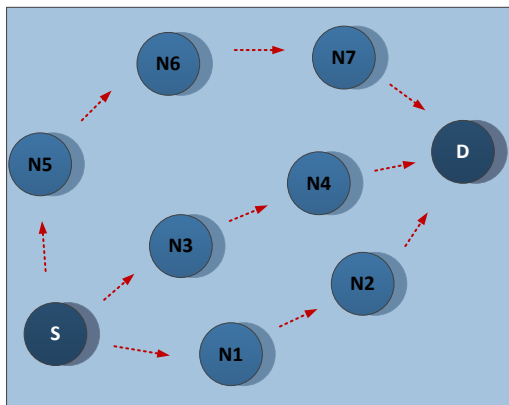


Fig.1. The RREQ Spreading by AODV.

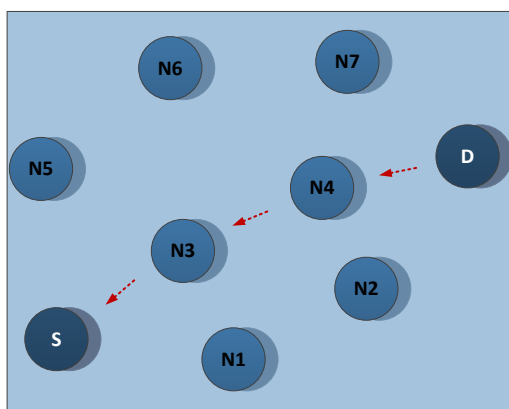


Fig.2. The RREP Spreading by AODV.

AOMDV is an extension of the AODV protocol which aims to discover multiple paths between the source and the destination in each route discovery

[6]. AOMDV is characterized by three innovative features compared to other reactive protocols multi-paths. Firstly, AOMDV generates very low additional costs of coordination. Second, it provides the disjunction of alternative routes. Finally, it computes paths alternating with minimal load.

AOMDV and AODV have several characteristics in common:

- They are both based on the concept of distance vector and use "hop-by-hop" routing approach.
- They find routes on demand using a route discovery procedure.

The main difference between these two reactive protocols is the number of lines found in each route discovery. AOMDV establishes multiple reverse paths to the intermediate and source node. These paths are traversed by several RREPs. In the AODV environment, the reverse path is unique (Fig.3, Fig.4).

The rationale and effectiveness of AOMDV protocol lies in the fact to ensure that multiple paths discovered are disjoint and without loop, and find paths using a route discovery based on flooding. AOMDV use the routing information already available in the underlying protocol AODV, thus limiting the overhead involved in the discovery of several paths. Indeed, AOMDV does not use special control packets. Moreover, the additional packets (RREPs and RERRs) deployed for the maintenance and the discovery of multiple paths as well as some additional fields in the packet routing control, are the only additional overhead compared to AODV.

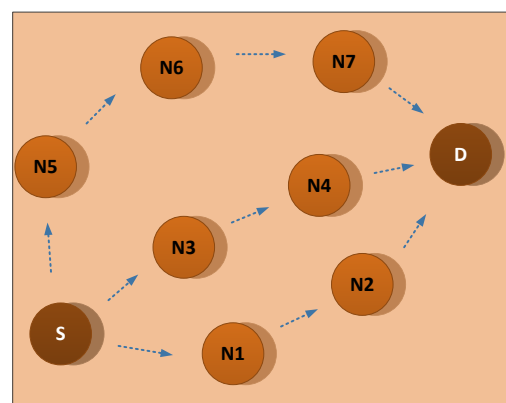


Fig.3. The RREQ Spreading by AOMDV.

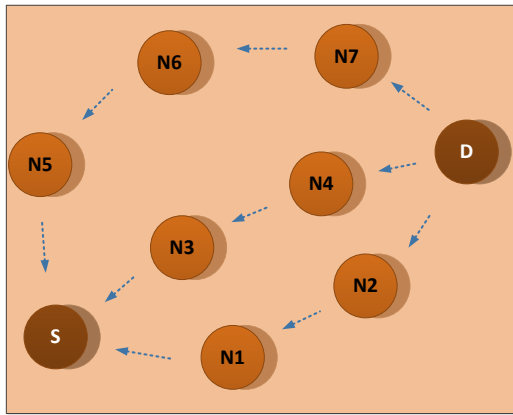


Fig.4. The RREPs Spreading by AOMDV.

B. MPEG-4

MPEG-4 is an ISO/IEC standard developed by Moving Picture Experts Group (MPEG) and standardized at the beginning of 1999 [7]. The MPEG group has developed several standards such as MPEG-1, MPEG-2, MPEG-7 and MPEG-21 (Fig.5).

MPEG-4 was developed to support tree principal's environments:

- The interactive media;
- The digital television;
- The interactive graphical applications.

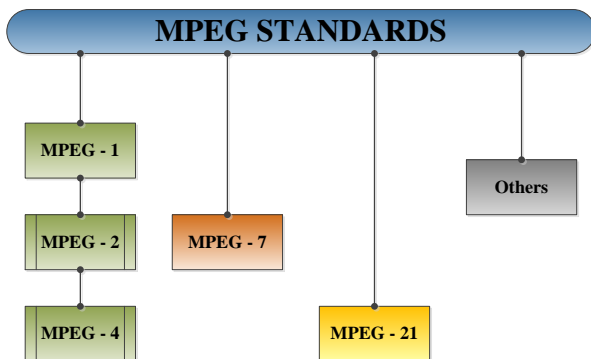


Fig.5. Some of MPEG standards.

MPEG-4 is designed to meet future requirements and interactive multimedia applications such as video on Internet where access to audio and live video. it is based on a coding system and compression sturdy which enables it to offer high quality video without requiring important throughputs. MPEG-4 encoding solution is flexible and appropriate to provide high-quality multimedia traffic on high speed networks. [9, 10].

These are probably the reasons for the great success of this technology. In fact, recent statistics published on 27 June 2012 by Palo Alto Networks, a specialist in computer network security, showed, according to an analysis of the environment of

application traffic of 2036 business between November 2011 and May 2012, as the amount of bandwidth occupied by reading applications live video has tripled in six months to achieve 13% of total consumption (Fig.6).

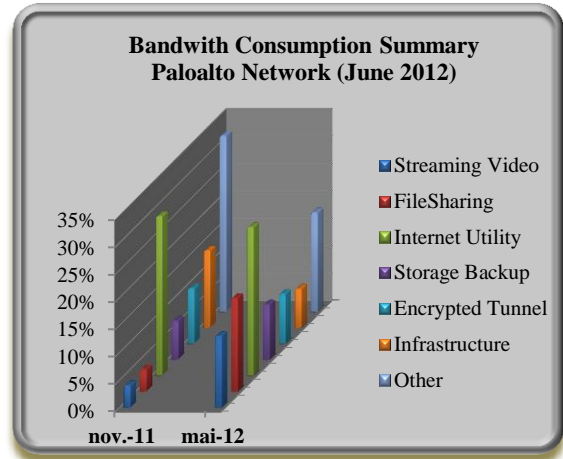


Fig.6. The total bandwidth Consumption of Internet traffic.

C. Related Work

A comparison of the performance of AODV and AOMDV has already been completed and was designed to study the behavior of these two protocols with respect to the tolerance of the network load. AOMDV performs better than AODV, viewpoint latency, packet loss rate and average time of transmission, although it generates more overhead routing [11, 12].

Lookout of speed and rate of times change (Gigue), AOMDV delivers more performance than AODV, and even another proactive protocol DSDV, when the mobility and network size are important [13]. By cons, AODV performs better than the two other protocols if the density of mobile environment is low

In another work, it has been shown that AOMDV is also better than AODV_Multipath and SMR, especially in a denser network with low mobility of nodes [14]. Although the overhead generated by these multi-paths routing protocols are more important than those of the mono-path protocols (such as AODV), network performance does not degrade significantly because the waiting time of data packets are more and more reduced when the size of network increases.

Another study comparing performance between AODV, AOMDV and DSR was conducted to determine the effect of multi-path routing when the network is static. This evaluation showed that AOMDV provides better performance in terms of

throughput and packet loss rate. While DSR is better in terms of the average transmission [15].

On live video traffic in Ad-hoc network, a study showed that the proactive routing protocol OLSR sends the MPEG-4 packets better than AODV and DSR when the network topology is dynamic or when the load traffic is important. AODV is more appropriate in a highly mobile environment due to its high ability to maintain routing packets when links break [16].

To manage the network load in the case of heavy traffic, such as MPEG-4, a new method has been proposed consisting in improving the AODV protocol to adapt to network congestion and adopting a concept alternating paths to avoid encumbered routs [17].

III. CONTRIBUTION

A. Simulation Environment

In order to calculate and measure the effect of multipath routing on video traffic performance we deployed the network simulator NS-2 version 2.34 for achieve our simulations [18]. We installed the AOMDV protocol and MPEG-4 codecs since they are not natively installed in this simulator. Subsequently, for study and analyze the real and major impact of multipath routing on video streaming traffic we used "Random Waypoint" [19] as a model of mobility simulation that will generate a very high mobility . As well, we fixed pause time of nodes in zero.

In our virtual model, nodes use omnidirectional antennas and move in an area of 1000m x 1000m with a variable density (10 to 100) and a constant speed of 30 m/s. We use CBR to transmit packets of 512 bytes [20].

B. Parameters of simulation

The main parameters of our simulation are shown in the following table (Tab.1):

Simulator	NS2
Routing Protocols	AODV and AOMDV
Mobility Model	Random Waypoint
Simulation Time (sec)	900
Pause Time (sec)	0
Simulation Area (m)	1000 x 1000
Number of Nodes	50, 60, 70, 80, 90, 100
Transmission Range	250 m
Speed (m/s)	30

Tab. 1. Parameters of simulation.

We studied and compared the performance of AODV and AOMDV in two modes: before and after generating an MPEG-4 traffic. For this we used four metrics:

- PDR (Packet Delivery Ratio),
- Delay,
- Overhead,
- Throughput.

C. Results and analysis

❖ Before generating an MPEG-4 traffic:

a) First metric : DPR

To determine and compare the quality of service provided by AODV and AOMDV, we calculated rates of delivery of packets over the network density. The results are displayed in Figure 7 and indicate that the rate of AODV are almost static at 80% when the network contains less than 40 nodes and begin, after that, to fall sharply and steadily to reach 10% at 100 nodes. The side of AOMDV, we see that these values are almost stationary and identical to those of the AODV for low densities. Thereafter they decreases from 75% to 40%, and take the form of a half-parabola in the range 50-100 nodes. These results are consistent. Indeed, the number of broken links and corrupted packets increases as the network becomes denser. This loss is more severe for AODV than AOMDV who can deploy other paths when routing packets transmission paths are damaged.

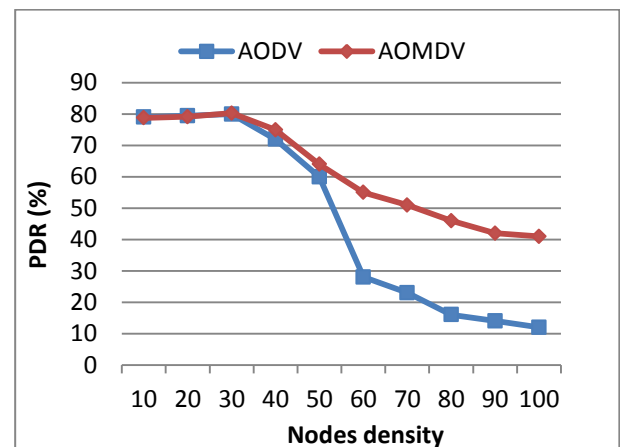


Fig.7. Calculation of PDR compared to the density of nodes.

b) Second metric : Delay

Figure 8 represents the transmission delays of packets by both protocols AODV and AOMDV compared to the density of nodes. AOMDV delays are almost static and varying slightly between 5 and 50ms on the whole interval of the nodes. AODV delays vary between 5ms and 2s. They are

approximately equal to 5ms when the number of nodes is less than 40. Then they increased to 220ms at 50 nodes before increasing to 1s at 60 nodes, beyond which they take the form of an exponential until reaching 2s at 100 nodes. This is explained by the fact that AODV is able, through his multi-path system, to maintain the same pace routing when the network becomes denser, while AODV, which must rediscover the network whenever a path is lost, generates further delay in transmissions gradually as the number of nodes increases.

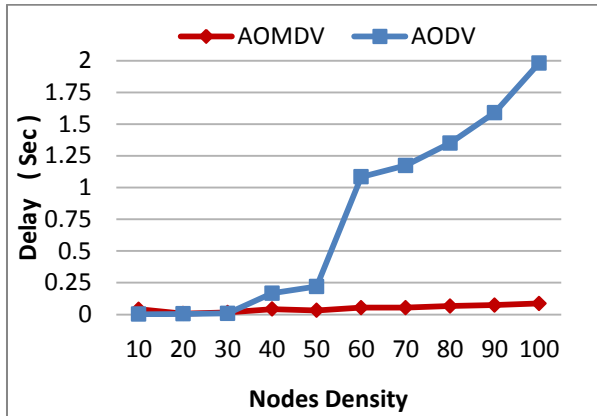


Fig.8. Calculation of delays compared to the density of nodes.

c) Third metric : Overhead

Figure 9 illustrate the overhead generated by AODV and AOMDV depending on the density of nodes. We note that the values of AODV are very negligible and static when the number of nodes is less than 30, but they grow rapidly and almost linearly thereafter to reach a value of 30 at 100 nodes. AOMDV values are also close to zero at low densities, and take the form of a half-parabola that grows slightly beyond the 40 nodes to reach a value of 13 at 100 nodes. The criterion of multi-path is also very crucial for reduce overhead despite the generation of additional packages (RREPs and RRERs) by AOMDV. In fact, when the network becomes denser, AODV must generate more routing packets to restore broken links, which results in increased the levels of overhead. Regarding AOMDV, and even with the deployment of multiple response messages and errors, his system of multi-path can reduce overhead.

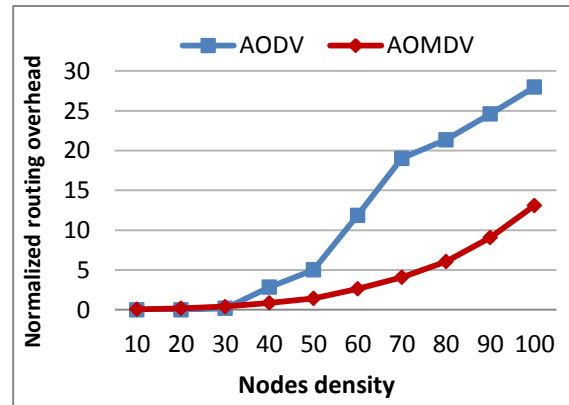


Fig.9. Calculation of overhead compared to the density of nodes.

d) Fourth metric : Throughput

Our fourth criterion of comparison is to calculate the throughputs of AODV and AOMDV relative to the network density. The simulation results are shown in Figure 10. When the number of nodes is less than 30, the values of AODV and AOMDV are quasi-stationary and almost identical. From 40 nodes, the throughputs of AODV are almost linear and vary slightly between 200 and 400 Kbps on the entire range of number of nodes, while the throughputs of AOMDV vary with a significant increase between 150 and 1000 kbps. This indicates that AOMDV transmissions require less time compared to AODV to successfully reach the destination. Indeed, thanks to its system of multi-path, AOMDV forwards packets quickly on another link when the primary path is corrupt, while AODV requires a lot of time to find a new routing path.

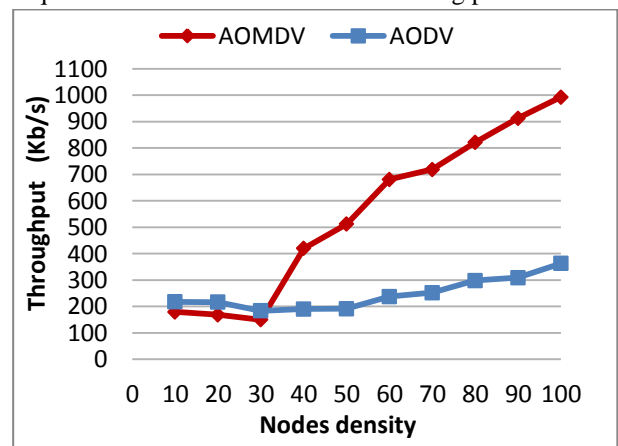


Fig.10. Calculation of throughput compared to the density of nodes.

❖ After generating an MPEG-4 traffic:

a) First metric : DPR

In Figure 11 we present the packet delivery rates of MPEG-4 by AODV and AOMDV compared to the density of nodes. About AODV, rates begin to fall from 40% at 20 nodes until achieve 5% to 50 nodes

beyond which they remain almost static. For cons, the rates of AOMDV decrease less sharply and continually, from 45% at 20 nodes to 9% at 100 nodes. Thanks to its multi-system routing, AOMDV is able to provide greater delivery rate of packets of MPEG-4 relative to AODV, despite their drop proportional to the density of the network.

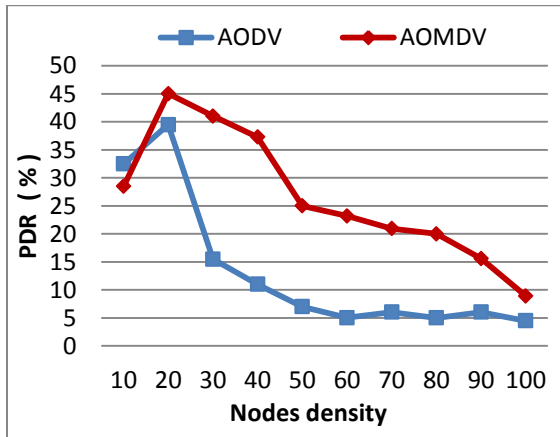


Fig.11. Calculation of PDR of MPEG-4 packets compared to the density of nodes.

b) Second metric : Delay

Figure 12 shows the delays of packets MPEG-4 by AOMDV and AODV, in relation to the density of the network. The AODV values vary between 145ms and 2.5s, while AOMDV provides values between 88ms and 3s. The AODV delays increase between 10 and 20 nodes and grows at 30 nodes. Beyond this value, they wiggle between 2 and 2.5 s. The AOMDV delays increases almost linearly between 10 and 50 nodes up to achieve 2.5 s. Then they decrease to 2s at 80 nodes before increasing again to reach 3s at 100 nodes. We note that the packets transmission delays of MPEG-4 by the two protocols are almost identical in emphasizing that the values of AOMDV are more regular and provide less change of direction sense. This can be explained by the fact that AOMDV offers more routing choice and, therefore, more stability than AODV (despite it is slight) .

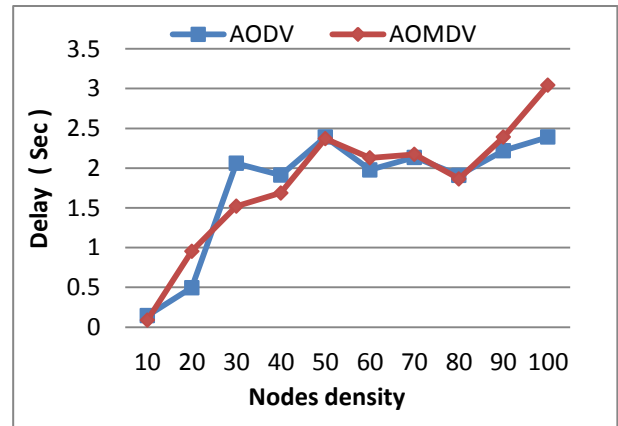


Fig.12. Calculation of delays of MPEG-4 packets compared to the density of nodes.

c) Third metric : Overhead

Figure 13 illustrates the routing overhead generated by AODV and AOMDV when transmitting packets MPEG-4 relative to the network density. We note that the values of AODV grow from 0 to 30, while those of AOMDV increase from 1.5 to 11. We find that in low density, the overheads generated by AODV and AOMDV are almost identical. But when the number of nodes exceeds 40, AODV generates more and more overhead than AOMDV when the density of the network increases. On the other hand, we note that the results are almost identical to those calculated before the packets generation MPEG-4. This indicates that the overhead does not depend on the type of traffic.

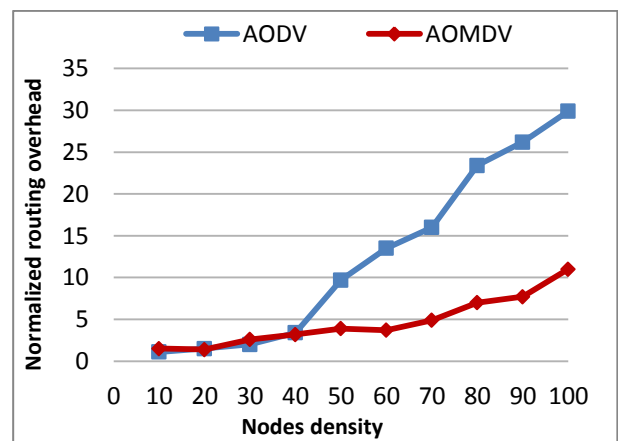


Fig.13. Overhead routing calculation of MPEG-4 packets compared to the density of nodes.

d) Fourth metric : Throughput

The throughputs of AODV and AOMDV measured after packet generation MPEG-4 are extrapolated according to the density of nodes in Figure 14. Between 10 and 20 nodes, the values of AODV increases of 145 to 331 kbps before decreasing

almost linearly and continuously until it reaches 163 kbps at 100 nodes. Regarding AOMDV, the values increases of 127 to 470 kbps between 10 and 20 nodes and then they increases up to 635 kbps to 40 nodes. These throughputs will decrease linearly up to 575 kbps before increasing again between 70 and 80 nodes to reach their maximum: 680kps. In the end, the AOMDV values will decrease up to 370 kbps at 100 nodes. We note that AOMDV provides very high throughputs compared to AODV to transmit MPEG-4 packets. The criterion of the multi-path AOMDV is a key factor for improving the delivery times of packets to destination. However, the MPEG-4 video packet is a particular traffic that can excessively disturb the quality of transmissions, which explains the irregular values recorded by AOMDV.

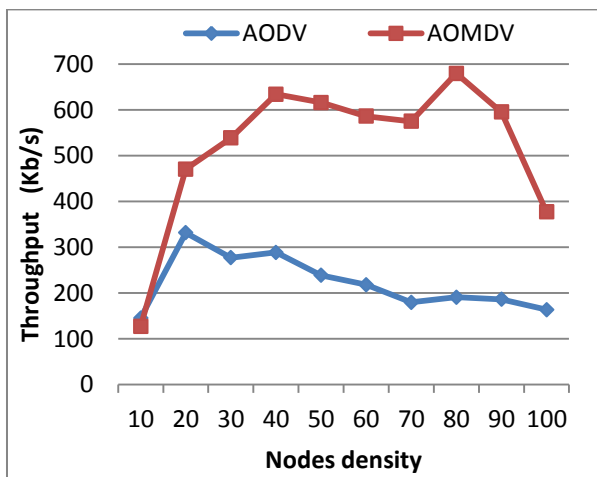


Fig.14. Throughput calculation of MPEG-4 packets compared to the density of nodes.

D. Comparison Table:

In the following table (Tab.2), we present the results of the performance of AODV and AOMDV before and after traffic generation MPEG-4 obtained in our simulations. We use three symbols of comparisons:

- " - " : lower
- " o " : equal
- " + " : higher
- " ++ " : much higher

Metric	AODV marking	AOMDV marking
Before generating MPEG-4 traffic		
PDR	-	+
Delay	-	++
Routing Overhead	-	+
Throughput	-	++
After generating MPEG-4 traffic		
PDR	-	+
Delay	o	o
Routing Overhead	-	+
Throughput	-	++

Tab. 2. Results comparison Table.

IV. CONCLUSIONS AND PERSPECTIVES

The routing protocols, that use the multi-paths, improve the quality of transmissions in Ad-hoc network. The traffic, such as streaming video, requires infrastructure and robust routing systems to ensure efficient and successful exchange data. In this paper, we studied and compared the performance of AODV (reactive routing protocol single-path) and AOMDV (reactive routing protocol multi-path) to manage traffic video MPEG-4 in a highly mobile network with variable density. To do this, we deployed the NS-2 simulator. We found that AOMDV has greatly increased the PDR of more than 100% and has significantly improved throughputs of over 130%. In a dense network (over 50nodes) AODV decreased largely the overheads to more than 170%. However, we observed that delays have not been influenced by the introduction of the criterion of multi-path routing of video packets. By cons, we found that the generation of traffic MPEG-4 has greatly increased delays and disrupted transmission of AOMDV. Indeed, the average delay increased from 46ms before generating the video traffic to an average of 1.8 s after generation.

The multipath routing has significantly improved the performance of a streaming video traffic compared to a single-path routing. However, several changes must be made at the system in order to achieve useful levels of quality data exchange, particularly at control packets forwarding. Our main objectives are to reduce transmission delay and increase the rate of the RDP of AOMDV in dense and mobile Ad-hoc network.

REFERENCES

- [1] E. M. Royer and T. Chai-Keong, "A review of current routing protocols for ad hoc mobile wireless networks," *Personal Communications, IEEE*, vol. 6, pp. 46-55, 1999.
- [2] C. S. R. Putta, K. B. Prasad, D. Ravilla, R. S. M. Nath, and M. L. R. Chandra, "Performance of ad hoc network routing protocols in IEEE 802.11," in *Computer and Communication Technology (ICCCT), 2010 International Conference on*, 2010, pp. 371-376.
- [3] M. Ikeda, M. Hiyama, E. Kulla, and L. Barolli, "Mobile Ad-hoc Network Routing Protocols Performance Evaluation Using NS-3 Simulator," in *Intelligent Networking and Collaborative Systems (INCoS), 2011 Third International Conference on*, 2011, pp. 14-20.
- [4] F. Qian, C. Zhongmin, Y. Jin, and H. Xunchao, "A Performance Comparison of the Ad Hoc Network Protocols," in *Computer Science and Engineering, 2009. WCSE '09. Second International Workshop on*, 2009, pp. 293-297.
- [5] C. Perkins, E. Belding-Royer, and S. Das, *Ad hoc On-Demand Distance Vector (AODV) Routing*: RFC Editor, 2003.
- [6] M. K. Marina and S. R. Das, "Ad hoc on-demand multipath distance vector routing," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 6, pp. 92-93, 2002.
- [7] J. Iso Iec, "International Organisation For Standardisation Organisation Internationale De Normalisation Iso/iec Jtc1/sc29/wg11 Coding Of Moving Pictures And Audio," 2009.
- [8] S. Vetrivel, M. Gowri, M. S. Sultana, and G. A. Dr, "AN OVERVIEW OF MPEG FAMILY AND ITS APPLICATIONS," *Indian Journal of Computer Science and Engineering*, vol. 1, pp. 240-250, 2010.
- [9] A. Puri, X. Chen, and A. Luthra, "Video coding using the H.264/MPEG-4 AVC compression standard," *Signal Processing: Image Communication*, vol. 19, pp. 793-849, 2004.
- [10] Y. Bo and K. W. Ng, "A survey on the techniques for the transport of MPEG-4 video over wireless networks," *Consumer Electronics, IEEE Transactions on*, vol. 48, pp. 863-873, 2002.
- [11] S. R. Biradar, M. Koushik, S. Subir Kumar, and C. Puttamadappa, "Performance Evaluation and Comparison of AODV and AOMDV," *International Journal on Computer Science and Engineering*, vol. 2, pp. 373-377, 2010.
- [12] Chintan Bhavsar, Mayur.M Vegad, and S. A. Bakhru., "Performance Comparison of Stability Enhanced AODV and AOMDV Protocols for MANETs," *International Journal of Emerging Technology and Advanced Engineering.*, vol. 2, pp. 279-284, 2012.
- [13] W. Deepinder Singh and P. Tripatjot Singh, "Performance Comparison of Single and Multipath Routing Protocols in Adhoc Networks," *International Journal of Computer Technology and Applications*, vol. 02, pp. 1486-1496, 2011.
- [14] P. Georgios, L. Vincent, M. Martin, and P. Bernhard, "Multi-path Routing Protocols in Wireless Mobile Ad Hoc Networks: A Quantitative Comparison," 2009.
- [15] C. Manveen Singh, J. Rambir, and Sandeep, "Simulation and Comparison of AODV, DSR and AOMDV Routing Protocols in MANETs," *International Journal of Soft Computing & Engineering*, vol. 2, pp. 375-381, 2012.
- [16] K. Kunavut and T. Sanguankotchakorn, "Performance evaluation of ad hoc routing protocols to deliver MPEG-4 traffic," in *Communication Technology (ICCT), 2010 12th IEEE International Conference on*, 2010, pp. 207-210.
- [17] R. B and M. D, "CA-AODV: Congestion Adaptive AODV Routing Protocol for Streaming Video in Mobile Ad Hoc Networks," *International Journal of Communications, Network and System Sciences*, vol. 01, pp. 322-328, 2008.
- [18] W. Peter, "The VINT Project, The Network Simulator - ns-2, URL: <http://www.isi.edu/nsnam/ns/>. Page accessed on," 2008.
- [19] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," *Mobile Computing, IEEE Transactions on*, vol. 2, pp. 257-269, 2003.
- [20] D. K. George, V. K. Srikanth, K. D. Son, and J. P. Gregory, "Multicasting Sustained CBR and VBR Traffic in Wireless Ad-Hoc Networks," 2009.

STOCK PRICES FORECAST USING RADIAL BASIS FUNCTION NEURAL NETWORK

Julia Fajaryanti

Faculty of Industrial Technology
Gunadarma University
Jalan Margonda Raya 100 Depok - Indonesia

Priyo Sarjono Wibowo

Faculty of Industrial Technology
Gunadarma University
Jalan Margonda Raya 100 Depok - Indonesia

Abstract— Neural Network has been implemented in various applications especially in pattern recognition. This power has attracted several people to use Neural Network for various systems. One of the neural network implementation in the field of finance or investments is forecasting stocks. Assuming that the prediction of the output system is deterministic, than the suitable Neural Network model to predict it is Multilayer Network. To get the solution, Multilayer Neural Network method with supervised algorithm is applied. The supervised algorithm used for stock price prediction is Radial Basis Function. This algorithm can supervise the networks by using previous stock price data, classifying them and putting weight on the networks. This journal illustrate how Radial Basis Function Neural Network method can be used to predict stocks. The result showed that Radial Basis Function Neural Network method is able to forecast and follow the movement of stock data used in the experiment.

Keywords: Stock Prices, Multilayer Neural Network, Radial Basis Function, Supervised

I. INTRODUCTION

The role of capital markets in the Indonesian economy began institutionalized. Currently one of the purchase of shares legitimate capital choices, in addition to other forms of capital such as money, land, and gold. Rational factors and various irrational factors be the deciding factor in purchasing shares. Rational factors commonly associated with the analysis fundamental. Fundamental analysis does not consider the pattern of movement of shares in the past but trying to determine the appropriate value for a stock.

In perfect capital markets and efficient, stock prices reflect all publicly available information on and information exchange that can only be obtained from certain groups. High and low stock prices influenced by many factors such as conditions and company performance, risk, dividend, interest rates, conditions economy, government policy, inflation, supply and demand as well as many more. Because anticipate possible changes in the factors above, the stock price can rise or fall. Prediction of the stock price will very useful for

investors to be able to see how the prospects of investing in the stock of a company come. Stock price prediction can be used to anticipate the rise and fall of stock prices. With the prediction of share price, it is very helpful for investors in decision making. There are two methods that prediction can use in this implementation, namely: conventional methods and Artificial Neural Network (ANN). In this journal, authors will implement Radial Basis Function Neural Network in financial application to forecast the stock prices.

II. STATE OF THE ART

There are many factors that influence the price of a stock. Most of these factors are included in the factors related to the social situation, where it is very difficult to estimate. However, there are things that can be used as a basis for estimating the price of a stock, one by studying the previous price of a stock, and then we can estimate the stock price in the future. This, of course can help decision-makers to the activities of buying and selling a stock [7]. For that we need a method to identify and study the movement of the stock price over time in order to estimate the price of a stock of a particular company.

It also takes a special device to do it, this is due to the limitations of human beings in the data processing activities that are so difficult and numerous in numbers. Thus, we use the computer to do it. To be able to work on it, the computer requires a learning process. In this process, computers are faced with a series of data that has been classified and will study the patterns of the data. Lessons learned include the adjustment to a predetermined pattern, or by studying the similarity of the pattern. It is accompanied by the development of computers, the better, in terms of speed, accuracy, and cost [8].

Neural network is a system that can be relied to do a complex quick count processing with higher speed compared to another organs. The use of neural networks requires an understanding of the way of thinking in the human brain. So many elements in the human brain, where they are connected

to each other, working together in order to process information. An important part of this paradigm is the simple structure of the information processing system. It consists of a large number of processing elements (neurons) that are interconnected and work together in order to solve a particular problem [9]. As this idea, many researchers start to make the neural network which is applicable to the computer machine.

According to [3] Neural Networks have shown to better predictive performance in predicting future stock prices movements by analyzing the past sequence of the stocks and Neural Networks can outperform the conventional statistical models [4,5,6]. Then raised some of artificial neural network models which has been applied in various areas. An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. Artificial neural network determined by three things. Determining the appropriate network architecture is the first step in building a well and according to the needs of artificial neural network. This following some of the network architecture is often used in the artificial neural network, such as single layer network, multi layer network, and recurrent network.

III. METHODOLOGY

Stock market prediction is a world of financial forecasting area which attracted much attention from various circles, especially the investors. Stock price prediction is very useful for investors to be able to see how the stock of a company's investment in the future. By using the prediction, it is helpful to investors in making decisions. Expected income perpetrators in the majority of the equity investment is a capital gain. This causes the process of forecasting the stock value is very important in stock investing. An attempt to determine the stock price should have been done by any financial analysts with the aim to obtain an attractive rate of return, although quite difficult for investors to "beat the market" and earn profits above normal levels.

The prediction of the stock price is the complex interaction between unstable market and unknown random processes factor. The data from stock price can be determined by time series. If we have daily data from a certain period, for example : $X_t(t = 1, 2, \dots)$ than the stock price for the next period (t+1) can be predicted (the timing used can be in hourly, daily, weekly, monthly or yearly). To get the good prediction, the inputs from several aspects of the share prices have to be input in Neural Network after that the weighing principal can be adapted to minimize the wrong prediction in the first future

steps.

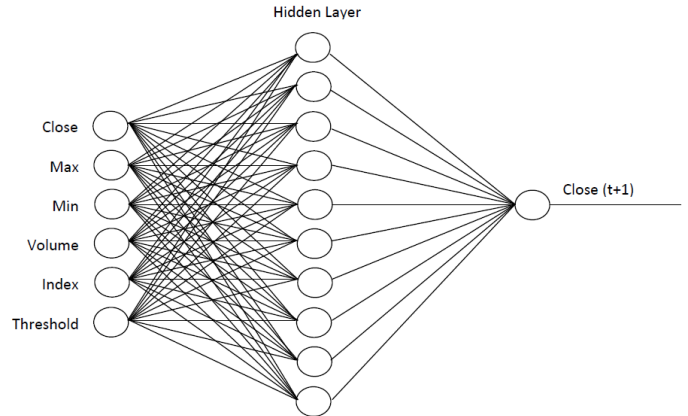


Figure 1. Artificial Neural Network to Forecast Stocks [11]

By using the final weighing, an action is done to done to minimize the total error for the next iteration. Due to that, the risk of Investors decision to sell or buy the stock can be minimized. The steps to be performed in a simulated stock price forecasting using Neural Network:

- Collecting data stock prices
- Determining the structure of the ANN to be built
- Conduct training and testing of neural network which is built using existing

Implementing Radial Basis Function to Forecast the Stocks using MATLAB

A. Implementing Radial Basis function to Forecast the Stocks using MATLAB

Before build a network, then we must determine the spread is to be used. Determination of trial and error spread is obtained. Process of trial and error will produce coating weights and biases. Spread that will be used RBF is that producing the greatest weight value of the bias layer. Furthermore, the process of building the network. We form a network that will be included in the RBF with the third layer, where the first layer (input) consists of one neuron with activation function based on radial radbas and the second layer (hidden layer) consists of n neurons with activation function purelin, and output layer consisting of one neuron and is represented in Matlab

```
net = newrbe(P, T, 5);
```

Weights are determined by Matlab, with the calculation of each weight has been determined through the function of tissue formation. To see the value of these weights are used the following command:

```
Bobot_Input = net.IW{1,1}  
Bobot_Bias_Input = net.b{1,1}  
Bobot_Lapisan = net.LW{2,1}  
Bobot_Bias_Lapisan = net.b{2,1}
```

At the output layer, the layer weights ($W_{2,1}$) and the weight refraction layer ($b_{2,1}$) was obtained from the simulation results in each hidden layer (A) which then solved using a linear function (purelin):

$$[W_{2,1} b_{2,1} * A_{1,1}; ones] = T$$

Input on the output layer (A1) and target (T) is linear, so that the weight of layers and layers of refraction weights can be calculated by

$$Wb = T/[P;ones(1,Q)]$$

Wb containing layer weight and the weight bias, with the refraction weight lies in the last column.

B. Simulation and output network on training data

From the results in training that can be simulated with the same input with data input training. Output simulation results are stored in a vector

```
a = sim(net,P);  
H = [(1:size(P,2))' T' a' (T'-a')];  
sprintf('%5d %11,2f %11,2f %11,2fn', H);
```

sprintf aims to print variable H' with the format listed above. Variable H will hold value:

- **(1: size (P, 2))'** means the loop will be performed starting from an initial value to the size of input training data.
- **T'** target of the original data.
- **a'** is the output network.
- **(T' - a')** is the error on each training data, that is the result of a reduction in the output target value network.

network output and the target were analyzed by linear regression using Postreg :

$$[m1,c1,r1] = postreg(a,T);$$

its produce

line gradient reversed (m1): $m1 = 0.96005$

constants (c1): $C1 = 343.07$

For the best fit line equation: $m1 + c1 = 0.96005+343.07$

connection coefficients (r1): $r1 = 0.97982$

If coefficient value (approaching 1), it showed good results for a match with the target network output.

C. Simulation and output network on testing data

From the results in training that can be simulated with the same input with data input training. Output simulation results are stored in a vector

```
b = sim(net,Q);  
L = [(1:size(Q,2))' TQ' b' (TQ' - b')];  
sprintf('%5d %11,2f %11,2f %11,2fn', L);
```

sprintf aims to print variable L' with the format listed above. Variable L will hold value:

- **(1: size (Q, 2))'** means the loop will be performed starting from an initial value to the size of input training data.
- **TQ'** target of the original data.
- **b'** is the output network.
- **(TQ' - b')** is the error on each training data, that is the result of a reduction in the output target value network.

network output and the target were analyzed by linear regression using

Postreg :

$$[m2,c2,r2] = postreg(b,TQ);$$

its produce

line gradient reversed (m2): $m2 = 0.9423$

constants (c2): $C2 = 423.98$

For the best fit line equation: $m2 + c2 = 0.9423+423.98$

connection coefficients (r2): $r1 = 0.87207$

if coefficient value (approaching 1), it showed good results for a match with the target network output.

IV. RESULTS AND DISCUSSION

In this section, authors analyzes the results of output program. Before analyze the result, authors test the system to get the result of prediction data. Please note that this test takes a sample of 247 data of stocks, as mentioned before, the data used in this journal is Telkom share data in the period August 3, 2009 until August 26, 2010. Then authors divide the data into two parts, the first 187 data means the data from period August 3, 2009 until May 31, 2010 used for training the data in system. And for the 60 data are from June 1, 2010 until August 26, 2010. As shown in the table, for the data training there is the biggest error that is equal to 3.92% in the data numbers 125. And for the data testing there is the biggest error that is equal to 7.5% in the first data. However, overall results show the accuracy of predictive data above 90%.

A. Result of Data Training

To train the data for the system, authors used 187 data. As shown in the table bellow, for the data training there is the biggest error that is equal to 3.92% in the data numbers 125. However, overall results show that the accuracy of predictive data is above 90%.

Table 1. Table Result of Sum of Data Training

Number	Closing+1	Prediction	Error	Error(%)	Accuracy
1	8800	8812.5	-12.5	0.125	99.875
2	8600	8762.5	-162.5	1.625	98.375
3	8750	8611.1	138.89	1.388	98.612
4	8800	8768.8	31.25	0.31	99.69
5	9000	8762.5	237.5	2.37	97.63
6	8900	9014.3	-114.29	1.14	98.86
7	8550	8812.5	-262.5	2.625	97.375
8	8750	8542.9	207.14	2.071	97.929
9	8700	8768.8	-68.75	0.68	99.32
10	8500	8645	-145	1.45	98.55
11	8350	8487.5	-137.5	1.375	98.625
12	8450	8370	80	0.8	99.2
13	8450	8418.8	31.25	0.31	99.69
14	8450	8418.8	31.25	0.31	99.69
15	8650	8418.8	231.25	2.31	97.69
16	8550	8590	-40	0.4	99.6
17	8600	8542.9	57.143	0.57	99.43
18	8650	8611.1	38.889	0.38	99.62
19	8400	8590	-190	1.9	98.1
20	8350	8428.6	-78.571	0.78	99.22
21	8250	8370	-120	1.2	98.8
22	8400	8300	100	1	99
23	8400	8428.6	-28.571	0.28	99.72
24	8550	8428.6	121.43	1.21	98.79
25	8450	8542.9	-92.857	0.92	99.08
26	8350	8418.8	-68.75	0.68	99.32
27	8450	8370	80	0.8	99.2
28	8450	8418.8	31.25	0.31	99.69
29	8300	8418.8	-118.75	1.18	98.82
30	8400	8400	0	0	100
31	8350	8428.6	-78.571	0.78	99.22
32	8400	8370	30	0.3	99.7
33	8750	8428.6	321.43	3.21	96.79
34	8650	8768.8	-118.75	1.18	98.82
35	8500	8590	-90	0.9	99.1
36	8550	8487.5	62.5	0.625	99.375
37	8650	8542.9	107.14	1.07	98.93
38	8700	8590	110	1.1	98.9
39	8550	8645	-95	0.95	99.05
40	8500	8542.9	-42.857	0.42	99.58

Table 1 (Cont.). Table Result of Sum of Data Training

Number	Closing+1	Prediction	Error	Error(%)	Accuracy
41	8600	8487.5	112.5	1.12	98.88
42	8650	8611.1	38.889	0.38	99.62
43	8650	8590	60	0.6	99.4
44	8550	8590	-40	0.4	99.6
45	8700	8542.9	157.14	1.57	98.43
46	8800	8645	155	1.55	98.45
47	8700	8762.5	-62.5	0.62	99.38
48	8650	8645	5	0.05	99.95
49	8700	8590	110	1.1	98.9
50	8700	8645	55	0.55	99.45
51	8700	8645	55	0.55	99.45
52	8650	8645	5	0.05	99.95
53	8600	8590	10	0.1	99.9
54	8650	8611.1	38.889	0.38	99.62
55	8500	8590	-90	0.9	99.1
56	8450	8487.5	-37.5	0.375	99.625
57	8350	8418.8	-68.75	0.68	99.32
58	8250	8370	-120	1.2	98.8
59	8400	8300	100	1	99
60	8250	8428.6	-178.57	1.78	98.22
61	8350	8300	50	0.5	99.5
62	8600	8370	230	2.3	97.7
63	8700	8611.1	88.889	0.88	99.12
64	8700	8645	55	0.55	99.45
65	8650	8645	5	0.05	99.95
66	8750	8590	160	1.6	98.4
67	8750	8768.8	-18.75	0.18	99.82
68	8800	8768.8	31.25	0.31	99.69
69	8750	8762.5	-12.5	0.12	99.88
70	8850	8768.8	81.25	0.81	99.19
71	8900	8900	0	0	100
72	8900	8812.5	87.5	0.875	99.125
73	9000	8812.5	187.5	1.875	98.125
74	9000	9014.3	-14.286	0.14	99.86
75	8950	9014.3	-64.286	0.64	99.36
76	9000	8850	150	1.5	98.5
77	9050	9014.3	35.714	0.35	99.65
78	8750	8750	0	0	100
79	9000	8768.8	231.25	2.31	97.69
80	9000	9014.3	-14.286	0.14	99.86

Table 1 (Cont.). Table Result of Sum of Data Training

Number	Closing+1	Prediction	Error	Error(%)	Accuracy
81	9000	9014.3	-14.286	0.14	99.86
82	9200	9014.3	185.71	1.85	98.15
83	9250	9350	-100	1	99
84	9300	9250	50	0.5	99.5
85	9300	9290	10	0.1	99.9
86	9200	9333.3	-133.33	1.33	98.67
87	9450	9350	100	1	99
88	9800	9530	270	2.7	97.3
89	9500	9500	0	0	100
90	9600	9464.3	135.71	1.35	98.65
91	9850	9612.5	237.5	2.37	97.63
92	10100	10100	0	0	100
93	9500	9500	0	0	100
94	9700	9464.3	235.71	2.35	97.65
95	9600	9600	0	0	100
96	9600	9612.5	-12.5	0.12	99.88
97	9500	9612.5	-112.5	1.12	98.88
98	9450	9464.3	-14.286	0.14	99.86
99	9550	9530	20	0.2	99.8
100	9600	9600	0	0	100
101	9500	9612.5	-112.5	1.12	98.88
102	9250	9464.3	-214.29	2.14	97.86
103	9350	9250	100	1	99
104	9450	9375	75	0.75	99.25
105	9500	9530	-30	0.3	99.7
106	9400	9464.3	-64.286	0.64	99.36
107	9350	9333.3	16.667	0.16	99.84
108	9500	9375	125	1.25	98.75
109	9400	9464.3	-64.286	0.64	99.36
110	9450	9333.3	116.67	1.16	98.84
111	9500	9530	-30	0.3	99.7
112	9450	9464.3	-14.286	0.14	99.86
113	9300	9530	-230	2.3	97.7
114	9250	9290	-40	0.4	99.6
115	9300	9250	50	0.5	99.5
116	9250	9290	-40	0.4	99.6
117	9350	9250	100	1	99
118	9300	9375	-75	0.75	99.25
119	9300	9290	10	0.1	99.9
120	9350	9290	60	0.6	99.4

Table 1 (Cont.). Table Result of Sum of Data Training

Number	Closing+1	Prediction	Error	Error(%)	Accuracy
121	9250	9375	-125	1.25	98.75
122	8950	9250	-300	3	97
123	8700	8850	-150	1.5	98.5
124	8550	8645	-95	0.95	99.05
125	8150	8542.9	-392.86	3.92	96.08
126	8450	8400	50	0.5	99.5
127	8350	8418.8	-68.75	0.68	99.32
128	8250	8370	-120	1.2	98.8
129	8300	8300	0	0	100
130	8600	8400	200	2	98
131	8750	8611.1	138.89	1.38	98.62
132	8600	8768.8	-168.75	1.68	98.32
133	8600	8611.1	-11.111	0.11	99.89
134	8400	8611.1	-211.11	2.11	97.89
135	8350	8428.6	-78.571	0.78	99.22
136	8600	8370	230	2.3	97.7
137	8350	8611.1	-261.11	2.61	97.39
138	8250	8370	-120	1.2	98.8
139	8050	8300	-250	2.5	97.5
140	8000	8025	-25	0.25	99.75
141	8200	7987.5	212.5	2.12	97.88
142	8300	8150	150	1.5	98.5
143	8200	8400	-200	1	99
144	8100	8150	-50	0.5	99.5
145	8100	8075	25	0.25	99.75
146	8050	8075	-25	0.25	99.75
147	8150	8025	125	1.25	98.75
148	8350	8400	-50	0.5	99.5
149	8200	8370	-170	1.7	98.3
150	8050	8150	-100	1	99
151	8100	8025	75	0.75	99.25
152	8100	8075	25	0.25	99.75
153	8050	8075	-25	0.25	99.75
154	8050	8025	25	0.25	99.75
155	7950	8025	-75	0.75	99.25
156	8000	8050	-50	0.5	99.5
157	8000	7987.5	12.5	0.12	99.88
158	7900	7987.5	-87.5	0.87	99.13
159	7950	7875	75	0.75	99.25
160	8100	8050	50	0.5	99.5

Table 1 (Cont.). Table Result of Sum of Data Training

Number	Closing+1	Prediction	Error	Error(%)	Accuracy
161	8100	8075	25	0.25	99.75
162	8050	8075	-25	0.25	99.75
163	7900	8025	-125	1.25	98.75
164	7800	7875	-75	0.75	99.25
165	7650	7650	0	0	100
166	7700	7680	20	0.2	99.8
167	7850	7700	150	1.5	98.5
168	8000	7825	175	1.75	98.25
169	7850	7987.5	-137.5	1.37	98.63
170	7650	7825	-175	1.75	98.25
171	7750	7680	70	0.7	99.3
172	7600	7525	75	0.75	99.25
173	7800	7662.5	137.5	1.37	98.63
174	7650	7650	0	0	100
175	7600	7680	-80	0.8	99.2
176	7650	7662.5	-12.5	0.12	99.88
177	7650	7680	-30	0.3	99.7
178	7700	7680	20	0.2	99.8
179	7550	7700	-150	1.5	98.5
180	7600	7600	0	0	100
181	7450	7662.5	-212.5	2.12	97.88
182	7100	7100	0	0	100
183	7200	7200	0	0	100
184	7500	7500	0	0	100
185	7600	7600	0	0	100
186	7750	7662.5	87.5	0.875	99.125
187	7450	7525	-75	0.75	99.25

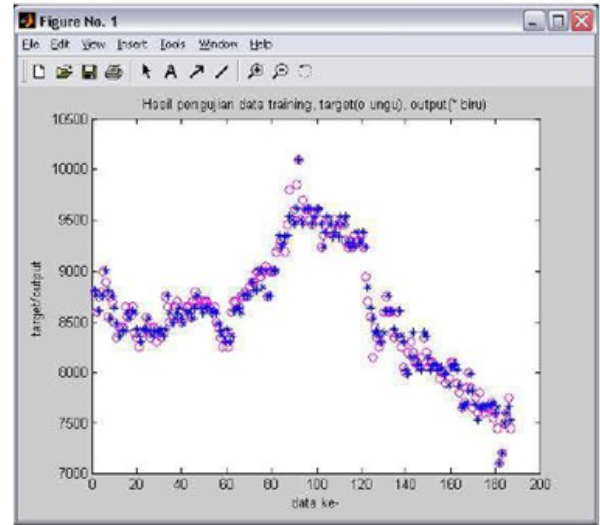


Figure 3. Comparison of target and output data training

B. Result of Data Testing

To train the data for the system, authors used only 60 data. The usage of data in data testing is less than in data training, because for train the data system must be learned more data to produce a more precise accuracy when testing the data. As shown in the table bellow, for the data testing there is the biggest error that is equal to 7.5% in the first data. However, overall results show the accuracy of predictive data above 90%.

Table 2. Table Result of Sum of Data Testing

Number	Closing+1	Prediction	Error	Error(%)	Accuracy
1	7850	7100	750	7.5	92.5
2	8000	7825	175	1.75	98.25
3	7900	7987.5	-87.5	0.87	99.13
4	7750	7875	-125	1.25	98.75
5	7950	7525	425	4.25	95.75
6	7850	8050	-200	2	98
7	7750	7825	-75	0.75	99.25
8	7850	7525	325	3.25	96.75
9	8000	7825	175	1.75	98.25
10	7950	7987.5	-37.5	0.375	99.625
11	7950	8050	-100	1	99
12	8000	8050	-50	0.5	99.5
13	8050	7987.5	62.5	0.625	99.375
14	7900	8025	-125	1.25	98.75
15	7900	7875	25	0.25	99.75
16	8000	7875	125	1.25	98.75
17	8000	7987.5	12.5	0.125	99.875
18	7950	7987.5	-37.5	0.375	99.625
19	7900	8050	-150	1.5	98.5
20	7700	7875	-175	1.75	98.25

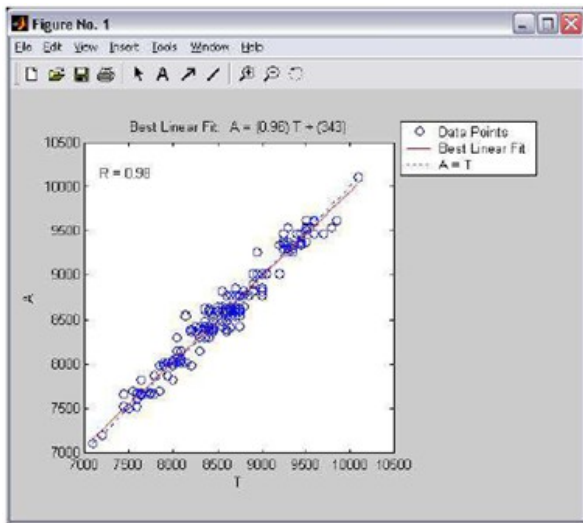


Figure 2. Performance of Linear Regression of target and output data training

Table 2 (Cont). Table Result of Sum of Data Testing

Number	Closing+1	Prediction	Error	Error(%)	Accuracy
21	7700	7700	0	0	100
22	7700	7700	0	0	100
23	7650	7700	-50	0.5	99.5
24	7750	7680	70	0.7	99.3
25	7850	7525	325	3.25	96.75
26	7900	7825	75	0.75	99.25
27	7750	7875	-125	1.25	98.75
28	7800	7525	275	2.75	97.25
29	7850	7650	200	2	98
30	7800	7825	-25	0.25	99.75
31	7900	7650	250	2.5	97.5
32	8100	7987.5	112.5	1.12	98.88
33	8100	8075	25	0.25	99.75
34	8150	8075	75	0.75	99.25
35	8100	8400	-300	3	97
36	8150	8075	75	0.75	99.25
37	8050	8400	-350	3.5	96.5
38	8200	8025	175	1.75	98.25
39	8250	8150	100	1	99
40	8250	8300	-50	0.5	99.5
41	8450	8300	150	1.5	98.5
42	8400	8418.8	-18.75	0.18	99.82
43	8000	8428.6	-428.57	4.28	95.72
44	8100	7987.5	112.5	1.12	98.88
45	8200	8075	125	1.25	98.75
46	8400	8150	250	2.5	97.5
47	8500	8428.6	71.429	0.71	99.29
48	8400	8487.5	-87.5	0.87	99.13
49	8450	8428.6	21.429	0.21	99.79
50	8400	8418.8	-18.75	0.18	99.82
51	8500	8428.6	71.429	0.71	99.29
52	8550	8487.5	62.5	0.625	99.375
53	8600	8542.9	57.143	0.57	99.43
54	9000	8611.1	388.89	3.88	96.12
55	8800	9014.3	-214.29	2.14	97.86
56	8700	8762.5	-62.5	0.625	99.375
57	8700	8645	55	0.55	99.45
58	8900	8645	255	2.55	97.45
59	9000	8812.5	187.5	1.87	98.13
60	9000	9014.3	-14.286	0.14	99.86

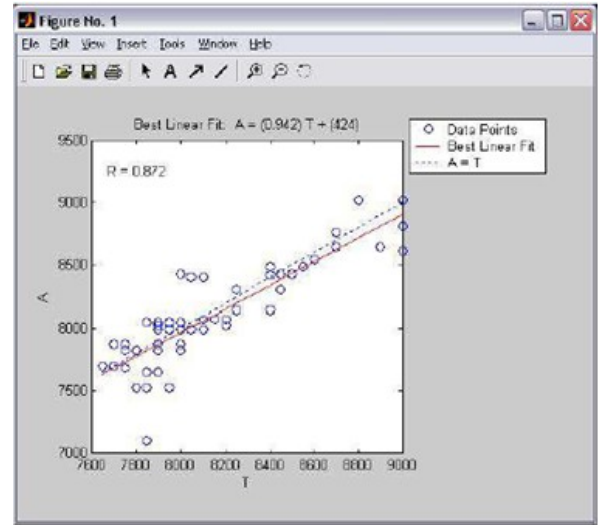


Figure 4. Performance of Linear Regression of target and output data testing

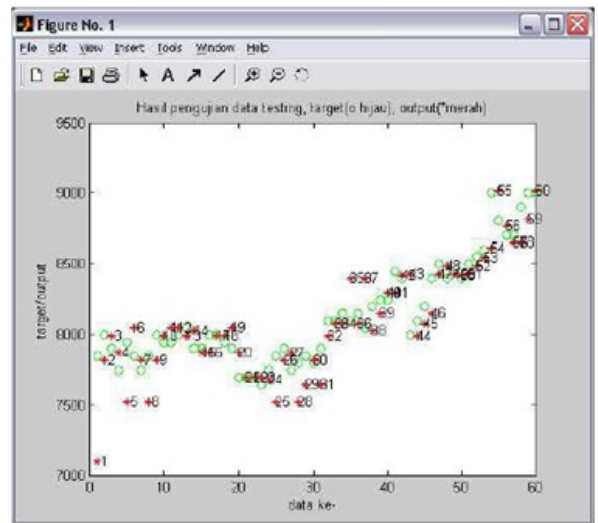


Figure 5. Comparison of target and output data training

	2005	2006	2007	2008	2009	2010
Indonesia	16.24%	55.30%	52.08%	-50.64%	86.98%	46.13%
Korea	53.96%	3.99%	32.25%	-40.73%	49.65%	21.88%
Malaysia	-0.76%	21.74%	31.82%	-39.33%	45.17%	19.34%
Singapore	13.37%	27.47%	16.07%	-49.17%	64.49%	10.09%
Japan	40.24%	6.92%	-11.13%	-42.12%	19.04%	-3.01%
Hongkong	5.73%	32.69%	39.31%	-48.27%	52.02%	5.32%
USA	-0.61%	16.29%	6.43%	-33.84%	18.82%	11.02%
UK	16.71%	10.71%	3.80%	-31.33%	22.07%	9.00%
China	-8.33%	130.43%	96.66%	-65.39%	79.98%	-14.31%
India	42.33%	46.70%	47.15%	-52.45%	81.03%	17.43%

Figure 6. Regional Indices year 2005 – 2010 [10]

This can also be caused by the increasing number of companies offering shares as a way to seek fresh funds in order to carry out the capital increase, thus increasing the diversification and production capacity. The number of these companies will continue to grow in line with economic growth in Indonesia and also increasing the demand for economic goods and services. The productivity of goods and services will increase in line with market demands. Thus, the competition will boost the performance of an enterprise that will ultimately increase the value of a company in the stock market.

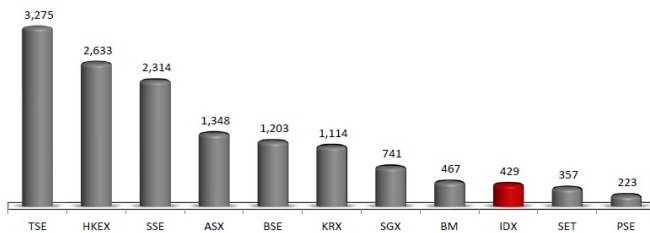


Figure 7. Regional Market Capitalization [10]

However, there are several other factors that can affect the movement of stock prices in Indonesia. The factors mentioned above are government policy, unclear regulations and enforcement, as well as the safety factor. These things can also affect domestic stock prices. Therefore, we need a reliable method to predict the stock market, which can provide assistance in times to buy or sell stocks.

Using neural networks to forecast stock market prices will be a continuing area of research as researchers and investors keep striving to outperform the market, with the ultimate goal of bettering their returns. It is unlikely that new theoretical ideas will come out of this applied work. However, interesting results and validation of theories will occur as neural networks are applied to more complicated problems. Continued work on improving neural network performance may lead to more insights in the chaotic nature of the systems they model. However, it is unlikely a neural network will ever be the perfect prediction device that is desired because the factors in a large dynamic system, like the stock market, are too complex to be understood for a long time.

V. CONCLUSION

There are some conclusions from the research that have been done, such as RBF Neural Network is fully influenced by the network architecture. To implement RBF Neural Network in this system, authors treat the data of stocks as time series, in which stocks data that existed at a particular period is used to predict or forecast the stocks in subsequent periods. And Neural Network learn the value of the past stock as the knowledge to determine future stock price. As shown in the result, the accuracy of predictive data in this system is almost 100%. It means that this system has good enough to implement Radial Basis Function Neural Network to forecast the stocks. Although neural networks are not perfect in their prediction, hope that one day we can more fully understand dynamic, chaotic systems such as the stock market.

VI. REFERENCES

- [1] F. Yan and M. Paindavoine, "Implementation of rbf neural network on embedded system: Real-time face tracking and identity verification." IEEE Transaction on Neural Network, 2003.
- [2] W. Setiawan, "Prediksi harga saham menggunakan jaringan syaraf tiruan multilayer feedforward network dengan algoritma backpropagation," 2008.
- [3] Rahib H. Abiyev and V. Hidayat Abiyev. " Differential Evaluation Learning of Fuzzy Wavelet Neural Networks for Stock Price Prediction." Journal of Information and Computing Science. Vol.7, No.2, 2012, pp. 121-130. 2012.
- [4] R.John, B. Guerard and Eli Schwartz. Regression Analysis and Forecasting Models. *Quantitative Corporate Finance*. DOI: 10.1007/978-0-387-34465-2_12
- [5] A. Vellido, P. J. G. Lisboa, J. Vaughan. Neural Networks in business: a survey of applications (1992-1998). *Expert System with Applications*. 1999, 17:51-70.
- [6] Y. Chen. B. Yang J. Dong, A. Abraham. Time-series forecasting using flexible neural tree model. *Information Sciences*. 2005, 174: 219-235
- [7] Gholamreza Jandaghi, Reza Tehrani, Davoud Hossenpour, Rahmatollah Gholipour, Seyed Amir Shahidi, Shadkam, Application of Fuzzy-neural Networks in multi ahead forecast of stock price, *African Journal of Business Management*. Vol 4(6), 2010. pp. 903-914. 2010.
- [8] Jayanta Kumar Basu, Debnath Bhattacharyya, Tai-hoon Kim, Use of Artificial Neural Network in Pattern Recognition. *International Journal of Software Engineering and Its Applications*. Vol 4, No. 2, 2010. pp. 23 - 34. 2010.
- [9] Dase R.K. and Pawar D.D., Application of Artificial Neural Network for stock market prediction: A Review of literature. *International Journal of Machine Intelligence*. Vol 2, Issue 2, 2010. pp. 14 - 17. 2010.
- [10] I Gede Nyoman Yetna, "Listing on the IDX", unpublished.
- [11] Budi Bambang DP, Rochani J Widodo, Iftikar Z. Sutralaksana, Moses L. Singih. "Teknik Jaringan Syaraf Tiruan Feedforward Untuk Prediksi Harga Saham Pada Pasar Modal Indonesia". *Jurnal Informatika* Vol.1, No.1. 1999.
- [12] R. Lawrence "Using Neural Networks to Forecast Stock Market Prices". 1997.

AUTHORS PROFILE



Julia Fajaryanti received the Graduation in Information Techniques (ST) and Master (MMSI) degree in Gunadarma University, Jakarta 2009 and 2010, respectively. Presently working as Lecturer (Information Technology) in Gunadarma University, Jakarta. Interested field of research are Artificial Neural Network, Android programming, and Image Processing.

Authored many research papers in International / national journals/conferences in the field of computer.



Priyo Sarjono Wibowo received the Graduation in Information (S. Kom.) and Master (MMSI) degree in Gunadarma University, Jakarta 2003 and 2009, respectively. Presently working as Lecturer (Information Technology) in Gunadarma University, Jakarta. Interested field of research are Artificial Neural Network, Android

programming, Java programming, and Visual Basic.

Model Based Framework for Estimating Mutation Rate of Hepatitis C Virus in Egypt

Nabila Shikoun
Ph.D. Student
Faculty of Engineering
Al Azhar University

Mohamed El Nahas
Prof. of Pattern Recognition
Faculty of Engineering
Al Azhar University

Samar Kassim
Prof. of Medical Biochemistry & molecular Biology
Faculty of Medicine
Ain Shams University

Abstract - Hepatitis C virus (HCV) is a widely spread disease all over the world. HCV has very high mutation rate that makes it resistant to antibodies. Modeling HCV to identify the virus mutation process is essential to its detection and predicting its evolution. This paper presents a model based framework for estimating mutation rate of HCV in two steps. Firstly profile hidden Markov model (PHMM) architecture was builder to select the sequences which represents sequence per year. Secondly mutation rate was calculated by using pair-wise distance method between sequences. A pilot study is conducted on NS5B zone of HCV dataset of genotype 4 subtype a (HCV4a) in Egypt.

Keywords: *Hepatitis C virus (HCV), Profile Hidden Markov Model (PHMM), Non-structure 5 B(NS5B), Phylogenetic tree, pair-wise distance.*

I. INTRODUCTION

Once hepatitis C virus (HCV) has been discovered, it has been an importance subject of research and clinical investigations as its major role in human disease has emerged. An estimated 170 million people (3% of the world's population) worldwide have hepatitis C virus (HCV) infection and creates a huge disease burden from chronic, progressive liver disease [1]. Hepatitis C is a predominant genotype found throughout the Middle East and parts of Africa, with high population prevalence in Egypt. Due to the world's constant effort to find treatment for this fatal disease; many researches and trials have been made [2]. HCV has become a major cause of liver cancer and one of the commonest indications for liver transplantation. HCV infection can be treated, but this is costly and requires long-term medical support and follow-up; current therapies are impractical for the majority of HCV carriers worldwide. The development of a protective vaccine remains, a distant goal [1].

The aim of this study is to identify evolution model which estimates the mutation rate of non-structure 5B (NS5B) which has more variation (mutation) than other zones in Hepatitis C Virus genotype 4 subtype a (HCV4a) in Egypt. Profile hidden markov model (PHMM), phylogenetic tree, a polynomial fitting and a least-square interpolation is used to outline the progression of the viral mutations over time. This model shall be used to predicate mutation rate of HCV in blood samples. The mutation model presents new therapeutic

targets as well as genomic information for designing vaccine candidates.

In this research the profile hidden markov model (PHMM) was identified as a technique to select sequence from many sequences observed in one year. A relationship between several sequences was deduced representing several years by estimating mutation rate using pair-wise distance method. The mutation model can predicate mutation rate for next year. Our approach introduces NS5B zone in HCV4a genome for 5 years from 2007 to 2010.

This paper is organized as follows: In section 2, related research for evolution of hepatitis C virus is presented. Section 3 describes HCV virus genome and explains the concept of the mutation rate. Section 4 presents the suggested method for evaluation of mutation rate. Section 5 presents the study data set and experimental results. Section 6 concludes the paper with future research directions.

II. RELATED RESEARCH

Several researches have been conducted to unravel information and useful patterns in a database for evolution of RNA and calculating mutation rate in HCV virus. Pybus et al. [3] developed Bayesian inference framework to estimate the transmission dynamics of HCV in Egypt from sampled viral gene sequences, and to predict the public health impact of the virus. Bruijne et al. [4] investigates the genetic diversity and evolutionary origin of HCV-4 in Amsterdam, The Netherlands used a molecular epidemiological approach. Phylogenetic analysis of the NS5B sequences (668 bp) was obtained from 133 patients newly diagnosed with HCV-4 infection over the period from 1999 to 2008. Xiong et al. [5] propose a stochastic model based on the branching process for estimation and comparison of the mutation rates in proliferation processes of cells or microbes. Barash et al. [6] introduce some Programs for RNA mutational analysis. These programs can be used for suggesting point mutations, investigating the effect of deleterious and compensatory mutations in allosteric ribozymes and riboswitches and analyzing regulatory RNA sequences by their mutational profile. Ribeiro et al. [7] measured the accumulative rate of mutations and fitted the model to the sequence data of HCV by estimating the median in vivo viral mutation rate.

III. HCV VIRUS GENOME

The HCV genome is an enveloped structure approximately 50 nm in diameter. HCV is a positive single stranded enveloped RNA virus belonging to the *Flaviviridae* family with an average length of 9600 bases and carries a single, long open reading frame (ORF) flanked by 5' and 3' non-translated regions. The ORF encodes a polyprotein of ~3000 amino acids that is processed into three structural proteins (Envelopes 1 and 2 and p7) and six non-structural proteins named NS2-NS5B [8 -9].

HCV is classified into eleven major genotypes (designated 1-11), many subtypes (designated a, b, c, etc.), and about 100 different strains (numbered 1, 2, 3, etc.) based on the genomic sequence heterogeneity. Genotype 4 (HCV4) is principally found in the Middle East and Africa, particularly Egypt, which represent more than 90% of infections due to genotype 4 worldwide [9]. Ray et al. [10] characterize the genotype distribution of HCV in Egypt. In their work, specimens were obtained from blood donors in 15 geographically diverse governorates throughout Egypt. The result showed that 111 (91%) were genotype 4, 1 (1%) was genotype 1a, 1 (1%) was genotype 1b, and 9 (7%) could not be typed.

Definition of Mutation Rate

A mutation is a change of the nucleotide sequence of the genome of an organism and virus. Mutation rate is a measure of the rate at which various types of mutations occur during some unit of time [11]. There are two units used in viral mutation rate. These units are related with the nature of the viruses replication. Viruses replication are one of two types: binary replication and Linear replication "stamping machine". In case of binary replication, mutation rates are expressed as substitutions per nucleotide per strand copying unit. For linear replication mutation rates are expressed as substitutions per nucleotide per cell infection unit. HCV mutation rate is expressed with the last units [12]. In this paper mutation rate as substitutions per nucleotide per cell infection unit time was adopted. Reasons of mutations are from unrepaired damage to DNA or to RNA genomes (caused by radiation or chemical mutagens), from errors in the process of replication, or from the insertion or deletion of segments of DNA by mobile genetic elements. The high mutation rate in HCV is the reason of persistence in the human host [11].

An accurate estimate of mutation rate of virus is very important to understand the evolution of the viruses and to combat them [12]. There are various statistical methods to estimate mutation rates. It's classified into three general approaches: linear regression, maximum likelihood, and Bayesian inference. Linear regression procedures calculate substitution rate by comparing directly genetic distance between two sequences with the interval separating their isolation times. These methods are fast and useful for visualizing new data sets. They can assist in model selection. But they make several limiting assumptions. Maximum likelihood (ML) approach defines methods that accommodate the time structure of temporally spaced sequences. ML methods are more sensitive and accurate than distance-based

methods of linear regression. In Bayesian statistical inference, substitution rate is obtained empirically from the frequency distribution of the parameters values sampled by the Markov chain Monte Carlo (MCMC) algorithm. Maximum likelihood and Bayesian methods utilize more information from the sequences and allow much more complex models of molecular evolution and demography to be investigated [13].

One of linear regression methods is based on pair-wise distance [13]. Pair-wise distance enables to estimate distances in terms of the number of nucleotide substitutions. In this paper two types of distances to measure the genetic distance between sequences are used: Jukes Cantor and Kimura.

Jukes & Cantor distance: assumes all changes between all nucleotide are equally [14]. This is defined in equation (1)

$$d_{xy} = -(3/4) \log_e(1 - 4/3 D) \quad (1)$$

d_{xy} = distance between sequence x and sequence y expressed as the number of changes per site,

D = is the observed proportion of nucleotides which differ between two sequences (fractional dissimilarity).

The 3/4 and 4/3 terms reflect that there are four types of nucleotides and three ways in which a second nucleotide

Kimura method: this method distinguishes between two types of differences when comparing a pair of nucleotide sequences. Transition type which gets distance difference between nucleotide both are purines or both pyrimidines ($T \leftrightarrow C$, $A \leftrightarrow G$). The second type is transversion where the difference distance between one of the two is a purine and the other is a pyrimidine ($T \leftrightarrow A$, $T \leftrightarrow G$, $C \leftrightarrow A$, $C \leftrightarrow G$) [15]. This method is defined in equation (2)

$$K = -\frac{1}{2} \log_e \left\{ (1 - 2P - Q) \sqrt{1 - 2Q} \right\} \quad (2)$$

Where P and Q are the fractions of nucleotide sites of transition and transversion types respectively between two compared sequences.

IV. EVOLUTION OF MUTATION RATE

The objective of evolution model is to estimate mutation rate using Profile Hidden Markov Model (PHMM). Where PHMM is a special type of left-to-right HMMs is commonly used to model multiple alignments. The architecture of PHMM was introduced by Krogh (1994) [16]. El Nahas et al. [9] define a tuple of HMM, its tasks, and the architecture of PHMM.

The first step is to collect sequences from Egypt for several years. The data collected from region NS5B sequences from HCV genotype 4 subtype a (HCV4a). Choice one sequence represent certain year from sequences collected to this year. This is done by using PHMM. The Baum Welch algorithm is generally accepted to estimate PHMM parameters. However, this algorithm assumes that the model length is known, which not the case in this work. Hence, we have to adapt the learning procedure to search for the optimal

model length. The second step gathering all choice sequences and get matrix distance between each two sequences. Drawing these genetic distances and fitting it to get polynomial mutation rate. In this work MATLAB bioinformatics tool box functions was uses. The procedure is detailed in the following steps:

Procedure Evaluate Mutation Rate

Begin

- 1- **Input** all sequences of each year of NS5B region of HCV4a.
- 2- **Apply Multiple Sequence Alignment (MSA)** to these sequences of each year. MSA performs by using a heuristic search known as progressive technique (also known as the hierarchical or tree method).
- 3- **Preprocess data to filter unknown symbols.** Sometimes, a characters 'r, g, w, m, n, y, k, s' is found in sequences which do not map to any of 4 nucleotides (A, C, G, T) so it is replaced by one of nucleotide by using MSA.
- 4- **Initialize structure for PHMM of MSA.** The initial model structure and length are defined using information derived from the alignment together with its prior knowledge of the general nature of proteins.
- 5- **Estimate the PHMM parameters from training sequences to each year.** All the parameters in the PHMM (i.e. the transition probabilities and the nucleotide distributions) are estimated from a set of aligned sequences to maximize the likelihood of the observed sequences in the family. The likelihood of observed sequences is defined as:
$$P(\text{sequences} | \text{model}) = P(\text{sequence } 1 | \text{model}) * \dots * P(\text{sequence } n | \text{model})$$
- 6- **Score the model.** Scoring is used to assign a score with respect to the model to any query sequence for each year, the better the score, the higher and the chance that the query sequence is a member (homologue) of the protein family represented by the model. Scores are computed using log-odd ratios for emission probabilities and log probabilities for state transitions.
- 7- **Computing the sequence Pair-Wise Distances.** That is by constructing distance matrix which computes distances between each sequence pair. The method to calculate pair-wise distances is 'Jukes-Cantor', which estimate Maximum likelihood of the number of substitutions between two sequences. Ignore sequence sites representing gaps. P is described with the method P-distance. Proportion of sites at which the two sequences are different. Note that P is close to 1 for poorly related sequences, and P is close to 0 for similar sequences.
- 8- **Construct a phylogenetic tree.** Using distance matrix computed in step 7 to build the phylogenetic tree. Where Phylogenetic trees

represent evolutionary relationships, or genealogy, among species. The neighbor-joining method used to build the tree. Assuming equal variance and independence of evolutionary distance estimates.

- 9- **Estimate the date of origin of the mutation.** Consider the pair-wise distances according to the Kimura method, which distinguishes between transitional and transversal mutation. Then, restrict analysis to the distance of each sequence from the reference year consider as the starting of the mutation. Plot the genetic distance versus the date of collection.
- 10- **Compute Progression of Viral Mutation.** Perform a polynomial fitting and a least-square interpolation to outline the progression of the viral mutations over time. Which finds the coefficients of a polynomial $p(x)$ of degree n that fits the data, $p(x(i))$ to $y(i)$, in a least squares sense. The result p is a row vector of length $n+1$ containing the polynomial coefficients in descending powers.
- 11- **Reroute the Phylogenetic Tree.** The rerouted tree better illustrates the progression of the NS5B(HCV4a) mutation starting with the early infections.

end

V. PILOT STUDY

The main objective of this pilot study is to estimate the mutation rate of region NS5B in HCV4a spread in Egypt. Genotype 4 (HCV4) is particularly principally found in Egypt, which represent more than 90% of infections worldwide [12]. The region of NS5B in HCV4 subtype a (HCV4a) is used for this purpose. A data set representing NS5B is collected and used for to identify its mutation rate. Then, the learning procedure described in preceding section is applied on this real world data set to identify the model.

A. Data Description

The dataset contains the genomic sequences of distinct non-structural proteins named NS5B genotype 4 subtypes 4a from Egypt (11 sequences) in 2007 year, (35 sequences) in 2008 year, (35 sequences) in 2009 year and (17 sequences) in 2010 year. The length of the sequences varied between 286 and 339 nucleotide (nt). The data is obtained from the site "<http://www.ncbi.nlm.nih.gov/protein>" and it is presented in Table 1 which contains virus sequence Number (Virus_seq_no) and genebank of each year.

The data set was found to contain the characters 'r, g, w, m,n, y, k, s' which are undefined as nucleotide, to overcome this problem and to replace these undefined sites with suitable nucleotide the following steps were applied

- 1- Determine sequence number and positions numbers which contain these characters in original sequences.

2- Apply global alignment to all sequences MSA, and determine sequence number, position number of these characters and the most character repeated

3- Replace these character with suitable nucleotide found in step 2.

TABLE 1: The Data Set (<http://www.ncbi.nlm.nih.gov/protein>)

Virus_seq_no	Genbank_2007	Virus_seq_no	Genbank_2008	Virus_eq_no	Genbank_2009	Virus_seq_no	Genbank_2010
2007_1	DQ911222	2008_1	EF694448	2009_1	AB470254	2010_1	FN668600
2007_2	DQ911173	2008_2	EF694425	2009_2	AB470018	2010_2	FN668570
2007_3	DQ911190	2008_3	EF694438	2009_3	AB470024	2010_3	FN668577
2007_4	DQ911228	2008_4	EF694446	2009_4	AB470055	2010_4	FN668587
2007_5	DQ911183	2008_5	EF694408	2009_5	AB470015	2010_5	FN668591
2007_6	DQ911164	2008_6	EF694420	2009_6	AB470046	2010_6	FN668593
2007_7	DQ911172	2008_7	EF694424	2009_7	AB470009	2010_7	FN668574
2007_8	DQ911208	2008_8	EF694455	2009_8	AB470048	2010_8	FN668588
2007_9	DQ911205	2008_9	EF694502	2009_9	AB470030	2010_9	FN668589
2007_10	AY548725	2008_10	EF694517	2009_10	AB470036	2010_10	FN668586
2007_11	DQ911206	2008_11	EF694456	2009_11	AB470019	2010_11	FN668583
2007_12	AY548723	2008_12	EF694413	2009_12	AB470060		
2007_13	DQ911231	2008_13	EF694525	2009_13	AB470057		
2007_14	DQ911163	2008_14	EF694501	2009_14	AB470021		
2007_15	DQ911207	2008_15	EF694513	2009_15	B470039		
2007_16	AY548722	2008_16	EF694498	2009_16	AB470028		
2007_17	DQ911218	2008_17	EF694396	2009_17	AB470008		
		2008_18	EF694486	2009_18	AB470014		
		2008_19	EF694450	2009_19	AB470037		
		2008_20	EF694418	2009_20	AB470027		
		2008_21	EF694439	2009_21	AB470052		
		2008_22	EF694434	2009_22	AB470043		
		2008_23	EF694442	2009_23	AB470049		
		2008_24	EF694440	2009_24	AB470026		
		2008_25	EF694493	2009_25	AB470006		
		2008_26	EF694499	2009_26	AB470053		
		2008_27	EF694445	2009_27	AB470013		
		2008_28	EF694463	2009_28	AB470023		
		2008_29	EF694505	2009_29	AB470005		
		2008_30	EF694476	2009_30	AB470007		
		2008_31	EF694475	2009_31	AB470056		
		2008_32	EF694433	2009_32	AB470017		
		2008_33	EF694478	2009_33	AB470059		
		2008_34	EF694459	2009_34	AB470252		
		2008_35	EF694524	2009_35	AB470033		

B. Experimental Results

Table 2 shows scores of PHMM model applied to all sequences of 2007 to 2010 years on NS5B of HCv4a. Scores show that the sequences 11, 9, 14 and 10 are more suitable candidate sequences of 2007, 2008, 2009 and 2010 years respectively.

TABLE 2: Scores of sequences in each year to PHMM model

	2007	2008	2009	2010
	Seq_11	Seq_9	Seq_14	Seq_10
200	173.2023	290.6751	305.7761	112.8878
210	189.8152	317.9683	318.7579	122.2278
220	206.9295	334.3079	328.1445	126.6255
230	223.6203	350.1565	342.7904	132.7961
240	240.3303	363.4668	356.4055	153.8358
250	254.2443	377.1032	369.6046	170.5305
260	270.6948	388.6816	385.3829	189.7923
270	288.0402	402.3539	394.7695	209.4087
280	305.6335	415.4973	405.1091	228.9011
290	321.3509	425.3251	418.0684	249.4041
300	337.0683	436.9035		269.9071
310	353.6787	446.3764		290.6723
320	365.4927	457.7152		301.4251
330	369.5766			

Number of Match states (length of the model)

Table 3 and Figure 1 show the scores of pair-wise distances using Juke-cantor correction expressed in a symmetric matrix.

TABLE 3: Pair-Wise Distances using Jack-cantor corrections

	2007	2008	2009	2010
2007	0	0.1077	0.1942	0.1009
2008	0.1077	0	0.184	0.0579
2009	0.1942	0.184	0	0.1956
2010	0.1009	0.0579	0.1956	0

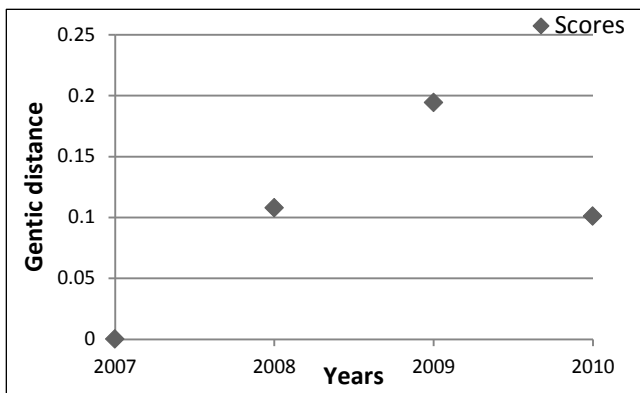


Figure 1 Construct score of genetic distances versus years

Figure 2 shows a phylogenetic tree of Table 3. The distance difference between 2008 and 2010 years is 0.0579 which is the least distance record in Table 3. Hence these two years are represented as on cluster. Year 2009 records high difference distance with respect to all years, so it is considered as one cluster as shown in Figure 2.

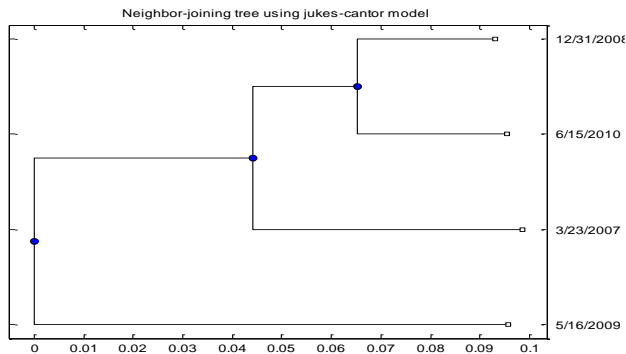


Figure 2 Neighbor-Joining Phylogenetic tree

Table 4 shows pair-wise distances using Kimura method. Estimate the date of origin of the mutation is 3/23/ 2007. Restrict analysis to the distance of each sequence to this year. Figure 3 plots the genetic distance versus this year, which shows that the minimum distance recorded is 0.0424 to 2009

year respect to 2007 year. The maximum distance recorded is 0.0589 to 2008 year with respect to 2007 year.

TABLE 4: Pair-Wise Distances using Kimura model

	2007	2008	2009	2010
2007	0	0.0589	0.0424	0.0515
2008	0.0589	0	0.0288	0.0584
2009	0.0424	0.0288	0	0.0397
2010	0.0515	0.0584	0.0397	0

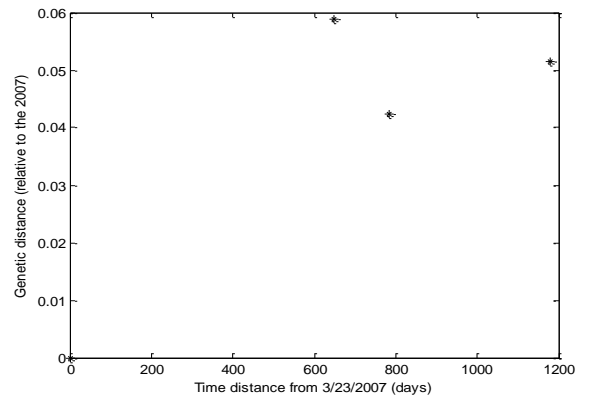


Figure 3 Genetic distances with respect to the year of 2007

Figure 4 shows a polynomial fitting of progression of viral mutation from scores in genetic distance in Figure 3. These scores increase approximately in a linear manner with time. The polynomial equation is found to be:

$$D = 0.00004493 X + 0.008826 \quad (5.1)$$

where 0.00004493 is the mutation rate (slope), 0.008826 is the intercept,

and D is any genetic distance (score) referred to year of 2007 and X is the time relative to the reference year .

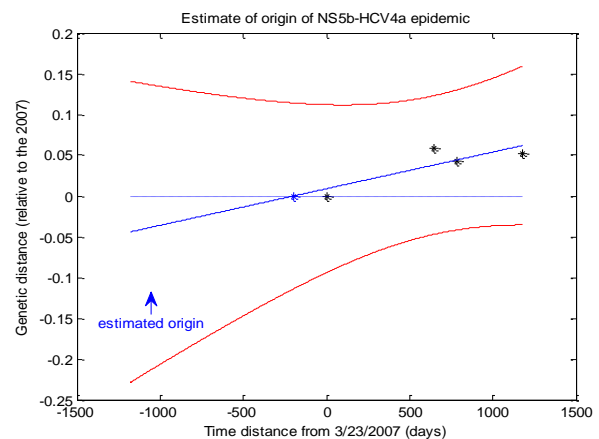


Figure 4 Progression of Viral Mutation with respect to the year of 2007

Figure 5 shows rerouting the phylogenetic tree of NS5B(HCV4a). This illustrates the progression of the mutation of the virus. 95% confidence interval of the estimated polynomial is also plotted in red color on the Figure 5.

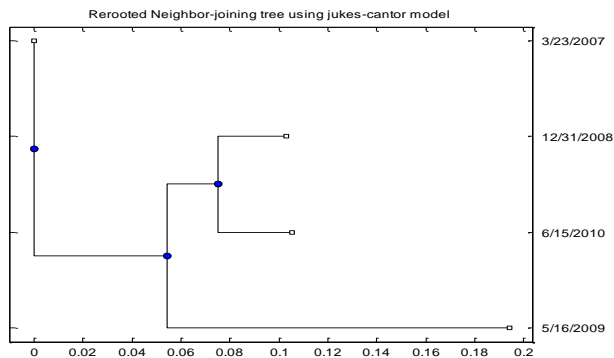


Figure 5 Rerouting Neighbor-Joining Phylogenetic tree

VI. CONCLUSION AND FUTURE WORK

In this paper, the model based framework for estimation of mutation rate of NS5B region of (HCV4a) in Egypt for several years is introduced. The framework is based on building phylogenetic tree from time tagged PHMM models. A learning process of PHMM model is used to select one sequence from all sequences collected in any year. Then scores based on genetic distance between any two sequences of selected sequences is calculated. Estimation of mutation rate of selected sequences is calculated by fitting the scores referred to the 2007 year. The mutation rate is estimated as 0.00004493 substitutions per nucleotide per year. This approach combines both the speed of regression methods and the accuracy of statistical methods for evaluation of virus mutation rate. The future work shall study the mechanism of the virus mutation between all sequences and between the positions.

REFERENCES

[1] P. Simmonds, "Genetic diversity and evolution of hepatitis C virus – 15 years on," *J. Gen. Viro.*, vol. 85, pp. 3173–3188, 2004.

[2] A. Hashad, K. Kamal, A. Fahmy and A. Badr, "Improving Virus C type 4 Interferon using Bioinformatics Techniques," *J. CIS*, vol. 3, no. 6, pp 885-895, 2012.

[3] O. G. Pybus, A. J. Drummond, T. Nakano, B. H. Robertson and A. Rambaut "The Epidemiology and Iatrogenic Transmission of Hepatitis C Virus in Egypt: A Bayesian Coalescent Approach," *Mol. Biol. Evol.*, vol. 20, no. 3, pp. 381–387, 2003.

[4] J. d. Bruijine, J. Schinkel, M. Prins, S. M. Koekkoek, S. J. Aronson, M. W. Ballegooijen, H. W. Reesink, R. Molenkamp and T. J. W. van de Laar, "Emergence of Hepatitis C Virus Genotype 4: Phylogenetic Analysis Reveals Three Distinct Epidemiological Profiles," *J. Clin. Mic.*, vol.47, no. 12, pp. 3832–3838, 2009.

[5] X. Xiong, J. M. Boyett, R. G. Webster and J. Stech, "A Stochastic Model for Estimation of Mutation Rates in Multiple replication Proliferation Processes," *J. Math. Biol.* vol. 59, no. 2, pp. 175–191, 2009.

[6] D. Barash and A. Churkin, "Mutational analysis in RNAs: comparing programs for RNA deleterious mutation prediction," *Briefings in Bioinformatics*, vol. 12, no. 2, pp. 104 -114, 2010.

[7] R. M. Ribeiro, H. Li, S. Wang, M. B. Stoddard, G.H. Learn, B. T. Korber, T. Bhattacharya, J. Guedj, E. H. Parrish, B. H. Hahn, G. M. Shaw and A. S. Perelson, "Quantifying the Diversification of Hepatitis C Virus (HCV) during Primary Infection: Estimates of the In Vivo Mutation Rate," *PLOS Pathogens*, vol. 8, no. 8, 2012.

[8] C. Combet, N. Garnier, C. Charavay, D. Grando, D. Crisan, J. Lopez, A. Dehne-Garcia, C. Geourjon, E. Bettler, C. Hulo, P. Le Mercier, R. Bartenschlager, H. Diepolder, D. Moradpour, J.M. Pawlotsky, C. M. Rice, C. Tre'po, F. Penin, G. Dele'age ., "euHCVdb: the European hepatitis C virus database *Nucleic Acids Research*, vol. 35, 2007.

[9] M. El Nahas, S. Kassim and N. Shikoun, "Profile Hidden Markov Model for Detection and Prediction of Hepatitis C Virus Mutation," *IJCSI*, vol. 9, no. 3, pp. 251-256, 2012.

[10] S. C. Ray, R. R. Arthur, A. Carella, J. Bukh and D. L. Thomas, "Genetic Epidemiology of Hepatitis C Virus throughout Egypt," *The Journal of Infectious Diseases*, vol. 182, pp. 698–707, 2000.

[11] <http://en.wikipedia.org/wiki/2013>.

[12] R. Sanjuán, M. R. Nebot, N. Chirico, L. M. Mansky and R. Belshaw, "Viral Mutation Rates," *J. of Viro.*, vol. 84, no. 19, pp. 9733–9748, 2010.

[13] A. Drummond, O. G. Pybus and A. Rambaut, "Inference of Viral Evolutionary Rates from Molecular Sequences," vol. 54, pp. 331-358, 2003.

[14] <http://bioinf.ncl.ac.uk/molysys/data/Distances.pdf>

[15] M. Kimura, "A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequence," *J. Mol. Evo.*, vol.16, pp.111-120, 1980.

[16] A. Krogh, M. Brown, I.S. Mian, K. Sjölander, P. Haussler, "Hidden Markov Models in Computational Biology," *J. Mol. Bio.*, vol. 235, pp. 1501-1531, 1994.

Exploring Tracer Study Service in Career Center Web Site of Indonesia Higher Education

Renny

Department of Accounting
Gunadarma University
Depok, Indonesia

Reza Chandra

Department of Information Systems
Gunadarma University
Depok, Indonesia

Syamsi Ruhama

Department of Informatics Management
Gunadarma University
Depok, Indonesia

Mochammad Wisuda Sarjono

Department of Information Systems
Gunadarma University
Depok, Indonesia

Abstract— Quality competence of worker the present do not meet labor market criteria and the low level of labor productivity, the lack of communication between the labor market with education, changing of socio-economic structure and global political influence labor market, the development of science and technology very rapidly lead to fundamental changes in terms of qualifications, competencies and requirements for entering the workforce. Tracer Study results can be used by universities to determine the success of the educational process that has been done towards their students. Therefore, universities need a technology services to support the optimization of the use of tracer study. One of that is the use of a website to facilitate the conduct tracer study. Most services tracer study provides information to college, like year graduated, got a job waiting period, the first salary to work, first job, the relevance of the curriculum to the work, and compliance with the major areas of work taken in college. Tracer study feature in Career Center Website affect the popularity website especially in traffic and rich file website.

Keywords— *career center, tracer study, traffic, popularity.*

I. Introduction

The success of the rapid changes in the workplace due to the globalization of the workforce, the revolution in technology and a variety of other disciplines requires anticipation and evaluation of the competencies required by the job.

Evaluation is necessary so that there is no gap between the world of higher education to the world of real work in the community. Some important shifts that occurred include the increase in unemployment of educated, both open and hidden unemployment, as a result of higher education massification, quality competence of workers do not meet labor market criteria and the low level of labor productivity, the lack of communication between the labor market with education,

changing of socio-economic structure and global political influence labor market, the development of science and technology very rapidly lead to fundamental changes in terms of qualifications, competencies and requirements for entering the workforce.

How big college graduates are able to act in accordance with the suitability of development education effort to do a search on graduates (Tracer Study). Tracer Study results can be used by universities to determine the success of the educational process that has been done towards their students.

Tracer study is tracking studies trace graduates / alumni conducted between 1-3 years after graduation and study aims to determine the outcome in the form of the transition from the world of higher education to the world of work, output of education is self-assessment against the control and acquisition of competencies, the educational process in the form of an evaluation of the learning process and the contribution of higher education to the acquisition of competencies and educational input in the form of further excavations to information sosiobiografis graduates.

If tracer studies conducted more than 3 years after graduation, tracer studies have several drawbacks, such as the period of retrospection bias due to information that was too far and the information obtained to be less relevant. If done immediately after graduation, the study called exit study, where the study was not able to see the whole process of transition optimally work for a short time after graduating likely unstable work situation or there may even be graduates who have not found a job.

Therefore, we need a technology services to support the optimization of the use of tracer study. One of that is the use of a website to facilitate the conduct tracer study.

Based on this background, the focus of this study is to look at optimizing the utilization of Tracer Study at universities in Indonesia which has a service career center website.

II. Theoretical Background

Tracer Study is an approach that enables higher education institutions to obtain information about possible deficiencies in the educational process and the learning process and can form the basis for planning activities for the improvement in the future [1].

A tracer study is a graduate or alumni survey that attempts to trace the activities of the graduates or previous students of an educational institution [2].

Tracer study enable the contextualization of graduates of a particular university through a system that is dynamic and reliable in order to determine their life path or movement. It also enables the evaluation of the results of the education and training provided by a particular institution and examines and evaluates the current and future career and employment opportunities / prospects of graduates [3]. Graduates' job titles, years of employment, nature of employment, income levels, and biographical data can be revealed through tracer studies [1].

A tracer study of graduates from the Department of Library and Information Studies at the University of Botswana [4]. The aim of the study was to determine graduates' characteristics, the relevance of their training to their tasks, and their perceptions of the curriculum of the Department of LIS at the University of Botswana. The study revealed that the graduates were employed in traditional library settings. The study also found that their training was relevant to the tasks that they performed, although they advocated the strengthening of the information technology component of the curriculum.

The main objectives of the tracer study were to: investigate the transition process from higher education to: shed light on the course of employment and work over a five year period after graduation; analyse the relationships between higher education and work in a broad perspective which includes the fulfilment of personal goals such as job satisfaction and objective measurement like job position, income, job security and the type of work; find out what factors are important for professional success of graduates taking into account personal factors like gender, work motivation, acquired qualifications during course of study and labour market conditions; evaluate on the basis of the experience and views of graduates, central aspects of the University, including resources, facilities and curriculum and get feedback for their improvement; and identify key aspects of the continuing professional education of graduates, and themes and kinds of courses, including extent, cost, location, reasons for participation, proposals for University courses [5].

In Nigeria a tracer study was done for the Nigerian Teachers' Institute (NTI) which launched its Nigeria Certificate in Education by ODL in 1990 in response to urgent need to train more teachers. The findings of the study were

that the performance of ODL graduates was as effective in the classroom as that of their peers who had studied in the traditional way. Their classroom teaching, lesson preparation, motivation of students, record keeping and communication in English was good. The students themselves rated the instructional materials provided quite highly. However the study revealed some dissatisfaction about the use of audio visual material. It was also thought that teachers needed to be better trained in the techniques of ODL . The Institute itself had improved its management and monitoring systems and efforts had been made to address these inadequacies [6].

A higher education institution (HEI) which strives to provide quality education should strive to fully understand the needs of its learners. One of the best ways to do so is through direct feedback from the learners themselves, particularly those who have successfully gone through and completed their study programmes with the institution. Having gone through the system and graduated from it, they are in a very good position to appraise the quality of education that they have received in terms of preparing them to become more holistic individuals in the workplace [7]

III. Methodology

The sample was 264 universities in Indonesia are included in the ranking of universities based on their activity on the internet, the rank 4ICU and webometrics. Measurements on the first stage is to check whether the college has its own website for the alumni, or herein after called Career Center. Career center is typically an administrative unit of an organization (e.g., school, business, or agency) that employs staff who deliver a variety of career programs and services.

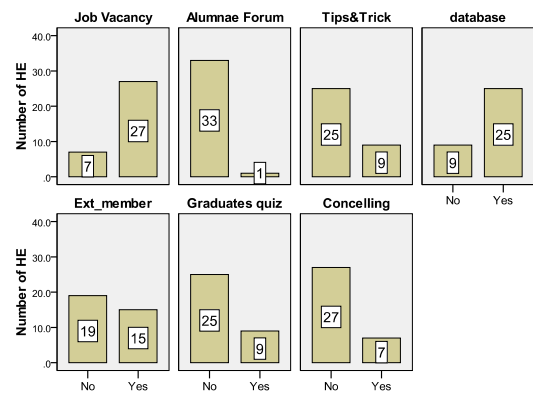


Figure 1. Career Center Website Features

The second stage of 264 samples, which had career center is 34 universities. From 34 universities that have tracer study only 9 universities. The third stage is to examine the features or services that are available in the career website, especially questionnaires alumni.

Observation and measurements conducted research variables in early February 2013. Description of the variables are presented graphically to show utilization the optimization

of Tracer Study in universities in Indonesia. With comparing tracer study service in career center website such as curriculum relevance with the implementation in corporate.

IV. Result and Discussion

A. Contents

Most services tracer study provides information on the year to go to college, year graduated, got a job waiting period, the first salary to work, first job, the relevance of the curriculum to the work, and compliance with the major areas of work taken in college.

Tracer study feature in Career Center Website affect the popularity website especially in traffic and rich file website.

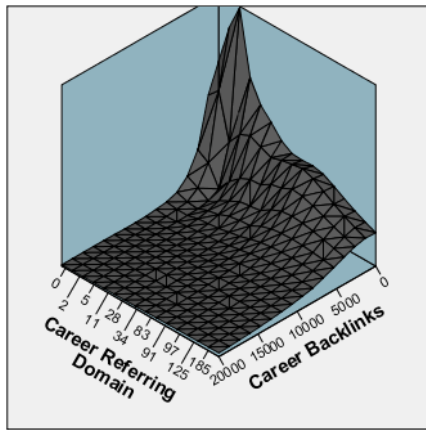


Figure 2. Popularity Career Center Website

Here are 2 examples of web page display face that features a questionnaire tracer or most complete service.



Figure 3. Questionnaires Tracer of Yogyakarta State University (UNY)

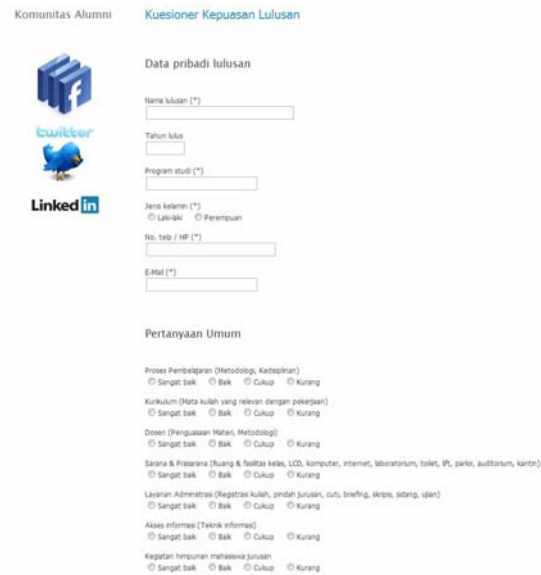


Figure 4. Questionnaires Tracer of Bunda Mulia University (UBM)

Tracer study situation in Indonesia today can be said to still be in the early stages. The analysis conducted by Fikawati and Shafiq [8] indicates that to date information and publications about the tracer study at universities in Indonesia is still very small. The results of the analysis also found that the tracer study in Indonesia vary greatly in terms of clarity of objectives, design and methodology. When compared with the development of tracer study in developed countries, the situation in Indonesia is lagging far enough. In Europe for example, networking tracer study has even produced a large study covering countries in Europe.

V. Conclusion

Universities in Indonesia have not been optimally utilize the Internet for their graduates. Some information published on the website graduates college, or not on a special site with its own sub domain. Most of the Tracer Study still off line. Variety of features or types of service they provide is still not complete. The most popular type of Tracer Study service is a type of work, the waiting period first job, first salary, entrepreneurship graduates and the relevance of the curriculum.

Questionare that are rarely available is question the number of employees in the workplace, job title, employment status, GPA, how many there are applications for employment sent first.

Tracer Study Service should be focused on the enrichment of question about department needs, such as curriculum, competency, skill (hardskill and softskill).

References

- [1] H. Schomburg, Handbook for Graduate Tracer Studies: Centre for Research on Higher Education and Work, Kassel: University of Kassel, 2003.
- [2] Millington, "Wikieducator," Open Education Resource Foundation, 1 June 2010. [Online]. Available: http://wikieducator.org/images/e/e1/PID_424.pdf. [Accessed 25 February 2013].
- [3] L.O. Aina and K. Moahi, "Tracer study of the Botswana library school graduates," *Education for Information*, vol. 17, no. 3, p. 215, 1999.
- [4] L. A. Latif and R. Bahroom, "OUM's Tracer Study: A Testimony to a Quality Open and Distance Education," *ASEAN Journal of Open and Distance Learning*, vol. 2, no. 1, pp. 35-47, 2010.
- [5] S. Zembere and N. Chinyama, "Association of African Universities (AAU)," 1996. [Online]. Available: <http://www.aau.org/studyprogram/notpub/ZEMBERE.pdf>. [Accessed 25 February 2013].
- [6] N. Boaduo, J. Mensah and S. Babitseng, "Tracer study as a paradigm for the enhancement of quality course programme development in higher education institutions in South Africa," University of North-West, Potchefstroom, 2009.
- [7] U. Abdurrahman, "Quality Assurance Procedures in Teacher Education: The case of the National Teachers," in *Towards a Culture of Quality*, Vancouver, Commonwealth of Learning, 2006, pp. 73-84.
- [8] S. Fikawati and A. Syafiq, "Tracer Study Report," CDC-UI, Depok, 2008.

Active Use of ICTs among the Elderly by Positive User Experience

Ayako Hashizume

Faculty of System Design,
Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

Masaaki Kurosu

Center for ICT and Distance Education,
The Open University of Japan
2-11 Wakaba, Mihama, Chiba-shi 261-8586, Japan

Abstract— Recent technological advances have drastically changed our daily life. Information and communication technology (ICT) devices are being used by a wide variety of people to achieve diverse goals in different situations. However, there is an identifiable gap between high-end users and low-end users depending on demographic traits, particularly age. Focusing on ICT usage, we conducted a field survey by using the contextual inquiry method, analyzed the data by applying the modified grounded theory approach, and then summarized the results in a category relationship diagram. We found that motivation, active involvement in communication, and literacy are three principal factors for the use of ICTs.

Keywords- *user experience (UX); elderly people; Qualitative approach; information and communication technology (ICT)*

I. INTRODUCTION

With the advancement and adoption of Information and communication technology (ICT), many devices such as personal computers (PCs) and cell phones have become more convenient and are more frequently used in daily life. Concurrent with these developments, Japan's population is aging rapidly; in 2005, Japan became the most aged society in the world. The percentage of the elderly in Japan (percentage of total population aged 65 and over) has increased from 5% in 1950, 7% in 1970, and 10% in 1985 to 23.1% in 2010. This is expected to rise to 30.5% in 2025 and 40.5% in 2050 [1]. Consequently, Japanese electronics manufacturers have developed cell phones and PCs specifically for elderly users [2] [3]; however, there is still a gap between high-end users and low-end users that can be related to demographic traits, particularly age. While use of ICT has consistently increased for most age groups, the elderly have resisted this trend. Internet usage and PC ownership is relatively low among the elderly compared with young and middle-aged people as most of those under 50 use ICT devices and the internet, but the number declines past the age of 60. According to the Ministry of Internal Affairs and Communications, only about 70% of people in their late 60s and about 40% in their 70s use ICT devices such as PCs and cell phones. In contrast, more than 90% of people between the age of 20 and 50 use these devices [4].

As reported in our previous studies [5] [6], we have conducted a quantitative survey questionnaire to assess the overall ICT usage trends. The responses revealed a generation gap regarding the use of ICTs. Compared with young people, the elderly use fewer ICT device functions and cannot operate commonly used functions effectively, particularly for ICTs. The current study focuses on the reasons for such differences among these age groups in relation to value criteria. In other words, we consider how design (appearance) and usability could affect the purchase of ICTs, particularly by the elderly. In our previous research, we found that there are differences in the relative importance of value criteria depending on age and gender, and that the elderly are not simply disregarding the design of devices (sensuousness aspect) but are placing more emphasis on ease of use (usability aspect). Moreover, we found that they will regard the design as important if there are no usability problems.

On the basis of conceptual analysis, Fig. 1 illustrates how changes in society may affect the lives of the elderly. Because a lack of communication and information appeared to be negative aspects of the everyday life of the elderly, we decided to focus on how the elderly can become more adept at using ICT-related devices and systems, especially cell phones. This is accompanied by an expectation that effective use of such devices and systems may decrease the negative aspects of their lives.

We adopted a qualitative approach using the contextual inquiry method as our research methodology [7].

II. METHOD

A. Survey Method

The results of the questionnaire described in the previous section showed that there is a difference in ICT device literacy depending on age; however, it was not clear why and how there is such a difference between two age groups, i.e., people in their 20s (younger generation) and people in their 60–70s (older generation). Therefore, we conducted interview-based research to identify the reasons. We adopted the contextual inquiry method and conducted interviews in the informants' residential area (home or nearby locations).

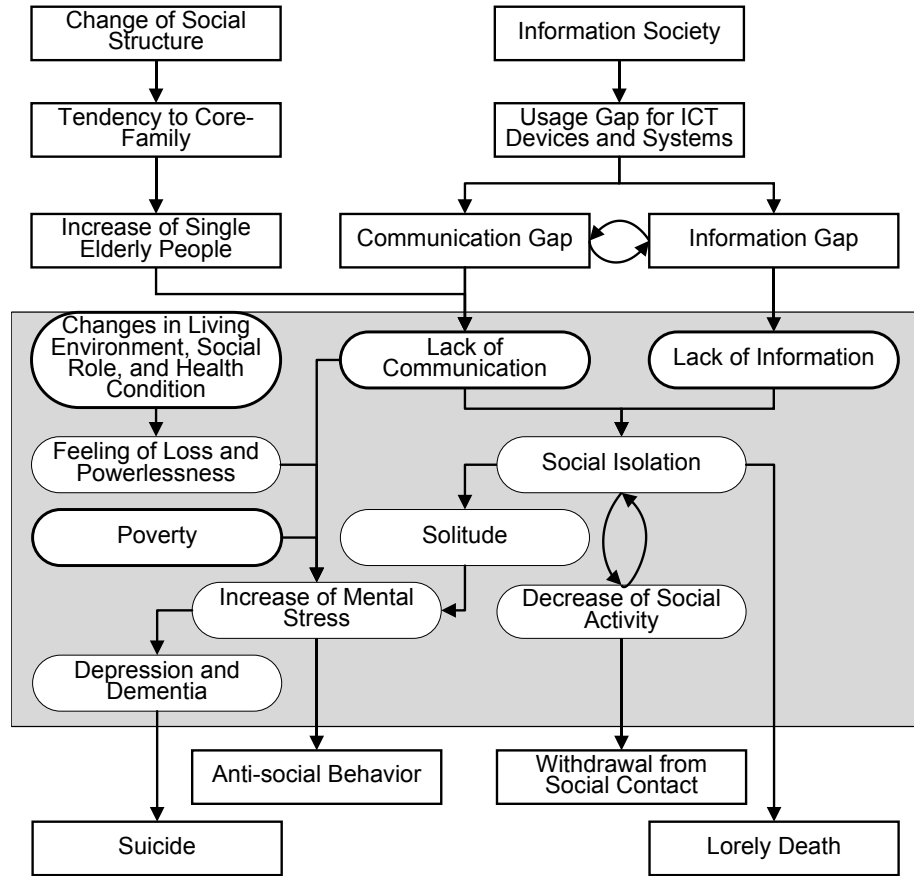


Fig. 1. Changes in society and the lives of the elderly.

We define an elderly person as someone over the age of 60, whereas the World Health Organization defines an elderly person as someone over the age of 65. We chose the age of 60 because this is the approximate age at which people usually retire, and retirement often marks significant changes in lifestyle and environment.

The interviews were conducted for 4 h with two sessions of 2 h each. There were 36 informants; 16 young people (average age was 25.5) and 20 elderly (average age was 69.26). There were an equal number of males and females. Half of the informants lived in an urban area and the rest in a rural area.

B. Instructions and Questions

First, we explained the informants that the research was for academic purposes and was not related to any sales activity. We then explained how long the interview was expected to last, their right to refuse to answer any question, how personal information would be treated, and the purpose of audio recording of the interview. We started the interview after obtaining each person's consent.

Because we adopted a semi-structured interview technique, the order of topics varied from session to session, and in some cases, questions were added as the interview process

progressed. We asked each participant to approximate their ICT usage and provide background information such as life history, current life situation, social relationship status, family membership, occupation, hobbies, value orientation, and view on their quality of life. We asked about their ICT usage including their usage history, current usage, functions used, differentiation of use among various devices and functions, contexts and situations for ICT use, and their action when they encounter a problem while using the ICT.

C. Analysis

The interview data was transcribed and analyzed using the modified grounded theory approach (M-GTA) and MAX-QDA software. M-GTA is a qualitative analysis method proposed by Kinoshita [8] [9]. It is based on the grounded theory approach originally proposed by Glaser and Strauss [10]. To analyze the interview data, we depicted the fragmented information affecting the use of ICT devices. All interview transcripts were analyzed using the processes described below, which are illustrated in Figs. 2–4.

1) *Step-1*: By using specific examples, we identified concepts from each participant's verbatim report. Then, we

clarified the concepts through concept correction and integration (Fig. 2).

2) *Step-2*: We re-examined the concept definitions and names. We then refined and extracted the final concepts (Fig. 3).

3) *Step-3*: We produced and extracted categories, grouped the identified concepts by these categories, and showed the relationships between them in an association chart, as illustrated in Fig.4.

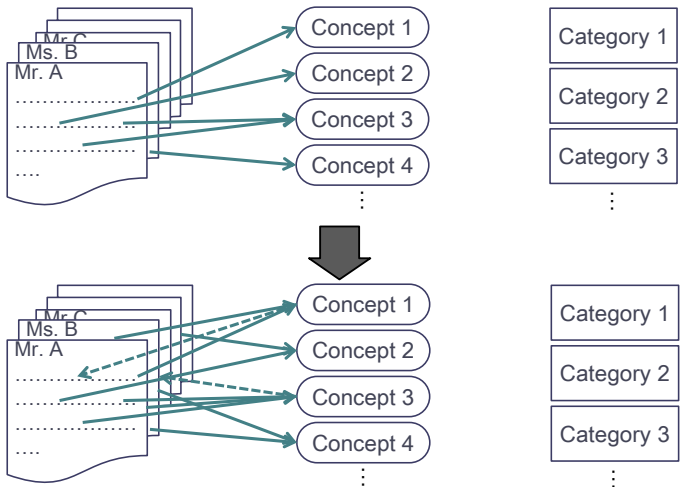


Fig. 2. Analytical Procedure using M-GTA: Step-1

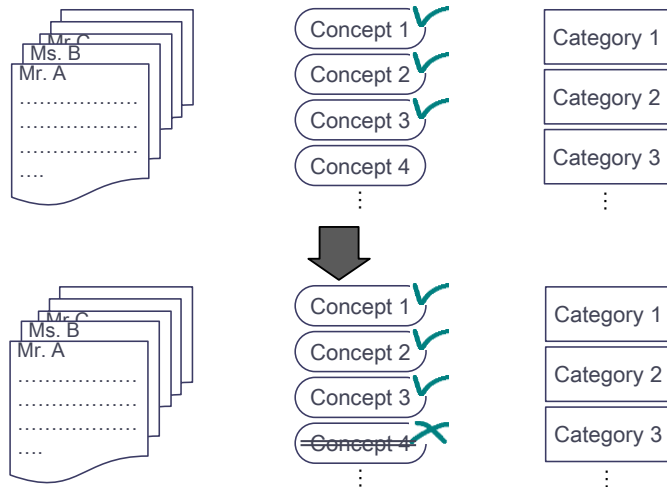


Fig. 3. Analytical Procedure using M-GTA: Step-2

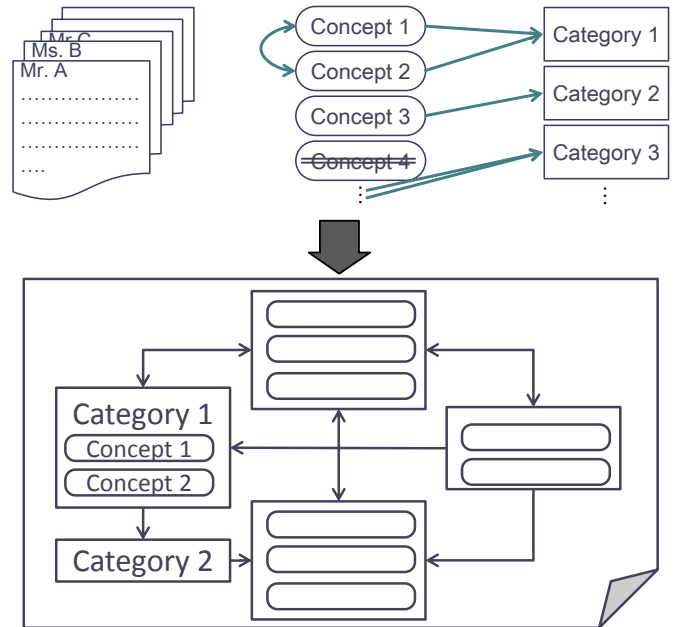


Fig. 4. Analytical Procedure using M-GTA: Step-3

III. RESULT

The results showed some differences in the use of ICT devices between the young and elderly. We produced and extracted concepts from the results of the M-GTA analysis. We identified 21 concepts from the data obtained through the analysis of the interviews with the elderly, and placed these concepts into eight categories. In the same manner, we produced and extracted concepts for young people, and placed 17 concepts into seven categories. Thereafter, we constructed the category relationship diagram shown in Fig. 5.

Among the different categories of relationships, “Emotional aspects of ICT usage” was affected by “Experience using ICT,” and it affected “Motivation for the use of ICT” and “Active attitude to communication.” In other words, a positive emotional experience played a key role in the active use of ICT devices. Furthermore, the category “Emotional aspects of ICT usage” included positive user experience concepts such as “Happiness and security experienced through usage” and “Realizing how convenient ICT devices are.” Other concepts such as “Troubles during usage” are negative user experiences.

An effective strategy to increase usage among the elderly must first increase motivation and then focus on increasing the level of literacy. For the elderly, it is especially important that the user experience is based on a positive user experience. Therefore, to increase the active use of ICT devices among the elderly, it is important to increase these positive user experiences, leading to higher levels of ICT literacy.

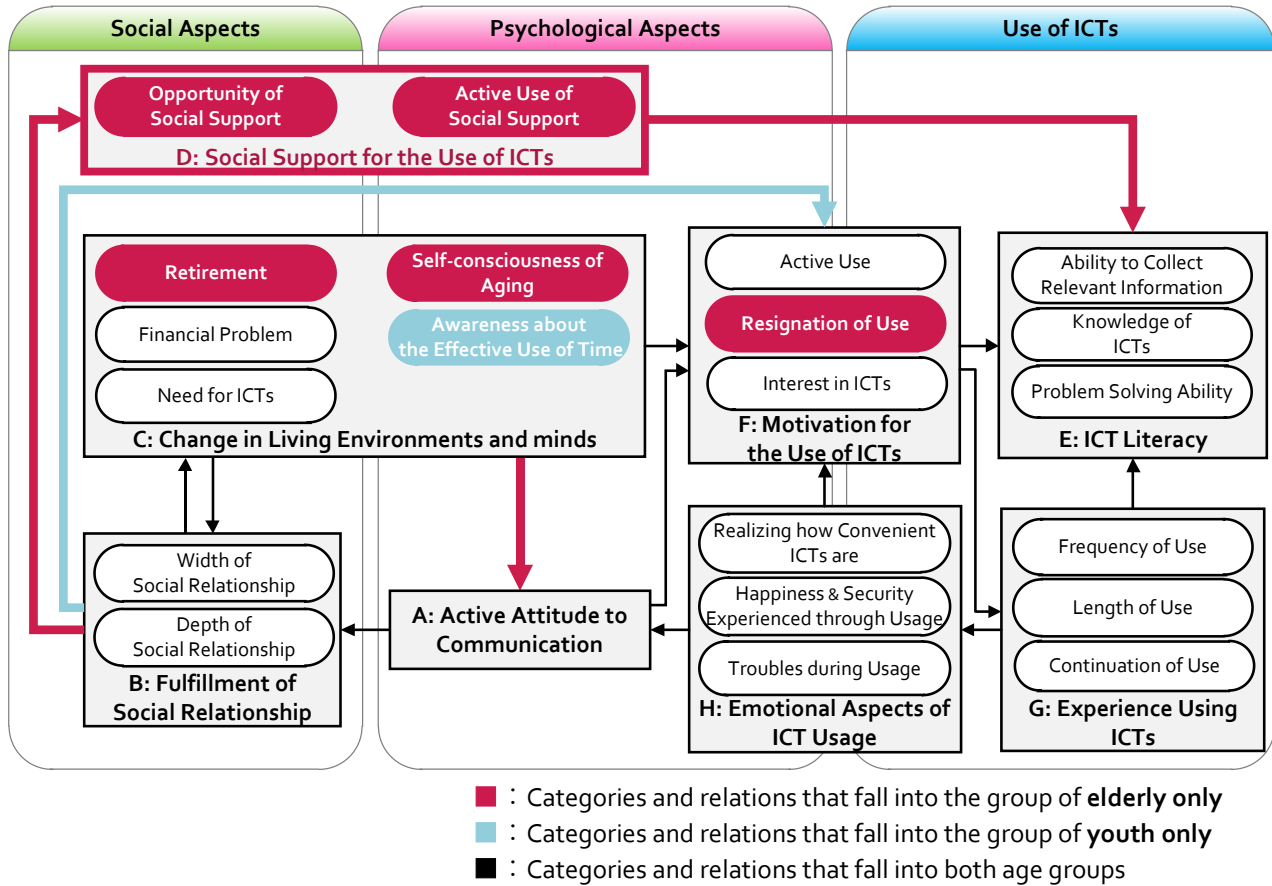


Fig. 5. Category Relationship Diagram

IV. CONCLUSION

The findings based on the data analysis are as follows: to increase ICT literacy among the elderly, motivation and actual usage experience are quite important. If the elderly have adequate social support and the problems they may be facing can be resolved, ICT literacy will be increased. The factors that motivate the use of ICT devices include an active attitude toward communication and positive user experiences. In contrast, the factors that discourage the use of ICT devices include a change of life circumstances due to retirement, decreased need to use ICT devices, and decreased motivation for active communication due to increased self-consciousness related to age. However, if retirement is regarded positively as the attainment of freedom, the elderly can be expected to possess active motivation. In addition, as they may have more free time than when they were working, interest in ICT devices may increase. In general, positive changes in life situation and self-awareness will positively influence social relationships. This fulfillment of social relationships is related to opportunities to receive social support and their willingness to accept such support.

The results of this study indicate that improving device usability is quite important, especially for elderly users. However, because there are different levels of ICT literacy among the elderly, user interfaces should be optimal for all literacy levels. In addition to deliberate considerations of usability, sensuousness aspect should also be considered. In Japan, many cell phones have been developed with the intention of increasing effectiveness and efficiency of use by the elderly. These devices were developed considering the physical characteristics of the elderly; however, a broader level of usability developed specifically for the elderly has not yet been completely optimized.

It is difficult to answer “why” questions by simply applying a quantitative method such as a questionnaire. By applying a qualitative approach such as contextual inquiry, it is possible to understand the user’s situational structure and context of use to provide a better understanding of the results from a quantitative approach. Such an overall approach will surely provide insight into the design of high-tech devices such as cell phones.

ACKNOWLEDGMENT

The authors would like to thank the Japan Society for the Promotion of Science (JSPS). This work was supported by a JSPS Grant-in-Aid for JSPS Fellows.

REFERENCES

- [1] Cabinet Office Japan, Annual Report on the Aging Society, 2011.
- [2] Irie R., et al., The Challenge to the Universal Design in the Development of the Cell Phone "Easy Phone (Raku-Raku Phone)," FUJITSU, 56(2), 2005 (in Japanese).
- [3] <http://www.fmworld.net/fmv/rakuraku/>
- [4] Ministry of Public Management, Home Affairs, Posts and Telecommunications "A Report on the Use of Communication System in 2009," 2010.
- [5] Hashizume A., Kurosu M., Kaneko T., Use of Cell Phone by Senior Users Compared to Young Users, The 12th World Multi-Conference on Systemics, Cybernetics and Informatics Proceedings, Florida, USA, pp. 244-247, 2008 (in English).
- [6] Hashizume A., Kurosu M., Kaneko T., The Choice of Communication Media and the Use of Mobile Phone among Senior Users and Young Users: Computer-Human Interaction, Springer-Verlag Berlin Heidelberg, pp. 427-436, 2008.
- [7] Beyer H., Holtzblatt K., Contextual Design: Defining Customer-Centered Systems, Morgan Kaufmann, 1998.

- [8] Kinoshita Y., M-GTA Grounded Theory Approach: Reproduction of a Qualitative Experimental Study, Kobundo Publishers Inc., 1999.
- [9] Kinoshita Y., M-GTA Grounded Theory Approach: Invitation for Qualitative Research, Kobundo Publishers Inc. 2003.
- [10] Glaser B. G., & Strauss A. L., The discovery of grounded theory, Aldine de Gruyter, 1967.

AUTHORS PROFILE

Ayako Hashizume received thea Ph.D. in Kansei Science from Tsukuba University, Japan, in 2011. Her research interests include psychology of emotion and Kansei, engineering in relation to HCD (Human Centered Design) and UX (User Experience), especially among elderly users. She was awarded the Kids Design Prospective Competition 2008 from the Japan Kids Design Association and the Best Design Prize at the Design Marathon 2009 from IAUD from the International Association for Universal Design.

Masaaki Kurosu is a professor at the Open University of Japan, since 2009. He is also a professor at the Graduate University for Advanced Studies as well as the President of HCD-Net (Human Centered Design Network). On the basis of his career as a usability professional in industry and academia, he proposed a concept of user engineering and is now proposing an ADA (Artifact Development Analysis) and a new concept of UX (User Experience). He is a member of IFIP/TC13, and ISO TC159/SC4/WG6. He is also a member of ACM SIGCHI, Usability Professionals' Association, IEEE Computer Society, Human Interface Society (Japan), Japanese Psychological Association, Japan Ergonomics Society, and Japanese Society for the Science of Design.

Feasibility Study of Millimeter Wave Transmission

Mamta Agiwal

Research Scholar at Jain University,
Bangalore

Dr. Fathima Jabeen

Professor and P G Coordinator, K S
School of Engineering and
Management, Bangalore

Kashif Ahmed

Assistant Professor, Dept. of EEE,
CMRIT, Bangalore

Abstract— Past few years have witnessed the stupendous growth in wireless communication networks. However much research on wireless communication has focused on Power consumption and the frequency reuse of spectrum in the range of 300 MHz–3 GHz, while the other direction could be in considering mm waves for transmission with a wave band of 30 GHz-300 GHz. Integrating of new technologies of optical fibers, mesh networks, improved CMOS platform and enhanced antenna design can open plethora of opportunities to use mm waves for transmission. The paper discusses advantages, limitations, possibilities and hardware developments to support transmission using mm waves.

Keywords—mm-wave; rain losses; mesh network; narrow beam and frequency reuse; CMOS platform.

I. INTRODUCTION

Wireless communications has emerged as one of the largest sectors of the telecommunications industry. Wireless data usage has increased at a phenomenal rate and demands the need for continued innovations in wireless data technologies to provide more capacity and higher quality of service.

Today there are billions of cell-phone users in the world. This enormous rise in wireless phone communication has been possible because of an enormous cost-reduction to cell phones despite their sophisticated hardware and software capabilities. The mobile phones today are also already equipped with sensing platforms, advanced digital imaging, advanced audio quality, HD video streaming etc that can be utilized for various applications. Faster mobile broadband connections, more powerful smart phones, connected tablets; networked laptops as well as new consumer and enterprise applications are all driving the wireless industry to provide new technical capabilities. Mobile communication has been one of the most successful technology innovations in modern history.

In order to meet this exponential growth, improvements in air interface capacity and allocation of new spectrum are of paramount importance. All cellular mobile communication requires ultra high frequency band of radio spectrum (collection of various types of electromagnetic radiations of different wavelengths is called as spectrum). The radio frequency spectrum is a limited natural resource. Variety of services like fixed communication, mobile communication, broadcasting, radio navigation, radiolocation, fixed and mobile satellite service, aeronautical satellite service, radio navigational satellite service etc work in the range of radio frequencies below 3 GHz.

This band of spectrum is therefore becoming crowded because of enormous growth of mobile services. However the millimeter wave spectrum at 30-300 GHz can be exploited for commercial applications, in order to meet the growing demand of data traffic at improved efficiency.

For network upgrade the operators require access to additional spectrum, because the capacity in the network is determined by the amount of spectrum. Spectrum is a finite, non-exhaustible common resource which influences the valuation, and some parts of the frequency band are more valuable than others. Almost all mobile communication systems today use spectrum in the range of 300 MHz–3 GHz.

A millimeter-wave mobile broadband (MMB) system has a candidature for the next generation mobile communication system. IT is the region in electromagnetic spectrum usually ranging from 10 millimeter to 1 millimeter. mm waves are longer than infrared waves, however they are shorter than radio waves .Millimetre-wave (mm-wave) band corresponds to 30-GHz ~300GHz, about 270-GHz bandwidth, which is ten times the bandwidth in Centimetre-wave band (3-GHz~30-GHz).Millimetres wave can be utilized for a variety of applications which involve large amounts of computer data transmission, wireless communications, and radar.

In this paper we discuss the feasibility of using radio frequency spectrum above 3 Ghz in supporting applications such as high speed data transmission and video distribution for wireless applications.

II. HISTORY OF MM WAVES

Millimeter wave technology goes back to the 1890's experiment on millimeter wave signals by J.C. Bose.

The early research paved the way for applications of mm wave technology in the field of Radio Astronomy. Satellite-based studies of the upper atmosphere, Climate, rainfall and vegetation patterns, and a host of other environmental concerns.

In the late 1970s, a millimeter wave radiometer began service aboard a NASA aircraft, where it monitored storm activity from an altitude of 60,000 feet. Scanning about 5,000 miles of atmosphere per hour, the device recorded the emitted and reflected energy of storms, including the almost infinitesimal amounts of energy emitted by moisture inside a storm.

This was followed by applications in the military. In the 1990's, the advent of automotive collision avoidance radar at 77 GHz marked the first consumer oriented use of millimeter wave frequencies above 40 GHz.

III. MORE FIRSTS

Georgia Tech scientists also achieved a number of firsts in millimeter characterization of clutter and targets - essential data for reliable millimeter radar systems. Since the 1960s, more than a dozen projects have provided millimeter measurements of the ocean, rain, snow-covered ground, desert, foliage and foreign military vehicles.

In the 1980s, researchers conducted a comprehensive study of the image-quality effects of atmospheric turbulence and precipitation on millimeter wave propagation.

More recently developing markets include consumer satellite communications that bring broadband Internet access to businesses and rural consumers, wireless broadband media transfer within the home, automotive radar for tasks such as adaptive cruise control and collision avoidance, and telecommunications links that are approaching the performance of optical fiber but at a fraction of the cost.

Millimeter wave security imaging, such as that used to screen airline passengers and personnel at other checkpoints, is undergoing deployment at airports and businesses, where it is used for loss prevention and inventory control. Systems are even commercially available for retail clothing shoppers to conduct body measurements to determine clothing sizes and recommend appropriate products and brands.

IV. ADVANTAGES OF MM WAVE TRANSMISSION

A. Huge Spectrum Availability

A 250 GHz bandwidth is available in the mm wave band (30 GHz-300 GHz). This is almost 1000 times higher than the frequency range used these days. The availability of such high carrier frequencies facilitates more data rates by using amplitude, phase or frequency modulations. It can also be reliable for data transmission at GBPS rates.

B. Small Component Size

For MM waves the wavelengths are shorter and therefore the frequencies are high. Therefore, the antenna systems required for the transmission can be of the millimeter size. This also enables densely packed communication link networks integrating high efficiency radiating elements at the millimeter scale, leading to compact, adaptive and portable integrated systems. Even arrays of antennas may be packaged within the area of a quarter for directionally transmitting and receiving radio signals.

C. Improvement In The Directivity

The other advantage of smaller antenna size and reduced packaging is the improvement in the directivity. About 25dBi of directivity can be comfortably achieved because the nodes have compact form factor as compared to wireless access point (this is because of small wavelengths).

Radar is an important use of millimeter waves, which takes advantage of another important property of millimeter wave propagation called beamwidth. Beamwidth is a measure of how a transmitted beam spreads out as it gets farther from its point of origin. In radar, it is desirable to have a beam that stays narrow, rather than fanning out. The use of millimeter-length has allowed engineers to overcome antenna problem. For a given antenna size, the beamwidth can be made smaller by increasing the frequency, and so the antenna can be made smaller as well.

D. Narrow Beam and Frequency Reuse

Millimeter wave links cast very narrow beams, as illustrated in Fig. 1. This allows the deployment of many independent links in close proximal distances.

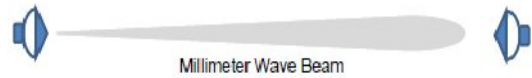


Fig. 1 Millimeter Wave Beam

Therefore along with the huge and unexploited bandwidth availability, smaller range and narrow bandwidth facilitate for a higher degree of frequency reuse. mm wave limits the propagation to a few kilometres, thus they are useful for densely packed communications networks. Furthermore the high oxygen absorption in mm range of frequencies gives a large frequency reuse factor. Example: For 60 GHz, the working range for a typical fixed service communications link is of the order of 2 km, and therefore the other link could be employed on the same frequency if it were separated from the first link by about 4 km. IN contrast, at 55 GHz, the working range for a typical fixed service link is about 5 km, but a second link would have to be located about 18 km away to avoid interference.

E. Technology Availability

Millimeter wave technology has a strong history and technological evolution behind it. Properties of millimeter wave propagation have been well researched and documented. Millimeter wave technology has reached a level of maturity comparable to older forms of radio technologies. Also; the carrier frequency is high, so expensive compound semiconductor technologies such as GaAs were the only choices earlier. But now some companies have demonstrated that the chipsets can be manufactured with silicon-based technologies.

V. LIMITATIONS OF MM WAVE TRANSMISSION

A. Atmospheric Gaseous Losses

When the mm waves are transmitted through the atmosphere they are absorbed by molecules of various gasses and water vapours. At the resonant frequencies of the gas molecules these losses are very high and the absorption results in high attenuation of signals. The transmission can be

effective if spectral regions between the absorption peaks are used for propagation.

B. Rain Losses

Rain greatly affects the mm wave propagation. The rain drops being same in size with the radio wavelengths causes large but slow changes in strength of radio signal. Example: A rain rate of 2.5mm/hr yields 1 db/km attenuation while a rate of 25mm/hr results in 10 db/km attenuation. During rainy season where rain rate high there can be loss of communication upto tens of dBs per km. In other words increasing rain factor reduces the availability of communication signals.

C. Foliage Losses

Foliage losses can change the attenuation rate substantially. Example: At 80GHz frequency and 10 meters foliage penetration, the loss can be about 23.5dB which is about 15dB higher compared to the loss at 3GHz frequency. The transition because of heavy foliage can be abrupt and leads to beam broadening (and depolarization) after transition has occurred. This can limit impairment for propagation of mm wave transmission. There can be a significant change in attenuation over the same transmission paths, under summer and winter conditions, i.e. with tree in leaf and without leaves.

D. Free-Space Loss and Limited Communication Range

Weakening of mm wave signal due to line of sight path through air is termed the free space path loss. Even a shorter distance leads to a high free space loss for mm waves. Small change in range causes 6dB of change in the attenuation. Obstacles like human body causes significant drop in received power of the signal. Moreover for long distances wireless mm wave signal nullifies gain of antennas. Also ability of signal to bend around edges of obstacle is very weak.

VI. POSSIBILITIES

A. Fiber Optics Links

Research is being done to transmit mm waves using fiber optic links. This will exploit the advantages of both optical fibers and mm-wave frequencies. Fig. 2 gives the architecture of mm-wave RoF system. Central Station (CS) and distributed Base Stations (BS) can be linked with optical fibers. BS can be designed to communicate with Mobile Terminals (MT) by wireless signals at mm-wave band.

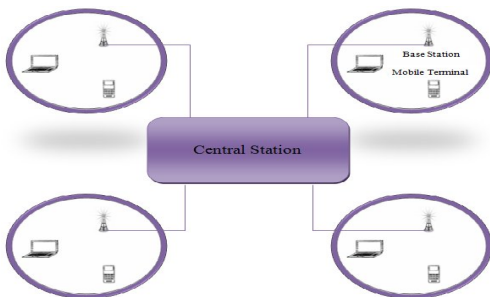


Fig. 2 Architecture of mm-wave RoF System

This system requires generation of low noise mm wave signals to overcome the effects of fiber chromatic dispersion. Base station can be made light wave to mm wave converter and the signal processing can be handled at the central station.

B. Mesh Networks

Outdoor mesh networks with multiGigabit links at relatively short ranges can be designed using mm wave technology. Such mm wave mesh networks can support a high-speed broadband connectivity. Networking can be based on multihops operating over lower frequency mm bands. The directional transmission can be incorporated to improve the connectivity of ad-hoc networks by establishing long-range links even without using smart beam steering.

Survival rate of millimeter-wave mesh is limited due to their severe weather conditions, like precipitation and humidity. Spatially correlated links of a mesh network should use routing protocols to route around the failures. This will increase dependability when compared against existing routing methods.

VII. HARDWARE DEVELOPMENTS

A. Multimeter Wave CMOS Platform

Since carrier frequency of mm wave is very high expensive GaAs technology was the only option until recently. Evolution of nm technology in CMOS has made possible the designing of low cost 24 – 60 GHz mm wave signals using silicon. Combining CMOS technology with FR4-based packaging technology ensures successful deployment of ultra high-speed and high capacity 60 GHz WPAN at minimal cost. IBM proved that silicon based technologies could be the solution for manufacturing chips at reduced cost, power and form factor than GaAs technology. For efficient data transmission at small cost and at low power consumption high speed coding techniques and signal processing can be used. The chip set including antenna can become small and affordable in near future.

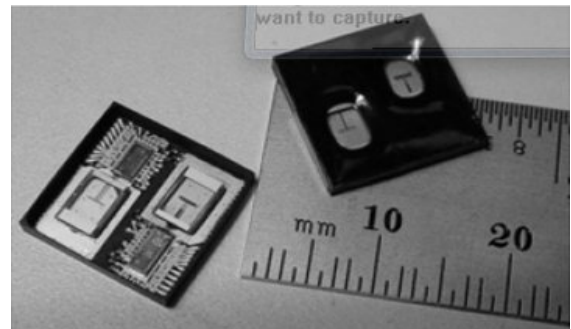


Fig. 3 IBM's mm-wave Transmitter and Receiver Chip

B. Antenna Technology

The major attractiveness of the millimeter-wave is the small wavelength. This allows deployment of many radiating elements in an array configuration which will occupy limited space. Compact multi-sector phased-array architecture can overcome range limitations of millimeter-wave signal

propagation. The sectored design can either be integrated on a one large panel or in a compact module containing an embedded filter and antenna phased array. Liquid Crystal Polymer has emerged as a promising low cost alternative for millimeter-wave module implementation. High Gain Adaptive phased array technology and multi sectored configurations can provide extended range and better elevation coverage. These can be exploited for commercial development of mm wave system.

VIII. CONCLUSION

The article discusses possibilities of using mm waves (30 GHz to 300 GHz) for wireless communication. Though mm waves have limitations of scattering and range their advantages cannot be overlooked. They have an enormous potential to be used for high data and high speed wireless communication. The developments in CMOS platforms and antenna design along with optical fiber and mesh network technology can make this high data transmission feasible.

REFERENCES

- [1] Michael Marcus and Bruno Pattan, "Millimeter Wave Propagation: Spectrum Management Implications," *IEEE Microwave Magazine*, pp. 54-62, June 2005.
- [2] R. Mudumbai, S. Singh, and U. Madhow, "Medium Access Control For 60Hz Outdoor Mesh Networks With Highly Directional Links," *INFOCOM 2009, IEEE*, pp. 2871-2875, 2009.
- [3] Theodore S. Rappaport, A. Annamalai, R. M. Buehrer, and William H. Tranter, "Wireless Communications: Past Events And A Future Perspective," *IEEE Communications Magazine*, pp. 148-161, May 2002.
- [4] David Wikner, "Millimeter-Wave Propagation Measurement Through A Dust Tunnel," *ARL-TR-4399*, March 2008.
- [5] "Understanding Millimeter Wave Wireless Communication," *Loea Corporation*, 2008.
- [6] Y. Katayama and B. Gaucher, "One-Minute Introduction to mmWave Technology and Applications," *IBM Research* [Online]. Available: <http://www.research.ibm.com/mmwave>.
- [7] J. Laskar, S. Pinel, D. Dawn, S. Sarkar, B. Perumana, and P. Sen, "The Next Wireless Wave Is A Millimeter Wave," *Microwave Journal*, vol. 50, no. 8, pp. 22, August 2007.
- [8] Wayne Pleasant, "Gigabit MM Wave Comm," *IEEE T802.16-02/02a*, March 2002.
- [9] T. Korakis, G. Jakllari, and L. Tassiulas, "Cdr-mac: A Protocol for Full Exploitation of Directional Antennas in Ad Hoc Wireless Networks," *IEEE Trans. Mob. Comput.*, vol. 7, no. 2, pp. 145-155, Feb. 2008
- [10] H. Koskinen, "Analytical Study of Connectivity in Wireless Multihop Networks Utilising Beamforming," *Proc. ACM MSWiM '06*, 2006, pp. 212-218.
- [11] R. R. Choudhury, X. Yang, R. Ramanathan, and N. H. Vaidya, "On Designing mac Protocols For Wireless Networks Using Directional Antennas," *IEEE Trans. Mob. Comput.*, vol. 5, no. 5, pp. 477-491, 2006
- [12] K. S. Yang, S. Pinel, I. K. Kim, and J. Laskar, "Low Loss Integrated Waveguide Passive Circuits Using Liquid Crystal Polymer System-on-Package (SOP) Technology for Millimeter-wave Applications," *IEEE Trans. On Advanced Packaging*, vol. 54, no. 12, pp. 4572-4579, 2006

Evaluation of Watermarking Approaches for Arabic Text Documents

Muhammed N. Kabir, Omar Tayan and Yasser M. Alginahi

IT Research Center for the Holy Quran (NOOR)
College of Computer Science and Engineering
Taibah University
P.O Box 344

Al-Madinah Al-Munawwarah, Saudi Arabia

Abstract — The embedding of digital watermark data in a cover text media, as in digital text watermarking, is used in numerous applications such as copyright protection, tamper detection, content-authenticity, steganography and other applications. Such issues have been largely studied for the security and protection of digital English-texts with relatively few citations found that address the specifics of other language characteristics. Moreover, the predominance of the text as a communications medium over the Internet suggests that more attention is required to protect online textual data in languages other than English. Hence, the purpose of this paper is to identify the properties of different watermarking applications for the case of Semitic languages with particular focus on Arabic-text documents by evaluating three invisible text-watermarking approaches; Kashida-based, space-based and sukun-based watermarking, the latter two of which present newly proposed watermark encoding schemes not found in the Arabic-text literature. This paper investigates the effect of two parameters on the watermarking scheme used, including; the word-group set-size, and, the number of bits embedded per set, before examining their consequent impact on the capacity and imperceptibility properties of the watermarking scheme on the host cover-text for different applications. Experimental results had illustrated the effect of the two encoding parameters on the resultant watermarking properties. It was found that by adjusting those variable watermark parameters, any target Arabic-text application could be optimized to achieve a desired capacity-ratio and level of imperceptibility.

Keywords-component; text watermarking, copyright protection, Kashida; diacritics, sukun.

I. INTRODUCTION

Digital watermarking of any media, such as text media, is considered a branch of steganography and its main objective is to provide copyright protection for intellectual property and prevent illegal copying and diffusing of documents.

Digital watermarking can be used in a wide range of applications such as, copyright protection, copy protection, source tracking, automatic monitoring and tracking of copyright material on the web and fingerprinting applications [1]. Both steganography and digital watermarking employ steganographic techniques to embed data; hence,

steganography aims for imperceptibility to human senses and digital watermarking tries to control the robustness as a top priority. In each case, data is embedded in the cover-text with no resultant degradation or access control to the transmitted data [2].

A watermarking system goes through three stages [3], starting from the generation and embedding of watermarks followed by the dissemination stage during which a document may be exposed to modifications during transmission, and finally, the extraction of watermarks as a means of copyright protection. Hence, this can be achieved by either making the watermark visible or invisible in the host cover-text.

In this paper, the evaluation of capacity and imperceptibility watermark properties will be investigated by adjusting the following parameters: set size for word-grouping and number of bits embedded per set. Watermarking capacity is the amount of information or length of embedded message that can be embedded in a cover media. Hence, high capacity is commonly obtained at the expense of either robustness strength or imperceptibility, or both.

Robustness is the ability to detect or extract a watermark after common signal processing operations which may occur between the time of embedding and the time of detection at the receiver side. Therefore, a digital watermark is considered to have a high robustness if it can be successfully extracted following any modification, manipulation or transformation of the encoded document. “Robust watermarks may be used in copy protection applications to carry copy and no access control information” [4]. Finally, imperceptibility is when both the original cover media and the watermarked media are perceptually indistinguishable and perceptible when the appearance of the watermark is recognizable within the encoded document.

Section 2 presents the background on related work, followed by the methodology in section 3. The results and discussion are presented in section 4, and finally, concluding remarks are given in section 5.

II. RELATED WORK

With the advancements in digital technology, the use of watermarking in text documents has seen an increase in the last two decades or so. The watermarking techniques from the literature can be divided into four main categories; image-based, syntactic-based, semantic-based and zero watermarking schemes. The image based techniques concentrate on the fine-tuning of document structures, such as line-shift encoding, word-shift encoding or characteristics coding. In syntactic watermarking techniques, watermarks are embedded in the syntactic structure of the text. In semantic watermarking techniques, the semantics and language properties are exploited to embed watermarks, and finally, zero watermarking techniques provide robust techniques which do not alter the text during the process of encoding the watermark [5]. In this paper, techniques pertaining to Arabic language will be discussed.

In the application of Arabic text steganography, [6] uses the possible extendable characters located within the word to hide secret data bits by inserting Kashidas (redundant Arabic character extensions) which have no effect on the Arabic text. This approach, [6], shows that using two Kashidas per word provides adequate capacity ratio and more security compared to other techniques in literature as claimed by the authors in [7 - 8].

The method in [7] utilizes the Arabic-language notation of dots (points) to hide the embedded bits. In the case of Semantic languages, the Persian language has 18 pointed characters from 32, whilst Arabic has 15 pointed characters from 28. A bit of '1' is hidden by shifting the dot above its normal position whereas hiding a bit of '0' is represented by keeping the dot position unaltered. This method provides a high capacity since each pointed character can hide a bit '0' or '1'. However, it is noted that retyping or applying OCR software causes this embedding information to be lost. Since this method is font-dependent, therefore, producing a negative robustness effect.

The approach in [8] presents a steganography method useful for Semantic-language scripts (languages such as Arabic, Urdu, Persian ... etc.). Such a method is classified as feature-coding based, which is implemented by exploiting the existence of the Kashida character. The proposed method uses pointed letters with a Kashida to represent bit 1 and un-pointed letters with Kashida to represent bit 0. This method is implemented in two ways by adding the Kashida before letters or after letters. The authors claim that this method offers security, capacity and robustness which are the three essential properties for hiding information. However, the security of such an approach could be reduced when using words with many possible Kashida insertion locations to represent secret bits, since this could reflect the appearance of hidden information in the cover text. The technique proposes additional security to mislead any trespassers by adding

extension characters before and after the letters within the same document.

Other techniques based on Kashida insert the Kashida according to specific rules such as the technique proposed by the authors in [9] in which the watermark key is predefined. Here, a Kashida is placed for a bit '1' and omitted for a bit '0'. By going through the document, Kashidas are inserted before a specific list of characters () until the end of the key is reached. If the end of the document is not reached then the key is re-embedded in a round robin fashion for the remainder of the document. Re-embedding the key circularly in a round-robin fashion (circular-embedding) ensures that all word-groups are included in the encoding to minimize the effect of isolated signal-processing operations or isolated attacks, and therefore increasing the robustness and security of the encoded document.

The use of diacritics is very important in Arabic writing in order to differentiate between words with similar spellings, since a single diacritic could completely change the meaning of the word. However, the use of such diacritics in standard written Arabic language is optional. The use of diacritics related to Arabic text is noticed in steganographic applications [10 - 11]. In such applications, diacritics are shown to have high capacity for security purposes, however, are rarely used, and are mainly found in online, formal and religious texts such as digital Quran texts. Therefore, their use in regular digital Arabic documents is not otherwise widely adopted, with their use in common/other documents resulting in security implications and vulnerability to increased attacks.

Other techniques found to be suitable for Arabic texts include zero-watermarking techniques, whereby, the publisher's key is generated by extracting properties from the text without any modifications on the cover-text; therefore, the watermark is not physically embedded in the text, but rather, the characteristics of the host document are used to generate new key. Zero-watermarking techniques have not been widely exploited and their use could definitely be applied to formal/religious and sensitive texts where no modifications to the text can be tolerated. Examples of zero-watermarking techniques are found in [5] and [12].

In this paper, three different embedding approaches will be applied on Arabic text documents in order to investigate the effect of varying the embedding parameters (variable bits embedded per set and set sizes) on different watermarking requirements. The three methods are, the Kashida-based technique proposed in [8], a simple proposed method which inserts spaces between and /or the end of words and finally, a new proposed diacritic-based watermarking technique which will be explained in the next section. However, the methods used in this work are based on using the Unicode standard.

III. METHODOLOGY

The main objective of this work is to investigate/evaluate the effect of chosen parameters (number of embedded bits/set, and set size) on the different watermarking properties that include; capacity and imperceptibility required for various host applications. The process of this evaluation is as follows:

- Three different watermarking methods are chosen for embedding within the host-document.
- The size of the watermark key is chosen to be 48-bits representing the Unicode for the sample three letter Arabic word ().
- A large size document was used to test these requirements so the watermark key can be inserted at least once, and circularly embedded if required.
- Each sample of the original document is divided into sets of words starting with 1 word/set, followed by 2 words/set up to 9 words/set. The size of the set refers to the number of words in the set within the cover-text.
- The number of embedded bits refers to the number of bits (from the publisher's key) to be inserted into each set.
- The embedding mode refers to the embedding method used such as spatial (where the key is embedded only once in the entire text) or circular (where the key may be re-embedded more than once throughout the document when the publishers key size < size of the host document). The results for the three methods are given in section 4.

The average percentage capacity is then calculated as: $(T_k/T_c)*100$, where, T_c is the total number of characters in a document, and, T_k is the number of Kashidas inserted into the document according to the watermark key used.

The average number of Kashidas per word is related to imperceptibility: a lower average number of Kashida's embedded per word results with higher imperceptibility, and therefore, the closer the perceptual similarity of the encoded document compared with the original cover-text. The three methods employed are now described as follows:

Method 1: here, the method in [8] was adopted in which a Kashida is inserted after pointed letters to represent a 'bit-1' and no Kashida is inserted at the identified position when representing a 'bit-0'.

Method 2: is a simple method for text document watermarking, which inserts spaces between and/or at the end of words in a set to represent a 'bit 1', with no space being inserted at the position when representing a 'bit 0'. This is similar to word-shifting for image based watermarking techniques [13]; however, the technique used in this work is based on inserting the space as a Unicode-value, and not shifting words horizontally as is the case in image documents.

Method 3: uses the Sukun diacritic, which has the following symbol (◌ْ); a circle-shape usually placed on top of letters. In this technique, the Sukun is removed from the top of a letter in the cover-text to represent a 'bit 0' and is kept unchanged to represent a 'bit 1'. The Sukun indicates that the consonant to

which it is attached is not followed by a vowel. The Sukun on any letter does not add any value to the character, since, in standard Arabic, writing a Sukun or omitting it produces the same pronunciation. This is the only Arabic diacritic which does not have an added effect in pronunciation of the word.

IV. RESULTS AND DISCUSSION

The three methods explained above were applied to a large Arabic document. Figures 1 – 3 show the results of the average allowable embedding capacity when applying the three methods respectively on Arabic-text documents. In addition, examples of output results are shown in Figures 4 – 6 respectively for the each of the three methods. The embedding mode used in each method is circular, which therefore occupies all usable positions in the host document to insert the watermark-bits while keeping the message bits intact (all in a sequence after each other with no skipping of identified allowable positions).

A. Results of Method 1.

From Figure 1, it is clear that as the number of words per set increases, the average percentage capacity/words decreases for a fixed number of embedded-bits per set. Moreover, it is noticed that the larger the set size (number of words grouped together per set) the higher the capacity. However, Figure 1 shows a smooth distribution of Kashidas throughout the document which results with medium to high imperceptibility since the Kashidas are unnoticed. Note that in Figure 4 (corresponding to results in Figure 1), the underlined words show the addition of Kashidas. For the sample document, it is clear that inserting 4 – 5 (or less) bits per set (the number of bits inserted corresponds to the number of possible Kashidas that could be inserted) produce an acceptable level of imperceptibility. Therefore, this method provides high capacity with acceptable imperceptibility, making this watermark approach less perceptible to human vision. Finally, a lower number of bits/set results with lower capacity, and higher imperceptibility levels compared to larger number of bits embedded per set. Using the circular embedding mode, has shown this method to be more robust and secure since the watermark key can be extracted if the document had experienced text-processing operations or isolated attacks. Finally, the characteristics and robustness of this technique had demonstrated its suitability to be used in copyright protection applications.

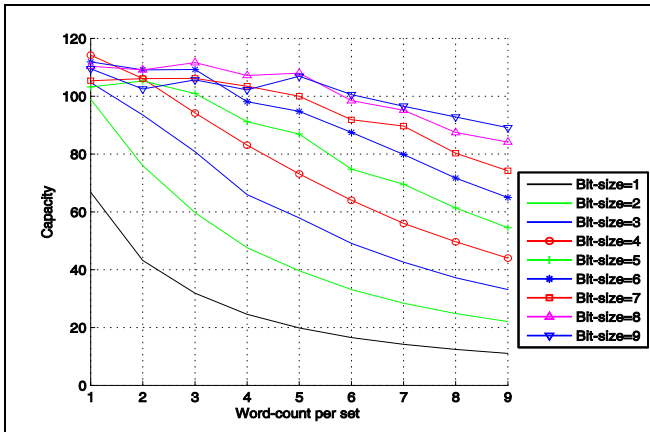


Figure 1: Capacity versus word-count per set for different embedded watermark bits using Method 1. (Note: bit-size refers to the no. of embedded bits/set.

B. Results of Method 2

The results and sample of this approach are shown in Figures 2 and 5 respectively for applying method 2 on Arabic-text documents. Figure 2 shows that for higher numbers of embedded bits, the behavioral trend follows a steady result, before gradually deteriorating due to the constraint that a space will always be inserted between words since the insertion position is always available. A lower number of embedding bits (< 4 per set) resulted with a lower capacity, whilst the higher the number of bits/word (> 4) produced higher capacity - showing a constant capacity for all set sizes used. On the other hand, in the case of method 1, Kashidas can only be inserted according to specific constraints/rules which results in skipping some locations in a word. This shows a high imperceptibility level for constant capacity requirements.

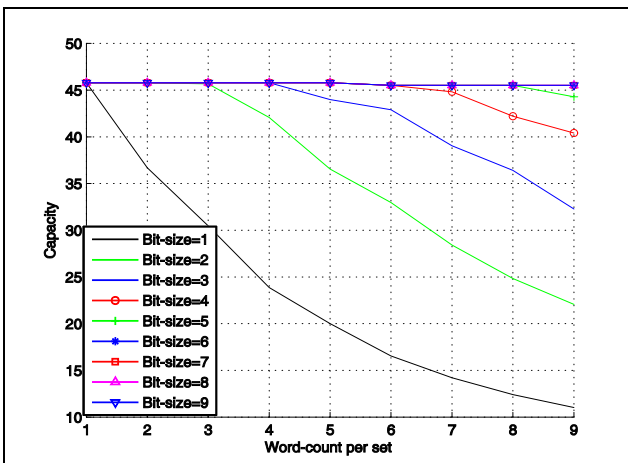


Figure 2: Capacity versus word-count per set for different bit sizes using Method 2.

C. Results of Method 3

From the results for method 3, shown in Figures 3 and 6, the capacity-level is low due to the fact that the availability of Sukun features in the text is very limited. Consequently, the capacity decreases as the number of bits and number of words/set increase. Figure 6, shows a small portion of the output results for the document used while testing this method. The results show a high imperceptibility-level compared to the other two techniques used. Therefore, the results show this method's applicability in particular Arabic diacriticized documents that include poetry text or classical Arabic documents.

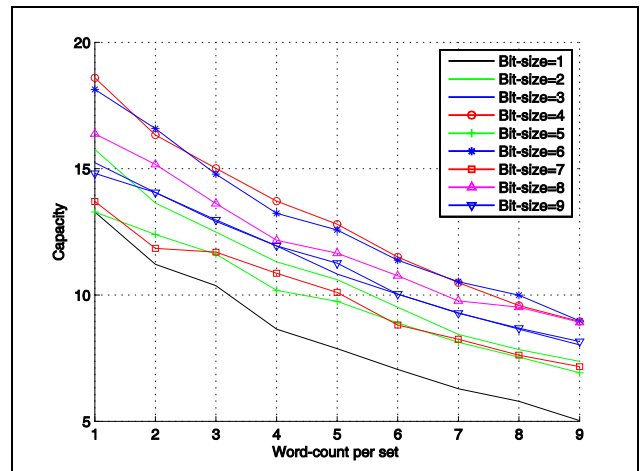


Figure 3: Capacity versus word-count per set for different bit sizes using Method 3.

V. CONCLUSION

In conclusion, Steganography/watermarking applications require high imperceptibility and capacity. However, a trade off exists depending on the technique used; since high imperceptibility was shown to produce low capacity as is the case for method 1 (Kashida-based) with higher number of bits embedded per set sizes. However, as explained in the results section, the number of embedded bits and/or number of words per set could simultaneously produce high imperceptibility and high capacity, (i.e. methods 2 and 3).

Finally, all methods investigated in this work produce high imperceptibility and capacity up to a certain level of embedded bits making them suitable for a wider range of applications such as copyright, tamper detection, steganography, ... etc.

ACKNOWLEDGMENT

The authors would like to thank and acknowledge the IT Research Center for the Holy Qur'an (NOOR) at Taibah University for their financial support during the academic year 2012/2013 under research grant reference number NRC1-126.

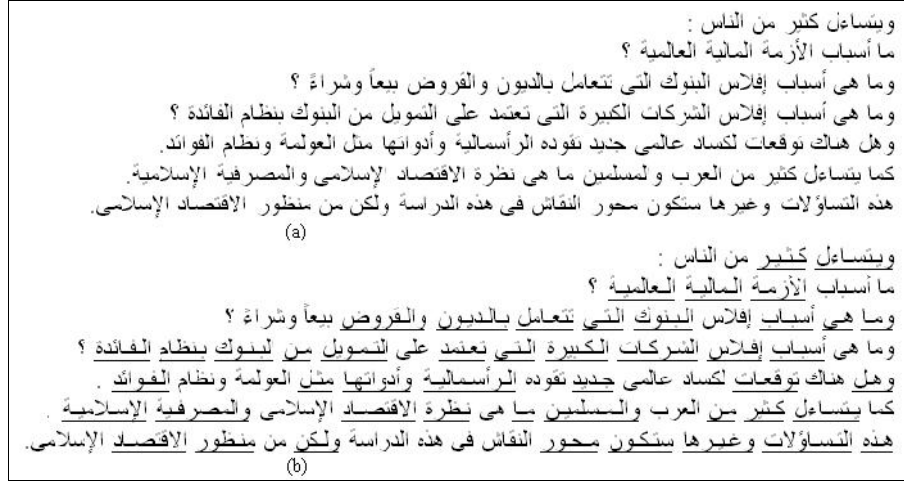


Figure 4: (a) A section from the original document used in testing (b) Results of applying Method 1.

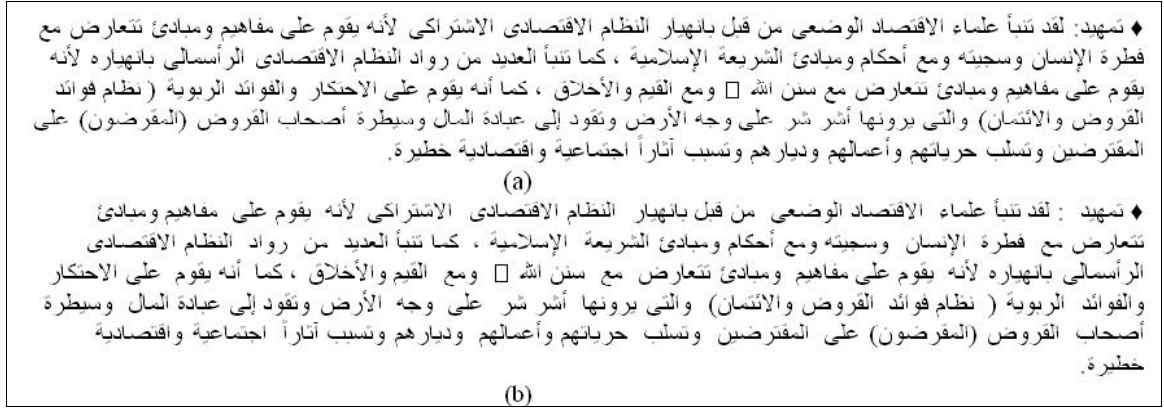


Figure 5: (a) A section from the original sample document used in testing, and, (b) Results of applying Method 2.

<p>[الباب الأول] : مَقَدِّمَاتُ ثَلَاثِ الْأَوَّلِي : فِي تَعْرِيفِ (الْأَصُولِ) وَ (الْأَحْكَامِ) •أَدْلَةُ الْفِقْهِ عَلَى الْإِحْصَالِ [19] وَصِفَةُ التَّوَجُّهِ لِإِسْتِدْلَالِ تَعْرِفُ ذِي : فِى أَصُولِ الْفِقْهِ ، مِنْ [20] أَتْرَكْهَا فَهِيَ الْأَصُولِي فَاعْلَمَنَّ وَالْفِقْهُ عِلْمٌ حُكْمٌ شَرَعِ اللهُ مِنْ [21] أَدْلَةُ تَفْصِيلُهُ فِيهَا رَكْنٌ •وَالْحُكْمُ : مُقْتَضَى حِطَابِ اللهِ [22] لِلْعَبْدِ تَكْلِيفًا بِأَلِ اسْمِيَاهِ إِنْ افْتَضَى الْحَرْمَ بِفِعْلٍ : يَجِبُ [23] وَغَيْرُ مُقْتَضَى لِحَرْمٍ : يُنْدَبُ وَمُقْتَضَى التَّرْكِ : حَرَامٌ إِنْ حُرِّمَ [24] بِهِ وَإِلَّا مَكْرُوهٌ عِلْمٌ وَالْعَقْوُ أَوْ مَا رَفَعَ الْخُنَاحُ</p>	<p>[الباب الأول] : مَقَدِّمَاتُ ثَلَاثِ الْأَوَّلِي : فِي تَعْرِيفِ (الْأَصُولِ) وَ (الْأَحْكَامِ) •أَدْلَةُ الْفِقْهِ عَلَى الْإِحْصَالِ [19] وَصِفَةُ التَّوَجُّهِ لِإِسْتِدْلَالِ تَعْرِفُ ذِي : فِى أَصُولِ الْفِقْهِ ، مِنْ [20] أَتْرَكْهَا فَهِيَ الْأَصُولِي فَاعْلَمَنَّ وَالْفِقْهُ عِلْمٌ حُكْمٌ شَرَعِ اللهُ مِنْ [21] أَدْلَةُ تَفْصِيلُهُ فِيهَا رَكْنٌ •وَالْحُكْمُ : مُقْتَضَى حِطَابِ اللهِ [22] لِلْعَبْدِ تَكْلِيفًا بِأَلِ اسْمِيَاهِ إِنْ افْتَضَى الْحَرْمَ بِفِعْلٍ : يَجِبُ [23] وَغَيْرُ مُقْتَضَى لِحَرْمٍ : يُنْدَبُ وَمُقْتَضَى التَّرْكِ : حَرَامٌ إِنْ حُرِّمَ [24] بِهِ وَإِلَّا مَكْرُوهٌ عِلْمٌ وَالْعَقْوُ أَوْ مَا رَفَعَ الْخُنَاحُ</p>
---	---

Figure 6: (a) A section from the original document used in testing (b) Results of applying Method 3.

REFERENCES

- [1] M. Thapa, S. K. Sood and M. Sharma, "Digital Watermarking: Current Status and Key Issues", International Journal of Advances in Computer Networks and its Security, vol.1, no. 1, 2011, pp. 327-332.
- [2] Digital Watermarking, <http://en.wikipedia.org/wiki/Digital_watermarking>, accessed 23rd of August, 2012.
- [3] L. Robert and T. Shanmugapriya, "A Study on Digital Watermarking Techniques," International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.
- [4] N. Veda Kumar, Y. David Soloman Raju, "A CDMA Watermarking Scheme For Subspace Projection Based On PS Sequence," International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 8, October – 2012.
- [5] Jalil Z., Farooq M., Zafar H., Sabir M., and Ashraf E., "Improved Zero Text Watermarking Algorithm against Meaning Preserving Attacks," World Academy of Science, Engineering and Technology, Vol. 46, 2010, pp. 592- 596.
- [6] Fahd Al-Haidari, Adnan Gutub, Khalid Al-Kahsah and Jameel Hamodi, "Improving Security and Capacity for Arabic Text steganography using Kashida Extensions," IEEE/ACS International Conference on Computer Systems and Applications (AICCSA 2009), 2009, pp. 396 - 399
- [7] M.H. Shirali-Shahreza, and M. Shirali-Shahreza, "A New Approach to Persian/Arabic Text Steganography," 5th IEEE/ACIS International Conference on Computer and Information Science, 2006, pp.: 310 - 315
- [8] A. Gutub, L. Ghouti, A. Amin, T. Alkharobi and M. Ibrahim, "Utilizing Extension Character 'Kashida' with Pointed Letters for Arabic Text Digital Watermarking," International Conference on Security and Cryptography, Barcelona, Spain, 2007, pp. 329- 332.
- [9] Yasser .M. Alginahi, Muhammad N. Kabir and Omar Tayan, An Enhanced Kashida-Based Watermarking Approach for Increased Protection in Arabic Text-Documents, 4th International Conference on Communications Security & Information Assurance (CSIA), Delhi, India, May 24 – 26, 2013. "Un-published"
- [10] J. A. Memon, K.Khowaja and H. Kazi, "Evaluation of Steganography for URDU/ARABIC Text," Journal of Theoretical and Applied Information Technology, 2008, pp. 232 – 237.
- [11] M. Aabed, S. Awaideh, A. Elshafai and A.Gutub, "Arabic Diacritics Based Steganography," IEEE International Conference on Signal Processing and Communications (ICSPC), 2007, Dubai, UAE, pp. 756 – 759.
- [12] Zunera Jalil, Anwar Mirza and Maria Sabir, "Content Based Zero-Watermarking Algorithm for Authentication of Text Documents," International Journal of Computer Science and Information Security, Vol. 7, No. 2, February 2010.
- [13] S. H. Low, N. F. Maxemchuk, J. T. Brassil, and L. O'Gorman, "Document Marking and Identification Using Both Line and Word Shifting," Proceedings of Infocom 1995, pp. 853 – 860, Boston, MA, April 1995.

AUTHORS PROFILE

Muhammad Nomani Kabir, received his PhD in Computer Science at the University of Braunschweig, Germany. He is an Assistant Professor at the Department of Computer Science, Taibah University in KSA. His research interests include issues related to Arabic document analysis (Security), numerical methods, crowd/traffic research and embedded systems.

Omar Tayan

He completed his joint BEng/BSc. degree in Computer and Electronic Systems from the University of Strathclyde, Glasgow in 2002 and his PhD in Computer Networks, Department of Electronic & Electrical Engineering from the same university in 2005 (graduation 2006). Omar worked briefly at Agilent Technologies-UK, and later at King Khalid University, Abha, KSA, as an Assistant Professor for nearly two years before he joined the College of Computer Science and Engineering (CCSE) at Taibah University in 2007, where he currently works as an Assistant Professor. He has been a consultant to the Strategic and Advanced Research and Technology Innovation Unit at the university for four years and is one of the founding members of the "IT Research Center for the Holy Quran and Its Sciences (NOOR)" at Taibah University, Al-Madinah Al-Munawwarah, Saudi Arabia. He has received a number of research grants in the fields of Computer Networks/Transportation Systems, and Information Security. Omar has numerous research papers from various disciplines to his credit including a published book titled: "Exemplar High-Speed Scalable Network Architectures: A Modern Approach to Networking". He is a member of the IEEE and IEEE-Computer Society and has research interests that include; performance modeling and evaluation, high-speed computer networks and architectures, software simulation techniques and queuing theory, Wireless Sensor Networks for Intelligent Transportation Systems, Networks-on-Chip (NoC), Optical Networks (ONs), Information Security for the Holy Quran and E-Learning technologies.

Yasser M. Alginahi, Senior member: IEEE, member of IEEE Computer Society, senior member IACSIT, earned a Ph.D., in electrical engineering from the University of Windsor, Canada. Currently, he is at the Dept. of Computer Science, College of Computer Science and Engineering, Taibah University. His current research interests are Arabic document analysis (Security), Pattern Recognition (OCR and Face recognition), and crowd/traffic management. He is a licensed Professional Engineer and a member of Professional Engineers Ontario, Canada (PEO). He has executed several scientific projects in the field of document analysis, pattern recognition and crowd/traffic research. He is an editorial board member of several international referred journals. He has over 40 research publications and technical reports to his credit.

Systematic Mapping Study On Security Threats in Cloud Computing

Carlo Marcelo Revoredo da Silva, José Lutiano Costa da Silva, Ricardo Batista Rodrigues, Leandro Marques do Nascimento, Vinicius Cardoso Garcia
Informatics Center, CIn
Federal University of Pernambuco, UFPE
Recife, Brazil

Abstract— Today, Cloud Computing is rising strongly, presenting itself to the market by its main service models, known as IaaS, PaaS and SaaS, that offer advantages in operational investments by means of on-demand costs, where consumers pay by resources used. In face of this growth, security threats also rise, compromising the Confidentiality, Integrity and Availability of the services provided. Our work is a Systematic Mapping where we hope to present metrics about publications available in literature that deal with some of the seven security threats in Cloud Computing, based in the guide entitled "Top Threats to Cloud Computing" from the Cloud Security Alliance (CSA). In our research we identified the more explored threats, distributed the results between fifteen Security Domains and identified the types of solutions proposed for the threats. In face of those results, we highlight the publications that are concerned to fulfill some standard of compliance.

Keywords: Security Threats, Cloud Computing, Systematic Literature Review, Security Domains, Compliance Issues.

I. INTRODUCTION

Cloud computing (CC), is a new concept that has the goal to make computational resources available as services on demand, in a short period of time and usage based cost. Cloud Computing is presented in three strategic business models: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). The aim of cloud computing models (CCM) is to cut operational costs and, more important, to allow IT departments to focus on strategic projects instead of being concerned only in keeping their datacenters working [Velte et al, "Cloud Computing, A Practical Approach", McGraw-Hill Osborne Media, 1st edition, 2009]. With such benefits, CC has become a world trend and an area of strong investments. According to Gartner [2], the financial investment on CC in 2016 will have a Global Compounded Annual Growth Rate (CAGR) of: IaaS: 41%, PaaS: 26.6% and SaaS: 17.4% in 2016 [2]. In this scenario, there is growing concern in relation to the security of services provided. In the same Gartner survey, the category Management and Security will have a CAGR of 27.2%. The security policies are present in the Quality of Service term (QoS), specified in the Service Level Agreement (SLA).

In fact, many solutions are being proposed in literature. However the resulting problems from Security Threats to Cloud Computing Models (STCCM) are even newer. Those threats compromise the CIA of the resources provided. Currently we may consider seven different threats: #1 Abuse and Nefarious Use of Cloud Computing, #2 Insecure Interfaces and APIs, #3 Malicious Insiders, #4 Shared Technology Issues, #5 Data Loss or Leakage, #6 Account or Service Hijacking and #7 Unknown Risk Profile [3]. One of the reasons why those threats are so challenging is because in cloud computing the computational resources are the result of homogeneous data centers. This characteristic means that there is not an individual and proper management for each data center, making harder the adoption of an efficient security model that fulfills the specifications of the security policies [4].

Currently there are several organizations motivated research in order to minimize STCCM, for example, the Cloud Security Alliance (CSA), an organization that arose in face of those concerns. But we may also mention other organizations, such as the National Institute of Standards and Technology (NIST), the European Network and Information Security Agency (ENISA), the OWASP [51] Foundation with the project OWASP-Cloud, and the Computer Emergency Response Team (CERT). One of CSA's lines of research is precisely the compilation of a guide defining STCCM.

An approach technique to detect deficiencies in a given theme is to present a Systematic Mapping (SM) of literature. A Systematic Mapping is a revision with a given degree of amplitude of primary studies, with the goal to identify evidences and lacunae that remain in the current literature, providing a systematic focus for future revisions [5]. The result is a general overview of the researched area, where is possible to evidence the results acquired over time, therefore, identifying trends [6] [7].

The aim of this study is to benefit from SM techniques and analyze works available in literature that deal with threats and elaborate metrics with the goal to identify which threats are being more considered in literature and what kinds of solutions are being proposed. In consequence, we pretend to observe which ones of those works care to comply with some compliance standard, that from our point of view we consider

able to reduce the problem related to the transparency between the security of the offered service and the client using it. Our work is structured as follows: in section 2 we describe our methodology and present the results in section 3. Section 4 is destined to answer our Research Questions and we develop our conclusions in section 5.

II. RELATED WORK

Concerns with Security Threats in Cloud Computing are quite recent, more precisely from 2008. In the last years the threats are receiving much attention by several researchers. In 2010, Farrell [12] alerts about problems of governance, risk and compliance of CC. In 2011, Hori et al [13] reports about security aspects for internal threats on CC. In the same year, Khorshed et al [14] propose two contributions: research in literature with focus on lacunae and challenges of threats, and defines an approach to prevention of attacks. In 2012, Ayala et al [15] identifies the threats and attacks and proposes solutions based in guides from NIST [16] and CSA [17]. In the same year, Yeluri et al [18] reports about experiences of Intel team with threats to security and resources control in CC. Also in 2012, Aqrabi et al [19] through a revision of literature and results obtained in simulations, proposes to identify the quests in adoption of security and compliance in CC. Nowadays, the theme of security threats in Cloud Computing is being well explored. We identified that 38% and 31% of the publications that we cataloged were made in 2011 and 2012, respectively, according to Figure 1. We were motivated to produce this work because, among the publications related to threats in current literature, we did not find one considering the type of solution proposed by the authors, that could identify which compliances were related in those publications.

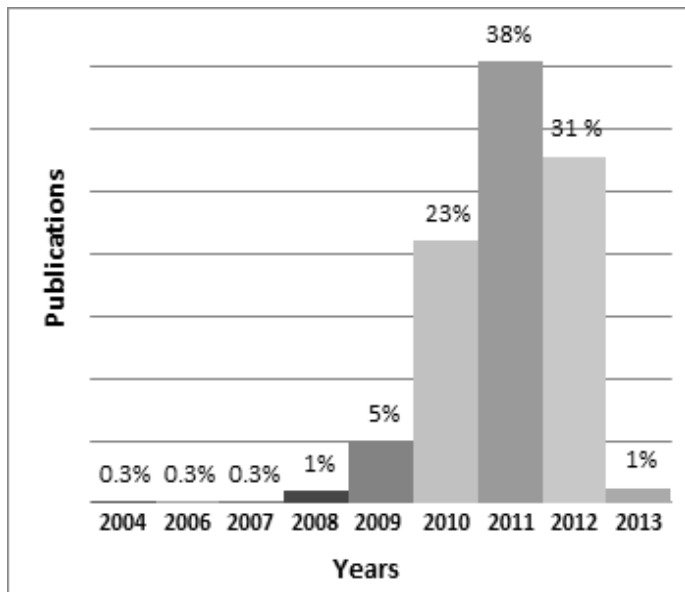


Figure 1. Percentage of publications spread by year.

III. SYSTEMATIC MAPPING

A. Research Questions

In this work we followed the orientations of Kitchenham [6] for elaboration of 4 Research Questions (RQ), with the goal to determine the content and conception of the systematic revision. Our work aims to answer the following RQ: (1) Which Security Threats to Models of Clouding Computing are being more addressed in literature? (2) Which Security Domains are being more explored by the Threats? (3) Which types of Solutions are being proposed in those approaches? (4) Among those approaches, which Compliances are involved?

B. Definition of Research and Primary Source

We defined Elsevier Scopus as the primary source of our work. Besides having a considerable number of publications, we observed that in Scopus a large number of indexations of works from other sources are available. In addition, its search engine is able to be refined with several functionalities in its filters. Other sources chosen were: IEEEExplore, ACM Digital Library, SpringerLink, Science Direct and Engineering Village.

As initial research, we searched in Scopus for works related to security on CCM, using the following filter rule: {[Non-compliance with security) OR (key-words for security)] AND [cloud computing solutions]} in the title or abstract or key-words in the article. Resulting on the following Search String: ("flaw" OR "risk" OR "threat" OR "vulnerabilit*" OR "unsafe" OR "untrust") AND ("security" OR "safe" OR "trust") AND ("cloud" OR "multi-tenant*" OR "*aas" OR "*" as a service" OR "* as-a-service"). Adding the results of all research sources, the total amounted to 1011 publications. Many of the occurrences were not in the research context and a manual refining based in the results or then triage had to be performed. We did not want to refine too much the Search String, because there was the risk of any relevant publication being excluded, we rather choose to leave the Search String wide open, leaving the refining in charge of a more detailed manual inspection.

C. Inclusion Criteria

From that, we started our triage process considering the following inclusion criteria:

- Security in Cloud Computing as the main theme.
- The publication should have some relationship with one of the seven threats.
- The publication should have a proposed solution.

D. Exclusion Criteria

- Duplication of publication.
- Journals not accessible online.
- Publications with related threat, but not active in cloud computing.
- Publications that only bring a revision or approach, without a proposal of solution.

E. Relevance Criteria

- Papers with well detailed solution proposal;
- Papers that have some kind of proposal validation, with statistical data, experiment, etc;
- Papers focused in fulfilling some compliance;

F. Screening of Publications

Each researcher applied the triage in a superficial way, based on the abstract of publications. When it was detected that at least a threat was applied, and some solution identified as a contribution, the publication was already considered. We found that for our research there were two cases where the superficial process was not enough, the first case when the abstract was too short, the other when it was not possible to extract from the abstract the proposal solution as a contribution. Those publications were allocated in separate for a more detailed future evaluation where they will be analyzed in introduction or in other chapters of the publication. This triage resulted in 661 publications according to Figure 2.

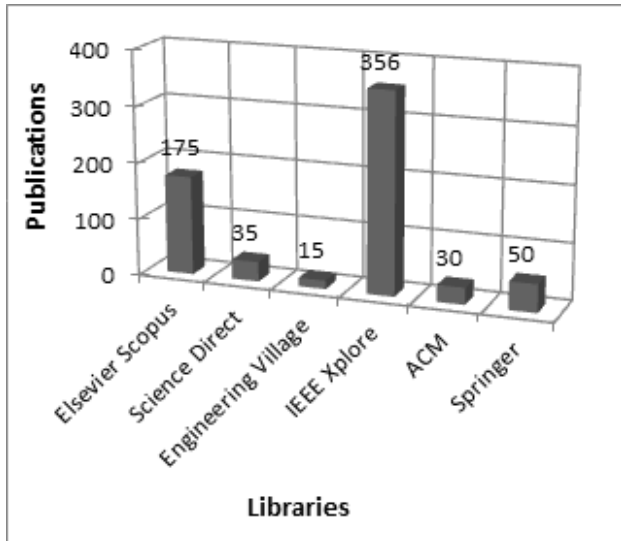


Figure 2. Publications spread by literary sources

G. Results Classification

In this phase, we made the analysis and classification of threats related to publications, the security domains involved, the type of solution proposed in each publication and if there is an approach aiming to fulfill some compliance standard. Posteriorly, the other authors interacted and analyzed the results related to the chosen classifications and reached the same conclusions.

IV. ANALYSIS AND RESULTS DISCUSSION

Here the questions of researches proposed in the protocol are answered.

A. Result Obtained From RQ1:

Despite each threat having a specific characteristic, nothing forbids it to act simultaneously with other threats in the same scene, resulting thus in several intersections between publications and threats. The seven threats are distributed in 661 publications according to Figure 3:

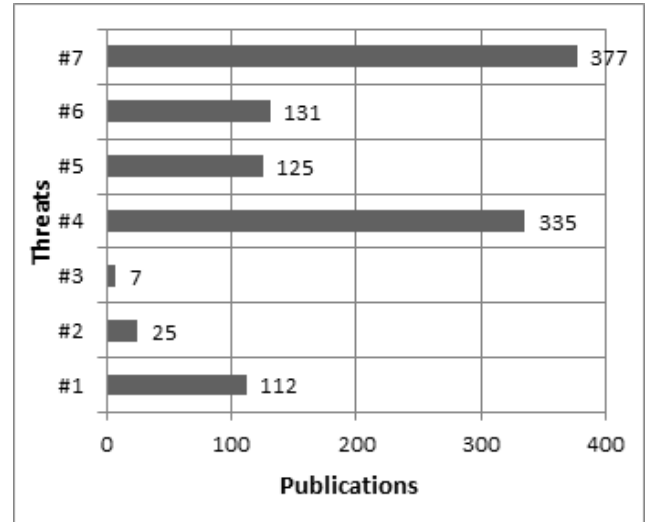


Figure 3. Distribution of publications by threats

1) Threat #1: Abuse and Nefarious Use of Cloud Computing

With 112 publications, it is an intermediary threat considering literary exploration. The consequence of this threat helps the growth of plagues like botnets, from which come problems like Distributed Denial of Service (DDoS), solves of Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA), storage of malicious files and botnet networks [3]. This threat evidences the fact that today it is very simple for any user to hire a cloud computing solution, it is even possible to get a free evaluation time, having only a valid credit card, which could come from a robbery or fraud. This ends up encouraging the action of malicious people to inject spam, malwares or even to practice illicit activities on the cloud [3]. There is only one proviso: until this moment, the version 2.0 of Top Threats Cloud Computing of CSA was not officially released, but a Survey was disclosed determining that instead of seven, there would be eight fails. This is because problems related to DDoS are being so explored that it was dismembered and became a distinct threat in order to ease the understanding of strategies for its prevention [8].

2) Threat #2: Insecure Interfaces and APIs

A very relevant area, but so far not explored enough in literature. We cataloged only 25 papers. There are thousands of available APIs to be consumed, being also possible to build combinations of other APIs, known as Mashups. Those interfaces have serious standardization problems [4], this makes hard to apply a consistent security policy and the

consequence is that many times access control, authentication, entry treatment, traffic of encrypted data, monitoring of activities, among other security aspects are neglected, offering a huge risk to cloud computing [3].

3) Threat #3: Malicious Insiders

This threat represents attacks of an active employee, ex-employee or business partner from the cloud provider that somehow has an authorized access and compromised the CIA of information stored in the cloud [10]. We consider this the hardest threat to be mitigated, having found only 7 publications related in literature. Despite being a very uncommon situation, its damage could be devastating [3], and it becomes even more critical when executed in environments without a straight access control of employees and without a well structured auditing that supports forensic analysis.

4) Threat #4: Shared Technology Issues

The second most explored threat in literature, with 335 publications found, focusing in IaaS models. Some components of this architecture were not projected for the scalability demanded from the model, making necessary to implement virtual machine monitoring to manage its resources [3]. Many times this layer does not have an adequate defense strategy and does not exert a good monitoring of network security. This is the scene where this kind of threat is more present.

5) Threat #5: Data Loss or Leakage

This threat happens when an exclusion, change or improper appropriation of some data in the cloud is made [3]. We considered this the most explored threat nowadays, because it represents a large number of the most recent publications. The cloud solutions for Storage and Bigdata are also having a strong growth. In consequence, the worry to provide CIA to data is also emerging; we found 125 publications in our research.

6) Threat #6: Account or Service Hijacking

Methods of phishing, fraud and vulnerability exploration, besides password credentials used in distributed ways, give amplitude to this problem [3]. The worry with kidnapping of accounts was the target of many cloud providers already consolidated in the market, such as Amazon [9]. There were found 131 publications in literature.

7) Threat #7: Unknow Risk Profile

It is the most explored threat in literature, with 377 selected publications. In cloud computing the abstraction regarding architecture details and maintenance responsibilities proportionate a greater security with obscurity by the cloud providers [3]. Details such as software version, failure fixes in order to avoid problems such as zero day, process that meet good security practices, among other aspects that are many times neglected by the cloud provider, falling in the large problem of the transparency of quality of service offered to its consumers.

B. Result Obtained From RQ2:

In this section we identified the Security Domains involved in each threat. We elaborated our classification based in the one proposed by [Mather et al, "Cloud Security and Privacy: A Enterprise Perspective on Risks and Compliance", O'Reilly Media; 1 edition, 2009], where eight different domains are enumerated. We classified them in a more granular way in order to get better visibility of results from our research, resulting in 15 Security Domains. The prevention measures of each threat may involve one or more domains, therefore also subject to intersections, according to the distribution displayed in Figure 4. Table 1 shows a brief description of each domain and the total amount of related publications.

TABLE I. FACET 1: SECURITY DOMAINS

Domain	Description	Score
Access Control	Intervene in user access, from what the user access is granted or denied to a given datum or resource. Covers practices as for example Single-Sign-On (SSO) and Role-Based Access Control (RBAC).	39
Accountability	Ensures the quality of information with regard to possible and undesired behaviours of a system or infrastructure in the cloud.	8
Anonymity	Refers to traffic of public data, not allowing the same to be intercepted, warranting anonymity in public or mixed clouds.	3
Applied Cryptography	Capacity of an emissor to make its data unreadable, with only the receptor being able to read the content.	27
Authentication	Verifies and validates a user identification.	16
Data or Database Protection	Technique for protection of informations stored either in bigdata or storage.	130
Digital Forensic	Technique of systematic inspections in computational resources in order to collect informations that may evidence a supposed crime committed. Presents itself as an excellent solution for problems related to inside threats in the cloud.	5
Identify Management	Is the management to establish and keep identity records applied to an access policy to each finality or service.	20
Integrity	Is the way to warrant that an information or behaviour cannot be changed by non-authorized people.	21
Intrusion Detection	Is the capacity to analyze a traffic or content that has the intention to compromise the integrity of a system or computational resource.	22
Formal Security Model	Overall, it is a scheme to specify and apply security policies.	203
Network Security	Guidelines to monitoring non-authorized or incorrect access in the network.	96
Privacy	It is the control of availability of a given information or resource in a public or shared environment.	73
Risk Analysis and Management	A set of policies to warrant that security processes happen in an efficient and continuous way over time.	81
Trust Model and Management	A set of policies that help to identify and estimate threats in a systemic way.	50

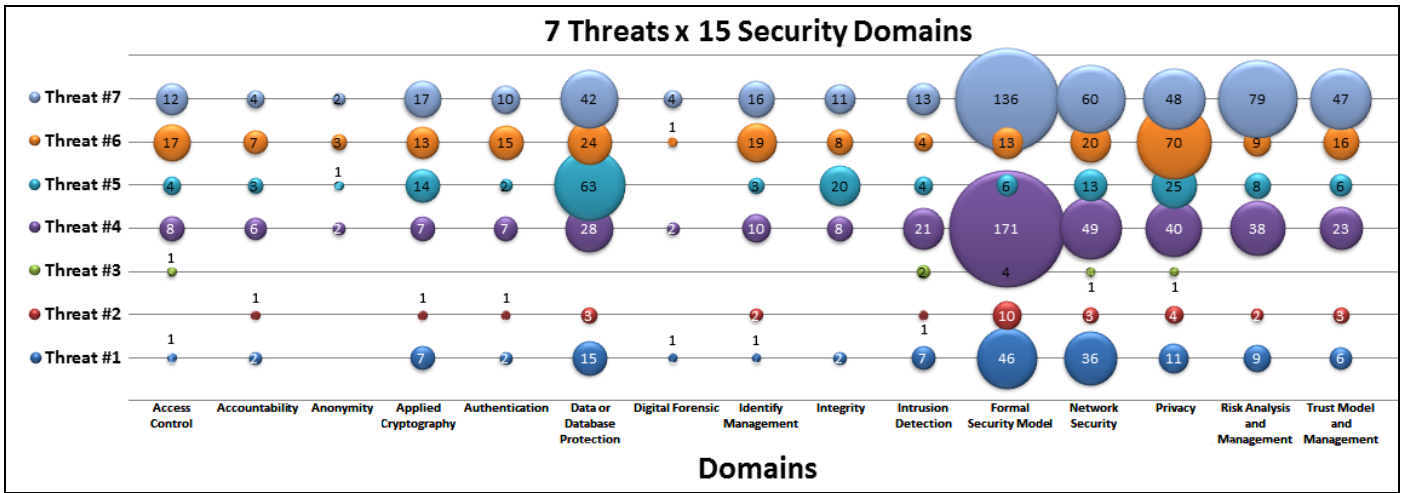


Figure 4. Distribution of publications by Threats and Domains

We detected the large number of combinations between “Access Control”, “Data or Database Protection” and “Privacy” in more recent publications, and intrinsically linked to Threat #5. Many concerns involve techniques such as “Granular Access Control” and “Granular Audits” in the fields of Storage and Bigdata.

C. Result Obtained From RQ3:

In this stage we identified and classified the proposals of each work. Unlike the previous metrics, in this one there were no intersections. We assumed that each publication would have a single proposal, in cases when there was more than one, we considered the more elaborate by the authors. We measured eight types of proposals, according to Figure 5.

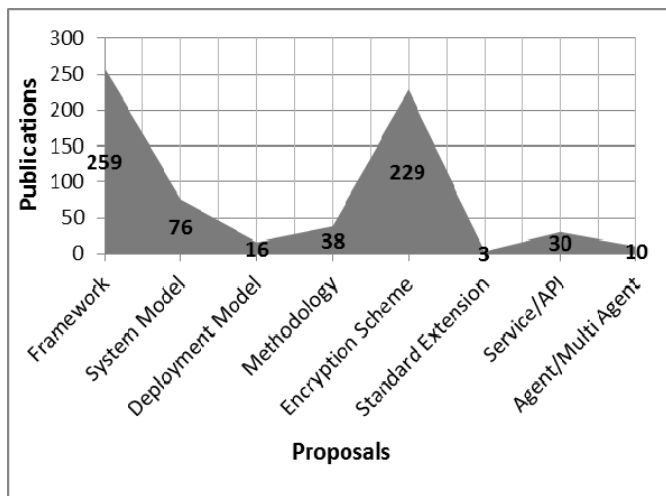


Figure 5. Distribution of publications by Proposals

D. Result Obtained From RQ4:

In this stage we analyzed and identified, among the selected publications, those that were concerned with some compliance standard. Compliance is the condition of someone or of a group of people or processes to be according to what is desired or previously established, the desired in question are the specification standards. In this stage there were 9 intersections, as for example, we observed a publication that focus in compliances from NIST and FCAPS. As the answer for the RQ we identified a total of 18 compliances according to Figure 6.

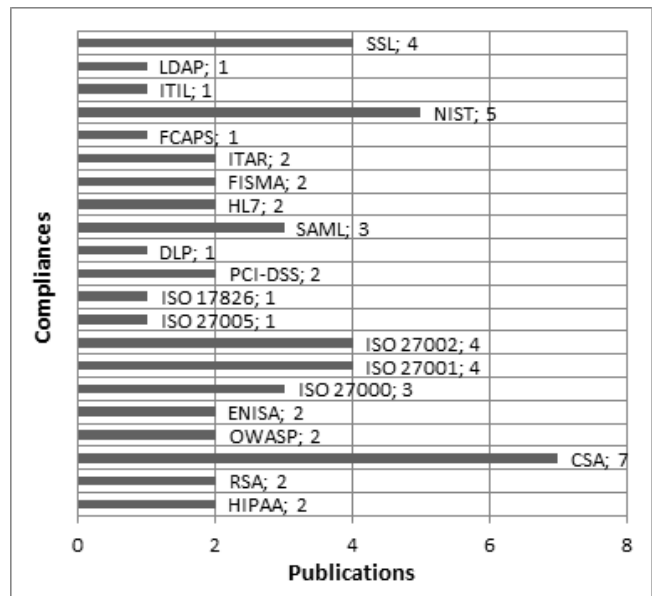


Figure 6. Distribution of publications by Compliances

1) *NIST*

Founded in 1901, is a non-governmental non-regulatory agency of USA trade. It has research in several areas, among them to promote standards to technological processes. After 15 revisions, it created the specification where it defines CC with 5 characteristics: self-service on demand, access to broadband network, resources pool, fast elasticity and measurable service. This definition was the milestone of standardization in the cloud. Other organizations such as CSA have this specification as the base for their researches [16].

2) *FCAPS*

It is an ISO standard, defining itself a model for network management composed by 5 levels: F: Fault, C: Configuration, A: Accounting, P: Performance and S: Security.

In level F are fixed the errors identified. It is also performed a management for prevention of future errors. In level C is performed a monitoring both in the network as well as in development changes. In this stage obsolete software or resources are removed from the network ecosystem and periodical updates of equipment and software are performed. Level A is dedicated to allocation and distribution of resources offered by the network, warranting that users receive resources according to the SLA. Level P is the management of performance where it is intended to identify problems and improvements. Level S is to ensure CID in all network resources [55].

3) *ITAR*

The International Traffic in Arms Regulation is a set of rules that control imports and exports of objects related to guns and ammunition [56].

4) *FISMA*

It is a federal law of the USA that recognizes the importance of information security in federal agencies data, demanding that each agency complies with security processes that control its assets. Compliance with FISMA was formalized by NIST in publication 800-53 [57].

5) *HL7*

Is a standard from the American National Standards Institute, ANSI, used to storage and handling of medical data. Any and all information related to patients, doctors and drugs is constructed from technical terms; this standard has the goal to universalize this communication [58].

6) *SAML*

It is a standard created by OASIS applied in the exchange of authentication and authorization of data between distinct security domains, based in protocols of Token exchange using XML, giving support to Web platforms and techniques such as SSO [59].

7) *DLP*

It is a technique to avoid, in time, incidents of violation or undue access to sensible data. The consequences may change, from access inhibition to the file or self-destruction of it [60].

8) *PCI-DSS*

It is a security pattern created by the Payment Card Industry Security Standards Council (PCI SSC), aimed to concerns with implementations in software that will do transactions with credit card. Its goal is to standardize the implementation and evaluate the providers of that software [61].

9) *ISO 17826*

It is a security standard of information published by the International Standards Organization (ISO) and by the International Electrotechnical Commission (IEC). Also called CDMI, it specifies the interface to storage and management of data in the cloud. This documentation is focused in developers or users of cloud storage [62].

10) *ISO 27005*

It is a standard for information security published by the ISO/IEC. Its goal is to provide orientation for management of information about security risks [64].

11) *ISO 27002*

It is a standard for information security published by the ISO/IEC. It has the goal to establish directives and general principles to implement keep and improve the management of information in an organization [65].

12) *ISO 27001*

It is a standard for information security published by the ISO/IEC. The rule focuses on the concerns with implementation, monitoring, improvement and revision of a Management System of Information Security (MSIS) [11].

13) *ISO 27000*

Is a standard for information security published by ISO/IEC. The rule is a standard about good practices in Management of Information Security, which brings companies to the higher international level of excellence in Information Security. [27]

14) *ENISA*

The European Network and Information Security Agency is an agency from the European Union. The agency has the goal to contribute for the development of a culture of information and network security for the benefit of citizens, consumers, companies and public sector organizations of the European Union. In consequence, it will contribute for the good functioning of the internal market of the European Union [20].

15) OWASP

The Open Web Application Security Project (OWASP) is an open source project for security of applications. The OWASP community has corporations, educational organizations and individuals from all over the world. This community works to create articles freely available, methodologies, documentation, tools and technologies that promote the good practices of security. The OWASP Foundation is a charity organization that supports and manages OWASP projects and its infrastructure. It is also a nonprofit registered trademark in Europe since June 2011 [51].

16) CSA

It is a nonprofit organization with the mission to promote the use of better practices, provide warranty of security in Cloud Computing, and provide education about cloud computing use to help to protect all kinds of computing. The Cloud Security Alliance is led by a wide coalition of industry professionals, companies, associations and other interested parties [3].

17) RSA

RSA is an algorithm for data cryptography, which owes its name to three teachers of the MIT (founders of the current company RSA Data Security, Inc.), Rivest, Shamir and Adleman. It is considered the most well succeeded implementation of asymmetric keys algorithms, and is based in classical theories of numbers. It was also the first algorithm to allow cryptography and digital signature, and one of the great inventions in public key cryptography [1].

18) HIPAA

It is the acronym for Insurance Portability and Accountability Act. It was approved by the American Congress in 1996, during the Bill Clinton government. It is a standard with the goal to protect data related to health, ensuring privacy and fraud prevention [63].

E. Relevant Works

The publications that we considered relevant are based in the criteria of relevance defined in the protocol. We selected three publications of threat #1, nine publications of threats #4 and #5, four publications of threat #6 and nine publications of threat #7, totaling 34 publications considered the most relevant in our research result. Curiously the result of our search reveals that none of the works related to threats #2 and #3 were concerned to fulfill any compliance.

1) Threat #1

TABLE II. COMPLIANCES INTO THREATS #1

Compliance	Proposal	Domain	Reference
ISO 27001	Standard Extension	Formal Security Model	Ristov et al [38]
ITAR	Framework	Formal Security Model	Wang et al [53]
ITIL	Framework	Formal Security Model	Kamer & Vranken [25]

2) Threat #4

TABLE III. COMPLIANCES INTO THREATS #4

Compliance	Proposal	Domain	Reference
ISO 27000, ISO 27001, ISO 27002.	Framework	Risk Analysis and Management	Zhao [36]
ISO 27001, ISO 27002.	Methodology	Authentication	Auty et al [39]
ISO 27001	Framework	Formal Security Model	Julich & Hall [40]
ISO 27002	Framework	Formal Security Model	Rebollo et al [41]
PCI-DSS	Framework	Trust Analysis and Management	Hizver & Chiueh [43]
PCI-DSS	System Model	Privacy	Kounelis et al [44]
HL7	Deployment Model	Formal Security Model	Mouleeswaran et al [50]
ITAR	Framework	Formal Security Model	Poolsappasit et al [52]
NIST	System Model	Privacy	Kim et al [54]

3) Threat #5

TABLE IV. COMPLIANCES INTO THREATS #5

Compliance	Proposal	Domain	Reference
HIPAA	Encryption Scheme	Access Control, Integrity Privacy, Applied Cryptography, Data or Database Protection	Li et al [28]
HIPAA	System Model	Access Control, Integrity Privacy, Data or Database Protection	Huemer et al [29]
RSA	Encryption Scheme	Formal Security Model, Applied Cryptography, Data or Database Protection	Saravanan et al [30]
RSA	Encryption Scheme	Formal Security Model, Applied Cryptography, Data or Database Protection	Lin et al [31]
ISO 17826	Standard Extension	Formal Security Model	Teckelmann et al [42]
DLP	Encryption Scheme	Formal Security Model, Applied Cryptography, Data or Database Protection	Basak et al [45]
LDAP	Encryption Scheme	Formal Security Model, Applied Cryptography	Zissis & Lekkas [22]
SSL	Encryption Scheme	Applied Cryptography, Authentication, Data or Database Protection	Mansukhani & Zia [23]
SSL	Framework	Formal Security Model, Data or Database Protection	Ahmed et al [24]

4) Threat #6

TABLE V. COMPLIANCES INTO THREATS #6

Compliance	Proposal	Domain	Reference
SAML	Framework	Access Control, Authentication, Identify Management	Lonea et al [46]
SAML	Framework	Identify Management, Trust Model and Management	Cabarcos et al [47]
SAML	Encryption Scheme	Identify Management, Applied Cryptography	Guerrero et al [48]
HL7	Encryption Scheme	Access Control, Risk Analysis and Management	Sharma et al [49]

5) Threat #7

TABLE VI. COMPLIANCES INTO THREATS #7

Compliance	Proposal	Domain	Reference
NIST, CSA	Methodology	Formal Security Model	Ayala et al [15]
NIST, FCAPS	Framework	Risk Analysis and Management	Sitaram & Manjunath [20]
NIST, FISMA	Framework	Risk Analysis and Management	Almorsy et al [21]
CSA, ENISA	Deployment Model	Risk Analysis and Management	Kao et al [26]
CSA	Framework	Identify Management, Risk Analysis and Management	Bhardwaj & Kumar [32]
CSA, OWASP	Framework	Risk Analysis and Management	Saripalli & Walters [33]
OWASP, ISO 27002	Service/API	Risk Analysis and Management	Chou & Oetting [34]
ENISA	Framework	Risk Analysis and Management, Formal Security Model	Liu et al [35]
ISO 27000, ISO 27005	Standard Extension	Risk Analysis and Management, Formal Security Model	Beckers et al [37]

V. CONCLUSION

Our work has the goal to catalog the state of the art of publications available in literature, that report approaches about security threats in CC. We hope to help researchers who want to engage in the field and want to propose some solution to those problems. With our protocol we identified 661 publications about the subject, where we can analyze the Security Domains involved. We also presented types of solutions proposed by the authors, and identified that some of those publication were concerned with the compliance of some standard. We presented those compliances and reference the respective publications to ease the work of the researcher that wants to explore a specific compliance. We identified that Threat #7 is the most explored in literature and, in consequence, the Domains of Risk Analysis and Management and Trust Model and Management have expressive results. We also identified many combinations of Domains related to Access Control, Applied Cryptography, Data or Database Protection and Privacy. This reflects in the recent growth of publications that report experiences in solutions for Storage

and Bigdata in the cloud. In this same scenario, we identified that Framework and Encryption Scheme are the most used solutions. Regarding compliances, the most present in publications are those indicated by CSA, ISO 27002, ISO 27001 and NIST. However we also found some works where its authors propose the extension of an ISO standard to solve a given problem. For future works, we are planning to investigate in more detail the obstacles of a given compliance to be inserted in CC scene.

VI. ACKNOWLEDGEMENTS

We would like to acknowledge all the anonymous referees for their very helpful comments and suggestions.

VII. REFERENCES

- [1] RSA, "RSA Cryptography Standard", June 2002, Available from: <ftp://ftp.rsasecurity.com/pub/pkcs/pkcs-1/pkcs-1v2-1.pdf>.
- [2] Networkworld, "Gartner: Cloud putting crimp in traditional software, hardware sales", July 2012, Available from: <http://www.networkworld.com/news/2012/071312-gartner-cloud-260882.html>
- [3] Cloud Security Alliance, "Top Threats to Cloud Computing V1.0", March 2010, Available from: <https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf>
- [4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, M. Zaharia, "Above the Clouds: A Berkeley View of Cloud", Electrical Engineering and Computer Sciences, University of California at Berkeley, February 10, 2009.
- [5] B. Kitchenham, "Procedures for Performing Systematic Reviews", Technical Report Software Engineering Group, Keele University, Australia, 2004.
- [6] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, "Systematic Mapping Studies in Software Engineering", in: 12th International Conference on Evaluation and Assessment in Software Engineering, Australia, 2008.
- [7] D. Budgen, M. Turner, P. Brereton, and B. Kitchenham, "Using Mapping Studies in Software Engineering", Proceedings of PPIG 2008, pp.195-204, Lancaster University, 2008.
- [8] Cloud Security Alliance, "Top Threats to Cloud Computing Survey Results Update ", July 2012, Available from: <https://cloudsecurityalliance.org/wp-content/themes/csa/download-box-topthreats-results-update-2012-1.0.php>
- [9] IRS asked Amazon EC2 to perform a C&A; Amazon refused. Available From: <http://news.qualys.com/newsblog/forrester-cloud-computingqa.html>
- [10] Federal Information Processing Standards Pub 199: Standards for security Categorization of Federal Information and Information Systems, Available from: <http://csrc.nist.gov/publications/fips/fips199/FIPS-PUB-199-final.pdf>
- [11] ISO 27001:2005, "Information Technology: Security Techniques, Systems Requirements", 2005, Available from: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=42103
- [12] R. Farrell, "Securing the cloud-governance, risk, and compliance issues reign supreme", in Information Security Journal: A Global Perspective, 2010, p.310-319.
- [13] Y. Hori, T. Nishide, K. Sakurai, "Towards Countermeasure of Insider Threat in Network Security", in Intelligent Networking and Collaborative Systems (INCoS), 2011, p.634-636.
- [14] M. T. Khorshed, A.B.M.S. Ali, S. A. Wasimi, "Trust Issues That Create Threats for Cyber Attacks in Cloud Computing", in Parallel and Distributed Systems (ICPADS), 2011, p.900-905.

- [15] I. D. C. L. Ayala, M. Vega, M. V. Lombardo, "Emerging Threats, Risk and Attacks in Distributed Systems: Cloud Computing", in *Lecture Notes in Electrical Engineering* Volume 152, 2013, pp 37-51.
- [16] National Institute of Standards and Technology, "The NIST Definition of Cloud Computing", P800-145, September 2011, Available from: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
- [17] Cloud Security Alliance, "Security Guidance for Critical Areas of Cloud Computing", November 2011, Available from: <https://cloudsecurityalliance.org/wp-content/themes/csa/guidance-download-box.php>
- [18] R. Yeluri, E. C. Leon, R. R. Harmon, J. Greene, "Building Trust and Compliance in the Cloud for Services", in *SRII Global Conference (SRII)*, 2012, p.379-390.
- [19] H. Al-Aqrabi, H. L. Lu, X. Jie, R. Hill, N. Antonopoulos, Z. Yongzhao, "Investigation of IT Security and Compliance Challenges in Security-as-a-Service for Cloud Computing", in *Object/Component/Service-Oriented Real-Time Distributed Computing Workshops (ISORCW)*, 2012, p.124-129.
- [20] European Network and Information Security Agency, ENISA, "Cloud Computing Risk Assessment", 2009, Available from: http://www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-risk-assessment/at_download/fullReport
- [21] M. Almorsy, J. Grundy, A. S. Ibrahim, "Collaboration-Based Cloud Computing Security Management Framework", in *Cloud Computing (CLOUD) IEEE International Conference*, 2011, p.364-371.
- [22] D. Zissis, D. Lekkas, "Addressing cloud computing security issues", in *Future Generation Computer Systems*, 2012, p.583-592.
- [23] B. Mansukhani, T. A. Zia, "An empirical study of challenges in managing the security in cloud computing", in *Proceedings of the 9th Australian Information Security Management Conference*, 2011.
- [24] M. Ahmed, X. Yang, S. Ali, "Above the Trust and Security in Cloud Computing: A Notion Towards Innovation", in *Embedded and Ubiquitous Computing (EUC), IFIP 8th International Conference*, 2010, p.723-730.
- [25] J. Kamer, H. P.E. Vranken, "The impact of server virtualization on ITIL processes", in *SciTePress*, 2011, p.643-649.
- [26] T. C. Kao, C. H. Mao, C. Y. Chang, K. C. Chang, "Cloud SSDLC: Cloud security governance deployment framework in secure system development life cycle", in *Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2012, p.1143-1148.
- [27] ISO 27000:2012, "Technology: Security Techniques, Information Security Management Systems Overview and Vocabulary", 2012, Available from: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=56891
- [28] J. Li, S. Singhal, R. Swaminathan, A. Karp, "Managing data retention policies at scale", in *Network and Service Management*, 2011, p.01-14.
- [29] D. Huemer, A. M. Tjoa, M. Descher, T. Feilhauer, P. Masser, "Towards a Side Access Free Data Grid Resource by Means of Infrastructure Clouds", in *Parallel Processing Workshops ICPPW '09*, 2009, p.198-205.
- [30] N. Saravanan, A. Mahendiran, N. Venkata Subramanian and N. Sairam, "An implementation of RSA algorithm in google cloud using cloud SQL", in *Research Journal of Applied Sciences, Engineering and Technology*, 2012, p.3574-3579.
- [31] I. C. Lin, H. L. Wang, "An improved digital signature scheme with fault tolerance in RSA", in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, 2010, p.9-12.
- [32] A. Bhardwaj, V. Kumar, "Cloud security assessment and identity management", in *Computer and Information Technology (ICCIT)*, 2011, p.387-392.
- [33] P. Saripalli, B. Walters, "QUIRC: A quantitative impact and risk assessment framework for cloud security", in *Cloud Computing (CLOUD)*, 2010, p.280-288.
- [34] Y. Chou, J. Oetting, "Risk assessment for cloud-based IT systems", in *International Journal of Grid and High Performance Computing (IJGHPC)*, 2011, p.1-13.
- [35] C. L. Liu, W. H. Chen, D. K. Tung, "Identification of critical security issues for cloud computing", in *Applied Mechanics and Materials*, 2011, p.272-276.
- [36] G. Zhao, "Holistic framework of security management for cloud service providers", in *Industrial Informatics (INDIN)*, 10th IEEE International Conference, 2012, p.852-856.
- [37] K. Beckers, H. Schmidt, J. Kuster, S. Fassbender, "Pattern-based support for context establishment and asset identification of the ISO 27000 in the field of cloud computing", in *Availability, Reliability and Security (ARES)*, Sixth International Conference, 2011, p.327-333.
- [38] S. Ristov, M. Gusev, M. Kostoska, "A new methodology for security evaluation in cloud computing", in *MIPRO Proceedings of the 35th International Convention*, 2012, p.1484-1489.
- [39] M. Auty, S. Creese, M. Goldsmith, P. Hopkins, "Inadequacies of current risk controls for the cloud", in *Cloud Computing Technology and Science (CloudCom)*, 2010, p.659-666.
- [40] K. Julisch, M. Hall, "Security and control in the cloud", in *Information Security Journal: A Global Perspective*, 2010, p.299-309.
- [41] O. Rebollo, D. Mellado, E. F. Medina, "A comparative review of cloud security proposals with ISO/IEC 27002", in *Proc. WOSIS*, 2011, pp.3-12.
- [42] R. Teckelmann, C. Reich, A. Sulistio, "Mapping of cloud standards to the taxonomy of interoperability in IaaS", in *Cloud Computing Technology and Science (CloudCom)*, 2011 IEEE Third International Conference, 2011, p.522-526.
- [43] J. Hizver, T. Chiueh, "Automated discovery of credit card data flow for PCI DSS compliance", in *Reliable Distributed Systems (SRDS)*, 2011, p.51-58.
- [44] I. Kounelis, J. Loschner, D. Shaw, S. Scheer, "Security of service requests for cloud based m-commerce", in *MIPRO*, 2012 Proceedings of the 35th International Convention, 2012, p.1479-1483.
- [45] D. Basak, R. Toshniwal, S. Maskalik, A. Sequeira, "Virtualizing networking and security in the cloud", in *SIGOPS ACM Special Interest Group on Operating Systems*, 2010, p.86-94.
- [46] A. M. Lonea, H. Tianfield, D. E. Popescu, "Identity management for cloud computing", in *Studies in Computational Intelligence*, 2013, pp.175-199.
- [47] P. C. Arias, A. L. Marín, R. G. Sánchez, F. M. Almenares, D. D. Sánchez, "A metric-based approach to assess risk for On cloud federated identity management", in *Journal of Network and Systems Management*, 2012, pp.513-533.
- [48] R. G. Sánchez, P. C. Arias, F. M. Almenares, D. D. Sánchez, "Trust-aware federated IdM in consumer cloud computing", in *Consumer Electronics (ICCE)*, 2012, p.53-54.
- [49] M. Sharma, B. Yan, C. Sam, D. Lirong, "Using Risk in Access Control for Cloud-Assisted eHealth", in *High Performance Computing and Communication & IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICES)*, 2012, pp.1047-1052.
- [50] S. K. Mouleeswaran, A. Rangaswamy, H. A. Rauf, "Harnessing and securing cloud in patient health monitoring", in *Computer Communication and Informatics (ICCCI)*, 2012, pp.1-5.
- [51] OWASP, "Cloud top 10 Security Risks", 2012, Available from: https://www.owasp.org/index.php/Projects/OWASP_Cloud_%E2%80%9090_10_Project/Releases/Initial_Pre-Alpha_List_of_OWASP_Cloud_Top_10_Security_Risks
- [52] N. Poolsappasit, V. Kumar, S. Madria, S. Chellappan, "Challenges in secure sensor-cloud computing", in *SDM'11 Proceedings of the 8th VLDB international conference on Secure data management*, 2011, pp.70-84.
- [53] W. Xin, H. Ting-lei; R. Zhi-jian, "Research on the anti-virus system of military network based on cloud security", in *Intelligent Computing and Integrated Systems (ICISS)*, 2010, pp.656-659.
- [54] A. Kim, J. McDermott, M. Kang, "Security and Architectural Issues for National Security Cloud Computing", in *Distributed Computing Systems Workshops (ICDCSW)*, 2010, pp.21-25.
- [55] ISO/IEC 10040:1998, "Information technology: Open Systems Interconnection, Systems management overview", 1998, Available from:

- http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=24406
- [56] U.S. Department of State, "Internation Traffic in Arms Regulations", 2012, Available from: http://www.pmddtc.state.gov/regulations_laws/documents/official_itar/2012/ITAR_Part_120.pdf
- [57] Federation Information Security Management Act, "Information Security", 2010, Available from: http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final_updated-errata_05-01-2010.pdf
- [58] Health Level Seven International, HL7, "Introduction to HL7 Standards", 1987, Available from: <http://www.hl7.org/implement/standards/index.cfm?ref=nav>
- [59] Oasis Open, "Security Assertion Markup Language (SAML) V2.0 Technical Overview", Available from: <https://www.oasis-open.org/committees/download.php/27819/sstc-saml-tech-overview-2.0-cd-02.pdf>
- [60] Shabtai, Asaf, Elovici, Yuval, Rokach, Lior , "A Survey of Data Leakage Detection and Prevention Solutions", Springer-Verlag New York Incorporated, 2012
- [61] Payment Card Industry Security Standards Council, PCI SSC, Available from: <https://www.pcisecuritystandards.org/documents/PCI%20SSC%20Quick%20Reference%20Guide.pdf>
- [62] ISO 17826:2012, "Information Technology: Cloud Data Management Interface (CDMI)", 2012, Available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=60617
- [63] Health Insurance Portability and Accountability, HIPAA, 1996, Available from: <http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf>
- [64] ISO 27005:2011, "Information Technology: Security Techniques - Information Security Risk Management", 2011, Available from: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=56742
- [65] ISO 27002:2005, "Information Technology: Security Techniques - Code of Practice for Information Security Management", 2005, Available from: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=50297

Survey of Server Virtualization

Radhwan Y Ameen
Department of Comp. Engineering
College of Engineering, Mosul University
Mosul, Iraq

Asmaa Y. Hamo
Dept. of Software Engineering
College of Computer Sc. and Mathematics, Mosul
University
Mosul, Iraq

Abstract— Virtualization is a term that refers to the abstraction of computer resources. The purpose of virtual computing environment is to improve resource utilization by providing a unified integrated operating platform for users and applications based on aggregation of heterogeneous and autonomous resources. More recently, virtualization at all levels (system, storage, and network) became important again as a way to improve system security, reliability and availability, reduce costs, and provide greater flexibility. Virtualization has rapidly become a go-to technology for increasing efficiency in the data center. With virtualization technologies providing tremendous flexibility, even disparate architectures may be deployed on a single machine without interference This paper explains the basics of server virtualization and addresses pros and cons of virtualization .

Keywords- *virtualization ,server ,hypervisor ,Virtual Machine Manager, VMM , para virtualization , full virtualization, OS level server.*

I. INTRODUCTION

Virtualization is a technique for hiding the physical characteristics of computing resources from the way in which other systems, applications, or end users interact with those resources. It introduces a software abstraction layer between the hardware and the operating system and applications running on top of it [9] [1]. This abstraction layer is called virtual machine monitor (VMM) or hypervisor and basically hides the physical resources of the computing system from the operating system (OS). Since the hardware resources are directly controlled by the VMM and not by the OS, it is possible to run multiple (possibly different) OSs in parallel on the same hardware. As a result, the hardware platform is partitioned into one or more logical units called virtual machines (VMs). "Virtuality" differs from "reality" only in the formal world, while possessing a similar essence or effect. In the computer world, a *virtual environment* is perceived the same as that of a *real environment* by application programs and the rest of the world, though the underlying mechanisms are *formally* different.

Virtualization was first developed in 1960's by IBM Corporation, originally to partition large mainframe computer

into several logical instances and to run on single physical mainframe hardware as the host. This feature was invented because maintaining the larger mainframe computers became cumbersome. The scientist realized that this capability of partitioning allows multiple processes and applications to run at the same time, thus increasing the efficiency of the environment and decreasing the maintenance overhead[15].

II. VIRTUAL MACHINE

A. Virtual Machine History

Virtual machines have been in the computing community since 1960s, systems engineers and programmers at **Massachusetts Institute of Technology** (MIT) recognized the need for virtual machines. In her authoritative discourse Melinda Varian [15] introduces virtual machine technology, starting with the compatible Time-Sharing System (CTSS).

IBM engineers had worked with MIT programmers to develop a time-sharing system to allow project teams to use part of the mainframe computers. Varian goes on to describe the creation, development, and use of virtual machines on the IBM OS/360 Model 67 to the VM/370 and the OS/390 [15]. Varian's paper covers virtual machine history, emerging virtual machine designs, important milestones and meetings, and influential engineers in the virtual computing community.

In 1973, Srodowa and Bates [14] demonstrated how to create virtual machines on IBM OS/360s. They describe the use of IBM's Virtual Machine Monitor, a hypervisor, to build virtual machines and allocate memory, storage, and I/O effectively. Srodowa and Bates touch on virtual machine topics still debated today: performance degradation, capacity, CPU allocation, and storage security.

Goldberg concludes "the majority of today's computer systems do not and cannot support virtual machines. The few virtual machine systems currently operational, e.g., CP-67, utilize awkward and inadequate techniques because of unsuitable architectures" [16].

Goldberg proposes the "Hardware Virtualizer," in which a virtual machine would communicate directly with hardware instead of going through the host software. Nearly 30 years later, industry analysts are excited about the announcement of hardware architectures capable of supporting virtual machines

efficiently. AMD and Intel have revealed specifications for Pacifica and Vanderpool chip technologies with special virtualization support features.

The 1980s and early 1990s brought distributing computing to data centers. Centralized computing and virtual machine interest was replaced by standalone servers with dedicated functions: email, database, Web, applications. After significant investments in distributed architectures, renewed focus on virtual machines as a complimentary solution for server consolidation projects and data center management initiatives has resurfaced [17].

Recent developments in virtual machines on the Windows x86 platform merit a new chapter in virtual machine history. Virtual machine software from Virtuozzo, Microsoft, Xen, and EMC (VMWare) has spurred creative virtual machine solutions. Grid computing, computing on demand, and utility computing technologies seek to maximize computing power in an efficient, manageable way.

The virtual machine was created on the mainframe. It has only recently been introduced on the mid-range, distributed, x86 platform. Technological advancements in hardware and software make virtual machines stable, affordable, and offer tremendous value, given the right implementation.

B. Virtual Machine Concepts

Goldberg R. P defined Virtual machines as : "A system...which...is a hardware-software duplicate of a real existing machine, in which a non-trivial subset of the virtual machine's instructions execute directly on the host machine..." [22,23]. While Goldberg R, June defined Virtual machines as: "A virtual machine is taken to be an *efficient, isolated duplicate* of the real machine. We explain these notions through the idea of a *virtual machine monitor*" (VMM).

See Figure 1.

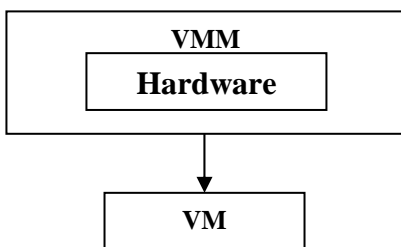


Fig. 1 The virtual machine monitor

As a piece of software a VMM has three essential characteristics.

First, the VMM provides an environment for programs which is essentially identical with the original machine;

second, programs run in this environment show at worst only minor decreases in speed; and last, the VMM is in complete control of system resources". [20] and Kreuter, D defined it as:

A virtual machine (VM) is an abstraction layer or environment between hardware components and the end-user. Virtual machines run operating systems and are sometimes referred to as virtual servers. A host operating system can run many virtual machines and shares system hardware

components such as CPUs, controllers, disk, memory, and I/O among virtual servers" [18].

C. Virtual Machine Types

Virtual machines are implemented in various forms. Mainframe, open source, para virtualization, and custom approaches to virtual machines have been designed over the years. Complexity in chip technology and approaches to solving the x86 limitations of virtualization have led to three different variants of virtual machines:

1. software virtual machines (see Figure 2), which manage interactions between the host operating system and guest operating system (e.g., Microsoft Virtual Server 2005);

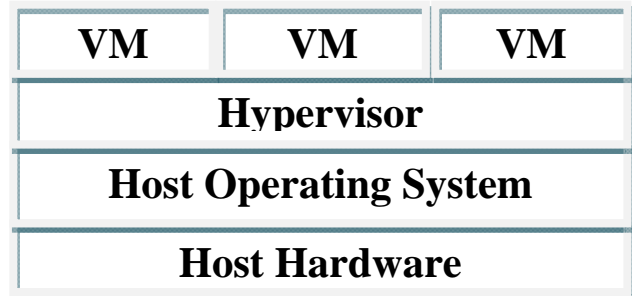


Fig. 2 Software virtual machines

2. hardware virtual machines (see Figure 3), in which virtualization technology sits directly on host hardware (bare metal) using hypervisors, modified code, or APIs to facilitate faster transactions with hardware devices (e.g., VMWare ESX);

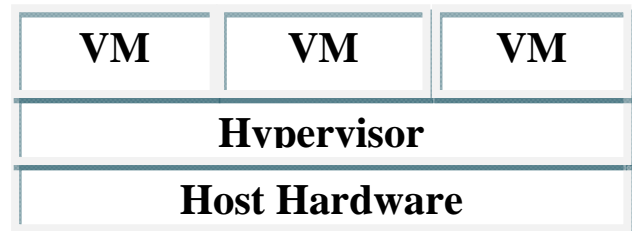


Fig. 3 Hardware virtual machines.

3. virtual OS/containers (see Figure 4), in which the host operating system is partitioned into containers or zones (e.g., Solaris Zones, BSD Jail).

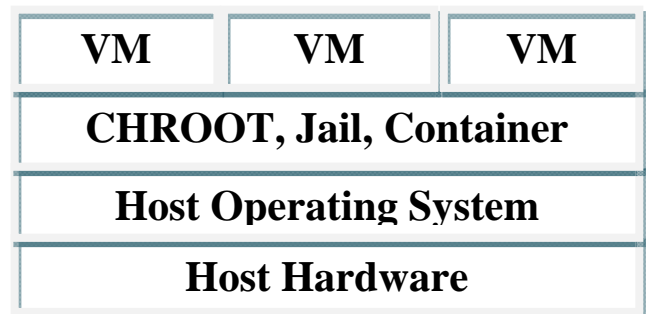


Figure 4: Virtual OS/containers virtual machines.

A simple UNIX implementation called *chroot* allows an alternate directory path for the root file system. This creates a “jail,” or sandbox, for new applications or unknown applications. Isolated processes in *chroot* are best suited for testing and applications prototyping. They have direct access to physical devices, unlike emulators. Sun Microsystems’ “Solaris Zones” technology is an implementation of *chroot*, similar to the FreeBSD jail design, with additional features. Zones allow multiple applications to run in isolated partitions on a single operating system [19].

Each zone has its own unique process table and management tools that allow each partition to be patched, rebooted, upgraded, and configured separately. Distinct oot privileges and file systems are assigned to each zone. [20].

III. VIRTUALIZATION

A. Definitions of Virtualization

There are many definitions of term virtualization as shown below:

Rune Johan Andresen defined it as : “*Virtualization* is a framework of dividing the resources of a computer into multiple execution environments. More specific it is a layer of software that provides the illusion of a real machine to multiple instances of virtual machines.” [4] [11]

While Susanta Nanda , Tzi-cker Chiueh defined it as : ”*Virtualization* is a technology that combines or divides computing resources to present one or many operating environments using methodologies like hardware and software partitioning or aggregation, partial or complete machine simulation, emulation, time-sharing, and many others” . [7]

IBM defined it as: “*Virtualization* is the creation of substitutes for real resources, that is substitutes that have the same functions and external interfaces as their counterparts, but that differ in attributes, such as size, performance, and cost.” [29]

Andi Mann defined it as : ”*Virtualization* is, at its foundation, a technique for hiding the physical characteristics of computing resources from the way in which other systems, applications, or end users interact with those resources. This includes making a single physical resource (such as a server, an operating system, an application, or storage device) appear to function as multiple logical resources; or it can include making multiple physical resources (such as storage devices or servers) appear as a single logical resource.” . [9]

G. Heiser defined it as:“*virtualization* allows a single computer to host multiple virtual boards (or virtual machines), each isolated from one another, with the possibility of running different operating systems. The main advantage is that, if a virtual board fails, the other ones are kept safe at a reasonable cost “[10].

William von H defined it as : ”*Virtualization* is simply the logical separation of the request for some service from the physical resources that actually provide that service” . [8]

Chaudhary V.,Minsuk Cha.,Walters J.P.,Guercio S.,Gallo S, defined it as :”Virtualization is a common strategy for improving the utilization of existing computing resources, particularly within data centers.”[3]

Amit Singh defined it as: ”Virtualization is framework or methodology of dividing the resources of a computer into multiple execution environments, by applying one or more concepts or technologies such as hardware and software portioning, time-sharing, partial or complete machine simulation, emulation, quality of service, and many others.”[2]

Joshua S. White,Adam W. Pilbeam defined it as: ”*Virtualization* is a mechanism permitting a single physical computer to run sets of code independently and in isolation from other sets”. [6]

Sahoo J., Mohapatra S., Lath R. defined it as: ”Virtualization is a technology that introduces a software abstraction layer between the hardware and the operating system and applications running on top of it.”[1]

TBD Networks defined it as: ”Virtualization is a technology that enables running two or more operating systems simultaneously on a single computer.”[5]

And Lawrence C. Miller, CISSP defined it as: “*Virtualization* is technology emulates real or physical computing resources, such as desktop computers and servers, processors and memory, storage systems, networking, and individual applications.”[25]

We define it as: “*virtualization* is a technology to divided or combined the resources of computer system between multiple operating systems or applications, to make illusion that each one access the real resources”.

B. Benefits of virtualization

There can be innumerable reasons how virtualization can be useful in practical scenarios, a few of which are the following:

- Server Consolidation. [9] [23] [8] [25] [29] [40]
- Application consolidation. [8] [25] [40]
- Sandboxing[8,].
- Multiple execution environments[8] [40]
- Virtual hardware. [8]
- Multiple simultaneous OS[13]. [8, [40].
- Debugging. [8]
- Software Migration. [8]
- Appliances[23] [8].
- Testing[23] [40].
- Better Use of Existing Hardware[9].
- Reduction in New Hardware Costs[9] [8][23][28] [36]
- Reduction in IT Infrastructure Costs[9] [23] [36]
- Reduced downtime[9] [28] [36].
- Simplified System Administration[9].

- Increased Uptime and Faster Failure Recovery[9]
- Simplified Capacity Expansion[9]
- Simpler Support for Legacy Systems and applications [12] [9] [23] [40].
- Simplified System - Level Development[9] [40].
- Simplified System Installation and Deployment[9].
- Simplified System and Application Testing Business Continuity and Disaster Recovery[9] [9] [23] [25] [28] [29].
- Business Agility[9].
- Resource sharing .
- Isolation [13]. [12].
- Increase Flexibility. [9] [13]. [40]
- Increase Availability[23] [26] [36].
- Increase Scalability[23] [26] [36] .
- Increase Hardware utilization[12]. [26]
- Increase Security[12]. [26] [.
- Load Balancing[36]
- brings hardware independence[13] [26].

C. Disadvantage of virtualization

- SPOF Single Point of Failure Problem.
- Overhead causing decreased performance has been the biggest con with virtualization.
- The management interface This can be a problem as it encumbers consolidation of several platforms into the same environment.
- Increase in Networking Complexity and Debugging Time. [1][8][9].

D. Types of virtualization

There are so many different types of virtualization, Mobile, Data, Memory, Desktop, Storage, Server, Network, Application, Grid, and Clustering as shown in fig 5.

1) Mobile Virtualization

VMware defined it as : Mobile Virtualization(MVP) is a thin layer of software that is embedded on a mobile phone to decouple the applications and data from the underlying hardware. It is optimized to run efficiently on low power consuming and memory constrained mobile phones. The MVP currently supports a wide range of real-time and rich operating systems including Windows CE 5.0 and 6.0, Linux 2.6.x, Symbian 9.x, eCos, μ ITRON NORTi and μ C/OS-II.

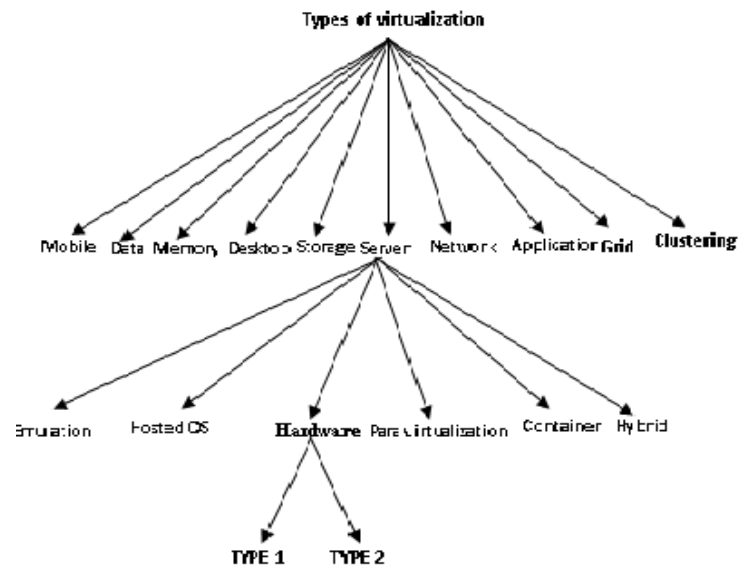


Fig. 5 Types of virtualization

2) Data Virtualization

Data virtualization from Andi Mann: "Data virtualization abstracts the source of individual data items including entire files, database contents, document metadata, messaging information, and more and provides a common data access layer for different data access methods such as SQL, XML, JDBC, File access, MQ, JMS, etc. This common data access layer interprets calls from any application using a single protocol, and translates the application request to the specific protocols required to store and retrieve data from any supported data storage method. This allows applications to access data with a single methodology, regardless of how or where the data is actually stored." [9]

3) Memory Virtualization

Carl A. Waldspurger ,Palo Alto define it as:

A guest operating system that executes within a virtual machine expects a zero-based physical address space, as provided by real hardware. ESX Server gives each VM this illusion, virtualizing physical memory by adding an extra level of address translation. Borrowing terminology from Disco [34], a machine address refers to actual hardware memory, while a physical address is a software abstraction used to provide the illusion of hard-ware memory to a virtual machine. We will often use "physical" in quotes to highlight this deviation from its usual meaning.[33]

4) Desktop Virtualization

William von H defined it as: The term "desktop virtualization" describes the ability to display a graphical desktop from one computer system on another computer system or smart display device.

Many window managers, particularly those based on the X Window System, also provide internal support for multiple, virtual desktops that the user can switch between and use to display the

output of specific applications. The X Window System also supports desktop virtualization at the screen or display level, enabling window managers to use a display region that is larger than the physical size of your monitor.[8]

5) Storage Virtualization

Li Bignag, Shu Jiwu, Zheng Weimin defined it as : "Storage Virtualization is the emerging technology that creates logical abstractions of physical storage systems. Storage Virtualization has tremendous potential for simplifying storage administration and reducing costs for managing diverse storage assets."[21]

6) Network Virtualization

Naga Dinesh define it as: "*Network Virtualization* :would provide abstraction layer that can decouple the physical network equipment from the delivered business services over the network to produce a more responsive and well-organized communications"[22]

7) Application Virtualization

Naga Dinesh Defined it as: "This type of virtualization allows the user to run the application using local resources without installing the application in his system completely".[22]

While Joshua S. White, Adam W. Pilbeam define it as : "provides smaller single application virtual machines that allow for emulation of a specific environment on a client system. For example a Java Virtual Machine allows disparate operating systems such as Windows and Linux to run the same Java program as long as they have the Java VM installed. This form of virtualization is limited in that it only provides single program isolation from the host, but is useful when testing programs out without installing them". [6]

8) Grid Computing

Andi Mann defined it as: "Like a cluster, a grid provides a way to abstract multiple physical servers from the application they are running. The major difference is that the computing resources are normally spread out over a wide network, potentially across the Internet, and the physical servers that comprise a grid do not have to be identical. Unlike a cluster, where each server is locally connected, is likely to be identical, and can handle the same processing requirements, a grid is made up of heterogeneous systems, in diverse locations, each of which may specialize in a particular processing capability. Much greater coordination is needed to allocate the resources to appropriate workloads." [9]

9) Clustering

Andi Mann define it as: "A cluster is a form of virtualization that makes several locally-attached physical systems appear to the application and end users as a single processing resource. This differs significantly from other virtualization technologies, which normally do the opposite, i.e. making a single physical system appear as multiple independent operating environments. A typical use case for clustering is to group a number of identical physical servers to

provide distributed processing power for high-volume applications, or as a "Web farm", which is a collection of Web servers that can all handle load for a Web-based application." [9]

10) Server Virtualization(machine , cpu)

The terms " server virtualization ", " machine virtualization "and "cpu virtualization" describe the ability to run an entire virtual machine, including its own operating system, on another operating system. The most common virtualization known in general is Server Virtualization.

a) Server Virtualization definitions:

Lawrence C. Miller defined it as: "*Server virtualization* creates "virtual environments" that allow multiple applications or server workloads to run on one computer, as if each has its own private computer".[25]

While Naga Dinesh define it as: "The technique of masking of server resources, which includes the identity and number of every existing servers, processors, and OS users is termed as server virtualization".[22]

HP define it as: "*server virtualization* refers to abstracting, or masking, a physical server resource to make it appear different logically to what it is physically. In addition, server virtualization includes the ability for an administrator to relocate and adjust the machine workload."[26]

VMware define it as: "virtualization enables one computer to perform the job of multiple computers, by sharing the resources of a single computer across multiple environments".[30]

Citrix system define it as: "the ability to decouple software from the hardware layer, allowing server workloads to be streamed onto any platform in any direction." [32]

And Darla Sligh define it as: "*Server virtualization* is a software-based tool enabling the division of computer resources and the sharing of multiple environments simultaneously." [31]

We define it as: "server virtualization is the ability to run many operating systems with isolation and independences on other operating system"

b) Server Virtualization types

Figure 6 show the traditional computer system without virtualization.

In x86 environments, there are several variations within software-layer abstraction of the server hardware, including these general categories:

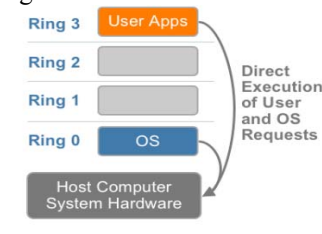


Fig. 6 x86 privileged level architecture without virtualization.

i. Emulation

Emulation is a virtualization method in which a complete hardware architecture may be created in software. This software is able to replicate the functionality of a designated hardware processor and associated hardware systems. This method provides tremendous flexibility in that the guest OS may not have to be modified to run on what would otherwise be an incompatible architecture. Emulation features tremendous drawbacks in performance penalties as each instruction on the guest system must be translated to be understood by the host system. This translation process is extremely slow compared to the native speed of the host, and therefore emulation is really only suitable in cases where speed is not critical, or when no other virtualization technique will serve the purpose. Examples of this approach are QEUM, Bochs, Crusoe, and BIRD. [6] [7][8]

Advantages:

- tremendous flexibility in that the guest OS may not have to be modified.

Disadvantages:

- performance penalties as each instruction on the guest system must be translated to be understood by the host system.

ii. Binary translation

With binary translation technology as shown in figure 7, the guest OS is not aware it is operating on virtualized hardware. The hypervisor manages the access of each guest OS to the physical hardware resources by masking the hardware from the guest OS. It emulates portions of the system hardware and provides the guest OS with the illusion of a standard physical server with well-defined hardware devices. The hypervisor ensures that any instructions from the guest OS that affect system parameters—such as privileged instructions to the CPU—are handled in a way that does not affect the operation of other guest operating systems or cause OS kernel faults. The hypervisor traps the instruction and performs necessary translations that make the guest OS think it has complete control over the server hardware. The critical issue of dynamical binary translation is its low performance efficiency and design complexity due to the incapability of classical trap-and-emulate virtualization with previous generation of x86 architecture. Examples of this approach is VMWare ESX [26][31][8][39] [38]

Advantages:

- provides the guest OS with the illusion of a standard physical server with well-defined hardware devices.
- No need to modified guest OS.

Disadvantages:

- low performance efficiency
- design complexity due to the incapability of classical trap-and-emulate virtualization

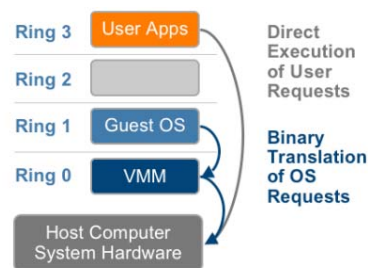


Fig. 7 The binary translation approach

iii. Hosted OS, application-layer abstraction virtualization

With Hosted OS, application-layer abstraction virtualization as shown in figure 8, another software-only approach uses a hypervisor layer that is hosted by an underlying OS. Because it resides as an application on top of the host OS, this type of abstraction inherits its hardware support and device compatibility from the host OS. This provides an advantage for customers who want to run an older, legacy OS on newer server hardware. However, the tradeoff for this hardware compatibility is the performance overhead required by the hypervisor layer. Typically, such hosted solutions are used in smaller, departmental environments rather than in large data center deployments because the hosted solutions often lack capabilities such as dynamic load balancing or clustering. Examples of this approach are Microsoft Virtual Server 2005 R2 and VMware Server (formerly VMware GSX)[26][31] [35][8] [37] [38].

Advantages:

- Virtualization product is installed onto the host desktop just as any other application
- The host desktop OS can continue to be used
- Uses the host OS's device drivers - the virtualization product supports whatever hardware the host does

Disadvantages:

- Slow performance. [36]

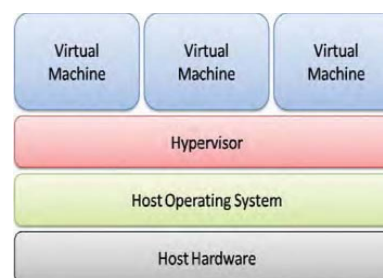


Fig 8 Hosted OS, application-layer abstraction virtualization

iv. Hardware-assisted virtualization (full virtualization, bare-metal virtualization)

With hardware-assisted virtualization (sometimes referred to as full virtualization) as shown in figure 9, the hypervisor is assisted by the processor hardware such as AMD-V or Intel VT-x processor virtualization technologies. In this scenario,

when the guest OS makes a privileged instruction call, the processor (CPU) traps the instruction and returns it to the hypervisor to be emulated. Once the operation is serviced by means of the hypervisor, the modified instruction is returned back to the CPU for continued execution. Hardware assistance reduces the software overhead required by the hypervisor. Hardware assistance from AMD-V and Intel VT-x technologies extends the x86 instruction set with new instructions that affect the processor, memory, and local I/O address translations. The new instructions enable guest operating systems to run in the standard Ring-0 architectural layer, as they were designed to do, removing the need for ring compression. Examples of this approach are Microsoft Hyper-v, Citrix Xen, Parallels Workstation, Virtual Iron and VMWare ESX Server [26][31][1][7][8][35][37][38]

Advantages:

- Performance .
- Products are distributed as appliances or server OSes.

Disadvantages

- Vendor publishes a hardware compatibility list (HCL) that dictates what hardware can be used with their virtualization product. [36]

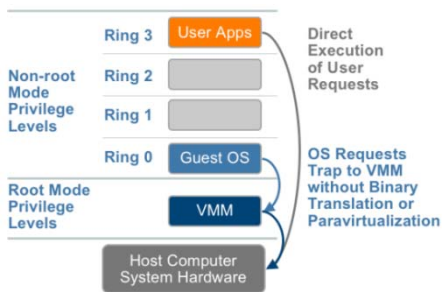


Fig. 9 The hardware assist approach

v. Paravirtualization

Paravirtualization as shown in figure 10, refers to a technique in which the guest OS includes modified (paravirtualized) I/O drivers for the hardware. Unlike a binary translation approach, the hypervisor does not need to trap and translate all privileged layer instructions between the guest OS and the actual server hardware. Instead, the modified guest OS makes calls directly to the virtualized I/O services and other privileged operations. Therefore, paravirtualization techniques have the potential to exhibit faster raw I/O performance than binary translation techniques. Some of the hypervisor implementations that use this method (Citrix XenServer, Red Hat Enterprise Linux 5, and SUSE Linux Enterprise) are unique in that they support paravirtualization when using a modified guest OS and hardware-assisted virtualization when the guest OS is not virtualization-aware. Device interaction in paravirtualized environment is very similar to the device interaction in full virtualized environment; the virtual devices in paravirtualized environment also rely on physical device drivers of the underlying host. Where paravirtualization differs is that it does not simulate hardware resources but instead

offers a special Application Programming Interface (API) to hosted virtual machines. Examples of this approach are Xen, Denali and User-Mode Linux (UML) [36] [26][31][1][6][3][7][35][37][38].

Advantages:

- significant performance improvements over other virtualization solutions

Disadvantages:

- The VM OS must be modified.

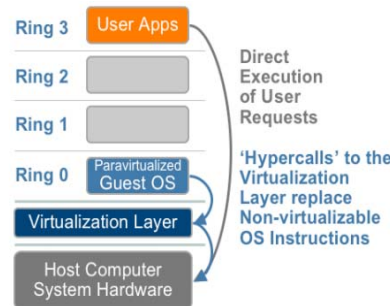


Fig. 10 The Paravirtualization approach

vi. Hosted OS, kernel-layer abstraction (OS Containers virtualization, Single Kernel Image (SKI))

Kernel-layer abstraction as shown in figure 11, refers to a technique in which the abstraction technology is built directly into the OS kernel rather than having a separate hypervisor layer. System - level virtualization is based on the change root (CHROOT) concept that is available on all modern UNIX - like systems.. The direct access to hardware could potentially provide greater performance than using a binary translation technology; however, because there is no separation between the hypervisor and the operating system, there is the possibility that resource conflicts may occur between multiple virtual machines. Virtual OS containers do not use hypervisors (or VMM), which is a software application that works to manage the logical separate of physical resource [40]. They use containers, or sandboxes, called chroot, to partition the host operating system into containers or zones (e.g., Solaris Zones, BSD Jail), so multiple applications can run in isolated partitions on a single operating system. [26][31][35][37] this concept implements virtualization by running more instances of the same OS in parallel. This means that not the hardware but the host OS is the one being virtualized[1] OS-layer virtualization tends to be more efficient and fails only by little to provide the same isolation [41]. Examples of this approach are FreeBSD 's chroot jails, FreeVPS, Linux VServer, OpenVZ, Solaris Zones and Containers, and Virtuozzo. [8][38].

Advantages:

- Performance
- Reduced disk space requirements, containers can use the same files

Disadvantages:

- The VM OS must be the same OS as the host OS. [36]

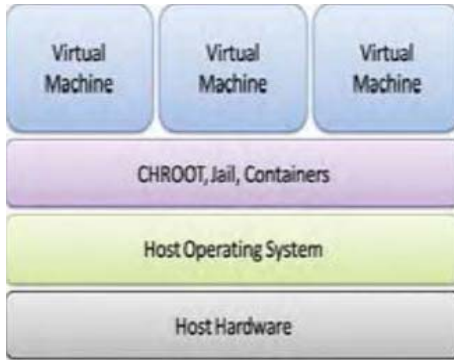


Fig.11 Containers virtualization

vii. *Native virtualization, Hybrid virtualization, a hybrid*

virtualization as shown in figure 12, is the newest form. It is a combination of full virtualization and paravirtualization and uses input/output (I/O) acceleration techniques. This compromise allows for an increase in speed (and indeed with hardware acceleration it can be very fast), but potential performance degradation can exist in an environment where the instructions are relying more heavily on the emulated actions rather than the direct hardware access portions of the hypervisor. It adds overhead and complexity. Examples of this approach are VMware and Microsoft Virtual PC [31] [6][35] [38]

Advantages:

- Performance

Disadvantages:

- Requires the underlying processor have virtualization extensions (examples: Intel-VT, AMD-V) to function.
- Older hardware that could otherwise be utilized by other virtualization architectures cannot be used. [36]

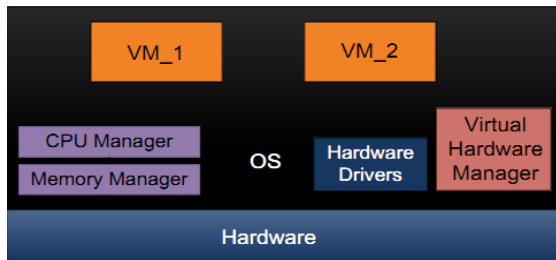


Fig. 12 The Hybrid Virtualization

IV. TYPES OF HARDWARE VIRTUALIZATION

- 1) Type 1 (native or bare metal).
- 2) Type 2 (hosted).

The kernel was known as the *supervisor* in mainframes; hence the term *hypervisor* was coined for the software operating above the supervisor.

Two types of hypervisors are defined for server virtualization:

Type 1 and Type 2 (see Figure 13,14). A Type 1 hypervisor, also known as a *native* or *bare metal* hypervisor,

type 1 hypervisors run directly on the system hardware. The following figure shows one physical system with a type 1 hypervisor running directly on the system hardware, and three virtual systems using virtual resources provided by the hypervisor.

A Type 2 hypervisor, also known as a *hosted* hypervisor, it run on a host operating system that provides virtualization services, such as I/O device support and memory management. The following figure shows one physical system with a type 2 hypervisor running on a host operating system and three virtual systems using the virtual resources provided by the hypervisor.[25][29]

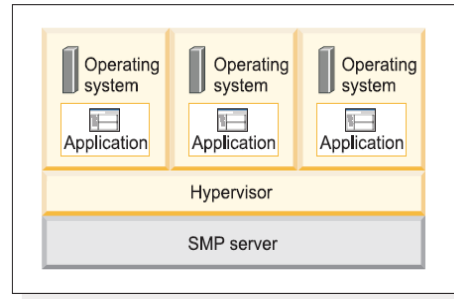


Fig. 13 Type 1 hypervisors.

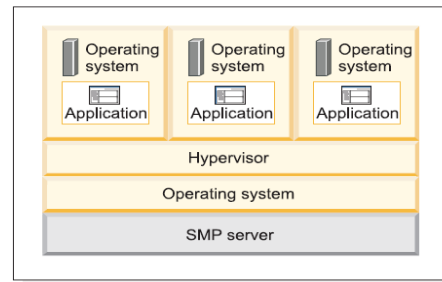


Fig.14 Type 2 hypervisors

c) *Advantages of server virtualization*

many researchers note the following benefits for virtualizing servers within data centers

- enabling automated data center operations[24]
- improving the speed of service delivery[24]
- supporting application configuration and availability. [29]
- Consolidation to reduce hardware cost
- reducing the need for physical servers[24].
- reducing server operational maintenance[28].
- reducing the operating expense][26].
- reducing provisioning and the deploying of new services[28]
- reducing disaster recovery times[24]][26]
- improving network and application security[27]

- reducing costs associated with the test and development of in-house applications[27] [29]
- enabling simple, responsive, utility-style „cloud computing“ infrastructure[27]
- reducing various testing and migration issues[26].
- Reducing (TCO) Total Cost of Ownership[27] [29]
- Improving Flexibility, High Availability and Performance[27] [29]

d) *Pitfalls of server virtualization*

Researchers indicate that improper employment of server virtualization can result in the following pitfalls

- overloading the server utilization infrastructure, which can introduce application latency; [31]
- increasing IT operational costs because of additional time and resources required for extensive research efforts; [31]
- magnifying failures because a hardware failure could impact multiple virtual servers and the applications they host;[31]
- introducing virtual machine sprawl, which may substantially increase the overall number of server operating images that need to be managed by system administrators; [31]
- enabling improper security processes because within the virtual server, the server administrator with access to the root ID can alter or disable security settings; thereby , exposing servers to security vulnerabilities;[31]
- exposing IT operations to network (traffic) uncertainties][31]
- requiring enhanced IT skill sets to manage more environments at once.[31]

REFERENCES

- [1] Sahoo J., Mohapatra S., Lath R., 2010 , "Virtualization: A Survey On Concepts, Taxonomy And Associated Security Issues", Second International Conference on Computer and Network Technology, pp. 222 - 226.
- [2] Amit Singh, 2004, "An Introduction to Virtualization", kernelthread.com
- [3] Chaudhary V., Minsuk Cha., Walters J.P., Guercio S., Gallo S., "A Comparison of Virtualization Technologies for HPC ", 22nd International Conference on Advanced Information Networking and Applications, 2008, pp. 861 - 868.
- [4] Rune Johan Andresen,2004," Virtual Machine Monitors", Enterprise Management Associates, Inc
- [5] TBD Networks," Network Virtualization", www.tbdnetworks.com, accessed on 02/02/2011.
- [6] Joshua S. White,Adam W. Pilbeam,2010," A Survey of Virtualization Technologies With Performance Testing", eprint arXiv:1010.3233
- [7] Susanta Nanda , Tzi-cker Chiueh, "A Survey on Virtualization Technologies", Department of Computer Science SUNY at Stony Brook
- [8] William von H,2008," Professional Xen Virtualization",Wiley Publishing, Inc.
- [9] Andi Mann,2006," Virtualization 101: Technologies, Benefits, and challenges", Enterprise Management Associates, Inc.
- [10] G. Heiser, 2008,"The role of virtualization in embedded systems",In IIES '08: Proceedings of the 1st workshop on Isolation and integration in embedded systems, pages 11–16, New York, NY, USA., ACM.
- [11] Rosenblum M. and Garfinkel T. , May 2005 , "Virtual machine monitors: current technology and future trends". Computer, 38(5):39–47.
- [12] Renato J. Figueiredo, Peter A. Dinda, and J. Fortes. 2003. "A case for grid computing on virtual machines". In ICDCS '03: Proceedings of the 23rd International Conference on Distributed Computing Systems, page 550, Washington, DC, USA. IEEE Computer Society
- [13] Alexandra Aguiar, Fabiano Hessel ,2010 , "Embedded Systems' Virtualization: The Next Challenge",IEEE.
- [14] Srodawa, R. J. and Bates, L. E. 1973. "An efficient virtual machine implementation".In Proceedings of ACM SIGARCH-SIGOPS Workshop on Virtual Computer Systems.
- [15] Varian, M. 1997. "VM and the VM community: Past, present, and future. Office of Computing and Information Technology", Princeton University, Princeton, NJ.
- [16] Goldberg, R. P. 1973." Architecture of virtual machines". Honeywell Information Systems, Inc., Billerica, MA.
- [17] Serjeant, A. 2005. "Building a case for server consolidation". VMWorld.Presentation (Oct. 20).
- [18] Kreuter, D. 2004. "Where server virtualization was born", Virtual Strategy Magazine (July 21).
- [19] Tucker A, Comay, D, 2004," Solaris zones:Operating system support for server consolidation". Sun Microsystems, Inc.
- [20] Jeff Daniels ,2009, "Server Virtualization Architecture and Implementation", www.acm.org/crossroads, Vol. 16, No.1
- [21] Li Bignag, Shu Jiwu, Zheng Weimin, February 2005, "Design and Implementation of a Storage Virtualization System Based on SCSI Target Simulator in SAN", Tsinghua Scienceand Technology, Vol. 10, Issue 1, pp 122-127.
- [22] Naga Dinesh,2011,"An approach to storage virtualization", Master of science, Texas A&M University-Kingsville
- [23] Jeanna N, Eli M, Todd D, Wenjin H, Jeremy B, Patrick F, Brendan J,2008," Running Xen: A Hands-On Guide to the Art of Virtualization", Prentice Hall.
- [24] Wolf, C., & Halter, E., (2005).," Virtualization from the desktop to the enterprise"., New York: Apress.
- [25] Lawrence C. Miller, CISSP,2012,"Server Virtualization For Dummies®, Oracle Special Edition",John Wiley & Sons, Inc
- [26] HP,2009, "Server virtualization technologies for x86-based HP Blade System and HP ProLiant servers technology brief, 3rd edition", Hewlett-Packard Development Company, L.P.
- [27] Ravi G Singh, (2007).," Server Virtualization and Consolidation - A case study"., IBM Corp..
- [28] VMware, (2006).," Reducing Server Total Cost of Ownershipwith VMware Virtualization Software", VMware, Inc.
- [29] IBM, (2005).," IBM SystemsVirtualization Version 2 Release 1"International Business Machines Corporation 2005
- [30] VMware,2009" Transform your business with virtualization", VMware inc
- [31] Darla Sligh .,2010. "PERSPECTIVES OF IT PROFESSIONALS ON EMPLOYING SERVER VIRTUALIZATION TECHNOLOGIES" ,Ph.D. , Robert Morris University
- [32] Citrix system.,2007" A pragmatic approach to server and data center consolidation",. PlateSpin Ltd.PlateSpin Ltd
- [33] Carl A. Waldspurger ,Palo Alto , 2002, "Memory Resource Management in VMware ESX Server",The USENIX Association.
- [34] Edouard Bugnion, Scott Devine, Kinshuk Govil, and Mendel Rosenblum., November 1997, "Disco: Running Commodity Operating Systems on Scalable Multiprocessors," ACM Trans-actions on Computer Systems.
- [35] Yang Yu,2007," OS-level Virtualization and Its Applications",Ph.D, Stony Brook University

- [36] Jean S. Bozman, Gary P. Chen, 2009,"Optimizing Hardware for x86 Server Virtualization",IDC.
- [37] Greg Shields,2008," Selecting theRight Virtualization Solution", Realtimepublishers.com, Inc.
- [38] VMware,2007" Understanding Full Virtualization, Paravirtualization, and Hardware Assist", VMware inc
- [39] Marshall, D., Reynolds, W., & McCory, D. ,2006. " Advanced server virtualization" , Auerbach Publishing, Inc
- [40] Qian Lin, Zhengwei Qi, Jiewei Wu, Yaozu Dong, Haibing Guan,2012" Optimizing virtual machines using hybrid virtualization", Elsevier Inc.
- [41] Soltész, Stephen; Poetzl, Herbert; Fiuczynski, Marc E.; Bavier, Andy; Peterson, Larry, 2007," Container-based operating system virtualization: a scalable, high-performance alternative to hypervisors", Proceedings of the 2007 conference on EuroSys. ACM Press.

A Proposed DK-PC Algorithm for Code Bloat Control in a Tree-Based Genetic Programming

Oghorodi, Duke Urhe-otokoh
Department of Computer Science,
College of Education, Warri
Delta State, Nigeria,

Asagba, Prince Oghenekaro
Department of Computer Science,
University of Port Harcourt,
River State, Nigeria

Abstract — This paper addresses the Genetic Programming (GP) issue of code bloat which is the uncontrolled growth of program codes without a commensurate improvement in the program fitness to solve a given problem. Code Bloat is a serious issue in GP as it consumes computer memory and processing time. Though, several reasons and solutions for code bloat control have been suggested in literature, yet no final solution has been found so far. Against this backdrop, we proposed the Delete lower and Keep higher fitness value Programs after Crossover (DKPC) algorithm which keeps the higher fitness value program and delete the lower value fitness value programs from memory. We tested the Boolean 6-multiplexer and Boolean 11-Multiplexer functions against the preparatory requirements using our proposed algorithm, and we got very impressive results; we observed that the algorithm was able to control bloat to a large degree. However, the algorithm performed better in the Boolean 11-multiplexer function than in the Boolean 6-multiplexer function. Both functions displayed almost the same behaviour; except that the Boolean 11-multiplexer exhibited higher performance result than the Boolean 6-multiplexer in terms of better program size reduction. To this extent, our algorithm performed better in bloat control, based on the benchmark problems used.

Keywords — Genetic Programming, Evolutionary algorithms, Code Bloat

1. INTRODUCTION

Genetic Programming, the most recently introduced of a group of four algorithms collectively known as the Evolutionary Algorithms is an offshoot of the Holland's Genetic Algorithm; developed by John Koza in 1990. Just like other Evolutionary Algorithms, it is an attempt to answer the central question in Artificial Intelligence of how do we solve problems using computers without explicitly instructing or telling the computer how to do it? Genetic Programming uses the principle of evolution such that optimal solution to problem is obtained from a gradual and a successive improvement to an initial guess at a possible solution over a period of time until a termination criterion is attained [8]. Therefore, setting a good initial solution based on requirements of the problem is very critical to obtaining an optimal solution to the problem otherwise the Genetic Programming algorithm will be abruptly terminated and so trapped in a local optimal [10] due to excessive growth of program codes otherwise referred to as code bloat.

Hence, preventing premature termination or premature convergence of the Genetic Programming run due to the issue of bloat is of a primary concern to GP researchers [13]. Genetic Programming therefore is a probabilistic, non-deterministic, optimization and a heuristic search technique [27], [30], [20]. It is a probabilistic and non-deterministic because it rarely gets a solution in precisely the form you contemplated and exactly the same result is obtained twice as anything can happen and nothing is guaranteed [15]. Hence, the technique is heuristic as it uses hunch or the rule of the thumb since there are obvious, straightforward and easy path at arriving at a solution, it is an optimization technique because it involves successive improvement of an initial solution until a best-so-far solution is obtained [4].

Genetic Programming researches have been so concentrated on the application of the algorithm to solving plethora real life problems that are usually with no well defined efficient solutions [9] but could be adequately measured and compared [34]. Such applications of the algorithm have generated impressive results that are humanly competitive.

In all of the researches in GP, application of GP to problem solving has received a lion share. Research efforts have also been on improving the understanding of fundamental aspects of the GP algorithm; and increasing the power and performance of Genetic Programming approach [13]. As at the moment, there are significant advances in the theory of Genetic Programming [28], but improving the power and performance of GP is less in the GP community [13]. Thus, research on bloat and bloat control therefore is an attempt on improving the performance of the GP algorithms.

In GP Literature, several Genetic Programming variants exist. These include: Linear Genetic Programming, Parallel Genetic Programming, Cartesian Genetic Programming, Traceless Genetic Programming, and Tree-based Genetic Programming. Tree-based Genetic Programming is the standard Genetic Programming technique introduced by John R. Koza and which is commonly used in the GP literature [15]. In Tree-based GP, the nodes of the tree are composed of functions and terminals that were identified from the problem to be solved. Thus, this paper addresses the issue of Code bloat in a tree-based Genetic Programming by specifically proposing a new method of bloat control.

II. BACKGROUND ON GENETIC PROGRAMMING ALGORITHM

The Genetic Programming algorithm is a class of the Evolutionary algorithms which includes the Evolutionary Strategies (ES), Evolutionary Programming (EP), Genetic Algorithm (GA) and lastly Genetic Programming (GP). These algorithms use the fundamentals principles of evolution, but however differ in their problem representation, and evolutionary operations. Basically, GP represents problems in variable length programs in any functional programming language and uses crossover and mutation operations. To solve a problem, the following Genetic algorithm must execute the following three steps [15]:

(1.) Generate an initial population of random compositions of functions and terminals of the problem (computer programs).

(2.) Iteratively perform the following sub-steps until, the termination criteria has been satisfied.

(a) Execute each Program in the population and assign it a fitness value according to how well it solves the program.

(b) Create a new population of computer programs by applying the following two primary operators. The operations are applied to the computer program in the population chosen with probability based on fitness.

(i) Copy existing computer programs into the new population.

(ii) Create new computer programs by genetically recombining randomly chosen parts of two existing programs.

(3.) The best computer program that appeared in any generation (that is, the best-so-far) is designated as the result of genetic programming. This result may be a solution (or an approximation) to the problem.

The application of the Genetic Programming algorithm to problem solving involves:

- Problem identification and definition
- Determination of high-level requirements of the problem
- Execute the GP algorithm on a standard benchmark problem
- Interpretation of results

Each of these stages is tasking and a time consuming process. A clear identification and definition of the problem to be solved constitutes the first and critical consideration in the application of GP algorithm to problem solving. A well identified and defined problem helps to further determines a suitable programming language to be used, followed by a second step of high-level requirement definition of the problem which is further subdivided into five steps criteria. [11], [17], [16] identify five steps criteria setting and regarded them as preparatory steps for running the genetic programming algorithm. These five criteria settings are:

- identification of the set of terminals,
- identification of the set of primitive functions,
- the fitness measure,
- the setting of parameters for controlling the run, and

- the method for designating a result and criteria for termination the run.

The first and the second steps is the determination of the terminal and function set of the problem. The terminal set is a collection of constants while the function set are operations to be performed on the constants. They form the external and internal nodes of a tree in a tree-based GP respectively. The terminal and function set define the search space in a GP problem [8] and so they must be properly defined to satisfy the properties of closure and sufficiency [14], [36]. The third step is the determination of the fitness measure. Fitness measure is the numerical value assigned to each program [8]. It is the measure of how good a program is able to solve a given problem [36]. It could be in terms of error [11], [3], time to balance or time to failure [14]. Fourthly is the determination of the control parameters. Two major parameters to be determined are population size and the number of Generations to be run. Other minor parameters are crossover rate, fitness reproduction rate, probability distribution over the potential cross over points, maximum tree size and maximum tree depth [14]. The last preparatory step is to determine at what point to terminate the process of evolution. It could be by maximum number of generation [11], loss of diversity [10], maximum allowable CPU time or total number of fitness evaluations reaching certain predefined limit [13]. The values obtained from the preparatory stage form the input into execution stage of the genetic programming process. [42], identified five major operations in the Genetic Programming execution stage. These are:

- Creation of initial population,
- fitness function evaluation,
- selection,
- genetic operations (reproduction, crossover mutation operations), and
- termination criteria and solution designation.

Figure 1 shows a pseudocode of the genetic programming execution stages. These operations form the essential components of the Genetic Programming algorithm.

```
START  
Create Gen: = 0  
Create Initial Random Population  
10 IF (Solution generated equal or approximate equal to  
Termination Criteria)  
    Print solution obtained as result  
STOP  
ELSE  
    Evaluate Fitness of each individual in population  
    i = 0  
20 IF i = M? THEN  
    Gen: = Gen + 1  
    Repeat Step label 10  
ENDIF  
ENDIF  
CASE Selection based on fitness = 1  
    Perform reproduction operation or  
    Perform mutation operation  
    Copy into the new population
```

```
ENDCASE
CASE Selection based on fitness = 2
  Perform Crossover
  Insert two offspring into new population
ENDCASE
I := I + 1
Repeat step label 20
```

Fig.1: A pseudocode of the genetic programming execution stages
Source: Modified from Koza's original GP algorithm [14].

III. RELATED LITERATURE ON CODE BLOAT

A. Theories of Code Growth.

Several reasons have been advanced for code growth or bloat. [3], classified the reasons for code bloat into two categories: Those that modify the program structure or genetic operator; and those that incorporate program size in fitness computation. [21], opined that code bloat is basically caused by introns (which are codes that does not contribute to fitness of programs but yet consume computational resources) and non-introns. In this paper, we did not precisely follow the code bloat classification (or taxonomy) as earlier stated above rather we presented a mixture of code bloat controls of these two taxonomies.

The earliest theory of code bloat is the Hitchhiking theory. [21], opined that introns is also known as the hitchhickers could attached to viable codes; and during a crossover operation that preserves viable codes, some of the introns are taken along with it and in so doing, introns are propagated throughout the population. Defense against crossover is the second theory in intron based code bloat theory. This theory sees code bloat as defense against crossover which has destructive effects on program fitness [33]. [21], observed that large amount of introns also known as inviable codes increases the number of crossover points. (Streeter, 2003) further observed that such increases in number of crossover points is more likely to result in a neutral crossover which has no destructive effect on program fitness. [34], [33] concluded that accumulation of more and more of these inviable codes are a possible cause of code growth. The third theory is the Removal Bias Theory. In order to guarantee the preservation of the individuals, the removal bias theory suggest that the sub-tree removed during crossover must not be larger than the inviable sub-tree area thereby placing a constraint or penalty for removing sub-tree that is larger than necessary; but there is no such constraint or penalty for inserting a large sub-tree [21], [34]. In this regard, the larger sub-tree that is inserted without constraint may lead to code bloat as more and more are inserted.

A number of scholars have worked on non-intron code theories. The Diffusion theory proposed by [18] assumes that code growth is direct result of evolution of program size and shape. According to this theory, there exists small-size fitness programs initially in the solution space; and as evolution progresses, more large-size programs will naturally evolve. As the Genetic Programming system progresses toward

convergence, more and more large-size programs are produce, thereby creating a situation for code bloat. [35], investigated the effect of fitness on bloat in Fitness causes bloat theory. They observed that there is a greater probability to find a bigger program with the same behaviour (that is, semantically equivalent) than to find shorter ones since the bigger programs evolve out smaller ones progressively. Earlier [19] concluded that in an attempt to improve program fitness, programs naturally tends to grow due to fitness pressure. [25] used the Pseudo-Hillclimbing theory, where children are rejected from joining the next generation if the fitness are not superior to that of their parents, and then a copy of its parent joins the next generation instead; this replicate large number of parents into the next generation thereby causing bloat as larger and larger parents are replicated.

B. Code Bloat Control methods

In Genetic Programming (GP), programs tend to bloat significantly in size when no control measure is applied [24]; consequently, a GP run is trapped in a local optimal. The control of code growth is important for reasons of breeding smaller solution that is easy to understand and which efficiently use memory resource [32]. Generally, code could slow the evolutionary process, that is, the rate at which new individuals (programs) can be evaluated [7], consumes memory and can hamper effective breeding [23], [25]; and excessively use CPU time [3]. Bloating is currently a hotly debated topic in the Genetic Programming; however, presently, there is no silver bullet (or final) solution to deal with the problem of bloat in GP [20]. In the absence of a widely-accepted generalized theory of bloat therefore, most effective methods for dealing with the problem of bloat can only be justified empirically [29]. Bloat is a common phenomenon in the evolutionary computation landscape, but most given attention in genetic programming as there is still absence of a widely-accepted theory of bloat in GP [29]. Several bloat control measures in GP literature attest to this. Some of the methods for controlling code growth include the use of Maximum tree depth restriction [3], dynamic maximum tree depth [31] parsimony pressure, [39] Pareto-based Multi-objective parsimony pressure, size fair crossover and mutations operators [37], Tarpeian method [38], Minimum Description Length method [40]. Double and Proportional Tournament, Baised-Multi-objective parsimony pressure (BMOPP), Waiting room, and death by size. Each of these is described as thus:

(1) Maximum tree depth restriction (that is, hard limit on program size) is the most common approach to bloat control in the literature [15]. It restricts breeding to only produce children less than some maximal tree depth [23]. Several approaches are used to restrict tree depth or program size.

- One approach is to check if the size of the resultant program is beyond the already set program size limit of 17. If the offspring depth size exceeds 17, it is rejected while the parents are returned back into their original state [15].
- Another approach is to try crossover operation several times until a valid-depth size is obtained [26].
- Lastly, yet another approach expand the tree depth dynamically until fitted individuals are discovered [41].

Setting a reasonable tree depth or limit is a difficult exercise. If the limit is too low, GP might not be able to find a solution and if is too high, evolution may slow down because of the immense resource usage and chances of finding small solutions will be very low [3]. However, the weakness of this method is that parents are returned back to pool if they violate the size limit constraints; otherwise it has proven to be a very successful method of code bloat control [25].

(2) [31], introduced the Dynamic Maximum Tree Size Depth as the modification of the [22] traditional setting of program hard limit. In the dynamic maximum tree size depth, a dynamic limit is set to at least, as high as to maximum depth of the initial random tree. Any new individual obtained from a crossover operation that breaks this limit is rejected and replaced by one of the parents instead, unless it is the best individual so far; the dynamic limit is increased to match the new best. In this case, the dynamic limit is never decreased and never surpasses the static limit.

(3) [37], introduced the size fair crossover as a means of size control in GP. In this method, the crossover point in the first parent program is selected randomly, and then the program size to be removed is calculated. The calculated program size is therefore used to determine the crossover point in the second parent so that the program size is not unnecessarily big. This is because keeping the size of tree small is generally an implicit objective in Genetic Programming [7].

(4) The Parsimony Pressure method which changes the selection probability by subtracting a value based on the size of each program from its fitness such that bigger programs have bigger subtractend. The bigger subtractend on larger programs lower their fitness and tend to have fewer children. That is, the parsimony pressure method uses the minimization of program size as a penalty by decreasing the fitness of programs by amount proportionate to their size. Parsimony pressure takes two forms: Parametric parsimony pressure and lexicographic parsimony pressure:

- The parametric parsimony pressure uses size metric and raw fitness to compute its final fitness of an individual. In this regard, it treats the individual's size as a linear factor in fitness. The problem with parametric parsimony pressure is that it is parametric, rather than being rank based; and it gives size an unwanted advantage over fitness; and
- The lexicographic parsimony pressure which optimizes fitness as a primary objective function and tree size as

secondary objective function in a lexicographic ordering [24].

(5) Another method is the Tarpeian method which acts on the selection probabilities of chosen programs that are longer-than-average by setting their fitness to zero. In other words, individuals with above-average-size are assigned very bad fitness with probability of W . This prevents such parent programs from being parents and so cannot cause bloat. The tarpeian method is overtly aggressive and usually sensitive to parameters; it rejects individuals by size before considering their fitness [25].

(6) The non-parametric methods of using two variations of tournament selection: double tournament and proportional tournament selection was also proposed by [25]. The double tournament uses two layers of tournaments in series on fitness and then size; while the proportional tournament determine randomly which out of fitness and size to use for the tournament.

(7) The Biased Multi-objective Parsimony Pressure (BMOPP) as a variation of the Pareto-optimization that combines lexicographic ordering, Pareto's dominance, and a proportional tournament features. This method is quite effective, easy to implement and combined with common evolutionary computation algorithms [29].

(8) The Waiting Room and Death by size method creates a pre-birth phase to all newly created individuals during the evolution. In this way, offspring programs are made to wait in the waiting room for certain period of time; and the larger offsprings are made to wait longer before they are permitted to enter the population to compete. Each individual program in the waiting room is assigned a value equal to the individual size. Next, the N children with the smallest queue value are removed from the waiting room and form the next generation. The Waiting Room concept came originally from the notion that larger children might take longer to be evaluated, and thus keeping them in the waiting room will let the fast, small children breed more rapidly through the population. The problem with the Waiting Room method is that it imposes significant computational complexity. The Death by Size method, fitness is used as objective such that some individuals with high fitness are selected to breed children while others with low fitness are selected to die and be replaced by the new children. Even though those low fitness program are disallowed to compete with other more fit programs, they are not actually deleted from memory and still lay claim to memory space as there are indication in Death by size algorithm that programs with low fitness are truly deleted. The death by size method is similar to other methods proposed in the past. [2], used the 'shrink' operator which removes a branch of a tree and replaces it with its terminal. [5] used 'trunc' (truncate) to shrink the size of the tree. [12], used the mutation operator called 'hoist' to select a node inside the tree and returns a copy of this subtree as a new individual. Recently, [1] introduced 'cut' in their new method of bloat control known as prune and plant method. In this method, a selected individual will have its branches 'pruned'

and substituted by terminal; and the pruned branch is ‘planted’ in the population as a new tree.

However, in all of these techniques for bloat control, the causes of the bloat are not completely eliminated. The only thing they succeed in doing is that they only delayed the occurrence of bloat as size of the trees will also increase and so make bloat inevitable. In this regard, we proposed the **Delete lower and Keep higher Fitness value Programs after Crossover (DKPC)** algorithm as this code bloat control measure uses element of hill climbing. It is therefore a greedy algorithm that works by deleting the parent’s programs lower fitness and retains the offspring program with higher fitness in the generation after crossover operation; otherwise deletes the offspring program and retain the parents program with higher fitness in the next generation. However, where both parents and offspring are equal in fitness, Keep the parent and delete the offspring. In this regard, the DKPC algorithm exhibits [6] natural selection by life and death in which any individual with favourable variation is preserved, and those with any unfavourable deviation of structure are destroyed. It is expected that in this DKPC-algorithm, every parent or offspring programs that do not contribute to fitness will not have the chance of unnecessarily bloating the overall program code as they are deleted from the very beginning. In this case, the computational resources are effectively used. Figure 2 shows the DKPC Algorithm.

1. Create initial population of trees
2. Set termination criteria
3. Start
 - 3.1 Select crossover points on the trees
 - 3.2 Perform crossover OR mutation.
 - 3.3 Compare parent fitness with offspring fitness.
 - 3.4 IF Parent fitness value is equal to OR greater then offspring fitness value,
 - Keep** the parent program
 - Delete** offspring program
 - ELSE
 - Keep** the offspring program
 - Delete** parent program
- 3.5 LOOP until termination criteria is met
4. Stop

Fig. 2: The DKPC Algorithm

IV. METHODOLOGY

In carrying out this research, we used software engineering and tree methodologies. For software products that have their feature sets redefined during development, the evolutionary methodology is most appropriate. Evolutionary methodology is iterative and is such that it evolves a final solution from an initial specification with well-defined and well-understood requirements by adding new features as the evolution progresses until a termination condition is met. A tree methodology composed of nodes. In this paper, we used the concept of tree where functions and terminals so identified were used as nodes to solve a given problem.

V. EXPERIMENTAL DESIGN

To test this DKPC algorithm, we used the Boolean 6-Multiplexer and Boolean 11-Multiplexer functions as benchmark problems. Generally, the Boolean N-multiplexer function consist of K address bits a_i and 2^k data bits d_i , where $N = K + 2^k$. In other words, the input to the Boolean multiplexer functions consists of the K+2k bits ($a_{k-1}, \dots, a_1, a_0, d_{2k-1}, \dots, d_1, d_0$). This means that, a Boolean 11-multiplexer is made of 3 address lines and 2^3 (8) data lines giving a total of 11 lines of the multiplexer; a Boolean 6-multiplexer is made of 2 address lines and 2^2 (4) data lines giving a total of 6 lines of the multiplexer while a Boolean 3-multiplexer is made of a 1 address line and 2^1 (2) data lines giving a total of 3 lines of the multiplexer. For these two benchmark problems, the same preparatory parameters were used in other to form the bases for comparison. We used best program fitness, size, and CPU time to compare the performances of the benchmark problems. Table 1 shows preparatory requirements for the benchmarks.

Table 1: Preparatory requirements for the benchmarks

Population size	4000
Generations	200
Maximum depth	$D_{max} = 17$
Maximum Initial depth	$D_{initial} = 5$
Probability of crossover	$P_c = 0.9$
Probability of mutation	$P_m = 0.1$
Selection method	Tournament
Tournament size	$T = 7$
Function set	{AND, OR, IF, NOT}
Terminal set	d_0, d_1, \dots, d_{k-1}

A. The Boolean Function Tree Structure

We have asserted earlier that in a Tree-based GP, the nodes of the tree are composed of functions and terminals. The functions (OR, AND, NOT, IF) are used to form the internal nodes while the terminals ($d_0, d_1, d_2, \dots, d_{11}$) are used to form the external nodes or terminal nodes. Having identified these functions and terminals as shown in Table 1, a typical Boolean multiplexer function tree is shown in Figure 3.

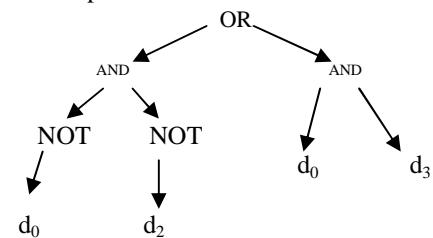


Fig 3: A typical Boolean Multiplexer function Tree

VI. RESULTS

Table 3 and Table 4 are results of Boolean 6-multiplexer and Boolean 11-multiplexer functions respectively. Figure 2 shows the line graphs interpretation of Boolean 6-multiplexer; while Figure 3 shows the line graphs interpretation of Boolean 11-multiplexer.

Table 3: Boolean 6-multiplexer function

Milestones during run	CPU Time (s)	Best program size	Best Program fitness
1.	18	52	45
2.	36	45	48
3.	54	60	28
4.	74	90	49
5.	90	130	50
6.	108	115	52
7.	126	95	61
8.	144	93	75
9.	162	93	76
10.	180	75	80

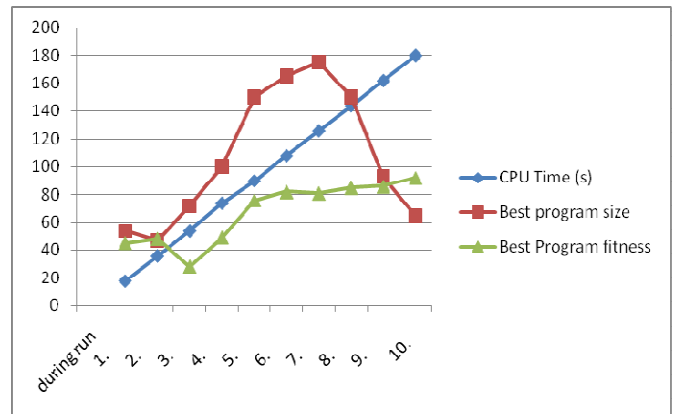


Fig. 3: line graphs interpretation of Boolean 11-multiplexer

VI. DISCUSSION AND FINDINGS

When the Boolean 6-multiplexer function and Boolean 11-Multiplexer function were executed against the preparatory requirements using our proposed algorithm, very impressive results were obtained. In the Boolean 6-multiplexer function, at various interval of run, the program size increased steadily as evolution progresses indicating a sign of code bloat, but after certain level of run, the program size dropped considerably as program fitness increases. The Boolean 11-multiplexer function also displayed almost the same behaviour; except that it exhibited better performance than the Boolean 6-multiplexer result in terms of better program size reduction. To this extent, it performed better in bloat control. This is in line with the ‘No-free lunch’ theorem that states that there is no single algorithm that provides omnibus solution to all problems.

VII. CONCLUSIONS

This paper Proposed a new method of code bloat control algorithm known as the DK-PC algorithm that Delete lower and Keep higher Fitness value Programs after Crossover. This algorithm was tested against the Boolean 6-multiplexer and Boolean 11-multiplexer functions, and we observed that the algorithm was able to control bloat to a large degree. However, the algorithm performed better in the Boolean 11-multiplexer function than in the Boolean 6-multiplexer function. It is therefore clear to us that bloat is inevitable, but what extent can it be control? Our algorithm has considerably reduced code bloat to the barest minimum. Our advice for further research is that, other benchmark problems like artificial ant problem, symbolic regression problems and so on can be used to further test the efficacy of this algorithm.

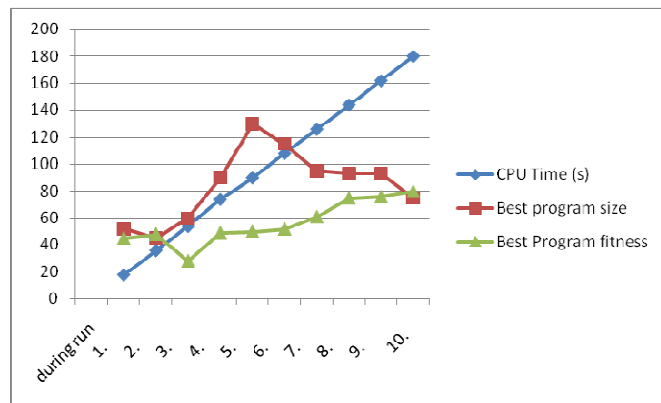


Fig. 2: Line graphs interpretation of Boolean 6-multiplexer

Table 3: Boolean 11-multiplexer

Milestones during run	CPU Time (s)	Best program size	Best Program fitness
1.	18	54	45
2.	36	47	48
3.	54	72	28
4.	74	100	49
5.	90	150	76
6.	108	165	82
7.	126	175	81
8.	144	150	85
9.	162	93	86
10.	180	65	92

REFERENCES

- [1] E. Alfaro-Cid, A. Esparcia-Alcazar, K. Sharman, F. Fernandez de Veg, and J. J. Merlo (2008), "Prune and plant: A new control method for genetic programming," Eighth International conference on Hybrid Intelligent Systems. IEEE doi 10.1109/HIS.2008.127
- [2] P. J. Angeline (1997), "Comparing subtree crossover with macromutation," In Proceedings of the Sixth International Conference on Evolutionary Programming, 1213, 101-111, Indianapolis, USA.
- [3] S. Bleuler, M. Brack, L. Thiele, and E. Zitzler (2001), "Multiobjective genetic programming: Reducing bloat using SPEA2," In *Proceedings of the 2001 Congress on Evolutionary Computation CEC2001*, pp 536-543, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea: IEEE Press.
- [4] D. S. Burke, K. A. De Jong, J. J. Grefenstette, R. Crawford-Marks, and L. Spector (2002), "Size control via size fair genetic operators in the pushGP genetic programming system," *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2002*, San-Francisco, CA: Morgan Kaufmann.
- [5] K. Chellapilla (1997), "Evolving computer programs without using subtree crossover," *IEEE Transactions on Evolutionary Computation*, 1(3):209-216.
- [6] C. Darwin (1996), "The origin of species. Oxford, UK: Oxford University Press. (Original work published in 1859).
- [7] E. D. de Jong, R. Watson, and J. Pollack (2001), "Reducing bloat and promoting diversity using multiobjective methods," In Proceedings of the Genetic and Evolutionary Computation Conference, Morgan Kaufmann, San Francisco, CA.
- [8] K. Dolan (2009), "Beginners' guild to genetic programming," Retrieved from: http://www.geneticprogramming.us/Further_Reading.html
- [9] D. Gupta and G. Sabina (2012), "An overview of methods maintaining diversity in genetic algorithms," *International Journal of Emerging Technology and Advanced Engineering*. 2(5). Retrieved from <http://www.ijetae.com>
- [10] S. M. Gustafson (2004), "Analysis of diversity in genetic programming," (Doctoral dissertation, University of Nottingham). Retrieved from
- [11] M. A. Keane, J. R. Koza, and J. P. Rice (1993), "Finding an impulse response function using genetic programming," In Proceedings of the 1993 American Control Conference. Evanston, IL: American Automatic Control Council. 3, 2345-2350.
- [12] J. K. E. Kinnear (1993), "Evolving a sort: lessons in genetic programming," In Proceedings of the 1993 International Conference on Neural Networks. 2, 881-888, San Francisco, USA, IEEE Press.
- [13] P. Kouchakpour (2008), "Population variation in canonical tree-based genetic programming," Doctoral thesis, University of Western Australia. Retrieved from: <http://robotics.ee.uwa.edu.au/theses/2008-Genetic-Kouchakpour-PhD.pdf>
- [14] J. R. Koza (1990), "Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems," Technical Report No. STAN-CS-90-314; Computer Sciences Department; Stanford University.
- [15] J. R. Koza (1992a), "Genetic programming: On the programming of computers by means of natural selection," Cambridge, MA: MIT Press.
- [16] J. R. Koza (1992d), "A genetic approach to finding a controller to back up a tractor-trailer," *Automatic Control Council*. 3, 2307-2311.
- [17] J. R. Koza (1992e), "A genetic approach to the truck backer upper problem and the inter-twined spirals problem," In *Proceedings of IJCNN International Joint Conference on Neural Networks*. Piscataway, NJ: IEEE Press. 4, 310-318.
- [18] W. B. Langdon, T. Soule, T. R. Poli, and J. A. Foster (1999), "The evolution of size and shape," *Advances in Genetic Programming* 3, chapter 8, pp. 163-190, MIT., Press, Cambridge, MA, USA.
- [19] W. B. Landon (1998), "The evolution of size in variable length representations," In *IEEE International Congress on Evolutionary Computations (ICEC 1998)*, pp. 633-638, IEEE press.
- [20] S. Luke (2000), "Issues in scaling genetic programming: Breeding strategies, tree generation, and code bloat," PhD Dissertation, University of Maryland, College Park, USA.
- [21] S. Luke (2000a), "Code growth is not caused by Introns," Late Breaking Papers at the 2000 Genetic and Evolutionary Computation Conference pp. 228-235, Las Vegas.
- [22] J. R. Koza (1992a), "Genetic programming: On the programming of computers by means of natural selection," Cambridge, MA: MIT Press.
- [23] S. Luke and L. Panait (2002), "Fighting bloat with nonparametric parsimony pressure," In *Parallel Problem Solving from Nature - PPSN VII (LNCS 2439)*. Juan Julian Merelo Guervos et al, eds. Springer Verlag, 411-421.
- [24] S. Luke and L. Panait (2002b), "Lexicographic parsimony pressure," *GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 829-836. Morgan Kaufmann.
- [25] S. Luke and L. Panait (2006), "A comparison of bloat control methods for genetic programming," In *Journal of Evolutionary Computation*. 14(3), 309-344.
- [26] P. Martin and R. Poli (2002), "Crossover Operators for a hardware implementation of GP using FPGAS and Handel-C," In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), New York, USA, July 2002, Morgan Kaufmann.
- [27] T. McConaghy, E. Iadislavleva, and R. Riolo (2010), "Genetic programming theory and practice 2010: An introduction," A paper presented In the Center of Studies of Complex Systems at the University of Michigan.
- [28] M. O'Neil and A. Brabazon (2009), "Recent patents on genetic programming," Retrieved from <http://www.benthamscience.com/cseng/samples/cseng2-1/0005CSENG.pdf>.
- [29] L. Panait and S. Luke (2004), "Alternative bloat control methods," In *Proceedings of the Genetic and Evolutionary Computation Conference- (GECCO)*.
- [30] R. Poli and W. B. Langdon (1998), "On the search of different crossover operators in genetic programming," In Proceedings of Genetic Programming; Madison, Wisconsin
- [31] S. Silva and E. Costra (2004), "Dynamic limits for bloat control: variations on size and depth," In Genetic and Evolutionary Computation Conference (GECCO), 666-677.
- [32] P. W. H. Smith (2000), "Controlling code growth in genetic programming," *Computing Techniques and applications*, 166-17. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.2268&rep=rep1&type=pdf>
- [33] Streeter, M.J. (2003) *The Root Causes of Code Growth in Genetic Programming*.
- [34] T. Soule (1998), "Code growth in genetic programming," PhD Dissertation; Computer Science Department, College of Graduate Studies, University of Idaho, USA.
- [35] O. Teytaud, S. Gelly, N. Bredeche, and M. Schoenauer (2005), "A statistical learning theory approach of bloat," http://hal.inria.fr/docs/00/04/19/72/PDF/antibloatGecco2005_long_version.pdf
- [36] W. Yan (2003), "Profitable, return enhancing' portfolio adjustments - An application of genetic programming with constrained syntactic structures," M.Sc. degree project of the University College, London. Retrieved from: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.416>
- [37] R. Crawford-Marks and L. Spector (2002), "Size control via size fair genetic operators in the pushGP genetic programming System," In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2002*, San-Francisco, CA: Morgan Kaufmann.

- [38] R. Poli (2002), "Covariant tarpeian method for bloat control in genetic programming," In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2002*, San Francisco, CA: Morgan Kaufmann.
- [39] R. Poli and N. McPhee (2008), "Parsimony pressure made easy, proceedings of the genetic and evolutionary computation conference (GECCO)," Atlanta, 1267-1274
- [40] G. Pandey and A. Dukkipati (2012), "Minimum description length principle for maximum entropy model selection," Retrieved from <http://arxiv.org/abs/1204.6423>
- [41] S. Silva and J. Almeida (2003), "Dynamic maximum tree depth - A simple technique for avoiding bloat in tree-based GP," In a Conference of Genetic and Evolutionary Computation (GECCO-2003); Retrieved from <https://www.ciscuc.uc.pt/publication/show/109>.
- [42] R. Stewart (2004), "Genetic programming: computers that program themselves," Retrieved from <http://www.robinstewart.com/personal/learn/gp/GeneticProgramming.pdf>

AUTHORS PROFILE



Prince Oghenekaro Asagba had his B.Sc. degree in Computer Science at the University of Nigeria, Nsukka, in 1991, M.Sc. degree in Computer Science at the University of Benin in April, 1998, and a Ph.D degree in Computer Science at the University of Port Harcourt in March, 2009. He is a Senior Lecturer and a visiting lecturer to several Universities in Nigeria since 2010.

His research interest include: Network Security, Information Security, Network Analysis, Modeling, Database Management Systems, Object-oriented Design, and Programming. He is a member of Nigeria Computer Society (NCS) and Computer Professional Registration Council of Nigeria (CPN).



Duke Urhe-otokoh Oghorodi received his Diploma in Data Processing (1988), B.Sc in Computer Science (1991), Master of Business Administration (1998), Postgraduate Diploma in Education (2005), M.Sc Computer Science (2009) from the University of Benin. He is currently pursuing a Ph.D programme in Computer Science at the University of Port Harcourt. He is a lecturer in the Department of Computer Science, College of Education, Warri, Delta State, Nigeria. His research interests include: Genetic Programming, Fuzzy Expert systems, programming. He is a member of Nigerian Computer Society (NCS) and Computer Professional Registration Council of Nigeria (CPN).

Attack Resilient & Adaptive Medium Access Control Protocol

Piyush Kumar Shukla
CSE, UIT, RGPV
Bhopal, India

Sarita Singh Bhadauria
ECE, MITS,
Gwalior, India

Sanjay Silakari
CSE, UIT, RGPV
Bhopal, India

Abstract— In IEEE 802.11, all nodes contending for the access to a medium needs to perform activities as per the specification of medium access control sub layer. It has been observed that when the number of node increases, it leads to the probability of collisions, which finally causes longer back-off values of the concerned collided nodes. The recent development within the field of computer networking enables everyone to access the Internet in the fastest manner using the Tablet, Mobile, Laptop or traditional Desktop. One common way to achieve Internet connectivity includes the use of Wi-Fi, which also forms the subject matter of present research. In the entire situations' basic requirement at the user's end- is a fair sharing of the available channel bandwidth, adequate quality of service (QoS), for which customers are paying lots of money.

Unfairness in the network performance indicates the presence of attackers or some kind of misbehavior by existing users. It has been observed that sometimes to get more share of available bandwidth, several legitimate users show greediness or selfishness, which results in injustice to the other users in the same Wireless Local Area Network. However, it is too difficult to understand about the type and behavior of misbehaving nodes in the common shared environment.

Another issue that requires attention is related to QoS that is if a user is availing better service than others within the network, then it will be appreciable. Otherwise, this is the matter of MAC misbehavior and needs to resolve.

This research is motivated by selfish node, which manipulates their working (differ from normal MAC protocols) in different ways to increase their share to occupying the access to the channel. This exploitation of the MAC layer protocol may be hidden from the upper layers and in this work a solution has been proposed to embark upon the problem at the MAC Layer itself.

A faired, Attacks Resilient and opportunistic Adaptive Medium Access Control protocol has been further modified, and its performance has been compared with existing CSMA/CA base on the Key performance Indicators, i.e. Throughput, Medium Access Delay, Collisions per frame and Fairness Index.

Keywords- Opportunist Mode, Attacks Resiliency, Adaptability, MAC Layer Misbehavior, Selfish Node, IEEE 802.11, DCF, back-off, Attacking Mode, Suspicious Mode.

I. INTRODUCTION

The IEEE 802.11 access scheme incorporates two access methods one is PCF to support real-time services by using the Point Coordination Function (PCF); the method operates primarily based on centralized control/scheduling and polling techniques. A more reasonable model of operation is known to be that of random access, Distributed Coordination Function (DCF) [6] for asynchronous, contention-based, distributed access to the channel. This research work is focused on DCF.

In IEEE 802.11 networks, it is extensively accepted that the back-off algorithm plays a crucial role in achieving a high aggregated throughput and a fair allocation of the channel to the stations. Thus, the back-off value should reflect the actual depth/level of contention for the channel.

MAC misbehavior is the crucial problem in the WLAN because it affects the normal working of the fair MAC protocol [7], which is not easy to detect. A novel approach has been described throughout this work to improve performance of MAC protocols by offering opportunist mode and also detecting the nodes those are misusing it.

The remaining paper is organized as follows: Section II introduces the System Model; Section III discusses regarding the MAC layer attack survey, Section IV presents a classification of nodes and Section V presents the different type of MAC layer misbehavior. Section VI presents operations of ARA-MAC. Section VII presents the Results based on N.S-2 .34 and also compares the performance of ARA-MAC with CSMA/CA base on Key Performance Indicator. Concluding remarks are given in Section VIII.

II. SYSTEM MODEL

In the next sections, we used the following system model and assumptions for ARA-MAC.

The IEEE 802.11 WLAN (AP and stations) works in the infrastructure mode using DCF (Distributed Coordination Function), which is the essential operation mode usually deployed.

The channel is sensed idle for DIFS (DCF Inter Frame Spacing) time when DCF delays frame transmissions right. It waits for an additional random time, back-off time, after which the frame is transmitted. The back-off time [9] is bounded by the contention window size CW . This applies to data frames in the basic scheme, and RTS frames in the RTS/CTS scheme.

The back-off time at each station is decreased as long as the channel is found unoccupied.

When the channel is busy, the back-off time is frozen. When the back-off time reaches zero, the station transmits its frame. If the frame collides with another frame (or RTS), the sender's time out which are waiting for the ACK (or the CTS) and computes a new random back-off time with a larger CW to retransmit the frame with lower collision probability. When a frame is successfully transmitted, the CW is reset to CW_{min}. The Network Allocation Vector (NAV) [8] of all other stations is set at the frame-duration field value in RTS/CTS and DATA headers. If Channel utilization is observed poor, then access point observed opportunist mode to every single station based on RTS counts.

- In all the simulations and calculations we have considered a single trusted Access Point.

- It is our basic assumption that if only user stations misbehave; then they do so in a rational way, meaning that misbehavior is motivated by a beneficial outcome in terms of obtaining a larger throughput, lower medium access delay. We also considered malicious misbehavior that aims at disrupting the functionality of the CSMA/CA and network.

- The detection system is implemented only at the AP. Thus, neither modification nor reconfigurations of wireless adapters have to be made in the user side. In addition, the solution is under the full control of the AP. All the users must have understanding to decode the information in beacon frames to enjoying the Opportunist mode, else they can still survive within the system.

III. LITERATURE REVIEW OF ATTACKS AT MAC LAYER

Vojislav B. Misić et al. [48] discusses security issues of networks. They have been compliant with the recent IEEE 802.15.4 standard for low rate WPANs and also a number of vulnerabilities at the MAC and PHY layer have been identified. Svetlana Radosavac et al. [42] considers node misbehavior in the MAC sub layer and its effects on the performance of the network layer.

Alvaro A. Cardenas et al. [19] revisits the problem of detecting greedy behavior in the IEEE 802.11 MAC protocol by evaluating the performance of two previously proposed schemes: DOMINO and sequential probability ratio test (SPRT). A survey on different attacks and their countermeasure in MANET has been provided by Bing Wu et al. [14]. S.Radosavac et al. [41] have considered the problem of detection and prevention of node misbehavior at the MAC layer, focusing on the back-off manipulation by selfish nodes. Jiahai Yang et al. [26] propose the architecture of a Coordinated Attack [26] Response & Detection System (CARDS).

Shafiullah Khan et al. [40] discusses various security risks, out of which denial of service (DoS) attack is the most severe security threat, as it can compromise the availability and

integrity of broadband wireless network. Lei Guan-G et al. [29] discussed the research area related to MAC layer misbehavior, which is based on the operating principles and the objective of misbehaving nodes. Pradeep Kyasan-ur et al. [34] presented modifications to the IEEE 802.11 Protocol to simplify the detection of such a selfish hosts. Svetlana- Radosavac et al. [43] presented a framework of study for the problem of MAC misbehavior detection. Their approach encompasses of an intelligent attacker who adapts its misbehavior strategy with the objective to remain undetected.

Szymon Szott et al. [46] deals with the problem of node misbehavior in ad-hoc networks in which a realistic approach is used to determine the impact of contention window manipulation and RTS/CTS cheating. Raja Gunasekaran et al. [37] discusses about selfish misbehavior due to waiting for smaller back-off intervals when compared to the other nodes in the same subnet. Rickshaw et al., S.A.A runmozhi et al. [36, 38] in their work, the DDoS attacks are introduced to the tactical mobile ad-hoc networks in both standby and moving cases.

Jahangir H. Sarker et al. [28] proposed utilization of the multiple power level system used for transmission to mitigate the attacking signals. Taimur Farooq et al. [47] discussed DoS attacks, which exploit the MAC layer vulnerabilities of IEEE 802.11 networks. Jaehyukchoi et al. [27] proposed a practical way to point out the misbehaving nodes without requiring access of hardware-level information in 802.11 WANS.

Vamshikrishna Reddy Giri et al. [50] investigates various types of MAC layer misbehaviors, and calculates their effectiveness in the form of their impact on important performance aspects, including throughput and fairness to other users. S. Szott et al. [39] suggests the use of the chi-square test approach for detecting back-off-related misbehavior in IEEE 802.11 based EDCA networks. Fei Shi et al. [21] proposes a method to avoid the occurrence of misbehavior. In their scheme, Local Most Trustworthy nodes (LMT node) are allowed to allocate the back-off value to the originator, rather than permitting the originator to choose himself the back-off values by itself.

Sangwon Hyun et al. [44] present a design of channel migration scheme to tone down wireless jamming attacks. By exploiting the multiple channels typically available on most of the wireless platforms, their scheme used a flexible and resilient approach to switch communication channels, which enables nodes to continue packet transmissions with their neighbors in the presence of jamming attacks. Iordanis Koutsopoulos [22] studied the competition and possible cooperation between two misbehaving adversaries who naturally hinder each other in wireless access. They considered two simple types of cooperation that require minimal coordination.

Ioannis Broustis et al. [23] identify an intellectual, low-power jamming attack that can take benefit of this behavioral trait, the placement of a low-power jammer in a way that it affects a single legitimate client which can cause a situation of starvation among all other clients. Jin Tang et al. [24] have developed a shuffle scheme to mitigate the short-period

fairness impact on the sample series, and investigate the proper shuffle period that can maintain the randomness in each node's back-off behavior while resolving the short-period fairness [31] issues. Matthias Wilhelm et al. [30] take the role of a wireless adversary and investigate one of its most powerful tools—radio frequency jamming. Mihui- Kim et al. [31] focuses the threats to fair scheduling [32] in WMNs resulting from node misbehavior and provide a generic verification framework to identify such a misbehavior.

S. Chen et al. [45] proposes a modest, unusual and difficult verifying scheme to deal with untruthful recommendation. They model the false recommendation problem as trials in reputation management court and present a simple, novel and effective scheme. E.G V. et al. [20] proposed the implementation of a technique based load balancing on cell breathing for mitigating the harmful effects of the jammer.

IV. NODE CLASSIFICATION

Nowadays, a lot of Internet applications are available, and we have improved our working by utilizing the Internet resources using enhanced and adaptive [2] the Internet protocols. Sometimes we have achieved high speed with less Security and vice-versa, but this is the time when we can emphasize designing a protocol which must have the capability of simultaneously resisting the effects of malicious nodes on performance of channel, as well as enhancing the dynamic behavior of MAC protocol to improve channel utilization [3] in different load scenario. It is very important to recognize and discard that attacking node after identification.

We have also provided the smaller contention window size of active nodes during under-load situation to improve channel utilization.

A. Normal Mode

To access a channel a node randomly selects a value between $[0, CW]$ (Initially $CW=CW_{min}$) under normal conditions, every node within the network operates according to basic working of a Binary Exponential Back-off Algorithm for CSMA/CA IEEE 802.11 WLAN, which doubles its contention window value up to a CW_{max} after each collision or unsuccessful transmission.

B. Opportunistic Mode

Access Point observes the common shared channel and will broadcast a message to all the nodes based on largest RTS counts to adapt the opportunist mode if the channel utilization [10] is found poor. AP senses the channel using the Fibonacci series to save the resources. This mode transforms nodes to a lower density zone; the consequences of those are fewer collisions and improved throughput on the network.

C. Suspicious Mode

A node found to be a cheater in terms of any MAC misbehavior and violating the rules of CSMA/CA categorized as suspicious node and puts into a suspicious mode. The Access point kept under the observation on these nodes by

running ARA-MAC attack detection technique. Nodes under suspicious modes follow basic BEB rules or opportunist mode rules depending upon its channel situation and its history table.

D. Attacking Mode

Access point declared a node as an attacker, after finding a node shifting from suspicious mode to attacking mode, after crossing the maximum allowable threshold. Activity of an attacker as a self-centre node is the basic motivating factor of this research. An attacker, in an attacking mode made collisions as well as it utilized unused slots when it works in egocentric or greedy mode.

This work initiates an in line research for Attack Resilience, Adaptive Medium Access Control Protocol [25] for WLAN 802.11.

V. TYPE OF MAC MISBEHAVIOR [1, 4]

- A node can adjust its back-off time to CW_{min} or less at all times.
- A node may scramble frames sent by other nodes in order to increase their CW [3].
- A node may delay CTS and ACK, or rejects RTS and DATA, so the sender doubles its CW and consequently, the malicious node [16] gets more chances to occupy the channel.
- A node may increase its data transmission time by enhancing the Network Allocation Vector (NAV) value to prevent [12] other nodes from contending during this period.
- A node may also transmit when it senses the channel idle before waiting DIFS time.
- A node only transmits RTS and never transmits data after getting CTS.
- A node pair consistently transmits RTS and CTS to each other because they are forever free to send, this behavior degrades fairness [44].
- Back-off manipulation (cheating with fixed Contention value) [5]
- CTS Scrambling
- DIFS value Reduction.
- Single Attacker and colluding attackers
- Adaptive Cheating
- Inter Layer attacks
- Choice of a short fixed value of a back-off counter. Here the node always waits or varies a short amount of time before accessing the idle wireless channel.

- The Attacker chooses a contention window, always resulting in a smaller back-off value.
- Choice of small DIFS, SIFS, etc.

VI. ARA-MAC UTILITY

Two new algorithms have been developed to enhance the performance of 802.11, which are basically suggesting some remarkable improvement in working of IEEE 802.11 WLAN based CSMA/CA, Medium Access Control Protocol. These algorithms are called AO-MAC (Adaptive & Opportunistic=Medium Access Control Protocol) and AR-MAC (Attack Resilient- Medium Access Control Protocol), which jointly called as ARA-MAC (Attack Resilient & Adaptive MAC Protocol) which improves the performance of network in two ways.

A. Scenario 1: Network under poor utilization

1)

Scenario 2: Network Under poor Utilization

Current scenario represents a network in which attacker nodes misbehave and degrade the performance of the channel. In such a situation, developed ARA-MAC protocol is useful and suggests a mechanism through which the attacker node can be identified and removed from the network.

B. Description of the ARA-MAC

DCF Based CSMA/CA work in a good manner, when the channel is properly loaded, but when the channel is underloaded, it has disadvantages that the nodes having a packet for transmission, but they have to wait even when the channel is idle. To remove this drawback, we have proposed adaptive behavior of the protocol as AO-MAC in 4.3 (a). In this, Access Point [16] periodically calculates the channel traffic, and if it finds that channel is under loaded then only it selects a node to go into an opportunist mode for a fixed duration of time. During the opportunist mode of operation, the node reduces its window size by a predefined value and starts its transmission. By doing so it reduces the waiting time given by BEB. In this manner, packets would be transmitted frequently, and hence channel utilization improves.

The selection of a node for opportunist mode is done based on largest RTS counts. At the end of the time period, the node comes back to normal mode of operation by resetting its window size as per the specification of 802.11 [14]

The Attack Identification algorithm shown in 4.3 (b) describes the attacker identification process. Access Point has monitored the transmission rate of every node in the network. For every received packet, it maintains a packet counter for every node. After every time period T, it calculates the transmission rate of the nodes using the value of respective packet counters. If the transmission rate for some node is found to be more than others, a counter named "Suspicious Counter (SC)" [25] is increased by 1 (it is initialized with 0). A non-zero value of SC signifies that the node is in Suspicious Mode.

After every interval T, the transmission rate is calculated and compared to the others, if the same happens to be case again the SC value is incremented. Otherwise, for every normal transmission over a period T value is decremented by 1. If the SC value reaches to the value of 3, the node is declared as an attacker and Access Point remove it from the list. As the working of 802.11 is followed, a node has to wait for a random period of time before getting the channel access for transmission.

Any node can be declared as a suspicious node if it receives an unfair number of packets within an observation period as compared to other nodes. Average Throughput [11] of all the nodes should be almost equal in all the nodes declared as an attacker node based on observations made by the Access Point.

1) AO-MAC: Algorithm for Opportunist Mode of Operation in WLAN 802.11

Input: Channel load

Output: Opportunist node list

Perform in every t second by the access point

1. Set Node=1
2. Calculate CHANNEL_LOAD
3. if CHANNEL_LOAD < CAPACITY, then
 - a. Set NODE_MODE = OPPORTUNIST for duration t
- Set CW = CWmin
- b. NODE = NODE + 1
4. EXIT
 - a. EXIT

2) AR-MAC: Attack Identification Algorithm

Input: Average number of packets transmitted.

Output: Attacker Identification and Penalty

1. Calculate Ave_P during OP
2. Set i=1
3. if (Ave_P + 10%) <= P_i then
 - a. Set i = i + 1
 - b. f I <= N, then go to step 3.
4. Else
 - a. If (i is in suspicious List), then.
 - b. Set Sci = Sci + 1
 - c. Else add node, i in suspicious List.
 - d. Set Sci = Sci + 1

3) Attacker Disassociation Algorithm

Input: Suspicious Counter Value

Output: Attacker Dissociation

1. Calculate Throughput
2. If $(T \leq (T_p - T_p * 0.4))$, then
3. Set $i = 1$
4. if (i in suspicious List) then
5. if $(SC \geq 3)$ then
6. Remove i from suspicious List
7. Add i in bannedNodeList
8. Set $i = i + 1$
9. if $(i \leq N)$ then
10. go to step 4
11. Exit

VII. RESULT & PERFORMANCE ANALYSIS

The performance of the novel protocol ARA-MAC has been analyzed using standard key performance Indicators Throughput, m access Delay, collisions, Fairness index.

A. SIMULATION PARAMETERS

Simulation has been performed: after simulating the scenario on the bed of Network Simulator- 2.34 using Linux Red Hat version -5 are already shown and publish in [25].

B. PERFORMANCE ANALYSIS

1) Average Throuhput

It is the measurement of the network performance [18]; throughput is total data packets/frames received per unit of time [25], which can be expressed through the ratio of the data frames successfully delivered at the destination for each flow to those generated by the source. It is generally represented in kbps.

The Table1 and Fig.1 show the throughput performance of the newly developed ARA-MAC compared with CSMA/CA in a WLAN (Wireless Local Area Network) 802.11 based network.

The figure shows that, the throughput of ARA-MAC is better than that of CSMA/CA in the absence of the attacker. This improvement is due to the opportunist mode offered by ARA-MAC to nodes, which reduces their waiting period, and hence they are ready to attempt sooner as compared to CSMA/CA. Table 1 and Fig.1 clearly this.

Furthermore, when the attacker [12] is introduced again ARA-MAC outperforms CSMA/CA. This is due to the attacker identification mechanism of ARA-MAC. The original CSMA/CA is unable to handle an attacker as a result of which channel utilization degrades. However, quick identification of the attacker and its removal from the network by Access Point gives an advantage in case of ARA-MAC.

In all these scenarios, it is observed that for a particular contention window size as the number of station increases, the throughput starts decreasing because more stations will go into waiting state and as a result of that total frame transmitted goes down, which impacts the throughput.

The next observation is based on contention window size [15]. The simulation runs for different window sizes, and throughput improves as window size increases. Because of finer window size, the selection of random waiting period spans across a finer range and hence the probability of getting same random number by more than one station gets reduced. This results in more successful transmissions, which contribute to this increase in throughput.

Even so, still ARA-MAC outperforms CSMA/CA in both scenarios. Fig.1and Table1 compares the simulated average throughput with the two schemes.

It shows that the novel ARA-MAC [25] scheme improves network throughput. Moreover, the discussed scheme increases collisions; its network throughput has significant improvement when the network load is low as well as heavy in both the cases (without and with Attacks), and the improvement becomes evident also with the original scheme, and proposed scheme has the Average Network Throughput of 0.873M mbps and 0.469 mbps (With Attacker) and 0.847 mbps and 0.823 mbps (without Attacker) respectively, However, when no. of nodes in the networks are 10. The improvement is $80.597\% = [(847-469) / 469 * 100]$ and $6.073\% = [(873-823) / 823 * 100]$ respectively when $CW_{min}=128$ and $CW_{max}=1024$.

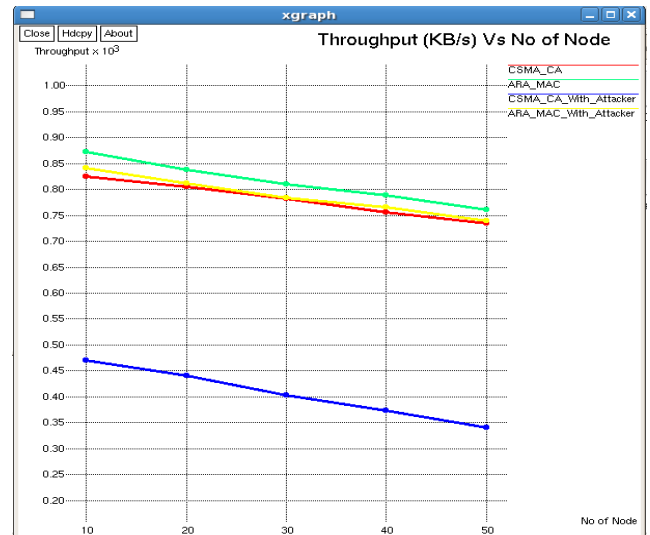


Fig1: Throughput vs. Number of nodes at $CW_{min}=128$, $CW_{max}=1024$

THROUGHPUT CONTENTION WINDOW CW _{MIN} =128, CW _{MAX} =1024				
N O D E	WITHOUT AN ATTACKER		WITH THE ATTACKER,	
	ARA -MAC	CSMA /CA	ARA -MAC	CSMA /CA
10	873	823	847	469
20	854	803	820	441
30	824	779	803	408
40	795	759	777	371
50	767	735	746	340

Table 1: Throughput vs. Number of nodes at CWmin=128, CWmax=1024

2) Medium Access Delay

It is the mean delay of all successfully delivered frames. It includes average queuing delay and contention delay [17]. It has been observed that the frame delay increases dramatically when active stations increases. Frames may be dropped either due to the buffer overflow or because of serious MAC layer contentions. Such frame losses may affect high layer networking schemes such as the TCP congestion control and network routing maintenance.

The major component of the medium access delay is contributed by back-off period. So to reduce the back-off waiting period; there should be some provisions in the protocol that takes care of it. Both ARA-MAC and CSMA/CA are using the concept of the binary exponential back-off (BEB) algorithm.

By looking at the figure and tables showing simulation results, it is concluded that ARA-MAC produces less delay as compared to CSMA/CA. The reason is the opportunist mode of ARA-MAC. Due to opportunist mode, a station comes out of waiting early and becomes ready for transmission. This reduces the back-off delay which means a reduction in the medium access delay. Consequently, As a result of this, queuing delay is automatically decreased. as a result, more frames will get the opportunity for transmission. In the absence of the attacker, ARA-MAC outperforms CSMA/CA due to this mechanism whereas in the presence of the attacker, the delay factor of CSMA/CA is severely affected, as can be seen. When contention window size increases the delay as well increases. The reason behind this is the larger span that is available for back-off.

Fig.2 and Table 2 compares the simulated average Medium Access Delay of the two schemes. It shows that the novel ARA-MAC [25] decreases the mean delay, in both cases (without and with Attacks), and the improvement becomes evident with an increased number of nodes. When the channel bit rate is 2mbps and load is 50%, which is 1mbps, the original scheme and proposed scheme have the average end-to-end of 11.53ms and 16.67ms, correspondingly when CWmin = 128 CWmax= 1024, without attacks. The improvement is 11.53% [= (26-23)/26*100] when numbers of nodes are 10. Similarly, with attackers, it can be calculated and found to be 16.67 % [= (30-25) /30*100] whereas the value of delay in CSMA/CA with attackers and ARA-MAC with Attackers is 11.53% and 16.67% correspondingly.

Values may be distinctive at different no. of nodes, obviously delay increases if nodes are more on the networks but the difference between performance with attackers, and without attackers can be only found at low loads and at nodes 4.53% [= (61-58)/61*100] when no. Of Similarly, with attackers, it can be calculated and found to be 11.49% [= (71-60) /71*100]

The reason is that when the Network load is less, the proposed scheme has a minor value of CWmin offered by Opportunist Mode, which leads to lesser contention and queuing delay: when the network load is heavy, although ARA-MAC has at its original larger minimal contention window size, but the collision is decreased, so that the average Medium Access Delay is decreased.

MEDIUM ACCESS DELAY AT CONTENTION WINDOW CW _{MIN} =128, CW _{MAX} =1024				
N O D E	WITHOUT AN ATTACKER		WITH THE ATTACKER,	
	ARA -MAC	CSMA /CA	ARA -MAC	CSMA /CA
10	23	26	25	30
20	30	32	32	38
30	38	41	40	48
40	47	50	49	59
50	58	61	60	71

Table2: Medium Access Delay vs. Number of nodes at CWmin=128, CWmax=1024 CWmax=1024

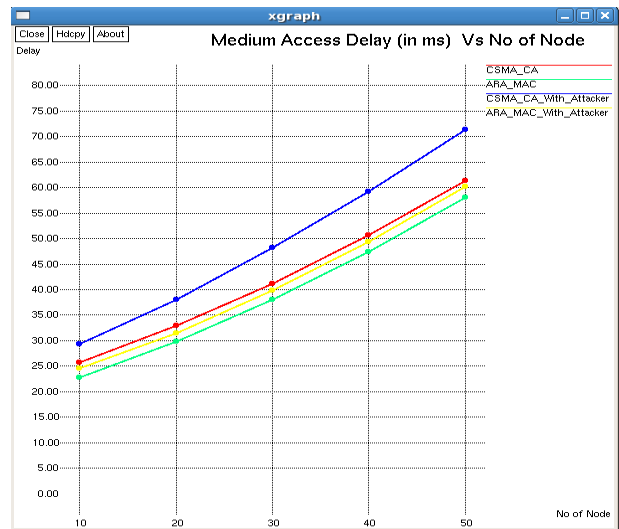


Fig2: Medium Access Delay vs. Number of nodes at CWmin=128, CWmax=1024

3) Collision Data Frames Vs Number of nodes

The collision ratio of both the protocols is almost same in the absence of the attacker as it's the attack resiliency feature of ARA-MAC that gives it the upper edge which is of no use in this scenario. Increasing window sizes are having a positive impact on the collision ratio as reflected from the figures. This is due to the large span of the back-off window that disperses the stations from being had identical timing.

Fig.3 (a) and Table 3(x) compares the simulated scenario for Collision per data frame with increasing number of nodes with the two schemes.

The original scheme and proposed scheme have the average collision per frame 55% and 75%, respectively, when CWmin = 128, CWmax= 1024, without attackers. The improvement is 55% [= (0.2-0.09) /0.09*100]. Similarly with attackers it can be calculated and found to be 75 % [= (0.42-0.24) /0.24*100].

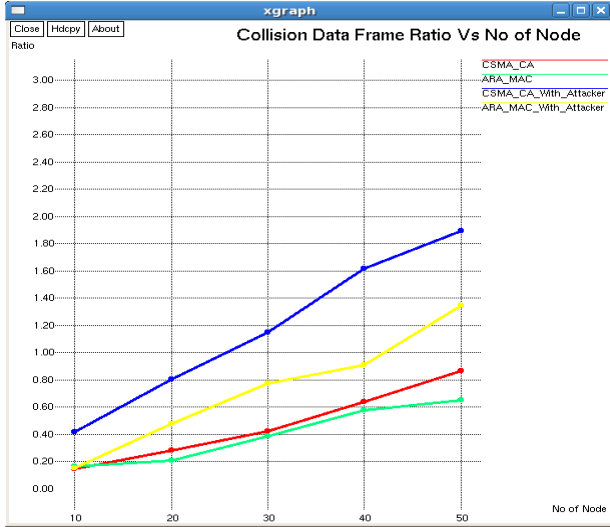


Fig.3 Collision Per Data Frame vs Number of nodes CWmin=128, CWmax=1024

COLLISION PER DATA FRAME AT CONTENTION WINDOW CWmin =128, CWmax =1024				
N O E	WITHOUT THE ATTACKER		WITH THE ATTACKER,	
	ARA -MAC	CSMA /CA	ARA -MAC	CSMA /CA
10	0.2	0.09	0.24	0.42
20	0.24	0.18	0.44	0.91
30	0.40	0.40	0.69	1.22
40	0.62	0.63	1.18	1.52
50	0.8	0.76	1.26	1.42

Table3: Collision Per Data Frame vs Number of nodes CWmin=128, CWmax=1024

4) Fairness Index

Ideally, throughput of each link within the network should not be deviated among all other links to maintain the Fairness of Medium Access Control Protocol. The standard BEB (Binary Exponential Back-off algorithm) [4] is sometimes unable to provide an adequate level of fairness, i.e. The Table 4 Fairness Index vs. Number of nodes at CWmin=128, CWmax=1024 back-off window size doubled after every collision and reduces to the minimum contention window after every successful transmission. Also, fairness drastically changes in presence of attackers. The well-known Jain's fairness index [5] is used to calculate fairness of throughput of

a different adaptive contention window algorithm based MAC protocols.

The fairness index (FI) is calculated at the end of every Observation Period (OP).

$$f(x_1, x_2, \dots, x_n) = \frac{(\sum_{i=1}^n x_i)^2}{n \sum_{i=1}^n x_i^2} \dots \dots \dots (1)$$

x_i = x_i = The throughput of contending flows (station)

n = The number of contending flows (stations)

If it is high ($0.9 \leq FI \leq 1$), it means all nodes are getting an almost proper share of the channel, and it implies that attackers are not present in the network and minor degradation in throughput is due to obvious collisions. There is no need to run the attacker identification algorithm in such a condition. The low value of FI (i.e. $0 \leq FI < 0.9$) nodes are getting an improper share of channel bandwidth and is a clear indication about the presence of the attacker or attackers in the network. And in this case, running attacker identification algorithm is a necessity. By using the concept of fairness index the load on the access point is reduced, and occasionally it has to run the algorithm.

In Figure 4 and Table 4 compares the simulated Fairness Index of the two schemes. It shows that the novel ARA-MAC [25] improves the Fairness Index, in both the cases (without and with Attacks), and the improvement becomes evident with an increased of nodes. When the channel bit rate is 2 mbps and load is 50%, which is 1 mbps, the original CSMA/CA and ARA-MAC have the average Fairness Index of .0.923 and .0.98, respectively when CWmin = 128 CWmax= 1024, with attacks. The improvement is 6.69% [= (0.98-0.923) /0.923*100=0.0669]. Similarly without attackers it can be seen by figure that it is almost same in both the schemes.

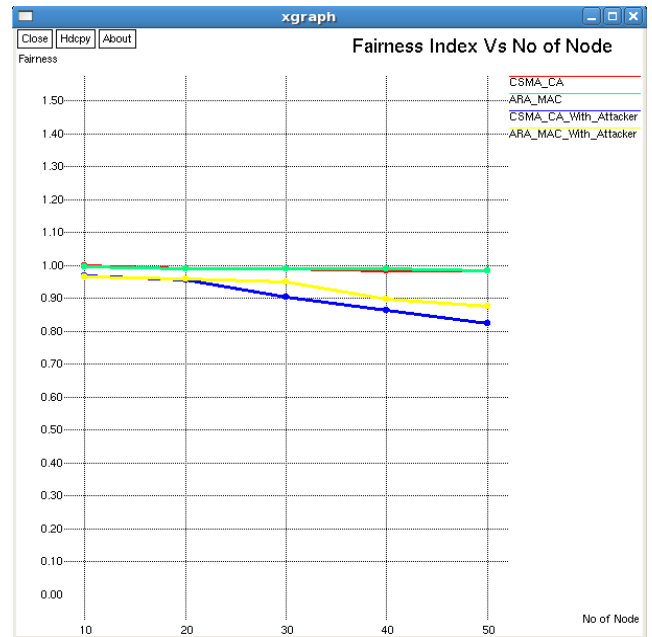


Fig. 4 Fairness Index vs. Number of nodes at CWmin=128, CWmax=1024

FAIRNESS INDEX AT CONTENTION WINDOW $CW_{MIN}=128$, $CW_{MAX}=1024$				
N O D E	WITHOUT THE ATTACKER		WITH THE ATTACKER,	
	ARA-MAC	CSMA/CA	ARA-MAC	CSMA/CA
10	1.0	1	0.96	0.97
20	0.992	0.99	0.92	0.91
30	0.991	0.99	0.90	0.87
40	0.98	0.99	0.88	0.87
50	0.98	0.99	0.85	0.83

Table4: Fairness Index vs. Number of nodes at $CW_{min}=128$,
 $CW_{max}=1024$

VIII. CONCLUSIONS & FUTURE WORK

In IEEE 802.11 WLANs, MAC layer misbehavior can lead to severe unfairness in channel utilization/bandwidth distribution. This can become a serious problem in a shared channel where individual users have to pay for network usage and hence may be motivated to cheat in order to increase their share of the medium.

In spite of its relevance, this topic is still relatively less explored in the research community, especially adaptability and attack resiliency to gather. In this paper, we have discussed MAC layer misbehaviors, classified the nodes in the channel, presented a method to utilize channel efficiently by adopting an opportunist mode, and if any node misused the opportunity or misbehave then provided the corresponding detection mechanisms.

In contrast with previous papers that have proposed modifications to the MAC protocol, thus requiring a modification of existing wireless cards, we have developed a solution that can be completely incorporated in the AP and uses at most transmitted/communicated data purely to analyze node behavior. An important feature is that a cheater has no way of knowing whether an AP is ARA-MAC enabled.

Using simulations, we have shown that ARA-MAC achieves high accuracy of detection in a variety of scenarios. The Ensemble system is resilient to several MAC layer misbehavior. The main feature of the proposed solution is its effectiveness and applicability to be static as well as real networks.

We believe that the scope of this paper goes beyond IEEE 802.11 networks; indeed, we provide a framework that can be adapted from the study of deceitful and detection techniques in any WLAN (Infrastructure less network) based on the shared spectrum.

As explained earlier the proposed work is comprised of two major stages, i.e. Adaptability and Attack Resiliency and thoroughly to view the predicament decision in entirety.

ARA-MAC has been designed to improve performance of CSMA/CA in terms of basic key performance Indicators, i.e. Throughput, Medium Access Delay, Collisions and Fairness Index, improve results shows the importance of the Novel designed algorithms.

For future work, we consider infrastructure less WLAN or Ad-hoc Network and also provide some solutions for the different types of MAC layer misbehaviors individually. We will try to incorporate some more intelligence in our approach for smart attackers.

We will also explore the effect of mobility in the proposed system. In regard to implementation, we will introduce these and other enhanced features to develop an intelligent driver for newly developed WI-Fi or Wimax, WSN wireless networks. Furthermore, the above proposed model can be converted into a single hardware chip.

It should also be noted that in this paper we have focused on attacks/misbehaves aimed at the traffic incoming/outgoing from the stations. In our future efforts, we will expand the set of attacks to new techniques that decrease the traffic arriving to the stations from the AP or vice-versa.

REFERENCES

- [1] Venkata Nishanth Lolla, Lap Kong Law, Srikanth V. Krishnamurthy, "Detecting MAC layer Back-off Timer Violations in Mobile Ad Hoc Networks," In Proceedings of the 26th IEEE International Conference on Distributed Computing Systems, pp. 63-72, 2006.
- [2] Piyush Kumar Shukla, Sanjay Silakari & Sarita Singh Bhadauria, "Efficient Model for Attack Verification in 802.11 WLAN using Filtering Mechanism based Medium Access Control Protocol: FMB-MAC," International Journal of Information Technology and Knowledge Management, July-December 2011, Volume 4, No.2, pp. 465-470.
- [3] IEEE Standard for Wireless LAN-Medium Access Control and Physical Layer Specification, 802.11,1999.
- [4] A. Akella, S. Seshan, R. Karp, and S. Shenker, "Selfish behavior and stability of the Internet: A game-theoretic analysis of TCP," In Proceedings of SIGCOMM'02, 2002.
- [5] Kyasanur, P. And Vaidya, N., "Detection and handling of MAC layer misbehavior in wireless networks," in The selected Journal of Dependable Systems and Networks, June 2003.
- [6] Piyush Kumar Shukla, Sanjay Silakari and Sarita Singh Bhadauria, "MAC Layer based Insubordination Mechanism for Differentiating and Penalizing/Rewarding Opportunist Nodes in 802.11 WLAN Network," , CiIT, International Journal of Networking and Communication Engineering, May 2011.
- [7] Andrew S. Tanenbaum, "Computer Networks,' Networks," Networks," PHI, Fifth edition.
- [8] M. Bernaschi, F. Ferreri, L. Valcamonici," Access point's vulnerabilities to DoS attacks in 802.11 networks," , Springer Science+Business Media, LLC, pp 1-11, 2006.
- [9] Frederico Cali, Marco Conti," Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," IEEE/ACM Transactions on Networking, Vol-8, No- 6, pp-785-799, 2000.
- [10] G. Bianchi, " Performance analysis of the IEEE 802.11 Distributed Coordination Function," IEEE Journal on Selected Areas in Communications, pp-535-547, March 2000.
- [11] L. Buttyan and J.P. Hubaux, "Stimulating cooperation in self-organizing mobile ad-hoc networks" ACM/Kluwer Mobile Networks and Applications, 8(5), October, 2003.
- [12] Alberto Lopez Toledo, and Xiaodong Wang, (2007) "Robust Detection of Selfish Misbehavior in Wireless Networks," IEEE journal on selected areas in communications, vol. 25, number. 6, pp. 1124-1134, August 2007.
- [13] T. Nandagopal, T. Kim, X. Gao, and V. Bharghavan, "Achieving MAC layer fairness in wireless packet networks," In Proceedings of MOBICOM'00, pp-87-98, 2000.

- [14] D. Seth, S. Patnaik, S. Pal, "A Fair Quality of service assured a MAC protocol for mobile Ad-hoc network and its performance evaluation," International Journal of Wireless and Mobile Networks, 2011.
- [15] Piyush Kumar Shukla, Sanjay Silakari & Sarita Singh Bhadauria et al. "Survey for Designing Attack Resilient and Adaptive Medium Access Control Protocol for Wireless Networks," CICSYN '09, First International Conference on Computational Intelligence, Communication Systems and Networks, 23-25, July 2009, pp. 178-183.
- [16] Lei Guang, Chadi Assi, Yinghua Ye, "DREAM: A system for detection and reactions against MAC Layer misbehavior in ad-hoc networks," Computer Communications, pp1841-1853, 2007.
- [17] Piyush Kumar Shukla, Sanjay Silakari & Sarita Singh Bhadauria et al. "An Adaptive Medium Access Control Protocol for Computer Network," CICSYN '09, First International Conference on Computational Intelligence, Communication Systems and Networks, 23-25, July 2009, pp. 184-189.
- [18] M. Bernaschi, F. Ferreri, L. Valcamonici," Access point's vulnerabilities to DoS attacks in 802.11 networks," , Springer Science+Business Media, LLC, pp 1-11, 2006.
- [19] Alvaro A. Cardinals, Svetlana Radosavac and John S. Baras "Performance Comparison of Detection Schemes for MAC Layer Misbehavior," INFOCOM, IEEE, pp. 1496-1504, 2007.
- [20] Eduard Garcia-Villegas Moises Gomez Elena López-Aguilera Jordi Casademont, "Detecting and mitigating the impact of wideband jammers in IEEE 802.11 WLANs," pp. 57-62.
- [21] Fei Shi, Jaejong Baek, Joosok Song Weijie Liu, "A novel scheme to prevent MAC layer misbehavior in IEEE 802.11 ad-hoc networks," Springer Science+Business Media, LLC, pp. 1-10, 2011.
- [22] Iordanis Koutsopoulos, "Competition and Cooperation in Wireless Access Misbehavior," IEEE, pp. 273-279, 2011.
- [23] Ioannis Broustis, Konstantinos Pelechrinis, Dimitris Syrivelis, Srikanth V. Krishnamurthy, Leandros Tassioulas, "A software framework for alleviating the effects of MAC-aware jamming attacks in wireless access networks," Springer Science+Business Media, LLC, pp. 1543-1560, 2011.
- [24] Jin Tang, Yu Cheng, Weihua Zhuang, "An Analytical Approach to Real-Time Misbehavior Detection in IEEE 802.11 Based Wireless Networks," IEEE INFOCOM, pp. 1638-1646, 2011.
- [25] Piyush Kumar Shukla, Sarita Singh Bhadauria and Sanjay Silakari, "ARA-MAC: A Qualifying Approach to improving Attack Resiliency and adaptability in Medium Access Control Protocol for WLAN 802.11, International Journal of Computer Applications, Volume 49–No.19, July 2012, pp. 1-10.
- [26] Jiahai Yang, Peng Ning, X. Sean Wang, and Sushil Jajodia, "Cards: A Distributed System for Detecting Coordinated Attacks," George Mason University, pp. 1-10, 2008.
- [27] Jaehyuk Choi, Alexander W. Min, Kang G. Shin, "A Lightweight Passive Online Detection Method for Pinpointing Misbehavior in WLANs," IEEE Transactions On Mobile Computing, pp. 1-15, 2010.
- [28] Jahangir H. Sarker, Hussein T. Mouftah, "Throughput and Stability Improvements of Slotted ALOHA Based Wireless Networks under the Random Packet Destruction DoS attack," IEEE, pp. 1-6, 2010.
- [29] Lei Guang, Chadi Assi and Abderrahim Benslimane, "MAC Layer Misbehavior In Wireless Networks: Challenges And Solutions," IEEE, pp. 1-9, 2008.
- [30] Matthias Wilhelm, Ivan Martinovic, Jens B. Schmitt, and Vincent Lenders, "Reactive Jamming in Wireless Networks— How Realistic is the Threat?," ACM, pp. 47-52, 2011.
- [31] Mihui Kim, Varagur Karthik Sriram Iyer, Peng Ning, "MrFair: Misbehavior-resistant fair scheduling in wireless mesh networks," Elsevier, pp. 299-316, 2011.
- [32] Nitin Vaidya, Anurag Dugar, Seema Gupta, and Paramvir Bahl, "Distributed Fair Scheduling in a Wireless LAN," IEEE, VOL. 4, NO.6, pp. 616-629, 2005.
- [33] Natalia Olifer, Victor Olifer "Computer Networks , " Wiley India Edition, Reprint 2010.
- [34] Pradeep Kyasanu et. al. "Detection and Handling of MAC Layer Misbehavior in Wireless Networks," UIUC Campus Research Board, pp. 1-10, 2008.
- [35] R. Kalaiarasi, Getsy S. Sara S. Neelavathy Pari and D. Sridharan, "Performance Analysis Of Contention Window Cheating Misbehaviors In Mobile Ad Hoc Networks," International journal of computer science & information Technology (IJCSIT) Vol.2, No.5, pp. 31-42, 2010 [130] Roberta Bragg, Mark Rhodes Ousley, Keith Strassberg "Network Security: The Complete Reference" Tata McGraw Hill, Tenth Reprint, 2009.
- [36] Roberta Bragg, Mark Rhodes Ousley, Keith Strassberg "Network Security: The Complete Reference" Tata McGraw Hill, Tenth Reprint, 2009.
- [37] Raja Gunasekaran, Vaidheyanathan Rhymend Uthariaraj, Umapathy Yamini, Rajagopalan Sudharsan, Selvaraj Sujitha Priyadarshini, "A Distributed Mechanism for Handling of Adaptive/Intelligent Selfish Misbehavior at MAC Layer in Mobile Ad Hoc Networks," Computer Science And Technology, pp. 472-481, 2009.
- [38] S.A.Arunmozhi1, Y. Venkataramani, "DDoS Attack and Defense Scheme in Wireless Ad-hoc Networks," , International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.3, pp. 182-18, 2011.
- [39] S. Szott, M. Natkaniec and R. Canonico, "Detecting back-off misbehavior in IEEE 802.11 EDCA," Wiley Online Library, pp. 1-4, 2010.
- [40] Shafiqullah Khan, Kok-Keong Loo, Tahir Naeem, Mohammad Abrar Khan, "Denial of Service Attacks and Challenges in Broadband Wireless Networks," IJCSNS International Journal of Computer Science and Network Security, VOL.8. No.7, pp. 1-6, 2008.
- [41] S. Radosavac, Alvaro A. Cardenas, John S. Barrows and George V. Moustakides, "Detecting IEEE 802.11 MAC layer misbehavior in ad-hoc networks: Robust strategies against individual and colluding attackers," Journal of Computer Security IOS Press, pp. 103–128, 2007.
- [42] Svetlana Radosavac, John S. Baras, George V. Moustakides, "Impact of Optimal MAC Layer Attacks on the Network Layer," ACM, pp. 135-144, 2006.
- [43] Svetlana Radosavac, George Moustakides, John S. Baras, Iordanis Koutsopoulos, "An Analytic Framework for Modeling and Detecting Medium Access Layer Misbehavior in Wireless Networks," ACM Transactions on Information and Systems Security, Vol. 11, No. 4, 2008.
- [44] Piyush Kumar Shukla, Sanjay Silakari & Sarita Singh Bhadauria, "Designing and Analysis Issues for an Attack Resilient and Adaptive Medium Access Control Protocol for Computer Networks: An Exclusive Survey ", IJCSIS, Vol. 1, No. 1, May 2009.
- [45] Shenlong Chen, Yuqing Zhang, Qixu Liu, Jingyu Feng, "Dealing with dishonest recommendation: The trials in reputation management court," Elsevier B.V. pp. 1-16, 2011.
- [46] Szymon Szott, Marek Natkaniec, and Andrzej R. Pach "Cheating on the CW and RTS/CTS Mechanisms in Single-Hop IEEE 802.11e Networks," SPRINGER, pp. 176–188, 2008.
- [47] Taimur Farooq, David Llewellyn-Jones, Madjid Merabti, "MAC Layer DoS Attacks in IEEE 802.11 Networks," PGNet, pp. 1-6, 2010.
- [48] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A Media Access Protocol for Wireless LANs," in Proc. of ACM SIGCOMM '94, 1994.
- [49] Vikram Gupta, Srikanth Krishnamurthy, and Michalis Faloutsos, "Denial of Service Attacks at the MAC Layer in Wireless Ad Hoc Networks," National Science Foundation, pp. 1-5, 2006.
- [50] V.R. Giri, and N. Jaggi, "MAC Layer Misbehavior Effectiveness and Collective Aggressive Reaction Approach" in 33rd IEEE Sarnoff Symposium, 2010, pp. 1 - 5, Princeton, April 2010.

A cloud computing architecture for e-learning platform, supporting multimedia content

Attlee M. Gamundani
University of Zimbabwe
Computer Science Department
P.O.Box MP 167
Zimbabwe

Email. amgamundani@gmail.com

Taurayi Rupere
University of Zimbabwe
Computer Science Department
P.O.Box MP 167
Harare
Zimbabwe

Email. taurai.rupere@gmail.com
trupere@science.uz.ac.zw

Benny M Nyambo
University of Zimbabwe
Computer Science Department
P.O.Box MP 167
Zimbabwe

Email. nyambo@science.uz.ac.zw

Abstract— E-learning based platforms that support multimedia content to enhance interactive learning demands large disk space. Despite research ground covered under e-learning circles, less attention has been devoted to solicit the best methods to address the disk space challenges at minimal cost. This research focuses on advancing a best architecture that meet the need for storage space when developing interactive multimedia e-learning based portals. Simulation was used using the CloudSim toolkit. Findings show that to precisely test the performance of viable architectures, there has to be a robust platform for such experiments. The main conclusions drawn from this research were that, there is room to improve on existing architectures to scale down on development costs so attributed to e-learning portals that are interactive in stature. Storage can be built from exiting personal computers through harnessing the cloud computing functionality designed as most of the personal computers are not fully being used by their owners. This research culminate by recommending the need to explore on best simulator packages that can be used to test the functionality of cloud computing based architecture for e-learning environments.

Keywords- E-learning, Multimedia, Architecture, Cloud Computing, Storage, CloudSim, Simulation, Interactive.

I. INTRODUCTION AND BACKGROUND

E-learning based platforms that support multimedia content to enhance interactive learning demands large disk space [1]. To curb the challenge of disk space and decongest the server(s) what should be done? In setups where there is a pool of networked computers or computers that can join a network for a dedicated period, what is happening to all the

idle space and processing power? What solution is there to the storage space challenge for e-learning platforms that supports multimedia content? A close glance at a simple e-learning structure as presented in Fig 1. below depicts the three main players in the e-learning ecosystem as the e-learning server, the network and the e-learning client.



Figure 1: E-learning System [1]

Paradoxical to this illustration, many depicts the existence of e-learning clients as the dormant player, there is in that light a single network and a single e-learning server, though scientifically it can mean two different words in different formats. There is little attention being paid in most researches on e-learning based systems on how best the e-learning clients can also contribute to the server space and thereby increase throughput to e-learning based systems.

As reference [2] posits that “Learning management systems (LMS’s) are already made available in SaaS model, but for most LMS providers, cost structure for providing LMS in SaaS model is currently governed by ‘old’ way of server infrastructure management (in-house or rented), which is then as-is passed on to the customer”.

It is on this background highlighted above, that part of this research work is built upon. There is cost involved in hosting e-learning systems, the stages at which such costs weather in monetary terms or otherwise needs to be traced and constituted individually to specific development stages of the e-learning life cycle. There is greater opportunity to explore the best ways in which certain costs could be minimized or avoided if proper attention is given across the spectrum. The rest of the paper is organized as follows: section 2 gives the problem definition and research questions; section 4 gives a brief literature about cloud computing architecture in e-learning environment followed by the methodology and research design. The results and analysis follows proceeded by the discussions, conclusions and recommendations.

II. PROBLEM STATEMENT

The main focus of this research following the background information outlined above, therefore was to explore the best resources available that can alleviate the challenges many e-learning system developers face especially for platforms that carry a substantial volume of multimedia content. The problems experienced in e-learning systems as a medium used to enhance learning and dissemination of information has particularly contributed to the formulated pilgrimage towards a solution that can best be implemented in e-learning systems. The search for such a solution was in light of addressing the challenges at the grass root level that is at the development stage of e-learning systems. Having observed that many institutions are able to launch and host e-learning based systems to enhance their learning criteria, there is a limitation of storage space.

In trying to address a challenge herewith presented reference [3], posits that “The process of E-learning is not a perfect one.....Course material used in e-learning sometimes is unattractive and non-compelling.” Also [4] highlighted that “Business divisions who have special and particular needs and mid-size and small companies continue to face stiff challenges and financial crunch to implement e-learning programs and solutions to fit their scale.”

To address this challenge, there was a motivation to look at the best cost cutting and resource efficient way in which e-learning systems can be developed and deployed at scales that best suits the business or institutions’ requirements. In particular, attention was given to storage capacity as one of the variables that demands some huge financial investments as sometimes a substantial volume of space is required, which in turn triggers sourcing of more servers to meet the ever increasing demand for space.

One of the features that strongly are a check list when testing the quality of e-learning systems as highlighted by[5] that “..... as the subject of the course is state of art and thus should be updated regularly. There should be high level

of interaction as well...”. This indicates multimedia content based e-learning systems are the solutions called for, however to host such e-learning platforms in a cost effective routine, what avenue can be followed?

III. RESEARCH QUESTIONS

The paper seeks to answer the following main research question: Is there an appropriate architecture that meets the needs of storage space and be used to host e-learning systems built with multimedia content to enhance interactivity in learning circles?” Sub questions emanating from the main research questions are:

- What are the other e-learning architectures that support multimedia content?
- How are the current e-learning architectures effective and assessed?
- Can we come up with a suitable architecture for multimedia e-learning systems with solve the need for storage space compared to existing one?

IV. LITERATURE REVIEW

A. E-learning architectures supporting multimedia content

E-learning involves a broad combination of processes, content and infrastructure that uses computers and networks to scale or improve learning parts including the management and delivery [6], [3], [7]. While architecture integrates various components, that consists the public side of interface[8], broken down into parts [9] that involves functionality, usability, performance, reuse, economic and technology [10].

The IMS LIP final specification by [11] does provide the following structures to support the implementation of "any suitable architecture" for learner privacy protection. However focus was more centred on addressing privacy principles as depicted in fig 2 below.

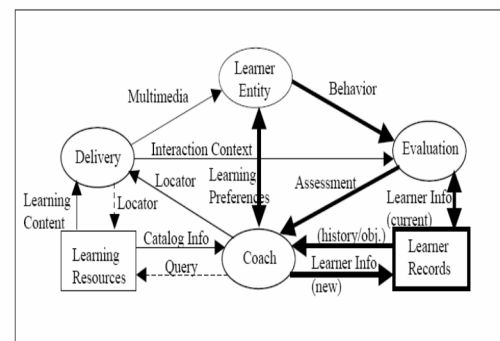


Figure 2: LSTA system components [11]

A close look at this architecture clearly show the display of learning resources and multimedia components for this setup is not being housed under a single shell. There is a possible danger in slowing delivery of the e-learning resources to end-users in such an architectural setup. It is much convenient to compute delivery latency at a go than at

different delivery levels. This is pointed out by reference [1] that “As with rapid growth of the cloud computing architecture usage, more and more industries move their focus from investing into processing power to renting processing power from a specialized vendor.”

From this nugget, the cloud computing architecture presented here is availed on rental from a specialised vendor. This pause an opportunity to look on how best users or organisations can be specialised vendors in own capacity to effectively make use of the available resources. Also [3] and [4] pointed out that “The technical standards for connecting the various computer systems and pieces of software needed to make cloud computing work still aren't completely defined”.

Well, cloud computing maybe presented as a viable architecture as indicated here, there is still a loophole that needs research attention in light of being able to define computer systems and pieces of software that will enable the cloud computing architecture a worth tool for use. Reference [4] posits that “Every decade or so, the computer industry's pendulum swings between a preference for software that's centrally located and programs that instead reside on a user's personal machine. It's always a balancing act, but today's combination of high-speed networks, sophisticated PC graphics processors, and fast, inexpensive servers and disk storage has tilted engineers toward housing more computing in data centers. In the earlier part of this decade, researchers espoused a similar, centralized approach called "grid computing." But cloud computing projects are more powerful and crash-proof than grid systems developed even in recent years.”

B. The effectiveness of E-learning architectures

A close glance at the following case study unveils some of the focus that the implemented e-learning systems architectures were inclined towards: Researcher [11] pointed out that “However, most of them are focusing on content management, meta-data specification, or other areas with little reference to security and privacy. For example: - The AICC focuses on practicality and provides recommendations on e-learning platforms, peripherals, digital audio and other implementation aspects. - The ARIADNE focuses mainly on meta-data specification of electronic learning materials with the goal of sharing and reusing these materials.”

It is easier to advance a solution that addresses how to host content alone based systems than multimedia content based systems. This present a great challenge to the e-learning design and implementation chores as a result, there is great need to explore this avenue in a holistic approach.

ADL-SCORM is mainly concerned with specifying how instructional content should be treated [11], [12]. Instructional content is mainly descriptive in nature, well if it embeds multimedia in it, the focus inclined by ADL-SCORM review are towards handling not hosting of such instructional content. As an additional requirement to produce a complete piece of work, the need to complement efforts made by advancing a possible hosting architectural setup that can both house the instructional content and provide elastic space for future storage needs, cannot be overemphasised.

Cloud computing delivers infrastructure, platform, and software that are made available as subscription-based services in a pay-as-you-go model to consumers. These services are referred to as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) in industries [1]. Being an emerging discipline, the cloud computing setup still passes to be high cost option and deterring to some users. It is noble to explore better ways in which this model can be used hence implemented on a low to no cost range [13], [14].

The current cloud computing architecture involves the existence of data centers that are able to provide services to the clients located all over the world. In this context, the cloud can be seen as a unique access point for all the requests coming from the customers/clients. It is from this setup that a motivation to conduct this research so as to identify the best ways in which the datacenter components of the architecture can be localized and make use of available resources to the maximum potential is obtained. Even if it seems not very reasonable, the cloud computing provides some major security benefits for individuals and companies that are using/developing e-learning solutions, like the improved improbability. It is almost impossible for any interested person (thief) to determine where is located the machine that stores some wanted data (tests, exam questions, results) or to find out which is the physical component he needs to steal in order to get a digital asset. As such, benefits presented by the e-learning architecture are recipe enough to aid exploration of best options to advance solutions hosted in the same fashion [1].

The fact that cloud computing has the most unique characteristic of virtualization makes possible the rapid replacement of a compromised cloud located server without major costs or damages. It is very easy to create a clone of a virtual machine so the cloud downtime is expected to be reduced substantially;

The other security benefits that accrue to cloud computing presented here are:

- centralized data storage – losing a cloud client is no longer a major incident while the main part of the applications and data is stored into the cloud so a new client can be connected easily and fast. Imagine what is happening today if a laptop that stores the examination questions is stolen [1].
- monitoring of data access becomes easier in view of the fact that only one place should be supervised, not thousands of computers belonging to a university, for example. Also, the security changes can be easily tested and implemented since the cloud represents a unique entry point for all the clients [1]. Hence it exposes one feature worthy pursuing- the data center for centralised control and management of resources [4], [12].

Upon analyzing these, it is expensive to secure servers, but why not use idle space in most of personal computers around say in an institution setup where there is potential to have a pool of computers that can join a network setup. How much cost can be reduced by exploiting such an avenue?

The following subsections will look into the tools that shall act as a basis for testing the proposed architecture for validation of this research.

C. The Resource Allocation algorithm

Following functionality of architectures reviewed, it is ideal to look at literature that explain the sole functionality of the resource allocation algorithm to ensure that there is sanity at the datacenter, by efficiently monitoring tasks in their various states of execution; the datacenter is in a position to precisely monitor its pool of resources. Resource allocation here is in a dynamic fashion, in that user requests are unpredictable and disk space requested varies with an application being developed.

We advanced a resource allocation algorithm that was tested and proven for its efficient management and monitoring of resources at the datacenter by [15], in their research on Dynamic Resource allocation in computing clouds through Distributed Multiple Criteria Decision Analysis.

```

ResourceAllocation ≡
Choose t in taskpool(self) with taskStatus(t) = unassigned
do
  Let n = placementNode(t) in
    If n ≠ undef then
      RequestResourceLock (n, t)
      potentialNode(t) := n
      taskStatus(t) := allocated
      setTimer(self, t)
    else
      Warning (“Cannot find suitable node”)

```

The execution being done by the algorithm above is straightforward in nature in that, a task is checked for its status from the task pool or table, a task that has an “unassigned” state, is picked and a Node is identified from the pool of nodes. Once a Node is allocated to a task and its confirmed it can execute the task requests with resources available (a functionality of the agent assigned at the node that constantly communicate with the datacenter agent of that node), the node to engage itself in a “Lock” state so that when the next round robin check of available nodes is done, the node is skipped. It is the node agent that has the functionality of creating virtualized processors for a dedicated period of the task’s execution processing cycles.

A task will only be considered executed if a commit confirmation status is reported back to the datacenter agent by the assigned node agent. Otherwise the datacenter agent reports a reschedule of the task to a node that can execute the task completely.

D. The Cloud sim simulator

What formulates the CloudSim architecture as a simulation tool, as illustrated by Fig 3 below, which displays the various layers of the simulator, suffice to select it as the right tool for this research task. It is from this dimension that one can safely indicate that the CloudSim tool is fully packed with structures, modules and the test environment that enable modeling of cloud computing architectures before their execution into real clouds. It is a tool that minimizes the risk of unqualified results or decision points. Simulation for this research using the CloudSim tool, will be done at the CloudSim simulation layer which provide support for modeling and simulation of virtualized setups, data center setups as well as dedicated management interfaces for virtualization variables such as memory, storage and bandwidth. The fundamental issues such as provisioning of hosts to VMs, managing application execution, and monitoring dynamic system state are handled by this layer. It is therefore at this layer that the efficiency of resource allocation criteria chosen is tested. Such implementation can be done by programmatically extending the core VM provisioning functionality [16], [17], [18].

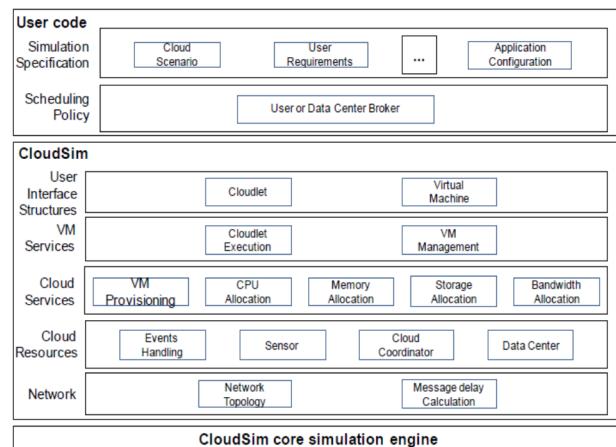


Figure 3: Layered CloudSim architecture by [16],[17],[18]

The top-most layer in the CloudSim stack is the User Code that exposes basic entities for hosts (number of machines, their specification and so on), applications (number of tasks and their requirements), VMs, number of users and their application types, and broker scheduling policies.

Host is a CloudSim component that represents a physical computing server in a Cloud: it is assigned a pre-configured processing capability (expressed in millions of instructions per second – MIPS), memory, storage, and a provisioning policy for allocating processing cores to virtual machines [16]. The existence of this feature in the CloudSim simulator makes it an ideal simulator for use in this research, where the Host will represent personal computers that have the capacity to lease space to datacenter building.

V. METHODOLOGY

The researchers proposed and came up with architecture for building a cloud computing IAAS architecture making use of the pool of available high computing power say in an institution setup where one can have the privilege of say more than hundred (100) computers in operation at any given instance. An algorithm for resource allocation is implemented that allocates resources in a cloud computing architecture setup at the datacenter. Performance of the resource allocator algorithm will give a fair analysis of the viability of the proposed architecture. There are variables that were held constant such as the network bandwidth, machine processing power and uptime and downtime of machines is presumed and set at predictable intervals.

The Cloudsim Simulator was used as the main tool for the simulation of the algorithm for resource allocation, which dynamically allocate resource, why dynamic? The selection of a dynamic approach to resource allocation is in light of the fact that, user demands and intervals for resource request and apportionment is not uniform; it varies from time to time.

A resource allocation algorithm that dynamically allocates resources is going to be advanced to the CloudSim tool; its performance shall form the basis for data collection in the form of graphical representation of variables being tested per given test case. It is an analysis of the graphical performance of each test case against defined variables that this research will be able to pass conclusions thereof.

A. The Proposed Architecture

The architecture being proposed by this research is in light of building a cloud computing IAAS, architecture. After such an initial design, what needs to be explored further is how the storage space is going to be managed? Where will it be managed from? Who will be managing storage? Why does one need the space?

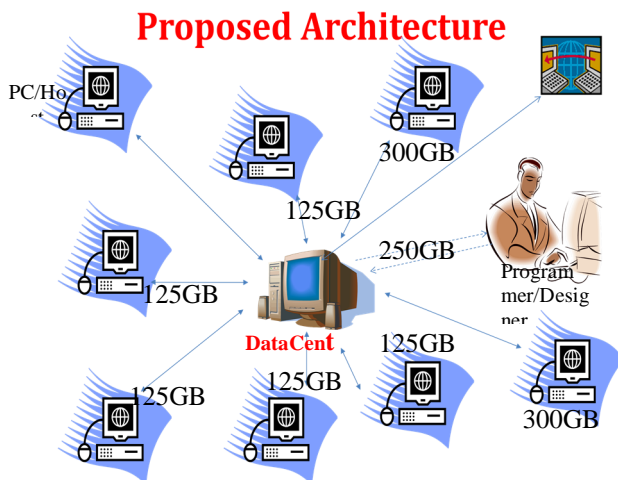


Figure 4: Proposed architecture diagram

As depicted in the diagram above, resources (i.e. disk space) is being forwarded to a central computer labeled "data center", which implies, this is the hub of this setup. Given such a scenario, one now need to strategize on how best to

manage this pool of storage space and subsequently be in a position to allocate it to requesting applications in this instance, users are specifically concerned about creating a pool of storage capacity and processing power to be in a position to host an e-learning portal that can incorporate multimedia content to enhance interactivity in the learning environment.

B. The Functionality of the Proposed Architecture

The functionality of this proposed architecture mainly revolves around the datacenter activities in acquiring, managing and allocating disk space to application requests in the operability of the e-learning system. It will be noble to have an algorithm that efficiently manages the pool of resources at the datacenter. Since the datacenter will be the main control board, the Nodes/personal Computers will communicate through the datacenter. It is the datacenter's responsibility to have an update status of each and every node so that at every request for resource allocation it can route the request to a node that has the necessary resources to host that request. Basically there are a set of activities that transpires at the node and the rest of core activities are centrally executed at the datacenter, all the same the monitoring and management of those activities is centrally done at the datacenter.

The Node activities that the datacenter monitors can be summarized as in Fig 5.

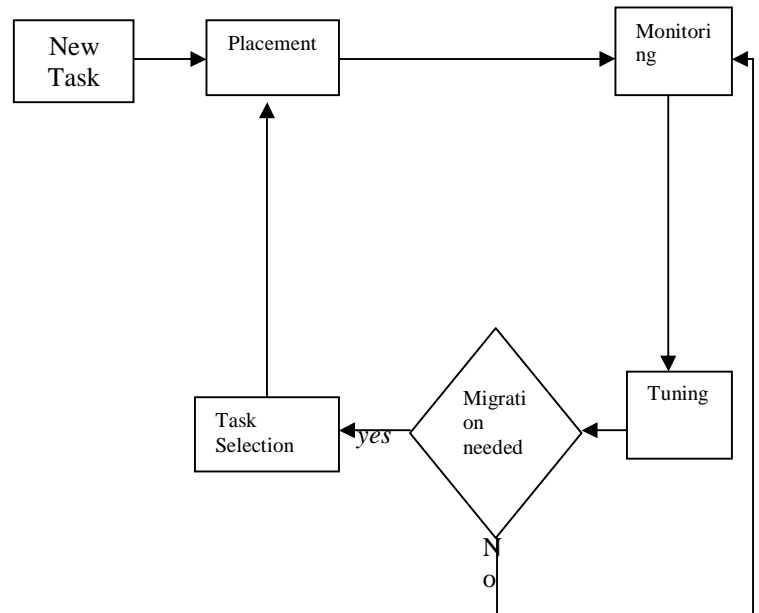


Figure 5: Node Activities

These activities will be monitored at the datacenter; it is not the node's responsibility to execute these processes. As a consequence of this setup, there is a possibility of a pool of resources at the datacenter that need to be monitored in terms of their state at any given instance. It is also the responsibility of the datacenter to populate its task bank with statuses of each task. Ideally the tasks can either be in a state of "unassigned" implying it will be waiting in the queue for execution. In another state, a task can either be "allocated," which implies it would have been allocated to a specific node, in the allocated

state, the node may be busy or taking forever to finish execution of its tasks, there may have to be “migration” decisions to reassign it to another node, if it is successfully find a processing slot, it will be in the “running” state. A successful processed task should a have “commit” state.

VI. RESULTS AND ANALYSIS

In this section we report on the findings from the simulation of the proposed architecture’s functionality. In the first instance, the results of the simulation are discussed. Findings are compared to the findings in the literature review. The graphical displays presented in the following review are the main output from the simulation done using the CloudSim simulator.

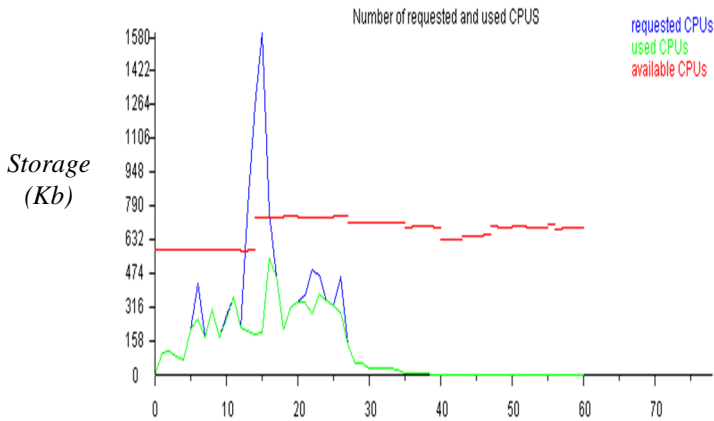


Figure 6: Graph of Requested and Used CPUs

As represented in the Figure 6, the nature of distribution patterns for requested CPUs, used CPUs and available CPUs projected over a period of an hour, is such that there may be a high demand for CPUs during the initial 10-20 minutes once the CPUs to the requested number are gathered, one will have a normal performance in resource scheduling at the datacenter for any demand for CPUs in this instance, the main agenda is to build up as much storage and processing power as possible.

Figure 7 shows the average machine usage per hour. It is clear that during the 10-20 minute peak period the average machine usage is up because of the high concentration of jobs at the datacenter.

Hosts

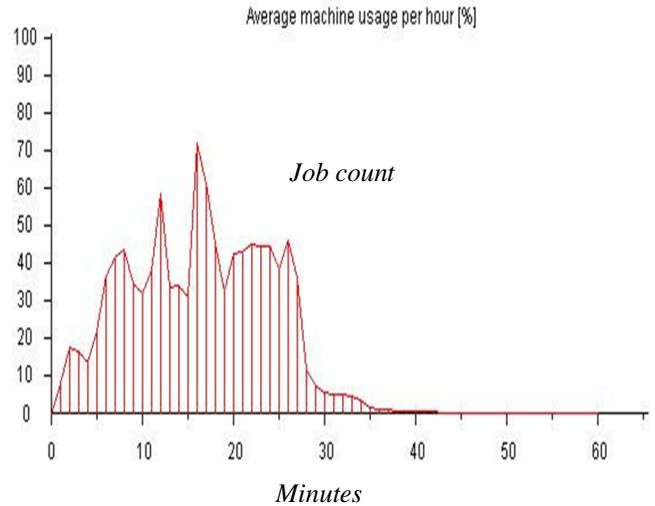


Figure 7: Graph of Average Machine Usage per hour

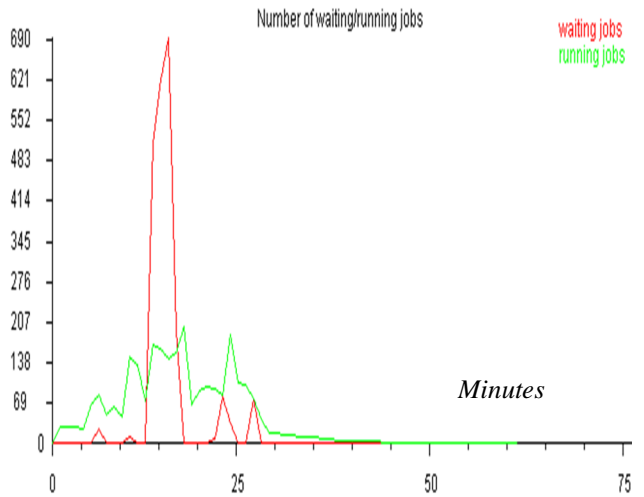


Figure 8: Graph of Number of Waiting /Running Jobs

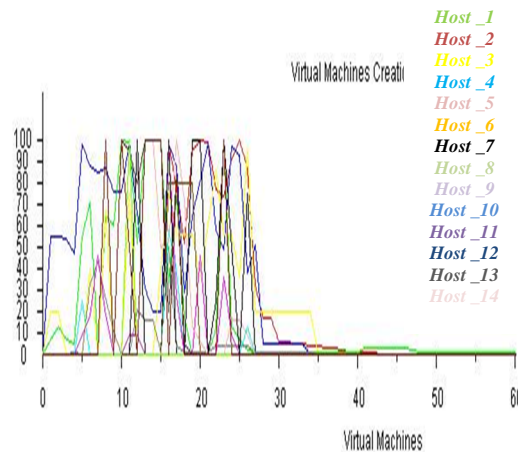


Figure 9: Graph of Virtual Machine Creation (%)

As in the previous graphs, the graph in Figure 8 depicts the same nature of functionality at the datacenter revealing almost the same trend. There is generally a sizeable number of waiting jobs during the same 10 -20minute peak range spelling a deficiency in availability of resources.

This display is somehow devoid of other previous representations, for the fact that, it now displays the concentration of creation of virtual machines, at Host machines, to meet the resource processing requests at the datacenter. It is not easy to follow the pattern formulated here as it is highly clustered in nature. Basically, Virtual machines are created as and when needed, at the same time, they are destroyed as and when they have finished processing jobs assigned to them. As virtual machines are highly depended on host machines, there is a general trend that such virtual machines' viability highly depends on the availability of a host, thus why it is important to observe the functionality of this variable, how it operates within the vicinity of the hosting machine.

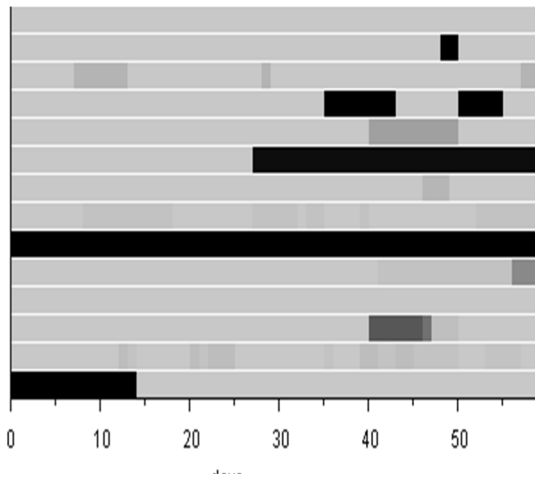


Figure 10: Graph of Percentage of failed CPUs per hour (%)

The representation is closely linked to the previous graph on virtual machine creation. The test for failed CPUs/Hosts is necessary to statically populate the datacenter control table so that hosts that are no longer functional are removed from potential pool of resources. It is one of the functionalities of the datacenter to continuously update its storage capacity, processing power and the need for further resources, so that queues are minimized. This improves the functionality of the resource allocator algorithm at the datacenter.

A. Analysis and Synthesis of Results

The simulation results from the expected functionality of a normal cloud computing architecture based on results of typical simulations using the CloudSim simulator are normal. The big challenge however, is the existence of testing metrics to quantify the results to reach a precise conclusion. From a qualitative analysis, the results are averagely in a functional order for the set of variables that were advanced for this typical simulation where concentration was mainly channeled towards the allocation, execution, virtualization and host

usage. If one produce an output for use to the e-learning community, one would strongly recommend further tests using any viable tools of the same architectural functionality advancing a variation of variables not only limited to what this research might have sampled in reaching its conclusion. The architecture may perform differently under different network setups, especially considering the fact that network bandwidth as one of the determinant factors of functionality that can have enormous contributions either negatively or to a certain screwed direction was not tested for lack of proper technology for a testing environment, the existing simulators for cloud computing that supports both a test for virtualization and network topology are expensive to implement and in most cases not easily accessible to the learning community.

If one were to pass a rating of the results of this simulation based on the qualitative analysis, one will graduate our confidence rate at 60%, implying the results reliability is set at the same rate, for a number of solid reasons, some already highlighted above. A simulation and a real world setup are two unique environments, simulations are easy to manipulate and results so produced will have a partiality towards the test variable being tested, this is a different scenario with real operational environments were all the environmental variables imposes their existence at untimed intervals.

Another factor which also worries could be the shift of trend of results is the latency rate, for the purposes of this research such a variable was not tested, however its contribution to the overall performance of the proposed architecture is one that cannot completely be ignored. Therefore, there is still room to improve the decision points reached by this research if better metrics are advanced and better tools for simulation are also used.

VII. CONCLUSIONS

After this research we made the following conclusions:

There are deterrence's to hosting of scalable e-learning platforms that need to be addressed fully and logically, that will see a better equipped e-learning environment. The e-learning community and a pool of researchers still have a long way to go in soliciting detailed information on cost effective ways e-learning platforms can be build, hosted, run and maintained.

There is still room to scale down on available options to cater for the low to no cost applications for the architectures available for hosting e-learning platforms. Research in the cloud computing architecture simulators is still very tentative and lacking on precision.

VIII. REFERENCES

- [1] P. Pocatilu, "Cloud Computing Benefits for E-Learning Solutions," *Oeconomics of Knowledge*, vol. 2, no. 1,1Q, 2010.
- [2] M. Gupta, "Cloud Computing & How It Impacts E-learning," *Tools & Technologies*, 24 June 2010.
- [3] J. Eklund, M. Kay and H. M. Lynch, "E-learning: Emerging Issues and Key Trends," *Flexiblelearning.net.au*, 2003.

- [4] R. Aaron, "Computing heads for the clouds.," *Bloomberg Businessweek.- Technology*, 16 November 2007. environments and evaluation of resource provisioning algorithms," *Software practice & experience Volume 41, Issue 9*, pp. 23-50, 2011.
- [5] H. F. O'Neil and R. S. Perez, "Issues in Synchronous versus Asynchronous E-learning Platforms," *Webbased Learning Theory Research and Practice*, 2006.
- [6] J. Gilbert, "E-Learning: The student experience," *British Journal of educational technology*, vol. 38, no. 4, pp. 560-573, 2007.
- [7] M. Al-Zoube, "E-learning on the Cloud," *International Arab Journal of e-Technology, Vol.1, No.2, June 2009*, 2009.
- [8] C. Bass, P. Clements and R. Kazman, *Software Architecture in Practice*, 2nd Edition, 2003.
- [9] M. Fowler, *Patterns of Enterprise Application Architecture*, 2003.
- [10] B. Grady, K. Phillippe, B. Kurt and R. Rich, "Software Architecture and Agile Software- An Oxymoron?," 2009.
- [11] K. El-Khatib, L. Korba, Y. XU and G. Yee, "Privacy and Security in E-learning," *October- December*, 2003.
- [12] W. Su-Chen, C. Bronwen and J. Alister, "Challenges of e-learning for University Instructors in Taiwan," 2008.
- [13] B. Soumya, M. Indrajit and P. Mahanti, "Cloud Computing Initiative Using Modified Ant Colony Framework," *World Academy of Science, Engineering and Technology*, p. 56, 2009.
- [14] Q. A. Janna and R. Lee, "The future of Cloud Computing," *Pew Internet & American Life Project Report*, 2010.
- [15] O. Y. Yaziq, M. Chris, F. Roozbeh, N. Stephen, G. Adel, G. Sudhakar and C. Yvonne, "Dynamic Resource Allocation in Computing through Distributed Multiple Criteria Decision Analysis," in *Cloud Computing (CLOUD), 2010 IEEE. 3rd International Conference*, 2010.
- [16] B. Rajkumar, R. Ranjan and N. Rodrigo, "Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities," in *High Performance Computing & Simulation, 2009, HPCS, '09. International Conference*, 2009.
- [17] N. C. Rodrigo, R. Rajiv, A. D. R. Cesar and B. Rajkumar, "CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services," 2010.
- [18] N. Rodrigo, R. Rajiv, B. Anton, A. D. R. Cesar and B. Rajkumar, "CloudSim: a toolkit for modeling and simulation of cloud computing

IX. AUTHORS PROFILE

Authors Profile ...

Attlee Munyaradzi Gamundani is a lecturer in Computer Science and done research in simulation and modelling applications.

Taurayi Rupere is a lecturer and researcher in computer science and draw much of his research in e-learning environments in both cloud and grid environment.

Munyaradzi Benny Nyambo is chairman and lecturer in Computer science from University of Zimbabwe and have carried quite a number of substantial research in mobile scheduling applications and networking algorithms.

A Survey of Predicting Close Value in Stock Market

Dharamveer Sisodia

Department of Computer Science & Engineering
SR Group of Institution, CSE campus
Jhansi, India

Beerendra Kumar

Department of Computer Science & Engineering
SR Group of Institution, CSE campus
Jhansi, India

Jitendra Kumar Gupta

Department of Computer Science & Engineering
SR Group of Institution, CSE campus
Jhansi, India

Dr. Saurav Shrivastava

Department of MCA
Bundelkhand University
Jhansi, India

Abstract— The goal of fundamental analysis is to decide the value of a stock based on the previously mentioned factors and to act on the assumption that the actual stock price will eventually reflect the determined value. Stock price forecasting is an important task for investment/financial decision making challenge. It receives considerable attention from both researchers and practitioners. Stock market is highly volatile, complex and dynamic area so stock/price forecasting is a considerable challenging issue. Several approaches have been used for forecasting stock price such as traditional and fundamental methods. In this paper we propose a hybrid combinatorial method of horizontal partition based decision tree and the genetic algorithm for the prediction of close values in the stock market.

Keywords- *Sstock market prediction, Genetic Algorithm, Decision Tree.*

I. INTRODUCTION

The prediction of stock market movement has been an issue of interest for centuries. Despite years of study and the latest technology, it seems that no method has been discovered that consistently works. Fundamental analysis usually works best over longer periods of time, whereas technical analysis is more suitable for short term trading. Neural networks offer the ability to process large amounts of data quickly and retain information learned from that data to be used again in the future. Technical analysis provides a slightly more concrete method of evaluation at the cost of locating and implementing the algorithms yourself.

Forecasting stock price or financial markets has been one of the biggest challenges to the AI community. Various technical, fundamental, and statistical indicators have been proposed and used with varying results. However, none of these techniques or combination of techniques has been successful enough. The objective of forecasting research has been largely beyond the capability of traditional AI research which has mainly focused on developing intelligent systems that are supposed to emulate human intelligence [1].

The difficulty with technical analysis is that a complete pattern is required to make an accurate prediction on the stock movement. Ideally, such a prediction should be made before the pattern was completed to enable prediction. For this task implement a Time-Delay Artificial Neural Network, called Midas. Like Metis, Midas takes input data including open, close, high and low prices per day for a particular stock ticker over a period of time, along with the corresponding trading volume for each day. This data is sufficient for implementing technical analysis through Metis and is therefore used as the input data for Midas as well [2].

Once data is pre-processed by Metis, locating all relevant technical indicators, it is passed to Midas as training data. This data is useful for training because we use the technical indicators to highlight areas in the data where Midas should be able to make a prediction. The goal of this process is to train Midas such that it can make these predictions before the technical indicator patterns are completed. The predictions made by Midas represent a predicted direction of movement at a given time interval. As such, it is very easy to check Midas' accuracy by simply running it on older data so we can see whether its predictions were correct.

A. Stock Market Prediction

It is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful prediction of a stock's future price could yield significant profit. Some believe that stock price movements are governed by the random walk hypothesis and thus are unpredictable. Others disagree and those with this viewpoint possess a myriad of methods and technologies which purportedly allow them to gain future price information [3].

B. Genetic Algorithm

GAs is commonly used as optimizers that adjust parameters to minimize or maximize some feedback measure, which can

then be used independently or in the construction of an ANN. In the financial markets, genetic algorithms are most commonly used to find the best combination values of parameters in a trading rule, and they can be built into ANN models designed to pick stocks and identify trades.

C. Decision Tree

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. If in practice decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Another use of decision trees is as a descriptive means for calculating conditional probabilities [4].

II. BACKGROUND

From the beginning of time it has been man's common goal to make his life easier. The prevailing notion in society is that wealth brings comfort and luxury, so it is not surprising that there has been so much work done on ways to predict the markets [1]. In the recent decade so many researchers have been done on neural networks to predict the stock market changes. One of the first efforts was by Kimmoto and his colleagues in which they used neural networks to predict the index of Tokyo stock market [5]. Mizuno and his colleagues also used neural networks to predict the trade of stocks in Tokyo stock market.

III. RELATED WORK

Past approaches to this problem first applied an artificial neural network directly to historical stock data using linear time series modeling [6]. This produced results which over fitted the training data and therefore rendered them unusable in real trading. In addition to omitting any preprocessing, the neural networks used only contained two layers, an input and an output layer. These linear techniques are now known to be provably insufficient for any nonlinear phenomenon including stock price movement.

One such attempt is by Chenowith [7] they relied on a single technical indicator called the average direction index (ADX), which identifies and quantifies trends by averaging the fraction of today's range above or below the previous day's range. The ADX is obtained through a feature selection component and used as input into two separate neural networks (Up and Down) whose results were then polled and applied to a rule base to predict the final market movement. Without knowing the exact predictive accuracy, it is difficult to quantitatively compare the system, which inevitably includes rules for trading which may be the actual cause of the

monetary gain achieved by the system rather than predictive accuracy.

Variants on the MLP theme are common, such as using a variety of different markets for training data [8] and using statistical analysis to assist in training [9].

Other approaches include using fuzzy neural networks, which is based on fuzzification, a technique that allows members of a set to have degrees of membership instead of binary membership as in classical set theory. The introduction of ambiguity allows data to be divided into several different technical indicators if necessary, hopefully producing more lenient and more accurate neural networks. However, results indicate that fuzzy neural networks have prediction accuracy virtually equivalent to classical neural networks [10].

Mahfoud and Mani [11] chose to use genetic algorithms to predict stock prices. Genetic algorithms are inspired by evolutionary biology and the concept of survival of the fittest. A large population of possible algorithmic representations for stock prediction is first created. Then, each member is executed and evaluated, keeping the algorithms which generate the best results and mixing their properties amongst other high scoring algorithms to obtain a new generation of algorithms in a Darwinian fashion. The process is repeated until the error has been reduced to an acceptable level.

In [12], review of data mining applications in stock markets was presented. [13], used a two-layer bias decision tree with technical indicators feature to create a decision rule that can make recommendations when to buy a stock and when not to buy it. [14] Combined the filter rule and the decision tree technique for stock trading. [15] Presented a hybrid model for stock market forecasting and portfolio selection.

IV. PROPOSED SCHEME

The proposed Efficient Prediction of Close Value using Genetic algorithm based horizontal partition decision tree The proposed methods is implemented using genetic Algorithm which includes the concept of decision tree. The scheme provides Stock market price index prediction is regarded as a challenging task of the finance. Genetic algorithms (GAs) are problem solving methods (or heuristics) that mimic the process of natural evolution. Unlike artificial neural networks (ANNs), designed to function like neurons in the brain, these algorithms utilize the concepts of natural selection to determine the best solution for problem.

Algorithm 1 : Genetic Algorithm

```
for all members of population
    sum += fitness of this individual
end for
for all members of population
```

```

probability = sum of probabilities + (fitness / sum)
sum of probabilities += probability
end for
loop until new population is full
do this twice
  number = Random between 0 and 1
  for all members of population
    if number > probability but less than next
probability then
  you have been selected
end for
end
create offspring
end loop

```

Algorithm 2: Horizontal Partitioned Based Decision Tree

- Define P_1, P_2, \dots, P_n Parties.(Horizontally partitioned).
- Each Party contains R set of attributes A_1, A_2, \dots, A_R .
- C the class attributes contains c class values C_1, C_2, \dots, C_c .
- For party P_i where $i = 1$ to n do
- If R is Empty Then
- Return a leaf node with class value
- Else If all transaction in $T(P_i)$ have the same class Then
- Return a leaf node with the class value
- Else
- Calculate Expected Information classify the given sample for each party P_i individually.
- Calculate Entropy for each attribute (A_1, A_2, \dots, A_R) of each party P_i .
- Calculate Information Gain for each attribute (A_1, A_2, \dots, A_R) of each party P_i
- End If.
- End For

- Calculate Total Information Gain for each attribute of all parties (TotalInformationGain()).
- $A_{BestAttribute} \leftarrow \text{MaxInformationGain}()$
- Let V_1, V_2, \dots, V_m be the value of attributes. $A_{BestAttribute}$ partitioned P_1, P_2, \dots, P_n parties into m parties
- $P_1(V_1), P_1(V_2), \dots, P_1(V_m)$
- $P_2(V_1), P_2(V_2), \dots, P_2(V_m)$
- \vdots
- \vdots
- \vdots
- $P_n(V_1), P_n(V_2), \dots, P_n(V_m)$
- Return the Tree whose Root is labelled $A_{BestAttribute}$ and has m edges labelled V_1, V_2, \dots, V_m . Such that for every i the edge V_i goes to the Tree
- NPPID3($R - A_{BestAttribute}, C, (P_1(V_i), P_2(V_i), \dots, P_n(V_i))$)
- End.

Algorithm 3: TotalInformationGain() - To compute the Total Information Gain for every attribute.

- For $j = 1$ to R do {Attribute A_1, A_2, \dots, A_R }
- Total_Info_Gain(A_j) = 0
- For $i = 1$ to n do {Parties P_1, P_2, \dots, P_n }
- Total_Info_Gain(A_j) = Total_Info_Gain(A_j) + Info_Gain(A_{ij})
- End For
- End For
- End.

Algorithm 4: MaxInformationGain() – To compute the highest Information Gain for horizontally partitioned data

- MaxInfoGain = -1
- For $j = 1$ to R do {Attribute A_1, A_2, \dots, A_R }
- Gain = TotalInformationGain(A_j)
- If MaxInfoGain < Gain then
- MaxInfoGain = Gain
- $A_{BestAttribute} = A_j$
- End If
- Return ($A_{BestAttribute}$)
- End For
- End.

V. PERFORMANCE MEASURES

$$MSE = \frac{\sum_{i=1}^n (A_i - F_i)^2}{n}$$

$$MAE = \frac{\sum_{i=1}^n |A_i - F_i|}{n}$$

$$MAPE = \frac{\sum_{i=1}^n |A_i - F_i| / A_i}{n} * 100$$

Where,
 A_i is the actual value of sample i
 F_i is a predicted value of sample i
 n is the number of samples

VI. CONCLUSION

The genetic algorithm can be used as an application to predict the close values in the stock market data and the performance factor of this algorithm is much better than ANN. Here proposed algorithm is based on the integration of genetic algorithm and the horizontal partition based decision tree and compare the performance factor of the proposed algorithm as compared to the existing decision tree and the proposed algorithm performs better for the stock market close value.

REFERENCES

[1] Mahdi Pakdaman Naeini, Hamidreza Tareimian, Homa Baradaran Hashemi “Stock Market Value Prediction Using Neural Networks”, International Conference on Computer Information Systems and Industrial Management Applications (CISIM), 2010, pp. 132-136.

[2] Szu Han Chang, Michael Mallon “Predicting the Stock Market through Preprocessing Historical Data via an Expert System as Input to a Time-Delay Artificial Neural Network”, Master Thesis Boston University Department of Computer Science April 2008.

www.szuhanchang.com/bucs/cs900/Masters%20Paper.pdf

[3] Shweta Tiwari, Rekha Pandit, Vineet Richhariya "Predicting future trends in stock market by decision tree rough-set based hybrid system with HHMM", International Journal of Electronics and Computer Science Engineering, ISSN- 2277-1956, Volume 1, Number 3, 2012, pp. 1578 – 1587.

[4] Definition of decision tree from answers.com: <http://www.answers.com/topic/decision-tree>

[5] T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural network," Proceedings of the International Joint Conference on Neural Networks, 1990, pp. 1-6.

[6] Fischer Black and Myron Scholes "The Pricing of Options and Corporate Liabilities" The Journal of Political Economy, Vol. 81, No. 3, May - Jun 1973, pp. 637-654.

[7] Chenoweth, Tim., et al. "Embedding Technical Analysis into Neural Network Based Trading Systems", Applied Artificial Intelligence, vol. 10, 1996, pp. 523-541.

[8] Roman, J. and Jameel, A. "Back propagation and recurrent Neural Networks in Financial Analysis of Multiple Stock Market Returns", Proceedings of the 29th Annual Hawaii International Conference on System Sciences, vol. -2, 1996, pp. 454-460.

[9] Wu, S. and Lu, R. P. "Combining artificial neural networks and statistics for stock market forecasting", Proceedings of the 1993 ACM conference on Computer, 1993, pp. 257 – 264.

[10] Rast, Martin "Improving Fuzzy Neural Networks using input Parameter Training", IEEE Fuzzy Information Processing Society - NAFIPS, 1998, pp. 55 – 58.

[11] Sam Mahfoud and Ganesh Mani "Financial Forecasting using Genetic Algorithms" Applied Artificial Intelligence, vol. 10, 1996, pp. 543- 565.

[12] Setty, Venugopal D., Rangaswamy, T.M. and Subramanya, K.N. "A Review on Data Mining Applications to the Performance of Stock Marketing", International Journal of Computer Applications, vol. 1, No. 3, 2010, pp.33-43.

[13] Wang, J-L., and Chan, S-H., "Stock market trading rule discovery using two-layer bias decision tree", Expert Systems with Applications, vol. 30, 2006, pp.605–611.

[14] Wu, M-C., Lin, S-Y., & Lin, C-H., "An effective application of decision tree to stock trading", Expert Systems with Applications, 31, 2006, pp.270–274.

[15] Huang, K.Y, and Jane, C.-J., "A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories", Expert Systems with Applications, 36, 2009, pp.5387–5392.

AUTHORS PROFILE

Dharamveer Sisodia

Mtech (Computer Science) Scholar
SR Group of Institution, Jhansi (UP),
Ph. +918878664378 , 08989004945
Email: dharamveersisodia2007@gmail.com

Biography: Dharamveer Sisodia has received MCA (Master of Computer Application) degree from Technocrats Institute of Technology, Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal (MP), India in 2007. He is associate member of Computer Society of India (CSI). His subjects of Interest include Computer Graphics, DBMS, and Networking. He has present one research paper in national conferences.



Beerendra Kumar

Assistant Professor (CS/IT),
Computer Science and Information Technology,
SR Group of Institution, Jhansi (UP),
Ph. +91 9918690789, 08005072495
Email: beerucsit@gmail.com

Biography: Beerendra Kumar has received B.Tech. (Bachelor of Technology) degree in Computer Science and Information Technology from Institute of



Engineering and Technology, Rohilkhand University, Bareilly (U.P), India in 2006. He has completed his M.Tech. (Master of Technology) in Computer Science from SCS, Devi Ahilya University, Indore, India in 2008. He is associate member of Computer Society of India(CSI). He has five years of teaching experience. His subjects of interest include Computer

Networking, Theory of Computer Science, Data Mining, Operating Systems and Analysis & Design of Algorithms. He has published five research papers in national conferences and four research paper in international journal. His research areas are Computer Networks, Data Mining, Secure Multiparty Computations and Neural Networks.

Jitendra Kumar Gupta

Assistant Professor (CS/IT),
Computer Science and Information Technology, SR Group of Institution,
Jhansi (UP),
Ph. +9109598335466
Email: jitendra1503@gmail.com

Biography: Jitendra Kumar Gupta has received Bachelor of Technology (B.Tech) in Information Technology from Institute of Engineering &



Technology (I.E.T.), Lucknow (State Govt. College of UP) in 2007. He has completed his Master of Technology (M.Tech) in Information Security from Maulana Azad National Institute of Technology (MANIT) Bhopal in 2009. He has Pursuing Doctor of Philosophy (Ph D) in Computer Science & Engineering. He is Associate member of Computer

Society of India (CSI). He has five years of teaching experience. His subjects' of interest include Data Networks, cryptography, Engineering and testing structured systems and Process Engineering. He has published five research papers in International conferences. His research areas are Computer Networks, Secure Multiparty Computations and Neural Networks, Network Security

Dr. Saurabh Shrivastava

Sr Lecturer
Department of Mathematical Sciences & Computer Applications
Bundelkhand University Kanpur Road, Jhansi (UP) 284 128 INDIA
Ph No + 91-510-2320496 (O)+9194155 04462
E mail : hanu.saurabh@gmail.com

Biography: Saurabh Shrivastava has received Post Graduate in Computer Applications (MCA) from Madhav Institute of Technology & Sciences, Gwalior in (1998). He has completed his Ph D (Computer Applications) from Bundelkhand University, Jhansi (2009). He is Member Board of Studies of the Department of Mathematical Sciences and Computer Applications. and Executive Council, Bundelkhand University, Jhansi and also Subject Expert/Board of Studies of Department of Statistics and Computers, Laxmi Bai National University of Physical University, Gwalior . He has Thirteen years of teaching experience.

Lecturer's Homepage Usage in Indonesian Private University

Differences based on gender and educational background

Dessy Wulandari AP

Department of Informatics Management
Gunadarma University
Depok, Indonesia

Abdus Syakur

Department of Informatics Engineering
Gunadarma University
Depok, Indonesia

M. Achsan Isa Al Anshori

Department of Informatics Management
Gunadarma University
Depok, Indonesia

M. Akbar Marwan

Department of Information Systems
Gunadarma University
Depok, Indonesia

Abstract— The use of personal homepage by lecturers is getting popular in universities, especially to support teaching and learning process in the classroom. This study observed a lecturer in Information Systems Department in Indonesian Higher Education. Only 55.53% of the 151 lecturers who actively use the personal homepage. The amount of content and the popularity of the personal homepage on average still lower than the number of web pages viewed, number of documents, referring domain, and total backlinks. Statistical test results showed that there were differences in the number of web pages, the number of documents, and the total backlinks between male and female lecturers. There are significant differences occur only in the total number of web pages and documents, while referring domain and total backlinks shows no significant difference seen from the education lecturer.

Keywords- Lecturer's Homepage; Indonesian Private University; Differences based on gender and educational background

I. Introduction

The rapid development of the internet has prompted changes in various sectors of life, including higher education. According to [1] the use of the internet for online education in the university consists of three phases, that are the experimental stage in the period from 1970 to the mid-1990s; stages of innovation until the turn of the 20th century, as well as stage of systematization and increasing scale until today. However, these steps may be different views of the respective universities, which depend on the ability or willingness of the university, including the various of learning technology facility in their respective colleges.

The convergence of the internet technologies and education to encourage the emergence of new concepts or terms such as e-learning, virtual class, or a variety of other learning technologies. One of the tools that can be utilized by lecturer is a personal website that can be used by lecturer for media information, including uploading learning materials or tasks for students. According [2] the use of web-based for learning management system as a support tool in higher education has become so popular with the rapid development of information technology. Is the use of web-based learning system has been implemented uniformly by lecturers in college? The research questions are at the core of this research is more focused on the use of a personal website.

Internet technology become enabled that allows teaching-learning process is not limited to space and time. The Internet has transformed higher education as raised by [3] that Information technology can be used to transform education. However, higher education transformation need to be initiatives supported by lecturer in using learning of technology facility that are already provided by the university. Lecturer as a key player in the learning process is not necessarily optimally utilize learning of technology.

This study aims to look at the utilization rate of personal website by lecturers in Department of Information Systems at one of the universities in Indonesia. Performance utilization of the website is viewed from several parameters: the number of web pages and the number of documents on the website of lecturers, and the popularity of a website is measured by the number of referring domains and the total

backlink Statistical test was also used to see the difference parameters of personal homepage viewed from gender and level of education.

II. State of The Art

[4] Stated that higher education has been a rollercoaster ride over the past few decades with the advent of the first personal computers in the 1980s and the Internet in the 1990s. According to [1] the new directions are needed which will allow us to make the selection of technology and pedagogy for future education which is more suitable for the public. The use of information technology to support teaching and learning become ubiquitous in higher education [5].

Pedagogical implications of open sharing that are embedded in almost every aspect of the conversion process of teaching materials [6]. Stated that previous researchers have reported that this new web-based training technology (which has its basis in computer-based training) have not been integrated the sound pedagogical practices into the authoring process when developing new tutorials [7].

[8] Stated that although IT has played an increasingly important role in contemporary education, IT remains significant resistance to the education field. An online instructor faces the challenge of handling a variety of intelligence to provide a quality learning experience for all types of learners [9]. Stated that learning facilitated and supported through ICT is not possible to make sense in the context of developing country [10].

According to [11] the difference in students' perceptions of e-learning tool among students who chose to take online courses and students who preferred to take part based campus. [12] stated that WBLT alone without other strategies, is not guaranteed to have a positive effect on student result. Many teachers were initially a bit skeptical about the web resources to be convinced of their potential in the taching and learning process [13].

According to [13] initial attempts have been made to use the framework for training academics in public and private Malaysian universities in the use of web resources for teaching and learning.

[14] Stated that perceived usefulness is a significant influence on attitudes toward the use of computers and behavioral intentions. Computer self-efficacy has a very positive effect on perceived ease of use and intention to use the teachers' acceptance of the

technology teacher [15]. According to [16], Overall, the majority of teachers rated Web Based Learning Tools (WBLT) as an easy to use and attractive to students studying success and promoted. [17] Stated that performance expectations, attitudes toward using technology, facilitating conditions, self-efficacy, and social influence have significant influence on behavioral intention in the use of web-based learning.

[18] Stated that it is the responsibility of the teacher to design properly for that the required learning result can be met. According to [2] most of the instructors techniques have a high awareness of the importance of web-based learning management system. The main interest in the correlated decision to adopt e-learning is the faculty confidence in using technology [9].

III. Methodology

The number of lecturers were observed in this study as many 151 lecturers enrolled in Information Systems Department. Observations on the first stage is to gather lecturer who has been using a personal website provided by Gunadarma University. Then, measured the amount of content and the popularity of any website. Total content consists of two indicators: the number of web pages, and number of documents indexed by Google. The format of the document consists of pdf, doc/docx and ppt/pptx. Popularity is measured by the number of referring domains and the total backlinks using majesticseo.com. To reduce fluctuations in results, the observations were made in the same week is the last week of January 2013. This study also presents the popularity of a website based on traffic monitoring system administered by the university.

Analysis of the data are more descriptive to see the patterns of using website seen from several aspects, such as gender, level of education, ladder, or other characteristics of the individual lecturers. The difference between male and female lecturer in the amount of content and the popularity of websites tested using independent sample t test, while differences in educational level or ladder tested using Analysis of Variance (ANOVA).

IV. Result and Discussion

A. Lecturer's Homepage Overview

Homepage of lecturer is used to display the profile lecturers lecturers such as level of education, teaching experience, research interests, work unit, or other personal information relating to the duties and role of the lecturer. The main menu consists of menu "Download" for the lecture material; "News & Links" for news or information from lecturer, especially for

students, including links to learning resources; "Publication" for a list of research and publications of teacher, as well as "Blog" is link to a blog address that is owned by the lecturer. Example of homepage of lecturer can be seen in the figure below.

Homepage of lecturer including teaching facilities are popular with students at Gunadarma University, a private university in Indonesia. Popularity can be seen from the usage per day that is seen from the traffic monitoring system, as shown in the figure below for daily usage and hourly usage.

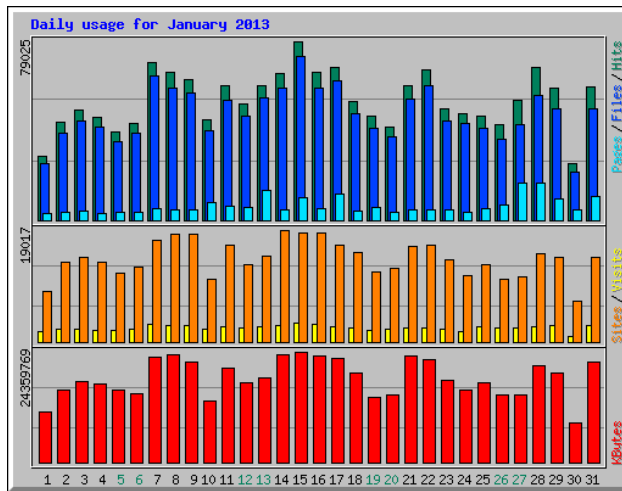


Figure 1. Daily usage for Januari 2013

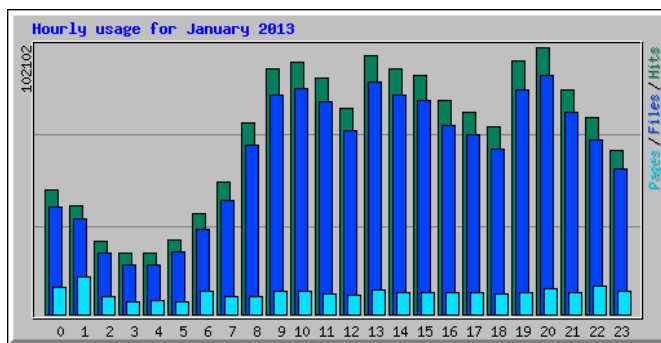


Figure 2. Hourly usage for January 2013

Homesite of lecturers including web-based teaching facility of the most visited websites other than subdomain of Gunadarma University, for example, department or library website. The number of visits or traffic to the homesite of lecturer is high, as shown in the table below.

TABLE 1. Monthly Statistics for January 2013

Total Hits	1648771	
Total Files	1459467	
Total Pages	198174	
Total Visits	73636	
Total Kbytes	575028162	
Total Unique Sites	214393	
Total Unique URLs	50468	
Total Unique Referrers	14147	
Total Unique Usernames	11	
Total Unique User Agents	23355	
	Avg	Max
Hits per Hour	2216	10129
Hits per Day	53186	79025
Files per Day	47079	72066
Pages per Day	6392	16286
Sites per Day	6915	19017
Visits per Day	2375	3221
KBytes per Day	18549296	24359769

Popularity of homesite from the perspective of students because of material lecture can be downloaded from the website. Lecturer rarely deliver course material in the form of paper or hardcopy. Moreover, students are relatively literate internet and prefer to take the course material is not restricted space and time in the classroom. The condition is in accordance with the statement of [17] which states that "The attitude of students have a positive impact on both behavioral intention and actual use of web-based learning system" However, popularity or student activity in visiting lecturers website does not guarantee academic success of teachers, as raised by [12] that our results show that whenever WBLT provided no positive effect on the overall exam performance. This research has not proven that statement, but more focused on the activities of homesite of lecturer, including comparison between male and female lecturer and education levels or other lecturer profile.

B. Lecturer's Homepage Productivity and Popularity

The number of lecturers who already take advantage of personal website as many as 84 people or 55.63%. The amount is relatively low when considering the program from lecturer of information systems. Lecturers have a background and expertise in the field of computer, so it can be a motivating factor in the use of technology by lecturer, as said by [8] that This study suggests that computer self-

efficacy has a great influence on the acceptance of technology to the teachers. Educational background was not a factor in the utilization of information technology in teaching and learning process, as raised by [9] that the academic background variables did not produce significant correlations with perception, and the decision to adopt e-learning.

The utilization level of homesite of lecturer is low depends socialization and institutional support in the use of the website to support the teaching and learning process provided by the lecturer. Until now, the use of the personal website is voluntary or private institutions do not require lecturer to use. This factor may be one low cause of personal website of utilization level by the lecturer. Another factor is the availability of the time and effort to complete and update the website. According to [16], a number of teachers noted that significant time was spent looking for the right WBLTs and preparing lessons.

Contribution number of web pages as 1927 with 1228 of documents number with details; 701 of the pdf document, 257 doc/docx, and 270 ppt/pptx. When compared with the number of lecturers as 151 people, content of personal website is very low with an average per person was 13 webpage and 5 documents per person. Distribution of web pages and documents were uneven, as shown in the following graph density.

Popularity of lecturer's website is generally less popular lecturer when seen from the number of referring domains and the total backlinks to the total as many as 4445 for referring domain and 5469 for total backlink or an average per person 29 and 36. Male lecturer is superior than female in the four parameters as shown in the figure below.

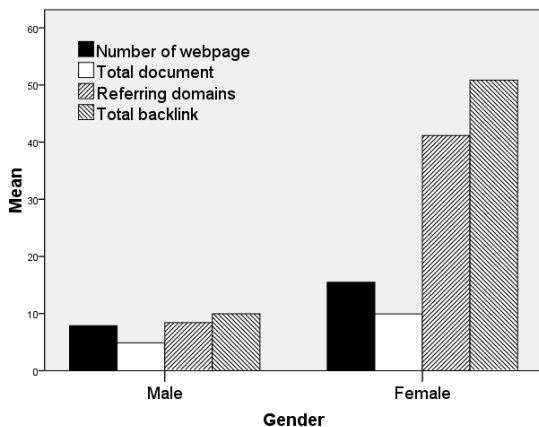


Figure 3. Gender differences in website use

Results of independent samples t test showed significant difference between male and female

lecturer for the number of webpages, total documents and total backlinks. [19] Stated that male Turkish prospective teachers were found to have more problematic use of the Internet than their female counterparts on the other hand no differences were found among South Korean teacher through sex.

Lecturers with doctoral education indicates the number of web pages, total documents, and total backlinks higher than lecturers with a Master degree referring just superior to the parameters of referring domain. Differences of website parameter by level of education can be seen in the figure below.

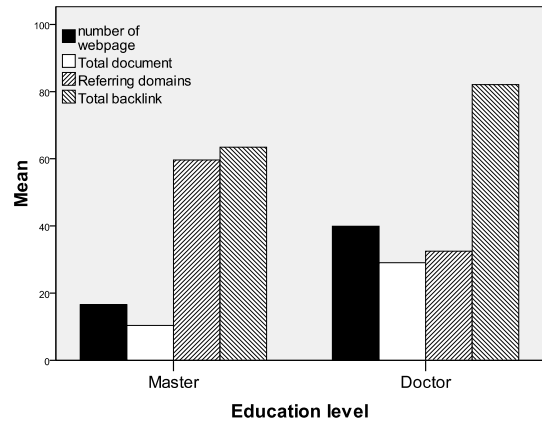


Figure 4. Website Performance based on educational level

Lecturer Doctoral more productive in providing content, either a web page or document format doc, ppt, or pdf. Observations based on the popularity of the website showed that doctoral lecturer better on the number of links, but referring domain lower than master lecturer. Difference performance of personal website based on the level of education can be seen from the ANOVA test results in the table below. The test results show that there are differences number of web pages and document number that significant seen from education level of lecturer, while referring domain and total backlinks do not show significant differences.

Tabel I. Anova

		Sum of Squares	df	Mean Square	F	Sig.
Size	Between Groups	14559.780	3	4853.260	10.700	.000
	Within Groups	66675.638	147	453.576		
Total		81235.417	150			

Totaldoc	Between	7860.563	3	2620.188	11.095	.000
	Groups					
	Within Groups	34714.788	147	236.155		
	Total	42575.351	150			
Referring Domains	Between	117000.389	3	39000.130	1.555	.203
	Groups					
	Within Groups	3687300.764	147	25083.679		
	Total	3804301.152	150			
Backlinks	Between	150383.647	3	50127.882	2.097	.103
	Groups					
	Within Groups	3513670.141	147	23902.518		
	Total	3664053.788	150			

v. Conclusion

The use of personal homepage in private universities is the object of this study is still voluntary. Lecturer in Information Systems Department are not required to use the personal homepage as a one of media information and individual publications to support the teaching-learning process in the conventional classroom. The voluntary alleged causes of the low level of utilization of personal homepages by lecturers. Nevertheless, personal homepages are not very useful from the perspective of the interests of students that can download course material, look for information or announcements from lecturer.

The amount of content and the popularity of personal homepage on average still low and showed a high diversity of the number of web pages viewed, total documents, referring domain, and total backlinks. Only few personal homepage that shows productivity and popularity, or uneven from population of lecturer that observed. These conditions indicate that the lecturer has not fully utilized the personal homepage as a personal branding for the benefit of its duties and functions as a lecturer in the university.

Utilization of personal homepage still shows a significant difference in views of gender and level of education. Female lecturer performing utilization of personal homepage higher for all parameters than male lecturer. The doctoral lecturer indicates the number of pages, total documents, and total backlinks higher than the master lecturer. Significant

differences in educational level only in the number of web pages, and number of documents.

References

- [1] Matthew Allen and Jane Long. Learning as Knowledge Networking: Conceptual Foundations for Revised Uses of the Internet in Higher Education. Proceedings of the World Congress on Engineering and Computer Science 2009 Vol I, WCECS 2009, October 20-22, 2009, San Francisco, USA.
- [2] Duygu Findik and Sevgi Özkan. Identifying Success Factors for WBLMS Use by Instructors of Engineering Departments. 40th ASEE/IEEE Frontiers in Education Conference, October 27 - 30, 2010, Washington, DC.
- [3] Dezhi Wu, Starr Roxanne Hiltz, and Michael Bieber. Acceptance of Educational Technology: Field Studies of Asynchronous Participatory Examinations. Communications of Association for Information Systems, Volume 26, Article 21, pp. 451-476, April 2010.
- [4] Ghulam Muhammad Kundi and Allah Nawaz. From objectivism to social constructivism: The impacts of information and communication technologies (ICTs) on higher education. Journal of Science and Technology Education Research Vol. 1(2), pp. 30 - 36, July 2010.
- [5] J. Coldwell, A. Craig, T. Paterson and J. Mustard. "Online Students: Relationships between Participation, Demographics and Academic Performance." *The Electronic Journal of e-Learning* Volume 6 Issue 1, pp. 19 - 30, 2008.
- [6] Monica Mawoyo and Neil Butcher. Sharing Existing Teaching Materials as OER: Key Considerations from Practice. Opening Up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge, Edited by Toru Iiyoshi and M. S. Vijay Kumar. The Carnegie Foundation for the Advancement of Teaching, The MIT Press, Cambridge, Massachusetts, 2008.
- [7] Thomas Janicki and Jens O. Liegle. Development and Evaluation of a Framework For Creating Web-Based Learning Modules: a pedagogical and systems perspective. JALN Volume 5, Issue 1, May 2001.
- [8] Min Gong, Yan Xu, and Yuecheng Yu. An Enhanced Technology Acceptance Model for Web-Based Learning. *Journal of Information Systems Education*; Winter 2004; 15, 4, pg. 365.
- [9] Mark Riha and Rebecca A. Robles-Piña. The Influence of Multiple Intelligence Theory on Web-Based Learning. MERLOT Journal of Online Learning and Teaching Vol. 5, No. 1, March 2009.
- [10] Z.T. Khalil, S. Kabir, and N.M. Patwary. Acceptance and Effectiveness Of E-Learning in a Developing Economy: Perspectives From Bangladesh. Proceedings of ICERI2012 Conference 19th-21st November 2012, Madrid, Spain.
- [11] Edith Galy, Clara Downey, and Jennie Johnson. The Effect of Using E-Learning Tools in Online and Campus-based Classrooms on Student Performance. Journal of Information Technology Education, Volume 10, 2011, p.209-230.
- [12] But, Jason; and Shobbrook, Rhys. 2012. Web-based Lecture Technologies and their Effects on Student Performance. Proceedings of the 2012 AAEE Conference, Melbourne, Victoria, Copyright © But, Jason; and Shobbrook, Rhys, 2012.
- [13] John Arul Phillips. Experiences in Finding and Using OER in Teacher Education Programmes: Pedagogical Approach and Challenges. Opening Up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge, Edited by Toru Iiyoshi and M. S. Vijay Kumar. The Carnegie Foundation for the

- Advancement of Teaching, The MIT Press, Cambridge, Massachusetts, 2008.
- [14] Kung-Teck, Wong; Rosma bt Osman; Pauline Swee Choo, Goh; and Mohd Khairezan Rahmat. Understanding Student Teachers' Behavioural Intention to Use Technology: Technology Acceptance Model (TAM) Validation and Testing. *International Journal of Instruction*, January 2013, Vol.6, No.1, p.89-104.
- [15] Jui-Ling Chang, Pang-Tien Lieu, Jung-Hui Liang, Hsiang-Te Liu, and Seng-lee Wong. A causal model of teacher acceptance of technology. *Educational Research and Review* Vol. 7(5), pp. 102-110, 5 February, 2012.
- [16] Robin Kay, Liesel Knaack, and Diana Petrarca. Exploring Teachers Perceptions of Web-Based Learning Tools. *Interdisciplinary Journal of E-Learning and Learning Objects* Volume 5, 2009.
- [17] Din Jong and Tzong-Song Wang. Student Acceptance of Web-based Learning System. *Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA'09)*, Nanchang, P. R. China, May 22-24, 2009, pp. 533-536.
- [18] Benson, Robyn and Samarawickrema, Gayani 2007, Teaching in context: some implications for e-learning design, in *ICT: providing choices for learners and learning*, ASCILITE, Singapore, pp. 61-70.
- [19] Aylin TUTGUN, Levent DENİZ, and Man-Ki MOON. A Comparative Study Of Problematic Internet Use And Loneliness Among Turkish And Korean Prospective Teachers. *TOJET: The Turkish Online Journal of Educational Technology* – October 2011, volume 10 Issue 4.
- [20] Ali Kamali. Antecedents of Adopting e-Learning: Toward a Model of Academic e-Learning Acceptance Culture. 2012 *Proceedings of the Information Systems Educators Conference*, New Orleans Louisiana, USA, v29 n1907.

COMPARATIVE SURVEY OF STEGANOGRAPHY TECHNIQUES

Rhythm walia, Om Pal

*Centre for development of advanced computing
Noida, India*

*Centre for development of advanced computing
Noida, India*

Abstract— Steganography in its core is about hiding message in such a manner that it is invisible to any inter-mediate party. Being undetected is the most important trait for any steganography technique. To be invisible a steganography technique should produce minimum distortion in the cover image. This paper contains a through description of the techniques and also a comparison between different steganography techniques for their security against various attacks.

I. INTRODUCTION

Steganography represents the art of covert communication. It embeds a confidential message into media called as a cover image. The goal is to embed in such a way that it reveals nothing neither the embedding nor the embedded message. While developing steganography systems it is very important to consider what the most appropriate cover work will be and how the stegogramme will reach its intended recipient. In modern life steganography has wide range of applications ranging from modern colour printers to digital watermarking and many other activities. The entire process of steganography is shown in fig (1). The process consists of two algorithms one for embedding and other for extracting

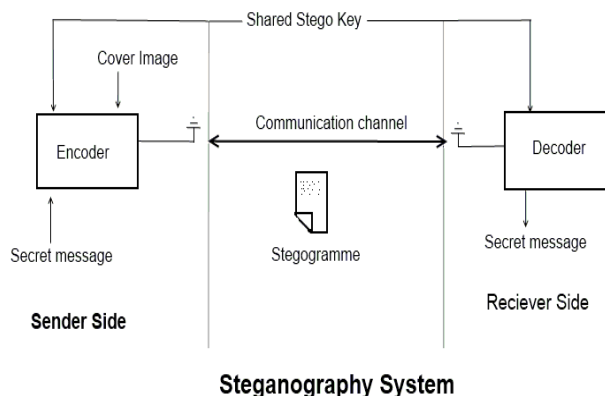


Figure 1.

Two inputs are required for embedding process:

- Secret message: usually a text file containing the message to be transferred, it can also be an image or an audio video clip.
- Cover work: usually an image used to construct stegogramme that contains secret message.

Now in the next step the secret message will be embedded in the cover work with as little distortion as possible using the stegosystem encoder. We can also use a key with stegosystem encoder to enhance security. The output of stegosystem encoder called stegogramme and the key are given to the stegosystem decoder where an estimate of the secret message is extracted. Steganalysis is the art of identifying stegogrammes that contain a secret message. It begins by identifying distortions that exist in the suspect file as a result of embedding message.

II. Basic steganography techniques

In order to understand how steganography works and to know about the magnitude of known embedding techniques we will discuss some steganography techniques and compare them on various parameters. We will learn that a simple algorithm is very easy to break while a cleverer algorithm is difficult to break.

A. LSB Substitution

Here the message data is inserted in the least significant bit (lsb) of the image [1]. Least significant bit represents the smallest bit in the binary sequence. Binary numbers can have only two values zero or one. If the message bit is 1 and lsb is 1 there is no change but if the message bit is 1 and lsb is 0, lsb is changed from 0 to 1. Changing the lsb does not have a huge impact on the final number as it is changed only by 1. If we consider 8 bit binary sequences and use them to represent the colour of the pixel of the image then it is clear that the colour will be change by +1 only which is very difficult to be noticed by naked eye. In fact the lsbs of each pixel value could be modified and the change will be still not be visible. So there is a huge amount of redundancy in the image data and each lsb pixel of image data can be changed to message data until the message is complete.

B. LSB matching

It is an improvement over LSB substitution technique [1]. It works by lowering the detection rate for some steganalysis. Even means 0-bit, odd means 1-bit. So, we simply add or subtract '1' from colour value to make it odd or even, depending on message bit. Here is a random distribution of value means a bit can take a value preceding it or following it by adding or subtracting a 1 respectively. If the value of message bit is 1 and the lsb bit is 0 we add a 1 to the lsb bit to make it 1 and we see that that the overall number gets increased by 1 otherwise we subtract a 1 from the lsb and the overall number gets decreased by 1. So we see that there is a 50% probability of change. We see that there is more randomisation in lsb matching (lsbm) as compared to lsb substitution.

C. EzStego

This technique [2] operates only on indexed image format like gif and png. For these image formats, each pixel value acts as an index to one of several colours in a pre-determined palette. For GIF images, each pixel is a single byte of information meaning that there are 256 possible colours for the image. We can calculate the number of colours available by saying that i -bit image gives 2^i colours in the palette. The plan for steganography was to perform an LSB style embedding approach in a similar fashion to the Hide & Seek method. However, as we have seen, this strategy either increments the entire pixel value by 1 or keeps it the same, which produces a problem for palette-based images. This is because the palette is not ordered in any particular manner, so indexing value 114 might produce a sky blue colour, where as the value 115 might produce a deep red colour. EzStego reordered the

original palette such that similar colours are placed next to each other before embedding. The embedding process works line-by-line on the image, and embeds the LSB of the sorted index within the pixel values according to the message bit stream.

D. Edge Based Steganography

This technique [3] is Edge adaptive based on lsb matching revisited embedding (EALMR). Lsb matching revisited (LSBMR) is an improvement over LSBM as it lowers the modification rate by embedding data as correlation between two pixels, two bits per pixels. In EALMR data embedding is done at the edges of an image. Edges are the locations in the image, where there is sharp change in visual property of the image. The technique take difference between two adjacent pixels as the parameter to define edge. If the difference is greater than a threshold, then both the pixels are taken as edge pixels. The technique is edge adaptive, as it chooses strong or smooth edges depending on the length of secret message.

III. Introduction to various comparison parameters

In this paper, various steganography techniques are compared on various parameters. These parameters measure capacity, image quality and their robustness against various attacks. A good technique is one who gives performs overall better. Following are the all comparison parameters.

- **Modification per pixel:** It is the Measure of how much data is hidden on an average per pixel. We want less modification to happen to embed a given data, so lesser is better. Unit is bits per pixel (bpp).

Formula: $(No. \text{ of pixel modified} \times \text{bits modified per pixel}) \div \text{total no. of stego pixel}$

- **Embedding Capacity:** It means how much data can be hidden inside an image. Its value should be high as we want more data to sent with a single image
- **PSNR Value:** It is the peak signal to noise ratio relative to cover and stego image. It is the most widely used parameter to check the image quality. High PSNR value is preferred. A value above 50 is considered good. If the two images are same, then PSNR value is infinite.

E. Steganalysis Attacks

Robustness against various steganalysis techniques is also a scale to rate steganography techniques. This is most important property a steganography technique should hold. In this sections, we will briefly explain steganalysis attacks used in this paper to test steganography techniques discussed above.

- **Visual attacks:** Hiding data produces many distortions in the image. Visual attacks check for the distortions visible to human eye. This distortion may be visible in raw stego image or in some other

processed form. One of the processing is to extract LSB plane from the stego image [2]. As most of the embedding takes place in LSB, there will be some distortions in LSB plane. However this approach works only when, there is some texture in the plane and not complete randomness. We will see later how LSB plane can be employed to detect steganography.

- **Structural attacks:** Data embedding changes structural statistics of the image, sometimes causing some patterns in the structural properties. Chi-square [2], Weighted Stego (WS) [4] and Sample pair analysis (SP) [5] are some popular structural steganalysis techniques.]

IV. Experimental Design

All the tests in the paper were carried on BOWS2 database consisting over 10000 images. All the images were grey scale spatial domain saved in *pgm* format. The resolution of all the images is 512X512. LSB Substitution, LSB Matching, PSNR and other helper functions were coded in Matlab. Whereas the code for Ezstego, EALMR, WS and SP was obtained from [6], [7], [8] and [8] respectively. The embedded message was randomly generated by a Pseudo random number generator.

V. Experimental Results

Tests were ran over 10000 stego images to compare various steganography techniques discussed above. Following are the results obtained both experimentally and theoretically.

E. Embedding capacity

For an image of resolution $m \times n$ maximum embedding capacity for LSBM, EzStego, LSB Substitution is $m \times n$ bits, since they all embed 1 bit in each pixel. How EALMR embed two bits in each pixels, so its embedding capacity is double than others.

F. Modification rate

Again for LSB substitution, LSBM and Ezstego is 0.5 bpp. This can be proved as follows:-

Let total no. of stego pixels: S

Probability that the LSB of current pixel is different than message bit: 0.5

Expected no. of modified pixels: $0.5S$
modification per bit (bpp): $(0.5 \times S \times 1) \div S = 0.5$

However since EALMR used LSBMR for embedding, its modification rate is 0.375 bpp. LSBMR makes pair of pixels to hide data. First pixel carries message bit and the odd-even relationship between pixels is chosen such that there is no need for modification in second pixel. In our experiments track of modified pixels were kept to get experimental results. From table (1) we can see that the theoretical values are well supported by experimental results. Note that the above value theoretical guessed is for maximum, i.e. 100% embedding.

G. PSNR Value

To be undetectable, a stego image should be similar to original cover image as much possible. PSNR value is the degree of similarity between two images. Its value is infinite for two exactly similar value. 10000 stego images, each with 10% embedding were compared with their respective cover images and then average PSNR value for each technique was calculated. From table (1) we can see that EzStego performs poor in terms of PSNR, results of LSB substitution is better but LSBM performs excellent. EALMR even being LSBMR based gives lower result than LSBM because of embedding two bits per pixel.

H. Visual Attacks

None of the steganography techniques listed above cause any visual distortion in the stego image which could be notable to human vision system. However since the embedding in done LSB, changes will be more visible in LSB plane. Look at fig (2)(b). This is the plane extracted from cover image. We can clearly see some smooth white and black areas. This belongs to pixels having same LSB. But embedding 1/0 bits in this region will cause inconsistency.

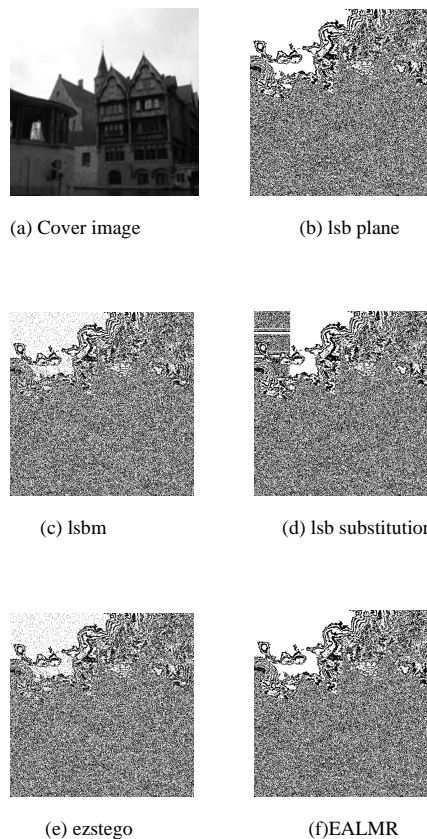


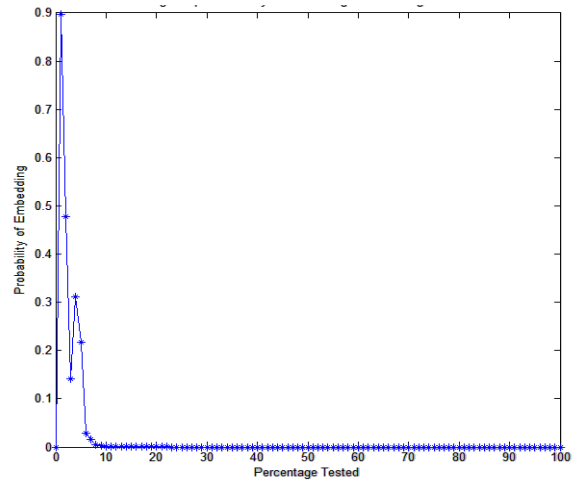
Figure 2. Comparing LSB planes of stego images obtained from various techniques with 10% embedding.

When we look at image (d) we can see a vertical patch across white area in the LSB plane. This is because LSB substitution hides sequentially causing this vertical patch. This clearly gives away the embedding. In case of both LSBM and Ezstego, this embedding is random. Thus these black pixels are spread out over smooth area. However these pixels may raise suspicions. However EALMR being edge adaptive does not hide anything in the smoother regions. Making original and stego LSB plane quite similar.

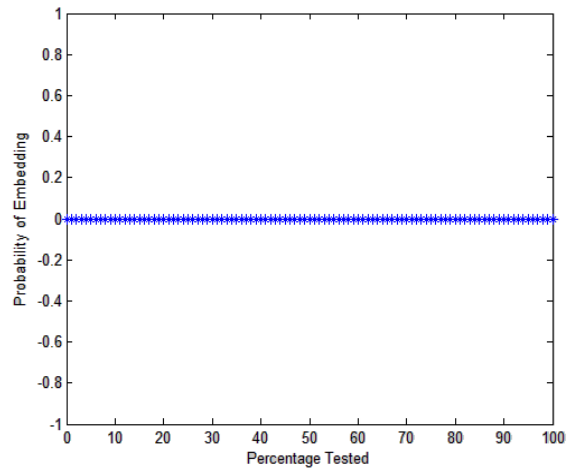
I. Structural Attack

This section covers structural attacks on the stego images. In this paper three structural steganalysis attacks have been covered.

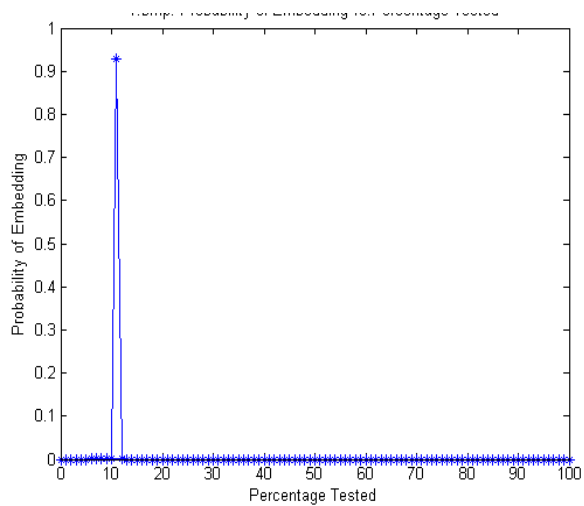
- **Chi-square Attack:** When we overwrite the least significant bits of an image with data from a message bit stream, we transform the values into each other. Let us consider an image where the first pixel value = 166. If we embed a 0, the value will remain the same, whereas if we embed a 1, the pixel value will change to 167. Now let us assume that a pixel value of 167 exists somewhere else within the image. If we embed a 0 now, the value will be decreased to 166, whereas embedding a 1 will leave the pixel value unchanged. This leads to the observation that all adjacent odd and even values can be changed into each other for all the pixels in the image. We get pair of values with fixed frequencies that Chi square can detect and break the stenography technique. See image (3). LSB Substitution and Ezstego with their similar embedding algorithm causes this effect and are detected with high probability by chi-square. LSBM and EALMR prevent this even distribution of adjacent odd-even numbers. Thus makes it impossible for chi-square to detect steganography.
- **Sample pair analysis:** Sample pair analysis make pairs of adjacent pixels and divide them into two classes, one with even value greater and the other with odd value. Now cardinality of both the sets should be same for an ordinary image. But data embedding changes this property in case of LSB Substitution. As we can see in table (1), Sample pair prediction is very close to 10% in case of LSB substitution and Ezstego. However preserve this property, thus SP gives very low results for them



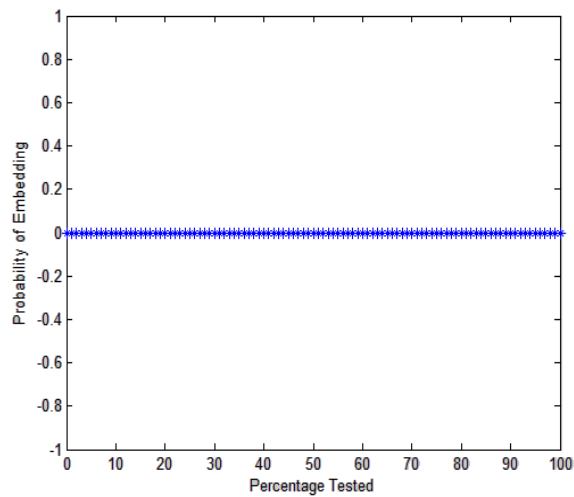
(a) Lsb substitution



(b) LSBM



(c) Ezstego



(d) EALMR

Figure 3. Probability of embedding predicting by chi-square for various techniques

Weighted Stego: Weighted stego try to guess the value of a pixel by using its neighboring pixel values. Difference between both expected and actual value is taken as steganography amount. Table (1) shows that both LSBM and EALMR are robust to WS then LSB Substitution, Ezstego, for which the value however wrong is big enough to raise suspicion. Table 1. Results obtained by running tests of stego images with 100% embedding for modification per pixel calculations and 10% embedding for rest of the parameter.

Comparison Parameters	Lsb Substitution	LSBM	EzStego	EALMR
Modification per pixel (bits per pixel) 100%	0.55	0.48	0.51	<u>0.38</u>
PSNR Value	51.3	<u>79.6</u>	34.4	64.2
Sample Pair (% of stegoed pixels) 10%	9.1	<u>1.2</u>	8.6	1.31
Weighted Stego (% of stegoed pixels) 10%	7.2	0.6	7.3	<u>0.3</u>

VI. Conclusion

LSB Substitution, although most simple technique is by far most insecure steganography technique. It creates distinct distortion in LSB plane. Distortion produced by Ezstego and

LSBM is more spread out, but still noticeable. If one wishes to use LSBM, one should pick cover image with total random LSB Plane. In terms of image quality LSBM produce best quality stego image, where Ezstego gives poor image. But it is still most popular tool for index based steganography. EALMR gives best overall results. It along with LSBM is resilient to structural attacks and produces no distortion in LSB plane. Moreover embedding capacity is double than any of the techniques discussed in this paper and has least modification rate.

REFERENCES

- [1] A. Ker, "Improved detection of LSB steganography in grayscale images" in *Proc. Information Hiding Workshop*, vol. 3200, Springer LNCS, 2004, pp. 97–115
- [2] A. Westfeld and A. Pfitzmann. "Attacks on Steganographic Systems", in *Lecture Notes in Computer Science*, vol. 1768, pp. 61-76, 1999
- [3] Weiqi Luo, Fangjun Huang, and Jiwu Huang, "Edge adaptive image steganography based on lsb matching re-visited," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 201–214, June 2010.
- [4] Andrew D. Ker and Rainer Böhme, "Revisiting Weighted Stego: Image Steganalysis". *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, Proc. SPIE Electronic Imaging*, vol. 6819, San Jose, CA, pp. 0501-0517, 2008.
- [5] Sorina Dumitrescu, Xiaolin Wu, Nasir D. Memon, "On steganalysis of random LSB embedding in continuous-tone images". *Proceedings of ICIP*, 2002, Rochester, NY, pp.641-644.

Single CCTA-Based Four Input Single Output Voltage-Mode Universal Biquad Filter

S. V. Singh

Department of Electronics and
Communication Engineering, Jaypee
Institute of Information Technology,
Sect-128, Noida-201304, India

R. S. Tomar

Department of Electronics
Engineering, Anand Engineering
College, Agra-282007, India

D. S. Chauhan

Department of Electrical
Engineering, Institute of
Technology, Banaras Hindu
University, Varanasi-221005 (India),

Abstract—In this paper a new single current conveyor trans-conductance amplifier (CCTA)-based four-input single-output (FISO) voltage-mode universal biquad filter is proposed. The proposed filter employs only single CCTA, two capacitors and two resistors. The proposed filter realizes all the standard filter functions i.e. low pass (LP), band pass (BP) and high pass (HP), band reject (BR) and all-pass (AP) filters in the voltage form, through appropriate selection of the four input voltage signals. The circuit does not require inverting-type input voltage signal(s) and double input voltage signal(s) to realize any response in the design. The filter enjoys attractive features, such as orthogonal electronic tunability of quality factor and pole frequency, low sensitivity performance. The validity of proposed filter is verified through PSPICE simulations.

Keywords-component; CCTA, Biquad, Universal, Filter

I. INTRODUCTION

In analog signal processing applications high performance continuous time (CT) voltage-mode filters have received a lot of attentions in the recent past because they are widely used in many applications such as video signal enhancement, graphic equalizer in hi-fi systems, dual tone multi-frequency (DTMF) for use in touch-tone dialing in the telephone market, phase locked loop and cross over network used in three way high fidelity loud speaker [1]. Depending on number of input and output signals, CT voltage-mode filters [2-15] can be categorized as single-input single-output (SISO), single-input multiple-output (SIMO), multiple-input multiple-output (MIMO) and multiple-input multiple output (MISO). The SISO filter can realize multi-function filter outputs by altering the connection way of the circuits, but altering the connection way can only realize single filtering output at a time. SIMO filters [2-5] simultaneously realize different filter functions (in general three or more) at different outputs, without changing the connection of the input signal and without imposing any restrictive conditions on the input signal. On the other hand MISO [6-9, 12-14] and MIMO filters [10,11,15] can realize multifunction outputs by altering the way in which input signals are connected. However, such filters can realize multifunction filter outputs when the input signals meet some restrictive conditions. MISO filter has single output and realize only one filter function at a time whereas MIMO filter has at least two outputs and may realize more than one filter functions at a time. In addition, the MISO and MIMO filters in comparison with SIMO filter may lead to a reduction in number of active elements used. In the literature, several voltage-

mode multifunction filters using different current-mode active elements have been proposed. However, most of the proposed circuits have two or more than two active elements [2-10]. Only very few circuits based on single active element are proposed in the literature [11-15]. The single active element based voltage-mode filter circuits [11-12] can realize all the standard filtering functions. The structure [11] realizes three input two output voltage-mode biquad by the use of single CCII, three resistors and two capacitors whereas the structure [12] realizes three input single output biquad filter by the use of single CDBA, four resistors and three capacitors. However, both the circuits require component matching conditions to realize at least two filtering functions. Another valuable single DVCCII based three input single output voltage-mode multifunction filter circuit [13] consists of only two resistors and two capacitors but it realize only three filtering functions (LP, BP and HP). In addition, all these three filters [11-13] also suffer from lack of electronic tunability. On the other hand voltage-mode circuits [14-15] provide the feature of electronic tunability. The CCTA-based four input single output voltage-mode filter[14] realizes all standard filtering functions and consists of two capacitors, two resistors as passive elements but no input is applied to high input impedance terminal in this circuit. Recently reported voltage-mode filter circuit [15] based on new active element as DVCCTA, consists of two capacitors and only one grounded resistors. However, it realizes only three filtering function (HP, LP, BP) by the use of two inputs and three outputs voltage input signals.

In this paper, new FISO voltage-mode biquad filter using single CCTA is proposed. The proposed voltage-mode filter employs two resistors and two capacitors as passive elements and can realize LP, BP, HP, BR and AP filtering functions. The filter enjoys attractive features, such as low sensitivity performance, orthogonal current control of pole frequency and quality factor. Moreover, one of the inputs in this circuit is applied at high input impedance terminal. The validity of proposed filter is verified through PSPICE, the industry standard tool.

II. CCTA AND PROPOSED VOLTAGE MODE FILTER

The CCTA [14,16] is a combination of second generation current conveyor (CCII) and operation transconductance amplifier (OTA). The block diagram of the CCTA is shown in Fig. 1. It consists of two input

terminals (X, Y) and two output terminals (-Z, -O). Port X is low input impedance terminal while port Y is the high input impedance terminal. Port -Z and port -O are two type of high output impedance terminals. The input-output current-voltage relationship between different terminals of the CCTA can be described by the following equations.

$$I_Y = 0, V_X = V_Y, I_{-Z} = -I_X, I_{-O} = -g_m V_{-Z} \quad (1)$$

where g_m is the trans-conductance of CCTA and g_m depends upon the biasing current I_s of the CCTA. The Schematic symbol of CCTA is shown in Fig. 1. MOS implementation of CCTA is proposed in Fig. 2. For a MOS CCTA, the g_m can be expressed as

$$g_m = \sqrt{\beta_n I_s} \quad (2)$$

where β_n is given by

$$\beta_n = \mu_n C_{OX} \frac{W}{L} \quad (3)$$

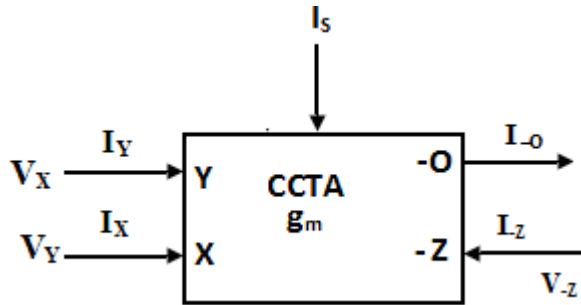


Figure 1. CCTA Symbol

where μ_n , C_{OX} and W/L are the electron mobility, gate oxide capacitance per unit area and transistor aspect ratio of NMOS, respectively.

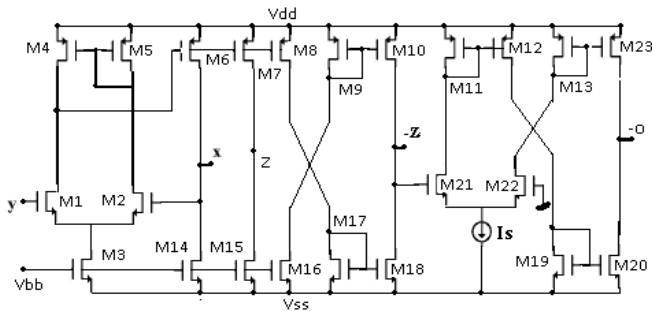


Figure 2. MOS implementation of CCTA

The proposed voltage-mode biquadratic filter is shown in Fig. 3. It uses only single CCTA, two capacitors and two

resistors. By routine analysis of the circuit in Fig. 3, the output voltage V_o can be obtained as

$$V_o = \frac{V_1 s^2 R_1 R_2 C_1 C_2 - V_4 s g_m R_1 R_2 C_2 - V_2 g_m R_1}{s^2 R_1 R_2 C_1 C_2 + s C_2 R_2 + g_m R_2} \quad (4)$$

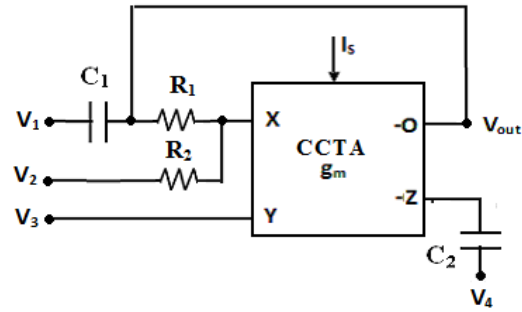


Figure 3. Proposed voltage-mode biquad filter

From (4) various filter responses in voltage form can be obtained through appropriate selection of input voltages.

- (i). HP response with $V_1 = V_{in}, V_2 = V_3 = V_4 = 0$
- (ii). Inverted LP response with $V_2 = V_{in}, V_1 = V_3 = V_4 = 0$
- (iii). Inverted BP response with $V_4 = V_{in}, V_1 = V_2 = V_3 = 0$
- (iv). BR with $V_1 = V_2 = V_3 = V_4 = V_{in}$, and $R_1 g_m = 1$
- (v). AP responses with $V_1 = V_2 = V_3 = V_4 = V_{in}$, and $R_1 g_m = 2$

It may be noted that realization of all the responses do not require any inverting-type input voltage signal(s) and double input voltage signal(s) to realize all the filtering responses in the design. The BR and AP realization require matching conditions and that are simple to satisfy through design, particularly in monolithic technologies, where inherently matched devices are realized. The filter parameters pole frequency (ω_o), the quality factor (Q) and bandwidth (BW) for the proposed circuit as shown in Fig. 3, can be expressed as

$$\omega_o = \left(\frac{g_m}{R_1 C_1 C_2} \right)^{\frac{1}{2}} = \left(\frac{\sqrt{\beta_n I_s}}{R_1 C_1 C_2} \right)^{\frac{1}{2}} \quad (5)$$

$$Q = \left(\frac{C_1 g_m R_1}{C_2} \right)^{\frac{1}{2}} = \left(\frac{C_1 R_1}{C_2} \sqrt{\beta_n I_s} \right)^{\frac{1}{2}} \quad (6)$$

$$BW = \frac{1}{C_1 R_1} \quad (7)$$

From (5) and (6), it can be remarked that Q and ω_o are orthogonally adjustable with simultaneous adjustment of g_m and R_1 such that product g_m/R_1 remain constant and $g_m R_1$ varies and vice versa. From (5) and (7), it can also be noted

that ω_o can also be electronically controlled without affecting BW. To see the effects of non idealities, the defining equations of the CCTA can be rewritten as the following

$$V_x = \beta V_y, I_{-z} = -\alpha I_x, I_{-o} = -\gamma g_m V_{-z} \quad (8)$$

where α , β and γ is transferred error values deviated from one. In the case of non-ideal and re-analyzing the proposed voltage-mode filter in Fig. 3, it yields voltage outputs as

$$V_o = \frac{V_1 s^2 \alpha R_1 R_2 C_1 C_2 - V_4 s \alpha \gamma g_m R_1 R_2 C_2 - V_2 \gamma g_m R_1 + V_3 [s \alpha \beta R_2 C_2 + g_m \gamma \beta (R_1 + R_2)]}{s^2 \alpha R_1 R_2 C_1 C_2 + s \alpha C_2 R_2 + \gamma g_m R_2} \quad (9)$$

In this case, the ω_o and Q are changed to

$$\omega_o = \left(\frac{\gamma g_m}{\alpha R_1 C_1 C_2} \right)^{\frac{1}{2}}, Q = \left(\frac{\gamma C_1 g_m R_1}{\alpha C_2} \right)^{\frac{1}{2}} \quad (10)$$

The active and passive sensitivities can be found as

$$S_{C_1, C_2, R_1, \alpha}^{\omega_o} = -\frac{1}{2}, S_{\gamma, g_m}^{\omega_o} = \frac{1}{2}, S_{\beta}^{\omega_o} = 0 \quad (11)$$

$$S_{C_2, \alpha}^Q = -\frac{1}{2}, S_{C_1, g_m, R_1, \gamma}^Q = \frac{1}{2}, S_{\beta}^Q = 0 \quad (12)$$

From the above results, it can be observed that all the active and passive are within ‘half’ in magnitude, ensuring good sensitivity performance. It can be also noted from (10) that $BW = \omega_o/Q$ is not effected by non ideal errors.

III. SIMULATION RESULTS

The PSPICE simulations are carried out to demonstrate the feasibility of the proposed circuit using CMOS implementation of CCTA as shown in Fig. 2.

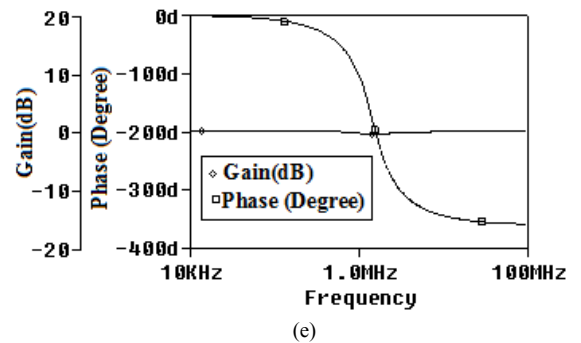
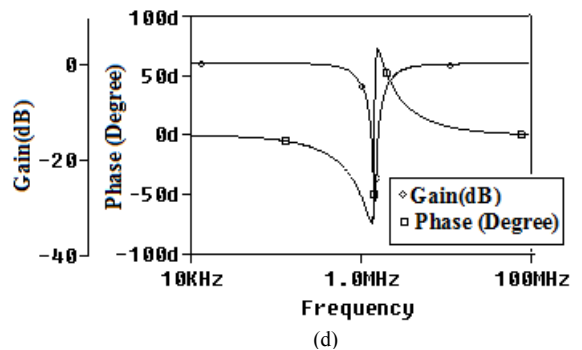
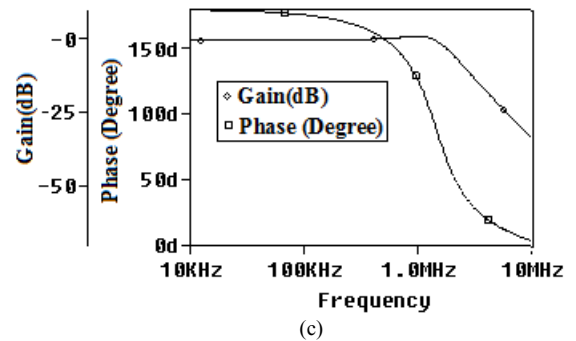
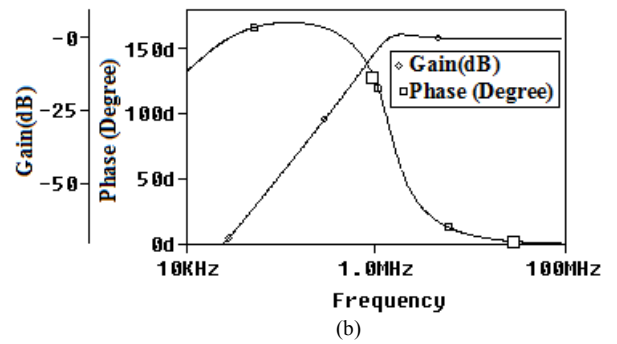
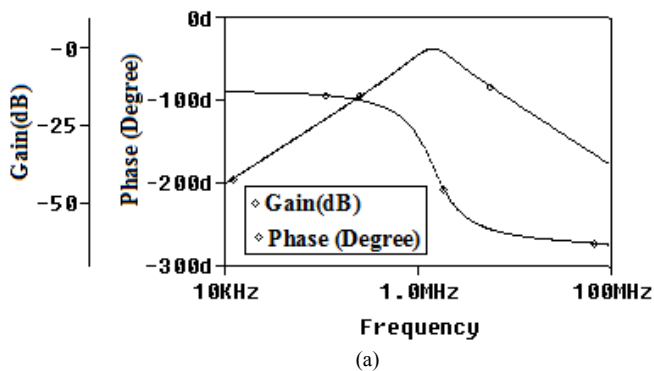


Figure 4. Simulated gain and phase response of (a) BP (b) HP (c) LP (d) BR (e) AP voltage-mode filter of Fig. 3

The simulations use a 0.35 μ m MOSFETs from TSMC whose model parameters are given in Table 1. The aspect ratio, for MOS transistors are specified in Table 2. The simulated gain and phase responses of the voltage mode BP, LP, HP, RN and AP for the proposed circuit of Fig. 3, designed with $R_1=R_2=2.86K$, $I_S = 80\mu A$, $C_1 = C_2 = 40pF$, $V_{DD} = -V_{SS} = 1.5V$ and $V_{bb}=0.96V$ are shown in Fig. 4. The simulated pole frequency is obtained as 1.42 MHz which is

very closed to designed value as 1.4MHz. The power dissipation of the proposed circuit for the design values is found as 1.22 mW that is a low value. Fig. 5 shows the variations in pole frequency with respect to I_S , without affecting the bandwidth of the filter. The bias current I_S is varied as 20 μ A, 60 μ A and 180 μ A and the pole frequency for a constant bandwidth at 1.39 MHz is found to be 1.08 MHz, 1.3 MHz and 1.67 MHz, respectively. Fig. 6 shows the gain responses of BP function, with different values of R_1 and I_S . From Fig. 6, it can be seen that quality factor can be electronically tuned by the bias current (I_S) without affecting pole frequency. Further simulations are carried out to verify the total harmonic distortion (THD). The circuit is verified by applying a sinusoidal voltage (V_{in}) of varying frequency and amplitude of 150mV. The THD measured at the HP output are found to be less than 3% while frequency is varied from 5 MHz to 80 MHz. The time domain behavior of the proposed current-mode filter is also investigated by applying a 60MHz sinusoidal input voltage signal with peak to peak amplitude of 300mV. Fig. 7 shows the time domain sinusoidal voltage input and corresponding HP output waveform for the proposed filter. Thus, both THD analysis and time domain response of HP output confirm the practical utility of the proposed circuit.

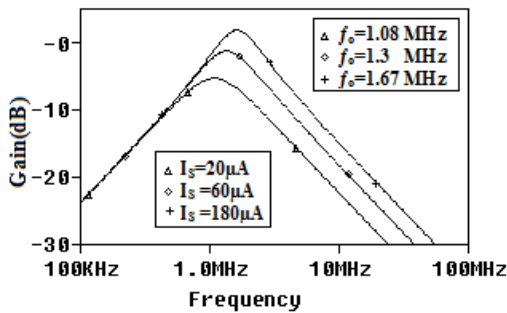


Figure 5. BP pole-frequency tuning

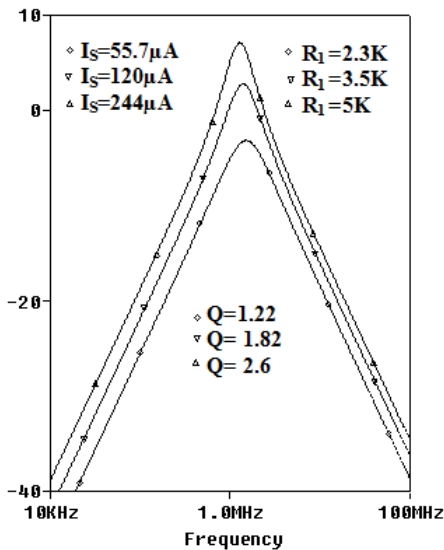


Figure 6: BP quality factor tuning of the circuit of Fig. 3

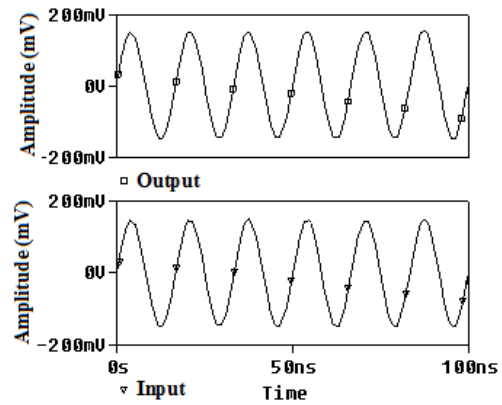


Figure 7. The time domain input waveform and corresponding response at HP output

Table 1. The SPICE model parameters of MOSFET for level 3, 0.35 μ m CMOS process from TSMC

NMOS	LEVEL=3 TOX=7.9E-9 NSUB=1E17 GAMMA=0.5827871 PHI=0.7 VTO=0.5445549 DELTA=0 UO=436.256147 ETA=0 THETA=0.1749684 KP=2.055786E-4 VMAX=8.309444E4 KAPPA=0.2574081 RSH=0.0559398 NFS=1E12 TPG=1 XJ=3E-7 LD=3.162278E-11 WD=7.046724E-8 CGDO=2.82E-10 CGSO=2.82E-10 CGBO=1E-10 CJ=1E-3 PB=0.9758533 MJ=0.3448504 CJSW=3.777852E-10 MJSW=0.3508721
PMOS	LEVEL=3 TOX=7.9E-9 NSUB=1E17 GAMMA=0.4083894 PHI=0.7 VTO=-0.7140674 DELTA=0 UO=212.2319801 ETA=9.999762E-4 THETA=0.2020774 KP=6.733755E-5 VMAX=1.181551E5 KAPPA=1.5 RSH=30.0712458 NFS=1E12 TPG=-1 XJ=2E-7 LD=5.000001E-13 WD=1.249872E-7 CGDO=3.09E-10 CGSO=3.09E-10 CGBO=1E-10 CJ=1.419508E-3 PB=0.8152753 MJ=0.5 CJSW=4.813504E-10 MJSW=0.5

Table 2. Aspect ratio of MOS

MOS -Transistors	Aspect Ratio [W (μ m)]/L(μ m)]
M1,M2	10/0.5
M3	27.25/0.5
M4,M5	5/0.5
M6-M10	8.5/0.5
M14-M18	44/0.5
M11-M13 and M19-M23	5/0.5

IV. CONCLUSION

A new voltage-mode FISO biquad filter using single CCTA has been proposed. The proposed filter offers the following advantages:

- I. Realizing of all standard filtering responses such as LP, HP, BP, BR, and AP in voltage form.
- II. Low sensitivity figures.
- III. The Q and ω_0 are electronically orthogonal tunable.
- IV. Both ω_0 and BW are orthogonally electronically tunable.

- V. Single active element.
- VI. No requirement of inverting-type input voltage signal(s) and double input voltage signal(s) to realize any response in the design.
- VII. Suitable for high frequency applications.
- VIII. $BW=\omega_0/Q$ of the proposed filter is not effected by non ideal errors.
- IX. Low power consumption.

REFERENCES

- [1] M. A. Ibrahim, S. Minaei and H. A. Kuntman, "A 22.5 MHz current-mode KHN-biquad using differential voltage current conveyor and grounded passive elements," *Int'l J. Electronics and Communication (AEU)*, vol. 59, pp. 311-318, 2005.
- [2] S. Maheshwari, S. V. Singh and D. S. Chauhan, "Electronically tunable low voltage mixed-mode universal biquad filter," *IET Circuits, Devices and Systems*, vol. 5, no.3, pp. 149-158, 2011.
- [3] R. Senani and V. K. Singh, "KHN-equivalent biquad using current conveyor," *Electronics Lett.*, vol. 31, no. 8, pp. 626-628, 1995.
- [4] S. V. Singh, S. Maheshwari, J. Mohan and D. S. Chauhan, "An electronically tunable SIMO biquad filter using CCCCTA," *Contemporary Computing, CCIS*, vol. 40, pp. 544-555, 2009.
- [5] M. Kumar, M. C. Srivastava and U. Kumar, "Current conveyor based multifunction filter," *International Journal of Computer Science and Information Security*, vol. 7, no.2, pp. 104-107, 2010.
- [6] J. W. Horng, C. G. Tsai and M. H. Lee, "Novel universal voltage-mode biquad filter with three inputs and one output using only two current conveyors," *Int'l J. Electronics*, vol. 80, pp. 543-546, 1996.
- [7] J. W. Horng, "High input impedance voltage-mode universal biquadratic filter with three inputs using DDCCs," *Circuits System Signal Processing*, vol. 27, no. 4, pp. 553-562.
- [8] I. A. Khan, M. T. Ahmed, and N. Minhaj, "A simple realization scheme for OTA-C universal biquadratic filter," *Int'l J. Electronics*, vol. 72, pp. 419-429, 1992.
- [9] H. P. Chen, S. S. Shen and J. P. Wang, "Electronically tunable versatile voltage-mode universal filter," *Int'l J. Electronics and Communications (AEU)*, vol. 62, pp. 316-319, 2008.
- [10] N. Pandey, S. K. Paul and S. B. Jain, "New high- input impedance voltage-mode universal biquad: multi input multi output," *J. Active and Passive Devices*, vol. 8, pp. 259-264, 2009.
- [11] H. P. Chen, "Single CCII based voltage-mode universal filter," *Analog Integrated Circuits Signal Processing*, vol. 62, pp. 259-262, 2010.
- [12] A. U. Keskin, "Multifunction biquad using single CDBA", *Electrical Engineering*, vol. 88, pp. 353-356, 2006.

- [13] S. Maheshwari, "High performance voltage-mode multifunction filter with minimum component count," *WSEAS Transactions on Electronics*, vol. 5, pp. 244-249, 2008.
- [14] N. Herencsar, J. Koton and K. Vrba, "Single CCTA-based universal biquadratic filters employing minimum components," *Int'l J. Computer and Electrical Engineering*, vol. 1, pp. 307-310, 2009.
- [15] N. Pandey and S. K. Paul, "Analog filters based on 0.25 μm CMOS differential voltage current conveyor transconductance amplifier (DVCCTA)," *IEEE Int'l Conference On Power Electronics (28-30 Jan. 2011)*, India, pp. 1-5, 2011.
- [16] T. Thosdeekoraphat, S. Summart, C. Saetiauw, S. Sanatalunai and C. Thongsopa, "Resistor-less current-mode universal biquad filter using CCTAs and grounded capacitors," *World Academy Of Science, Engineering and Tecnology* vol. 69, pp. 559-563, 2012.

AUTHORS PROFILE

S. V. Singh was born in Agra, India. He received his B.E. degree (1998) in Electronics and Telecommunication from NIT Silchar, Assam (India), M.E. degree (2002) from MNIT Jaipur, Rajasthan (India) and Ph.D. degree (2011) from Uttarakhand Technical University. He is currently working as Assistant Professor in the Department of Electronics and Communication Engineering of Jaypee Institute of Information Technology, Noida (India) and has been engaged in teaching and design of courses related to the design and synthesis of Analog and Digital Electronic Circuits. His research areas include Analog IC Circuits and Filter design. He has published more than 20 research papers in various International Journal/Conferences.

R. S. Tomar was born in Aligarh, India. He obtained his B. E (1995) from Bombay University, M. E. (2004) from Agra. He is currently associated with Anand Engineering College, Agra. His research areas include designing of microwave and analog circuits. He has published no. of papers in National and International Conferences.

D. S. Chauhan was born in Dholpur, India. He obtained his B.Sc Engg.(1972) in Electrical Engineering at I.T. B.H.U., M.E. (1978) at R.E.C. Tiruchirapalli (Madras University) and PH.D. (1986) at IIT/Delhi. His brilliant career brought him to teaching profession at Banaras Hindu University where he was Lecturer, Reader and then has been Professor till today. He has been director KNIT Sultanpur in 1999-2000 and founder vice Chancellor of U.P.Tech. University (2000-2003-2006). Later on, he has served as Vice-Chancellor of Lovely Professional University (2006-07) and Jaypee University of Information Technology (2007-2009). Currently he has been serving as Vice-Chancellor of Uttarakhand Technical University. He has supervised 24 Ph.D., one D.Sc. He has authored two books and published and presented 170 research papers in international journals and international conferences. His research areas include Analog IC Circuits and Control Systems design.

RELIABLE MULTIPATH ROUTING PROTOCOL (RMRP) FOR MOBILE AD HOC NETWORKS USING ADAPTIVE VIDEO COMPRESSION

S.P. Swornambiga,
Assistant Professor,
Department of Computer Applications & Software Systems,
C.M.S. College of Science and Commerce,
Coimbatore,
India.

Dr. Antony Selvadoss Dhanamani,
Associate Professor and Head,
Research Department of Computer Science,
NGM College,
Pollachi,
India.

Abstract - This paper presents a reliable multipath routing protocol for mobile ad hoc networks using adaptive video compression. Mobile ad hoc network is the kind of wireless network which consists of mobile nodes and has the characteristic of deploying anywhere anytime. An adaptive video compression mechanism is deployed. Multipath routing mechanism is adapted from the Ad hoc On-demand Distance Vector (AODMV) routing protocol and the received signal strength is measured for the discovered available paths. The path with the maximum received signal strength is selected and the packets (compressed video packets) are sent through the path. The performance metrics such as packet delivery ratio, throughput, drop, jitter are taken into account for comparison with AODMV. Simulation results proved that the proposed RMRP outperforms AODMV in all performance aspects.

1. Introduction

The technological and theoretical advances in the study of wireless communications lead to a rapid introduction of wireless ad hoc networks to a wide spectrum of applications, from scientific monitoring to military and rescue operations. A mobile ad hoc network (MANET) is a communication group which is formed by wireless mobile hosts without an established infrastructure or centralized control [1]. Due to its fast and easy deployment and self-organizing capabilities, it has been shown to have great potential in many fields, such as rescue missions and military applications. There has been an increasing demand for multicast applications in MANETs, such as interactive video conferences. Such real-time multicast applications require a QoS guarantee in terms of bandwidth, delay, delay jitter and/or packet loss probability.

Quality of service (QoS) for multimedia has become a critical issue which is closely related to resource. Many routing protocols have been developed for ad hoc networks [2]. They can be classified according to different criteria. The most important is by the type of route discovery. It enables to separate the routing protocols into two categories: proactive and reactive. In reactive protocols, e.g. Dynamic

Source Routing (DSR [24]) and Ad hoc On demand Distance Vector routing (AODV [25]), the routing request is sent on-demand: if a node wants to communicate with another, then it broadcasts a route request and expects a response from the destination. Conversely, proactive protocols update their routing information continuously in order to have a permanent overview of the network topology (e.g. OLSR [23]).

Another criterion for ad hoc routing protocol classification is the number of routes computed between source and destination: multipath and single path routing protocols. Unlike its wired counterpart, the ad hoc network is more prone to both link and node failures due to expired node power or node mobility. As a result, the route used for routing might break down for different reasons. To increase the routing resilience against link or/and node failures, one solution is to route a message via multiple disjoint paths simultaneously. Thus, the destination node is still able to receive the message even if there is only one surviving routing path. This approach attempts to mainly address the problems of the scalability, mobility and link instability of the network. The multipath approach takes advantage from the large and dense networks.

2. Literature Review

The notion of QoS is tied back to the well-developed theory of effective bandwidth [3], [4], [5] and its dual concept of effective capacity [6], [7], [8]. For scalable video transmission, a set of QoS exponents for each video layer are obtained by applying the effective bandwidth/capacity analyses on the incoming video stream to characterize the delay requirement [9], [10]. The problem of providing statistical delay bounds for layered video transmission over single hop unicast and multicast links was considered in [9]. Cooperation among mobile devices in wireless networks has the potential to provide notable performance gains in terms of increasing the network throughput [11], [12], [13], [14], [15], extending the network coverage [16], [17], [18], decreasing the end-user communication cost [19], and decreasing the energy consumption [20], [21], [22].

3. Proposed Work

3.1. Adaptive Video Compression

It is possible to achieve the advantages of both wavelets based and motion based compression, using a combined framework. The proposed methodology of video compression which is presented in this paper consists of three steps as following:

- Model creation,
- Wavelet based frame creation, and
- Combination of the model frame and the wavelet frame.

Model Creation

In the first step, the image is divided into (8 X 8 pixels were selected practically) blocks (gridding) and the motion vector is extracted for each block. Optical flow as an efficient video motion extraction is utilized to extract the motion vector of each frame. Optical flow presumes gray level (intensity) constancy between two consecutive frames is given below.

$$\sum W^2(x)[E_x u + E_y v + E_t]^2$$

Where E is pixel intensity with gradients toward spatial (Ex, Ey) and temporal (Et) dimension, u is the velocity toward the x-axis, v is the velocity toward the y-axis and finally w is a weighted window. Subsequently, the next frame is predicted according to the estimated motion vector of each block which is named predicted model frame. By warping each block in accordance with its motion vectors of it, the predicted model frame is constructed.

Wavelet Based Frame Creation

In the second step, the wavelet decomposition is applied to each block of the main frame (which should be predicted) and is used as error compensator of the predicted model frame. The level of wavelet coefficient quantization is computed according to the complexity of the motion vectors of each block achieved by computing of entropy. The entropy of motion of each block is defined as follows:

$$En = -\sum p \ln p$$

Where p contains the histogram counts of each block motion amplitude. Fuzzy C-means (FCM) is utilized to classify four levels of quantization (2, 4, 6 or 8) of wavelet coefficients based on the complexity of each block motion, while lower level of quantization means less motion complexity. As mentioned before, according to the motion complexity of each block the numbers of clusters are assigned to be 2, 4, 6 or 8. In the non-uniform quantization, sampling rate is increased in regions with greater data density and vice versa.

Combination

Two dimension correlations between each block of the predicted model frame (based on the motion) and corresponding block of the main frame (which should be predicted) is calculated using the following formula

$$\frac{\sum_{K=1}^K \sum_{L=1}^L (M(k,l) - \bar{M})(I(k,l) - \bar{I})}{\sqrt{\left[\sum_{k=1}^K \sum_{l=1}^L (M(k,l) - \bar{M})^2 \right] \left[\sum_{k=1}^K \sum_{l=1}^L (I(k,l) - \bar{I})^2 \right]}}$$

Where M and I are the predicted model frame and main frame respectively. K and L denote to the size of each block which are considered to be 8. If the 2D calculated correlation is more than 0.5, an optimized combination between the block of the predicted model frame and corresponding block of wavelet frames is performed. Otherwise, it means the block of predicted model frame is not a good representation of the corresponding block in the main image. Finally, ROI mask is applied on the wavelet frame or optimized combination of wavelet frame and predicted model frame to select the appropriate data which is necessary to be saved. If there is an overlap between white region and the computed blocks of wavelet frame or model frame, these blocks will be saved.

If the calculated correlation between the block of model frame and the block of the main frame is more than 0.5, optimized combination is performed. In the optimization framework, the difference between the reconstructed block and main corresponding block is minimized.

3.2. Multipath Routing

RMRP is an extension to the AOMDV [1] protocol for computing multiple loop-free and link disjoint paths. The routing entries for each destination contain a list of the next-hops along with the corresponding hop counts. All the next hops have the same sequence number. This helps in keeping track of a route. For each destination, a node maintains the advertised hop count, which is defined as the maximum hop count for all the paths, which is used for sending route advertisements of the destination. Each duplicate route advertisement received by a node defines an alternate path to the destination. Loop freedom is assured for a node by accepting alternate paths to destination if it has a less hop count than the advertised hop count for that destination. Because the maximum hop count is used, the advertised hop count therefore does not change for the same sequence number. When a route advertisement is received for a destination with a greater sequence number, the next-hop list and the advertised hop count are reinitialized. To find node-disjoint routes, each node does not immediately reject duplicate RREQs.

Each RREQs arriving via a different neighbor of the source defines a node-disjoint path. This is because nodes cannot be broadcast duplicate RREQs, so any two RREQs arriving at an intermediate node via a different neighbor of the source could not have traversed the same node. In an attempt to get multiple link-disjoint routes, the destination replies to duplicate RREQs, the destination only replies to RREQs arriving via unique neighbors. After the first hop, the RREPs follow the reverse paths, which are node disjoint and thus link-disjoint. The trajectories of each RREP may intersect at an intermediate node, but each takes a different reverse path to the source to ensure link disjointness.

It is possible to forecast the link quality and discard the links with the lower signal strengths from the route selection using the received signal strength from the physical layer. When a sending node broadcasting RTS packet, it piggybacks its transmissions power, P_t . On receiving the RTS packet, the intended node measures the signal strength received which holds the following relationship for free-space propagation model.

$$P_r = P_t \cdot (\lambda / 4\pi d)^2 \cdot G_t \cdot G_r$$

Where λ is the wavelength carrier, d is distance between sender and receiver, G_t and G_r are unity gain of transmitting and receiving omni-directional antennas, respectively.

By the above said method, the received signal strength is measured for the discovered available paths. The path with the maximum received signal strength is selected and the packets (compressed video packets) are sent through the path.

4. Simulation Settings and Performance Metrics

The simulation is done using NS2. Network Simulator 2 (NS2) is used to simulate RODRP and AOMDV routing protocol; 50 mobile nodes starting from IP address 192.168.1.1 to 192.168.1.50 move in a 1000 x 700 meter rectangular region for 200 seconds simulation time. The channel capacity of mobile nodes is set to the value is .5 mps. We use the distributed coordination function (DCF) of IEEE 802.11 for wireless LANs. It has the functionality to notify the network layer about link breakage.

We assume each node moves independently with the mobility speed 0.5 m/s. All nodes have the transmission range of 250 meters. The simulated traffic is Constant Bit Rate (CBR) with initial energy between 1.5 joules. The simulation settings are also represented in tabular format as shown in Table 1.

Throughput, Delivery ratio, packet drop, overhead and jitter delay are the performance metrics.

Table 1. NS2 Simulation Settings

No. of Nodes	50
Terrain Size	1000 X 700 m
MAC	802.11b
Radio Transmission Range	250 meters
Simulation Time	100 seconds
Traffic Source	CBR (Variable Bit Rate)
Packet Size	512 KB
Mobility Model	Random Waypoint Model
Speed	0.5 m/s

5. Results and Discussion

From the Fig.1 it can be clearly observed that the proposed RMRP achieves better throughput than that of AOMDV. Also the packet delivery ratio got better which is shown in Fig.2. From the Fig.3 and Fig.4 it can be seen that the packets drop and jitter delay got reduced reasonably in RMRP than the conventional routing protocol AOMDV.

Table.1 Simulation Results

	Throughput (No. of Packets)			Packet Delivery Ratio	
	AOMDV	RMRP		AOMDV	RMRP
20	52478	62345	20	0.628903	0.933902
40	53743	72345	40	0.570892	0.956077
60	54861	82345	60	0.845248	1
80	55357	92086	80	0.340727	0.856716
100	46924	91766	100	0.435932	0.607249
	Packets Drop (No. of Packets)			Jitter Delay (Seconds)	
	AOMDV	RMRP		AOMDV	RMRP
20	5478	2245	20	8.493277	0.653378
40	5743	2145	40	8.185072	0.80231
60	5868	2445	60	3.756587	0.014944
80	5356	2686	80	19.07075	5.689394
100	4923	1866	100	17.18173	10.50514

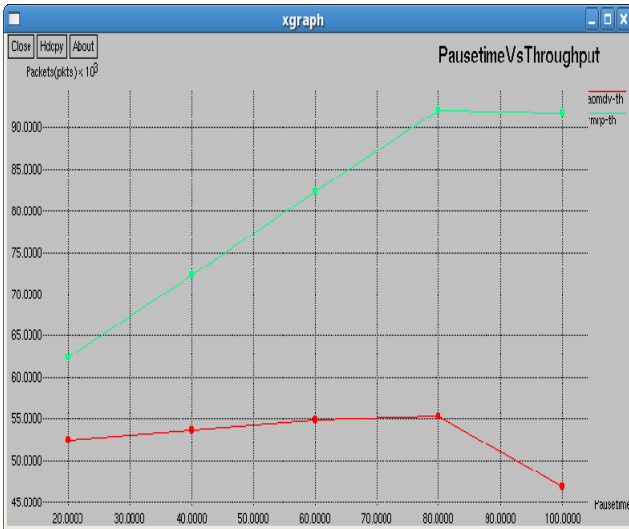


Fig.1 Pausetime Vs Throughput

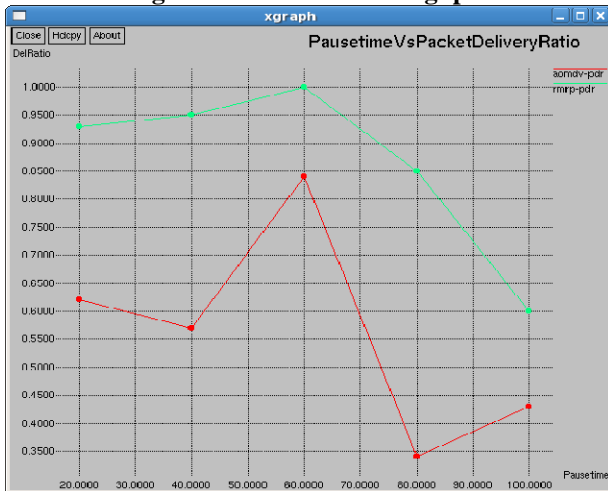


Fig.2 Pausetime Vs Packet Delivery Ratio



Fig.3 Pausetime Vs Packets Drop

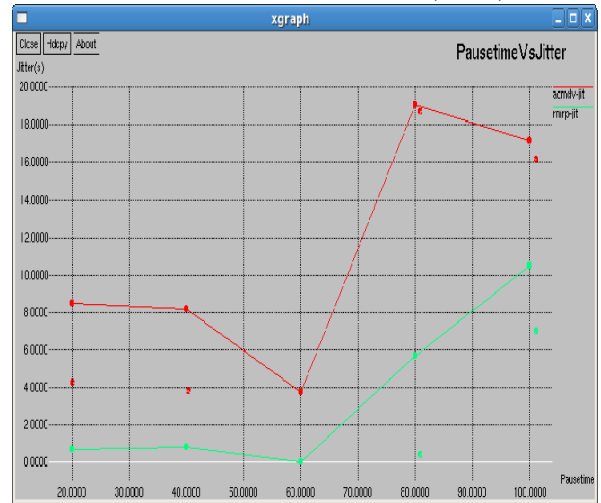


Fig.4 Pausetime Vs Jitter

6. Conclusion

This research article presented a reliable multipath routing protocol for mobile ad hoc networks using adaptive video compression. An adaptive video compression mechanism is used for better transmission of video data. Multipath routing mechanism is taken from the Ad hoc On-demand Distance Vector (AOMDV) routing protocol and the received signal strength is measured for the discovered available paths. The path with the maximum received signal strength is selected and the packets (compressed video packets) are sent through the path. The performance metrics such as packet delivery ratio, throughput, drop, jitter are taken into account for comparison with AOMDV. From the simulation results it is proved that the proposed RMRP outperforms AOMDV in all performance aspects.

7. References

- [1] J.P. Macker, M.S. Corson, Mobile ad hoc networking and the IETF, ACM SIGMOBILE Mobile Computing and Communications Reviews 2 (2) (1998) 9–14.
- [2] M. Abolhasan, T. Wysocki, E. Dutkiewicz, A review of routing protocols for mobile ad hoc networks, Ad Hoc Networks 2 (1) (2004) 1–22.
- [3] A. Kumar, D. Manjunath, and J. Kuri, Communication Networking: An Analytical Approach. Morgan Kaufmann, 2004.
- [4] C.-S. Chang, Performance Guarantees in Communication Networks. Springer-Verlag, 2000.
- [5] C.-S. Chang and J. Thomas, “Effective Bandwidth in High Speed Digital Networks,” IEEE J. Selected Areas in Comm., vol. 13, no. 6, pp. 1091-1100, Aug. 1995.
- [6] D. Wu and R. Negi, “Effective Capacity: A Wireless Link Model for Support of Quality of Service,” IEEE Trans. Wireless Comm., vol. 2, no. 4, pp. 630-643, July 2003.
- [7] D. Wu and R. Negi, “Effective Capacity-Based Quality of Service Measures for Wireless Networks,” Mobile

Networks and Applications, vol. 11, no. 1, pp. 91-99, June 2006.

[8] D. Wu, "Providing Quality-of-Service Guarantees in Wireless Networks," PhD thesis, Carnegie Mellon Univ., 2003.

[9] Q. Du and X. Zhang, "Statistical QoS Provisionings for Wireless Unicast/Multicast of Layered Video Streams," Proc. IEEE INFOCOM, Apr. 2009.

[10] X. Zhang and Q. Du, "Cross-Layer Modeling for QoS-Driven Multimedia Multicast/Broadcast over Fading Channels in Mobile Wireless Networks," IEEE Comm. Magazine, vol. 45, no. 8, pp. 62-70, Aug. 2007.

[11] H. Wu, C. Qiao, S. De, and O. Tonguz, "Integrated Cellular and Ad Hoc Relaying Systems: iCAR," IEEE J. Selected Areas in Comm., vol. 19, no. 10, pp. 2105-2115, Oct. 2001.

[12] H. Luo, R. Ramjee, P. Sinha, L.E. Li, and S. Lu, "UCAN: A Unified Cellular and Ad Hoc Network Architecture," Proc. ACM MobiCom, Sept. 2003.

[13] R. Bhatia, L.E. Li, H. Luo, and R. Ramjee, "ICAM: Integrated Cellular and Ad Hoc Multicast," IEEE Trans. Mobile Computing, vol. 5, no. 8, pp. 1004-1015, Aug. 2006.

[14] L. Popova, T. Herpel, W. Gerstacker, and W. Koch, "Cooperative Mobile-to-Mobile File Dissemination in Cellular Networks within a Unified Radio Interface," Int'l J. Computer and Telecomm. Networking, vol. 52, no. 6, pp. 1153-1165, Jan. 2008.

[15] M. Peng and W. Wang, "Investigation of Cooperative Relay Node Selection in Heterogeneous Wireless Communication Systems," Proc. IEEE Int'l Conf. Comm., May 2008.

[16] J.-C. Chen, S.-H. Chan, J.-Y. He, and S.-C. Liew, "Mixed-Mode WLAN: The Integration of Ad Hoc Mode with Wireless LAN Infrastructure," Proc. IEEE GlobeCom, Dec. 2003.

[17] Z. Dawy, S. Davidovic, and I. Oikonomidis, "Coverage and Capacity Enhancement of CDMA Cellular Systems via Multihop Transmission," Proc. IEEE GlobeCom, Dec. 2003.

[18] J. Chen, S. Li, S.-H.G. Chan, and J. He, "WIANI: Wireless Infrastructure and Ad-Hoc Network Integration," Proc. IEEE Int'l Conf. Comm., May 2005.

[19] K.-J. Peng and Z. Tsai, "Distortion and Cost Controlled Video Streaming in a Heterogeneous Wireless Network Environment," Proc. IEEE Int'l Symp. Personal, Indoor and Mobile Radio Comm., Sept. 2006.

[20] F. Fitzek and M. Katz, Cooperation in Wireless Networks: Principles and Applications. Springer, 2006.

[21] Q. Zhang, F.H.P. Fitzek, and V.B. Iversen, "Design and Performance Evaluation of Cooperative Retransmission Scheme for Reliable Multicast Services in Cellular Controlled P2P Networks," Proc. IEEE Int'l Symp. Personal, Indoor and Mobile Radio Comm., Sept. 2007.

[22] M. Ramadan, L. Zein, and Z. Dawy, "Implementation and Evaluation of Cooperative Video Streaming for Mobile Devices," Proc. IEEE Int'l Symp. Personal, Indoor, and Mobile Radio Comm., Sept. 2008.

[23] T. Clausen, P. Jacquet, IETF Request for Comments: 3626, Optimized Link State Routing Protocol OLSR, October 2003.

[24] D.B. Johnson, Y. Hu, D.A. Maltz, IETF Request for Comments: 4728, The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4, February 2007.

[25] C. Perkins, E. Royer, Ad-hoc on-demand distance vector routing, in: Second IEEE Workshop on Mobile Computing Systems and Applications, 1999, pp. 90-100.

[26] B Narasimhan, R Vadivel. "Secured Reliable Multipath Routing Protocol (SRMRP) using Trust Computation and Carrier Sense Multiple Access with Collision Intimation (CSMA/CI) for Heterogeneous IP-based Mobile Ad-hoc Networks", International Journal of Computer Applications, Vol.60, No.10, pp:12-16, December 2012.

8. Profile of the Authors



S.P.Swornambiga is currently working as Assistant Professor in the Department of Computer Applications, Software systems, and Information Technology, CMS College of Science & Commerce, Coimbatore with a teaching experience of 10 years. She is the Research scholar in Manonmanium Sundaranar University, Tirunelveli, Tamilnadu. Her research area includes networks, and network security.



Dr. Antony Selvadoss Thanamani is currently working as Associate Professor and Head in the Research Department of Computer Science in NGM College, Pollachi with a teaching experience of about 25 years. His research interest includes knowledge management, web mining, networks, mobile computing, and telecommunication. He has guided 41 M.Phil scholars, attended 15 conferences, presented 35 papers, published about 8 books and many papers.

NON-PREEMPTIVE MULTI-CONSTRAIN SCHEDULING FOR MULTIPROCESSOR WITH HOPFIELD NEURAL NETWORK

Abdellatif H. ALI

ECE Department Modern Academy, Cairo, Egypt

ABSTRACT: In this paper, task scheduling for non-preemptive multi-constrained multi-processor systems presented. The proposed model based on discrete Hopfield Neural network augmented with a methodology for weighting constrains to form overall network energy function. The network augmented with a layer to handle network re-initialization, based on min-max algorithm, case of local minima trapped without an acceptable solution. The proposed neural network solution does not require a predetermined scheduling length. Constrains included in the study are: task time, precedence, resources conflict, task dead time, and favoring tasks of the same setup to run on the same processor to suit reconfigurable hardware.

KEYWORDS: Preemptive, Multi-constrains, Multi-processors, Scheduling, Hopfield, Neural Network.

I. INTRODUCTION

Multiprocessor have become powerful computing mean for many applications such as information processing, database systems, weapon systems, weather forecasting, image processing, and real-time high speed simulators [17]. The multiprocessor hardware both fixed configuration and reconfigurable is the trend of the new developed hardware to cope with the high demands to processing power. The utilization of this powerful hardware highly dependent on proper scheduling of the problem tasks over the processing resources available. It is well known that scheduling is the most challenging problem of parallel computing [16].

Scheduling a problem tasks over multiprocessors, normally, is subject to constrain(s). Scheduling tasks of a multi-constrain problem has wide range of applications such as computer operating systems, digital communications, industrial control, weapons to targets assignment, Unmanned Aerial Vehicles, fighter aircrafts, and operation research. Such a problem is well known to be NP hard type. A tremendous effort required to find out an acceptable solution due to the growth of the problem complexity exponentially with problem size. Many approaches explored by researchers in fields such as AI, operation research, and neural networks to solve such problems. The aim was to find out a pseudo optimal solution in an efficient manner. The use of genetic algorithms recently have taken a great attentions of researchers [17][18][19].

The neural network approach is considered superior giving the potential of hardware realization, owing to their parallel architecture. Moreover, the mapping of algorithms based on neural network to run on parallel platforms is more straightforward than others.

One of the neural network architectures utilized in solving scheduling problem is Hopfield. The Hopfield network is a recurrent neural network. The recurrent neural networks can capture more dynamic characteristics from a problem than the feedforward networks [20]. The Hopfield architecture introduced in [1],[2] by Hopfield and Tank. In those papers, they presented solution to the traveling salesman problem with large number of cities. In [3] Gene and others demonstrated the use of the K-out-of-N rule [4] to solve complex, nonlinear resource allocation problem to control weapons allocation to counter offensive threats. In [5-6] Carlos and Zoubir presented handling of the precedence constrains in real time multiprocessor scheduling. In [7] Ruey-Maw and Yueh-Min combined the characteristics of the Hopfield neural network structure and stochastic simulated annealing algorithm in so called mean field annealing technique to solve job scheduling problem of a multiprocessor with multiprocess. Same concept on the same problem under multi-constrain presented in [8]. In [9], Hopfield Neural network augmented by an additional layer of neurons to solve Internet network routing problem. In [10] the Hopfield architecture augmented with a layer to solve the economic dispatch and unit commitment for electric power industry.

The Hopfield neural network is well known to have the potential of being trapped in local minima's [5]. Such a problem is well known to exist in search schemes for NP complete type of problems [11]. To solve an optimization problem by Hopfield network one has to follow the following steps:-

- Finding a network topology that represents the problem. That is, the final state of the network neurons could be interpreted easily as a solution to the problem.
- Finding energy function that if optimized, minimized, it will correspond to the best solution to the problem. The targeted function, combines both constrains and cost functions. In multi-constrain type of a problem arises the problem of finding out the proper weight for each constrain to combine in a single energy function.
- Computation of network synaptic neurons weights and neurons thresholds based on the energy function.

In this study, a methodology for task scheduling over multiprocessor to non-preemptive multi-constrained problem is proposed. The proposed methodology includes a simple analysis step to compute constrains weights, re-initialization algorithm based on Min_Max algorithm [15] case of local minima trap or looking for better solution, and a neuron firing

sequence biased towards neurons with higher momentum to change. The constrains used in the study are task time, precedence, resources conflict, task dead time, and favoring tasks of the same setup to run on the same processor, beside the cost function of the schedule length.

This paper is organized as following: Section 2 introduces the Hopfield network. Section 3 contains the problem description and associated energy function. Section 4 presents the network synaptic weights and neurons threshold. Section 5 explains neurons firing sequence. Section 6 presents the network re-initialization. Section 7 includes tests and results followed by the conclusion.

II. HOPFIELD NEURAL NETWORK

The Hopfield network is a fully connected single layered feedback type neural network. The Laipunov function, or the energy function, Eq. (1), which is subject to optimization by the network was introduced in [12].

$$E = -\frac{1}{2} \sum_i \sum_j V_i W_{ij} V_j + \sum_j \theta_j V_j \quad (1)$$

Where: V_i is neuron state and W_{ij} is the synaptic weights connecting neuron V_i to V_j and θ_j is the threshold of neurons j . The neuron states, in discrete Hopfield, are binary values 0, or 1, which complies with problem requirement. The network, normally, starts with neuron states distributed uniformly random between 0, and 1. The update rule of the neurons in iteration $n + 1$ is as follows:-

$$V_j^{n+1} = \begin{cases} 1 & \text{if } net_j > 0 \\ V_j^n & \text{if } net_j = 0 \\ 0 & \text{if } net_j < 0 \end{cases}$$

where

$$net_j = \sum_i W_{ij} V_i^n - \theta_j$$

The network found to evolve to a stable state giving the following conditions [13]:-

- There is no self-feedback, in other words, the diagonal elements of the weight matrix are all equal zero.
- The synaptic weight connecting neuron i to neuron j is the same as the one connecting neuron j to neuron i , in other words, the connection matrix is symmetric. Welbing and YI Xlong in [14] eased the later condition.

III. PROBLEM DESCRIPTION AND ENERGY FUNCTION

We consider the problem of N tasks to be scheduled over M processing units with some known upper limit of the scheduling time T. The scheduling time could fit, or over fit the case. Each task time is known a priori as well as the set of

applied constrains. The Hopfield topology that well suits such a problem is a three dimensional architecture [7]. Each dimension represents a major factor in the problem: task, process, and time consequently. A neuron in such topology V_{ijk} represents the state of assignment of a task i to a processor j in time slot k . The synaptic weight matrix is 6-dimensional W_{ijkxyz} is the weight connecting neuron V_{ijk} to neuron V_{xyz} .

The energy function includes beside the problem given constrains and cost function: preventing a task from executing simultaneously on different processors, preventing a processing unit from running multiple tasks at the same time, and migration of a task over different processing units.

The following is the proposed energy function: -

$$E = \frac{C_1}{2} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^T \sum_{x \neq i} V_{jk} V_{xk} + \frac{C_2}{2} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^T \sum_{z=1}^T V_{jk} V_{yz} + \frac{C_3}{2} \sum_{i=1}^N (\sum_{j=1}^M \sum_{k=1}^T V_{ijk} - S_i)^2 + \frac{C_4}{2} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^T V_{ijk} G_{ik}^2 H(G_{ik}) + \frac{C_5}{2} \sum_{j=1}^M \sum_{k=1}^T (\sum_{i=1}^N V_{ijk} - 1)^2 + \frac{C_6}{2} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^T \sum_{x=1}^M \sum_{y=1}^M \sum_{z=1}^T P_1(i, x, k, z) V_{ijk} V_{xyz} + \frac{C_7}{2} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^T \sum_{x=1}^M \sum_{y=1}^M \sum_{z=1}^T P_2(x, i, j, y) V_{xyz} V_{ijk} + \frac{C_8}{2} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^T \sum_{z=1}^T P_3(i, k, z) V_{ijk} V_{xyz} + \frac{C_9}{2} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^T V_{ijk} P_4(k) + \frac{C_{10}}{2} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^T \sum_{x=1}^M \sum_{y=1}^M \sum_{z=1}^T P_5(i, x) V_{ijk} V_{xyk}$$

Where: -

S_i Is the time required for task i ,
 $C_1 \dots C_{11}$ Are the weighting factors for constrains/cost function.
 $G_{ik} = K - d_i$, d_i Is the dead time for task i ,
 $H(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$,
 $P_1(i, x, k, z) = \begin{cases} 1 & \text{if } (i \text{ precedes } x \text{ and } k \geq z) \text{ or } (x \text{ precedes } i \text{ and } z \geq k) \\ 0 & \text{otherwise} \end{cases}$,
 $P_2(i, x, j, y) = \begin{cases} 1 & \text{if } type(i) = type(x) \text{ and } j \neq y \\ 0 & \text{otherwise} \end{cases}$,
 $P_3(i, k, z) = \begin{cases} 1 & \text{if } ((k + S_i) \leq z) \text{ or } ((z + S_i) \leq k) \\ 0 & \text{otherwise} \end{cases}$,
 $P_4(k) = \frac{k-1}{T}$,
 $P_5(i, x) = \begin{cases} 1 & \text{if } \exists \text{ resource Conflict between } i, x \\ 0 & \text{otherwise} \end{cases}$

The terms of the equation above can be explained as following consequently: -

1. Processor unit i cannot execute more than one job

2. If job started on a processor unit it should continue, and complete on the same processor, prohibit task migration.
3. Tasks must have execution time equal to the required time S.
4. No assignments for a task after its dead time.
5. No more than one task shall be allowed to execute at some specific time on same processor unit.
6. Enforce the precedence constrains required.
7. Favor tasks of the same type to run on the same processor core.
8. Enforce non-preemptive task assignment.
9. Favor early schedule
10. Resolve use of exclusive resources between tasks.

IV. NETWORK SYNAPTIC WEIGHTS AND NEURONS THRESHOLD

The above mentioned energy function could be mapped to setting synaptic weights W and neurons threshold θ as following: -

$$\begin{aligned}
 W_{ijkxyz} = & -c_1(1-\delta(x,i))\delta(y,j)\delta(z,k) - c_2\delta(x,i)(1-\delta(y,j)) \\
 & - c_3\delta(x,i) - c_{11}(1-\delta(x,i))(1-\delta(y,j))\delta(k,z)P_6(i,x) - \\
 & c_5\delta(y,j)\delta(k,z) - c_6(1-\delta(i,x))(P_1(i,x,k,z) \\
 & - c_7(1-\delta(i,x))P_2(x,i,j,y) - \\
 & c_8\delta(i,x)(1-\delta(k,z))P_3(i,k,z) + c_{10}\delta(k,z)P_5(i,x) \\
 \theta_{ijk} = & -c_3(2S_i - 1) - c_5 + c_4G_{ik}^2H(G_{ik}) + c_9P_4(k)
 \end{aligned}$$

where,

$$\delta(x,y) = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases}$$

In [7], [10] the same problem studied for preemptive tasks under constrains number: 1,2,3,4,5,10 and no bases explicitly mentioned for computing c_i values. In this study, we propose methodology for calculation based on equal share of the initial total synaptic energy assuming uniform random distribution initialization.

The above mentioned energy elements could be seen as hard elements that if violated the solution provided will be of no value, and optimization/cost function. The same set could be classified based on their effect on network weights to: synaptic, threshold, and both. So, let us denote the following sets: -

- $\eta_H = \{1,2,3,5,6,8,10\}$ The set of hard elements that affects the synaptic weights.
- $\eta_O = \{7\}$ The set of optimization elements that affects the synaptic weights.
- $\xi_H = \{3,4\}$ The set of hard constrains that affects neurons threshold weights.

- $\xi_O = \{9\}$ The set of optimization constrains that affects neurons threshold weights.

Our primary focus is to find weights for constrains in η_H such that their shares in the initial synaptic energy are likely equal. So, let us assume C_i^c is the count of links set by constrain i , $i \in \eta_H$. A constrain share in energy is proportional to the number of weights set by that constrain. Therefore, the weight factor will be $C_i = \alpha \frac{1}{C_i^c}$ where

$\alpha > 0$, for valid contribution to the energy, and $C_i^c > 0$. The probability of a set link to contribute to the initial energy is the equal to 0.25 which is the probability of its two terminal neurons being ON. Then, the mean over large number of trails

t is $\frac{1}{t} \sum_{i=1}^t \gamma_i \cong 0.25\alpha$. Consequently, the total initial synaptic energy for hard constrains will be $\cong 0.25\alpha n$ where $n = |\eta_H|$. The rest of weighting factors are set as following:-

$\eta_O : C_7$ Not to exceed the hard constrains: $C_7 = \min\{C_i\} \forall i \in \eta_H$,

$\xi_H : C_4$ To resist firing after deadline, then:

$$C_4 = l \sum_{\forall i \neq 4} C_i \text{ where } l = \sum_{i=1}^N S_i. \quad (4)$$

$\xi_O : C_9$ Not to exceed the least hard constrain, then: $C_9 = \min\{C_i\} \forall i \in \eta_H$. (5)

V. NETWORK NEURONS FIRING SEQUENCE

After initialization, the network runs according to eq. 2. At any typical state, the network neurons could be divided into three sets: neurons have the tendency to change state from zero to one, we call them the hot set H, neurons have the tendency to change from one to zero, we call them the cold set C and others keep their state.

The most widely used network firing sequence [1-2] is uniform random choices to prevent network from looping in a sequence of states. In this study, we adopt a biased random firing methodology. The choice is random biased towards the neurons of the highest momentum to the change both in hot and cold sets.

The choice of the next neuron to change is as following: let us assume H^S is the set of hot neurons sorted descending on the value of net_j and C^S is the set of cold neurons sorted ascending on the values net_j then, firing of neurons alternated between the cold and hot with firing index as:-

For hot: $b_H = \min(L(1 + r^2|H^S|), |H^S|)$.

For cold: $b_C = \min(L(1 + r^2|C^S|), |C^S|)$.

Where: -

r Random variable uniformly distributed between 0, and 1.

L Floor integer function.

$||$ Set cardinality.

The network comes to a stable state when both sets are null. The stable state of the network provides a solution which is a local minima in the energy surface. Looking for more solutions to choose from a network re-initialization required.

VI. NETWORK RE-INITIALIZATION.

Network re-initialization required to look more satisfactory solutions. The most popular re-initialization is another random draw. The random re-initialization may produce a state close to a former initialization which maximizes the probability to end in same local minima. The initialization adopted in this study based on min-max [15] to assure start in farthest point in the hyperspace from all previous initializations, to maximize the probability of ending up with different solution. The procedure gives an indicative measure to the coverage of the space by the least distance between initializations. To summarize the re-initialization procedure let us assume we had n number of trials with unsatisfactory solutions and a set of historical neurons initialization vectors set: $H = \{h_1, h_2, \dots, h_n\}$ then we follow the following steps:-

$h_{n+1}(1) = \overline{h_n(1)}$, (complement of)

$\forall k = 2 \rightarrow NMT$ Find the : distance set ,and the minimal distance set as following:-

$$D = \left\{ d_i : \sum_{j=1}^{k-1} |h_i(j) - h_{n+1}(j)| \quad \forall i = 1 \rightarrow n \right\},$$

$H_m = \{h_i : d_i \leq d_j \quad \forall d_j \in D\}$, then set current bit according to the following rule:-

$$h_{n+1}(k) = \begin{cases} 0 & S > 0 \\ \overline{h_n(k)} & S = 0 \\ 1 & S < 0 \end{cases},$$

where

$$S = \sum_{\forall i \in H_m} h_i(k) - \sum_{\forall i \in H_m} |h_i(k) - 1|$$

Figure 1 shows the overall process of the proposed network working in a multi- constrains scheduling problem.

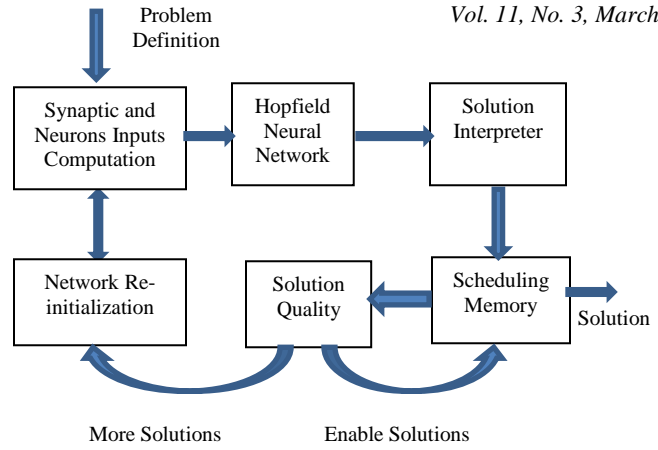


Fig. 1. Hopfield Neural Network Scheduler

VII. TESTS AND RESULTS

In this section, the network model tests on variety of task scheduling problems. One of these problems presented next followed by summary over others.

Figure 2 shows a precedence diagram for sixteen-task problem taken from [11]. The tasks are marked with “task number/ required time units/task type group”. The tasks are to be scheduled on two processing elements. The network formed and set to initial state according to the former model. The result of network run is presented in fig. 3. Fig. 4 shows the number of violations, for hard constrains, per network iteration. From the figure we can see that the network came to a valid solution, after 24 iterations.

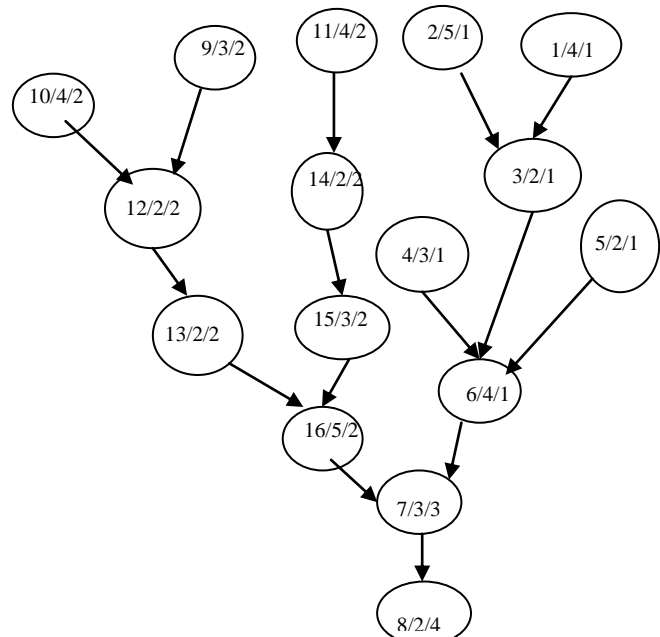


Fig. 2. The precedence diagram for the 16 task problem

Timer Units															
P1	T10	T11	T14	T9	T15	T12	T13	T16							
P2	T5	T4	T1	T2	T3	T6							T7	T8	

Fig. 3 Solution provided for the 16 tasks problem

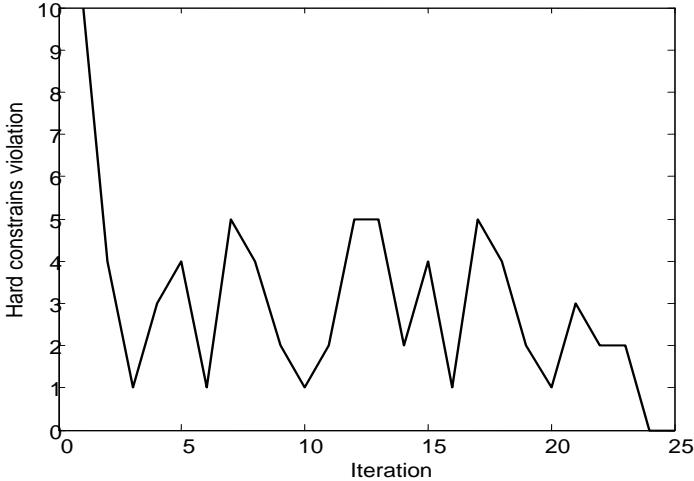


Fig.4. Network iterations in finding the solution

In the following table we summarize the results of running the network on 10 problems. The problems parameters (the number of tasks, number of processing elements, and the task processing time) taken from random normal distribution draws with mean and standard deviation as follow: (10, 2), (4, 1), and (5, 2.5). The precedence constrains generated using random generator as sequence of pairs (task1 proceeds task2) with rejection to invalid choices. The given scheduling time span computed by distributing the tasks equally on the processors and assuming that the lengthiest jobs will run on the same processing element then adding more 25%. Table 1 presents the results with only precedence constrains compared to the heavy brute search.

Table 1 Results of ten problems with only precedence constrains

Index	No. of Taks	NO. of processors	Given Schdlnng Time	Schdlnng length of first valid Slutn	Iterations of to first valid Solution	Trials count to global minima proposed: random	Minimum schedule length
1	9	4	24	14	1	-	-
2	6	4	15	13	1	-	-
3	10	4	30	20	1	1:3	18
4	9	4	17	10	1	-	-
5	8	3	22	15	2	2:2	11
6	9	4	26	16	2	-	-
7	10	5	25	18	3	3:5	14
8	12	3	31	22	1	2:5	19
9	8	4	24	14	1	-	-
10	11	5	23	17	1	1	14

From the table we easily infer that the networks converged to a solution within few numbers of iterations. In

50 percent of the time the network came to a deep local minimum. In 50 percent of the time the network came to the global minima. The cases in which the network did not converge to the global minima or deep local minima the solution is found to be within few re-initialization trials of the network.

Table 2 shows the results of running another 10 problems with constrains. Constrains generated, also, based on random uniform distribution, rejecting invalid values. The number of constrains is a random number drawn from normal distribution of mean equal to 1.5 times the number of tasks and standard deviation of 50% of that mean. The proposed firing sequence is faster overall 25 percent compared to random firing.

The network took more iteration to reach the first valid solution compared to precedence only cases. The probability of reaching global minima in the constrained problems appears to be higher than the unconstrained problems which points to the fact that constrains eliminate many of the valid unconstrained solutions.

Table 2. Results of constrained Problems

Index	No. of Tasks	No. of proce ssors	No of const rains	Given Sched uling Time	Schdl lngth of first valid Slutn	Iterat ions Of first valid Slutn	Trials to gbl mna Prpsd : Rrnd	Mini mum sched ule lengt h
1	12	3	13	27	23	16	-	-
2	13	4	15	25	19	32	-	-
3	9	4	3	23	15	3	-	-
4	12	2	8	36	26	70	-	-
5	12	2	8	30	18	7	-	-
6	9	2	10	23	15	3	1:3	14
7	12	3	8	27	22	18	-	-
8	18	3	14	26	21	91	-	-
9	12	4	10	14	9	2	-	-
10	20	2	23	31	22	21	-	-

VIII. CONCLUSION

In This paper a task scheduling model for non-preemptive multi-constrained multi-processor problems based on Hopfield Architecture presented. Methodologies for computing constrains weighting factor, and network re-initialization proposed. The proposed neural network solution dose not requires a predetermined scheduling length. The constrains used in the study are task time, precedence, resources conflict, task dead time, and favoring tasks of the same setup to run on the same processor beside the cost of schedule length.

The results of running the network on different problems show convergence to a valid solution within a fewer number of iterations. A selection of the shortest scheduling length from the first few solutions will be very close to the global minima or the global minima itself.

The major problem of this approach is network size and the simulation computations time. The hardware realization with simple and dense processing elements resolves this problem.

IX. REFERENCES

1. Hopfield, J. "Neural Computation of Decisions in Optimization Problems" *Biological Cybernetics*, 52, pp. 141-152, 1985.
2. J. Hopfield and D.W. Tank "Computing with neural Circuits a Model" *Science*, Vol. 233, pp. 625-633, Aug. 1986.
3. Gene A. Tagliarini and others "optimization using neural networks", *IEEE transaction on computers* vol. 40, No. 12, Dec 1991.
4. E. W. Page and G. A. Tagliarini "Algorithm development for Neural Networks" *Iee Transaction for Neural Networks* Vol. 1, pp 4-27 Mar.,1990.
5. Carlos Cardeira , Zoubir Mammeri "Neural Networks for Multi-processor real time scheduling" *Proceedings., Sixth Euromicro on Real-Time Systems*, pp 59 – 64,, 1994
6. Carlos Cardeira , Zoubir Mammeri "Handling the precedence constrains with Neural Network Based Real Time Scheduling", *Proceedings of Ninth Euromicro on Real-Time Systems*, pp 207 – 214, 1997.
7. Ruey-Maw Chen and Yueh-Min Huang "multi constraint Task Scheduling in Multi-processor by The Neural Network", *Tenth IEEE International Conference on Tools with Artificial Intelligence*, pp 288 – 294, 1998.
8. Xiuli Wang and Tihua Wa "Solving Multi-processor Job Scheduling with resource and Timing Constrains using Neural Networks". *Proceedings of IEEE TENCON '02 Conference on Computers, Communications, Control and Power Engineering*, pp 616 - 619 vol.1, 2002.
9. Flipe Araujo and Bernardete Rbeiro and Luis Rodrigues "A Neural Network for the shortest path computation" *IEEE Transaction on Neural Networks* Vol. 12, No. 5, Sept. 2001.
10. Simi P. Valsam and K.S. Swarup "Hopfield Neural Network approach to the solution of the economic dispatch and Unit Commitment" , *IEEE ICISIP 2004*.
11. Sin Ming Loo and B. Earl Wells, "task Scheduling in finite-Resource, Re-configurable Hardware/Software Co design Environment", *INFORMS journal on Computing*, pp. 1-23, 2005.
12. G. Bilbro, R. Mann, T. Miller W. Snyder, D.E. Van Den Bout and M. White , "Mean Field annealing and Neural Networks" *Advances in Neural Information Processing System* , Morgan-Kautmann PP. 91-98,1989.
13. Cohen, M. and Grossberg, S. "Absolute Stability of goal pattern formation and Parallel memory storage by Competitive neural networks " *IEEE transaction on Systems, Man and Cybernetics*, Vol. 13,pp.815-26, Sept/Oct. 1983.
14. WelBing Gao and YI Xlong "Absolute Stability of Asymmetric Hopfield Neural Network", *IEEE International Joint Conference on Neural Networks*, pp. 2195 - 2198 vol.3, 1991.
15. W. A. Porter and A. H. Abouali "Vector Quantization for Multiple Classes " *Information Sciences Journal*, Jan 1998.
16. H. El-Rewini, T. G. Lewis and H. H. Ali "Task scheduling in parallel and distributed Systems", *Prentice-Hall International Edition*, 1994.
17. MOSTAFA R. MOHAMED, MEDHAT AWADALLA "HYBRID ALGORITHM FOR MULTIPROCESSOR TASK SCHEDULING",

Comparative Analysis between Split and HierarchyMap Treemap Algorithms for Visualizing Hierarchical Data

Aborisade D. O. (Corresponding Author)

Department of Computer Science, College of
Natural Sciences, Federal University of
Agriculture, Abeokuta, (FUNAAB) Ogun State,
Nigeria.

Oyelade, O. J.

Department of Computer and Information
Sciences, Covenant University, Ota, Ogun State,
Nigeria.

Obagbuwa, I. C.

Department of Computer Sciences,
Lagos State University, Lagos, Nigeria.(LASU)

Oladipupo, O. O.

Department of Computer and Information
Sciences, Covenant University, Ota, Ogun State,
Nigeria.

Obembe O. O.

Department of Biological Sciences, Covenant
University, Ota, Ogun State, Nigeria.

Ewejobi, I. T.

Department of Computer and Information
Sciences, Covenant University, Ota, Ogun State,
Nigeria.

Abstract

We carried out comparative analysis between Split treemap algorithm and a more recently introduced treemap algorithm called HierarchyMap. HierarchyMap and Split are Treemap Visualization methods for representing large volume of hierarchical information on a 2-dimensional space. Split layout algorithm has been developed much earlier as an ordered layout algorithm with capability to preserve order and reduce aspect ratio. HierarchyMap is a newer ordered treemap algorithm developed to overcome certain deficiencies of the Split layout algorithm. The two algorithms were analyzed to compare their rate of complexity. They were also implemented using object-oriented programming tool and compared using a number of standard metrics for measuring treemap algorithms. Their implementation shows that

HierarchyMap and Split although maintain the same level of data ordering and usability but HierarchyMap algorithm has better aspect ratio, better readability, low run-time, and less number of thin rectangles compared to Split treemap algorithm. Since aspect ratio is an important metric for determining the efficiency of treemaps on 2-D and small screens, and the result of the analysis shows that HierarchyMap is better efficient than Split treemap algorithm, we conclude that HierarchyMap is more efficient than Split treemap algorithm.

Keywords: Treemap algorithm, Aspect ratio, HierarchyMap, 2-D space, Data Visualization.

I. Introduction

Hierarchical structure is a structure comprising a series of ordered class of elements or entities within a particular system. Hierarchical structures such as Family structure, a University Structure and Manual Directory have been found to be very useful in representing information in almost all systems of life. Arranging information in hierarchical structures has also been observed to be more useful in bringing out meaning in the system being represented to the observer more than other known ways of representing it. Representations in hierarchical structures also help to clearly reveal the relationship between the components in the system. Data that are modeled into such structures are referred to as Hierarchical data. It was later observed that Hierarchical structure is only efficient for representing small and manageable data items. Efforts geared towards improving the Visualization of hierarchical data especially when voluminous data items are involved brought to mind the concept of Treemaps in early part of the Nineties by [2]. Treemap involves turning a tree into a planar space-filling map. Treemap visualization method maps hierarchical information into a rectangular 2-dimensional display in a space-filling manner such that 100% of the designated display space is utilized. [3]. It is described as space-filling visualization method capable of representing large hierarchical collections of quantitative data [5]. It works by dividing the display area into a nested sequence of rectangles whose areas correspond to an attribute of the dataset, effectively combining aspects of a Venn diagram and a pie chart. With the development of algorithm for early treemaps like Slice and dice, and Cluster and ordered treemap algorithms like Strip, Split and HierarchyMap, very large volume of data

sets can be visualized on a 2D space like a computer screen with little or no difficulty. In this paper, a comparative analysis is made between a recently developed ordered algorithm called HierarchyMap and using metrics such as readability, aspect ratio, run time, and number of thin rectangles. The remaining sections are organized as follows; Section two reported the review of related literature, section 3 analyses the complexity of the algorithms and compares the two treemap algorithms (Split and HierarchyMap) using standard treemaps metrics, while section 4 discusses the implementation and results based on standard treemap metrics.

II. Related Works

From the time the idea of Treemaps was first conceived and original treemap developed to solve the problem of space usage by using the full display space to visualize the contents of the tree, many algorithms have been introduced to display hierarchical information structures [2]. These treemap algorithms in the order of their introduction and successive improvement include Slice and Dice, Cluster, Squarified, Pivot by Split Size, Pivot by Middle, Split Strip, and HierarchyMap treemap algorithm [8]. Of great importance to this paper are the ordered treemap algorithms like Pivot by middle, Pivot by Split Size, Strip, Split and HierarchyMap treemaps algorithms. The idea that lead to algorithms for ordered treemaps is that it is possible to create a layout in which items that are next to each other in the given order are adjacent in the treemap [6]. Treemap algorithm where the first step is to choose a special item, the pivot, which is placed at the side of rectangle R. In the second step, the remaining items in the list are assigned to three large rectangles that make up the rest of the display area. Finally, the algorithm is

then applied recursively to each of these rectangles [5]. This algorithm has some minor variations, depending on how the pivot is chosen. There are three pivot-selection strategies; the first is the algorithm where the pivot with the largest area is chosen. The motivation for this choice is that the largest item will be the most difficult to place, so it should be done first [5]. The alternate approaches to pivot selection are pivot-by-middle and pivot-by-split-size. Pivot-by-middle selects the pivot to be the middle item of the list i.e. if the list has n items, the pivot is item number $n/2$, rounded down. The motivation behind this choice is that it is likely to create a balanced layout. In addition, because the choice of pivot does not depend on the size of the items, the layouts created by this algorithm may not be as sensitive to changes in the data as pivot by size. Pivot-by-split-size selects the pivot that will split the list into approximately equal total areas. With the sub-lists containing a similar area, they are expected to get a balanced layout, even when the items in one part of the list are a substantially different size than items in the other part of the list. The Strip treemap algorithm is a modification of the existing Squarified Treemap algorithm [4]. It works by processing input rectangles in order, and laying them out in horizontal (or vertical) strips of varying thicknesses. It is efficient in that it produces a layout with better readability than the basic ordered treemap algorithm, and comparable aspect ratios and stability [5]. The inputs in a Strip treemap are the subdivision of rectangle R and a list of items that are ordered by an index and have given areas. As with all treemap algorithms, the inputs are a rectangle R to be subdivided and a list of items that are ordered by an index and have given areas. A current strip is maintained, and then for each rectangle, a check is done to know if adding the rectangle to the current strip will

increase or decrease the average aspect ratio of all the rectangles in the strip. If the average aspect ratio decreases (or stays the same), the new rectangle is added. If it increases, a new strip is started with the rectangle [5]. The result is the Split treemap which, like the Pivot, is a partially ordered algorithm. It produces a layout where the natural ordering of the data set is roughly preserved, while in most cases producing better aspect ratios than the Pivot and the Strip treemaps [6].

III. Method

Algorithms Complexity Analysis

This section describes the two treemap algorithms (Split and HierarchyMap) and their complexity analysis, as it helps to compare algorithms to see which one is better.

Split Algorithm:

Inputs to the algorithm are an ordered list, $L = \{l_1, l_2, \dots, l_n\}$ of items to layout and a rectangle, R , in which the items are distributed. Weight $w(L)$ is defined to be the sum of the sizes of all the elements in the list. The algorithm follows a recursive process, where L is split into two halves, L_1 and L_2 , such that $w(L_1)$ is as close as possible to $w(L_2)$. Noting that the ordering of the elements must not be changed. L_1 and L_2 are both ordered, and all the elements of L_1 have an index less than those of L_2 to give. $w(L_1) \approx w(L_2) \approx w(L)/2$ and $\forall l_i \in L_1, \forall l_j \in L_2: l_i \leq l_{i+1} \leq l_j \leq l_{j+1}$

$\alpha(R)$ is then defined to be the area of a rectangle R . The rectangle R is split, either horizontally or vertically depending on whether the width is bigger than the height, into two sub rectangles, R_1 and R_2 such that their areas correspond to the size of the elements of L_1 and L_2 , that is ;

$$\frac{\alpha(R1)}{\alpha(R)} = \frac{w(L1)}{w(L)} = \frac{\alpha(R2)}{\alpha(R)} = \frac{w(L2)}{w(L)}$$

Hence, recursively layout the contents of $L1$ and $L2$ in $R1$ and $R2$ according to the algorithm [6].

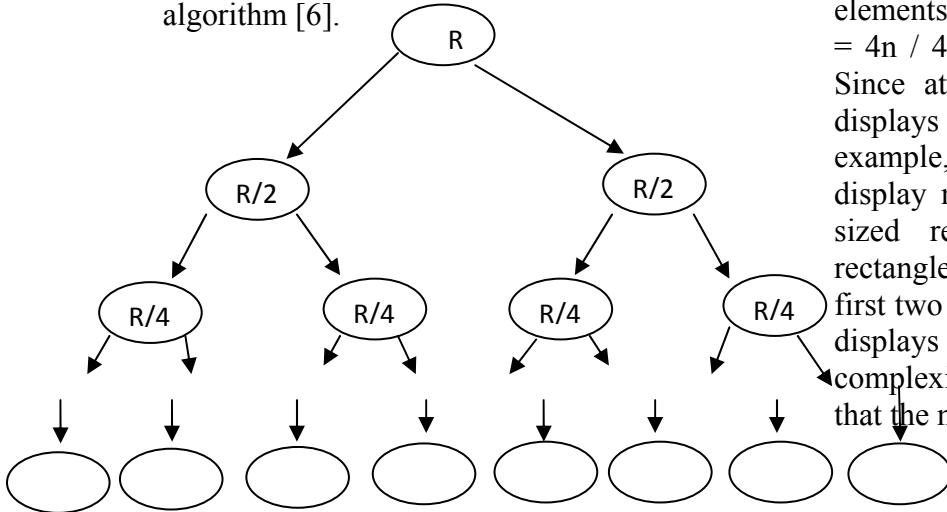


Figure 1: Split treemap recursion model

Here the Split treemap algorithm is modeled by a recursive tree where each circle represents a node (or rectangle R in which the items are distributed.) and the number written in the circle indicates the items (I_1, I_2, \dots, I_n) to layout. The first node stands for the original rectangle R to be sub-divided in layouts. The arrows indicate recursive calls made between nodes. Since the algorithm follows a recursive process, where L is split into two halves, $L1$ and $L2$, such that $w(L1)$ is as close as possible to $w(L2)$. The call to the next row shows the division of the first set of into 2 halves (i.e. $n / 2$). This is indicated by the two arrows at the top. In turn, each of these also makes calls to the next row for further sub-division of $n / 4$ each, and so forth until all the items are displayed. If the total the total number of items to be displayed in figure1 is taken to be n , and the total number of items in each level of the tree is n . The first row contains only one call the next row with an array of size n , so the total number of elements is n .

The second row has two calls to the next level (row) each of size $n / 2$. But $n / 2 + n / 2 = n$ and so again in this row the total number of elements is n . In the third row, we have 4 calls each of which is applied on an $n / 4$ -sized rectangle, giving a total number of elements equal to $n / 4 + n / 4 + n / 4 + n / 4 = 4n / 4 = n$. So again we get n elements. Since at each level of the tree rectangle displays the items from the input values. For example, the left node in level 1 has to display $n / 2$ elements. It splits the $n / 2$ -sized rectangle into two $n / 4$ -sized rectangle, calls recursively to display those first two nodes from the left in level 3), then displays all. This argument shows that the complexity for each row is $\Theta(n)$. And since that the number of levels in the is $\log(n)$.

We have $\log(n)$ rows and each of them is $\Theta(n)$, therefore the complexity of Split treemap algorithm is $\Theta(n * \log(n))$.

HierarchyMap treemap Algorithm

Inputs to the algorithm as ordered data in tree-like form. $infotree(treedata\ nodes)=T=\{t_1, t_2, t_3, \dots, t_n\}$ and a 2-D space divided into four equal rectangles.

Step 1: If the number of hierarchical items to be displayed is zero (i.e. $T=0$), then no display.

Step 2: If the number of hierarchical items to be displayed is 1 (i.e $T=1$), then

Set 2-D space to the item **Step 3:** If the number of items is greater than 1, divide the rectangular 2-D space into four equal sizes and recursively divided each of the resultant item into fours until all items in the list are exhausted. Such that $\forall t_i \in T_1, \forall t_j \in T_2, \forall t_k \in T_3, \dots, \forall t_n \in T_n : t_i \leq t_{i+1} \leq t_j \leq t_{j+1} \leq t_k \leq t_{k+1} \leq \dots \leq t_n \leq t_{n+1}$.

Step 4: an attribute of the each hierarchical item corresponds to an area of each of the nested rectangles defined as $\text{area}(R)$ in such a manner that their areas correspond to the size of the elements of $T_1, T_2, T_3,$ and T_4 where $\text{area}(R_1) \approx \text{area}(R_2) \approx \text{area}(R_3) \approx \dots \text{area}(R_n)$ [8].

total number of items to be displayed is n , and the total number of items in each level of the tree is $0.5n$. For example, the first row contains only one call. The second level with items of size n and hence has total number of elements is $0.5n$. The third level has two calls to the next level (row) each of size $n / 4$. Since $n / 4 + n / 4 = 0.5n$ and so again in this row the total number of

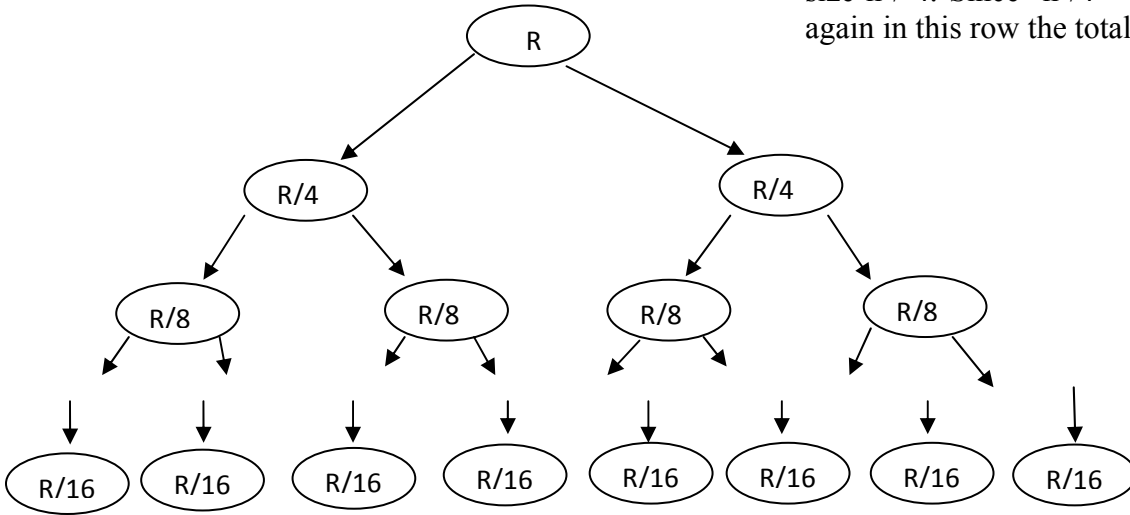


Figure 2: HierarchyMap treemap recursion model

In a similar manner, HierarchyMap algorithm is represented by a recursive tree in figure 2 where each circle represents a node (or rectangle R in which the items are distributed.) and the number written in the circle indicates the items (l_1, l_2, \dots, l_n) to layout. The root node stands for the original rectangle R to be sub-divided in layouts. The arrows indicate recursive calls made between nodes. HierarchyMap recursively processes the display of the items on rectangular space by sub-dividing the first rectangle R into four parts. $T_1, T_2, T_3,$ and T_4 where $\text{area}(R_1) \approx \text{area}(R_2) \approx \text{area}(R_3) \approx \dots \text{area}(R_n)$. The call to the next row shows the division of the first set of into 4 parts (i.e. $n / 4$). This is indicated by the two arrows at the top. In turn, each of these also makes calls to the next row for further sub-division of $n / 8$ each, and so forth until all the items are displayed. If the total the

elements is n . In the fourth level, we have 4 calls each of which is applied on an $n / 16$ -sized rectangle, giving a total number of elements equal to $n / 16 + n / 16 + n / 16 + n / 16 + n / 16 + n / 16 + n / 16 + n / 16 = 0.5n$, giving us $0.5n$ again. Since at each level of the tree, rectangle displays the items from the input values. For example, the left node in level 2 has to display $n / 4$ elements. It splits the $n / 4$ -sized rectangle into two $n / 8$ -sized rectangle, calls recursively to display those first two nodes from the left in level 3), then displays all. This argument shows that the complexity for each row is $\Theta(0.5n)$. And since that the number of levels in the is $\log(n)$. We have $\log(n)$ rows and each of them is $\Theta(0.5n)$, therefore the complexity of Split treemap algorithm is $\Theta(n * \log(0.5n))$ which is approximately equal to the complexity of the split treemap derived earlier as $\Theta(n * \log(n))$. But the constant multiplier in the HierarchyMap makes the difference. Since the constant multiplier is 0.5 , it means that it grows more slowly than

that of Split treemap and it is better and capable of quickly displaying items on rectangular space. This shows that the derived algorithm for Split is worse than that of HierarchyMap.

Implementation and Other Analysis Metrics

This section shows the implementation of the two algorithms (Split and HierarchyMap) and compares them on the basis of the standard treemap algorithm metrics like Aspect Ratio, Ordering, Readability, number of thin rectangles, Run time, and Usability. The behavior of each of the algorithm is observed with respect to the standard metrics when the treemap displays no (zero) item (Fig. 3a and Fig. 3b), displays between 10-15 items (Fig 4a and Fig 4b), displays between 20-25 items (Fig 5a and Fig 5b), displays between 30-60 items (Fig 6a and Fig 6b). Further discussion of these results is found in the remaining part of this Section.

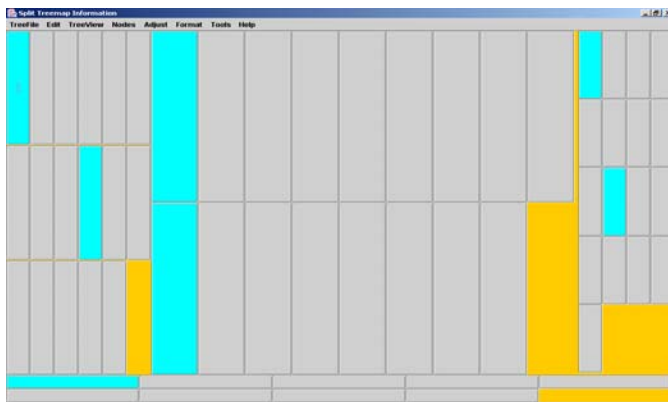


Figure 3a: Split treemap implementation with no item displayed (Aspect ratio is 2.92)

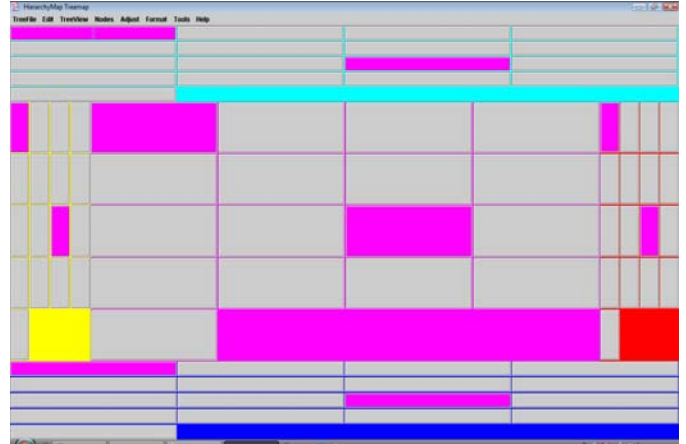


Figure 3b: HierarchyMap showing nested rectangles with no item displayed and Aspect Ratio of 1.72)

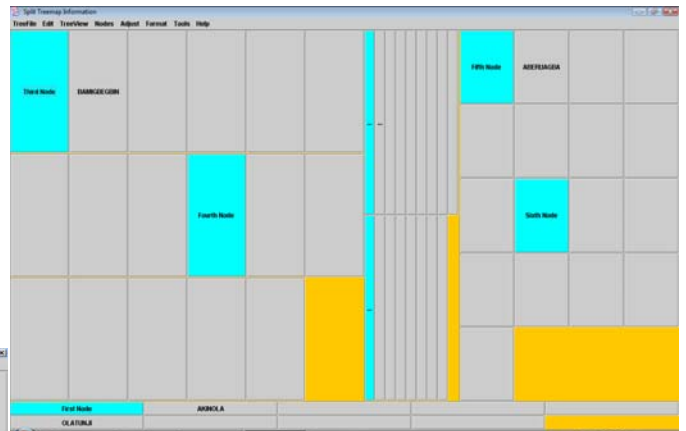
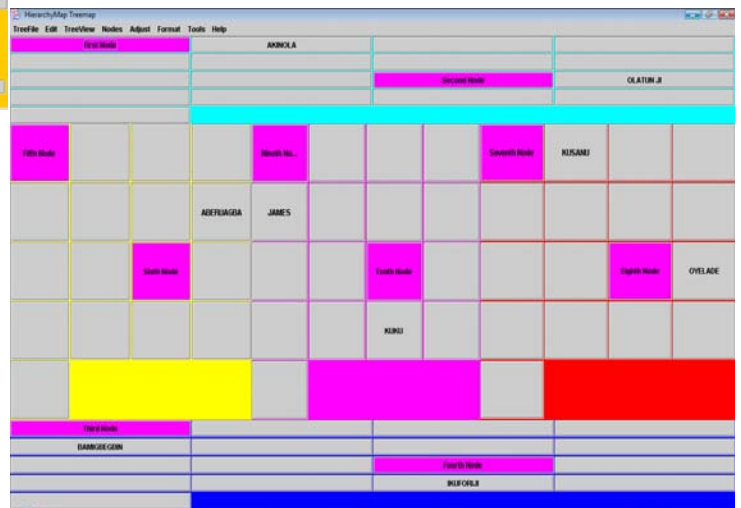


Figure 4a: Split Treemap with an average of 10 and 15 items giving Aspect Ratio 1.72



side. It is also defined as $\text{Max}(\text{Width}/\text{Height}, \text{Height}/\text{Width})$ of a rectangle. The lower the aspect ratio of a rectangle, the more nearly square it is. The aspect ratio for the two algorithms were determined using the same set of data. The Height/Width of each of the rectangles generated by each of the Treemap algorithm program are calculated (in cm). The result of the calculated values are added together and divided by four to get the average height and average width. The results of the calculated aspect ratios are represented in Figure 7 below.

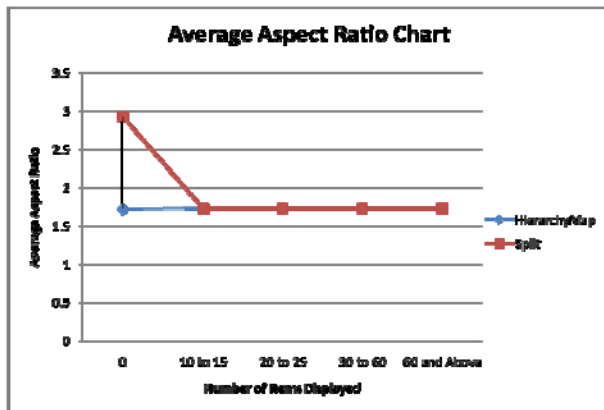


Figure 7: The graph plotted Average Aspect Ratio against Number of Items represents the relationship between Aspect ratio and the Number of rectangles generated in HierarchyMap and Split Treemap Algorithm.

The graph shows that HierarchyMap Treemap Algorithm has an Aspect ratio of 1.73 while Split Treemap Algorithm has Aspect ratio of about 2.92 when no rectangle is displayed. Both treemap algorithms maintain Aspect ratio of 1.73 when number of rectangles displayed are between 10, 60 and above in their treemaps. Hence, HierarchyMap is observed to have better aspect ratio than Split treemap.

4.0.2 Ordering

Ordering is a metric that determines the ability of the algorithm to create a layout

in which items that are next to each other in a given order are placed adjacent to each other (Berderson et al., 2002). Implementation of HierarchyMap and Split treemap algorithms as indicated above in the treemaps diagrams show that the two algorithms maintain items in the ordered manner.

4.0.2 Readability

Readability describes the measure of the number of times a user eye will have to change direction when scanning the treemap in order (Berderson et al., 2002). This test is used to measure how easy it is to locate a particular information between the layouts generated by the Split and HierarchyMap algorithms. In this experiment, twenty (20) persons (users) were carefully selected to scan through the treemap generated from the implementation of the two algorithms to locate a particular information. The time taken each of them was presented in Figure 8.

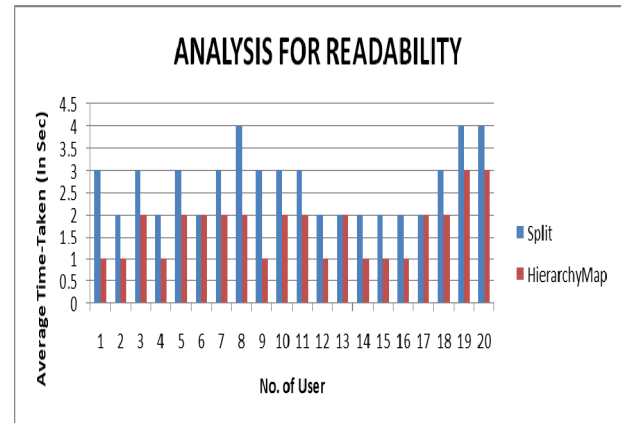


Figure 8: Analysis for Readability: Average time is plotted against the number of users for both Split and HierarchyMap

The graph shows that in HierarchyMap, readers use less time in most cases to locate information compared to Split treemap where more time is used in most cases by users to locate information of their choice on the treemap. This shows that HierarchyMap

has better readability than Split. This reflects the property of the Split layout, which changes direction more often than the HierarchyMap layouts that use several sub-lists instead of two. The results from the test indicate a slightly worse readability for the Split layout. HierarchyMap gives better readability because of the pivot. Assigning a pivot and then splitting the list in two, four, and then several parts generates a more consistent layout than the Split layout, which splits the list into two parts. Since the layout direction can alter between horizontal and vertical every time the list is split, the HierarchyMap algorithm is more predictable, since all the four sub lists will be laid out in the same directions, whereas the Split layout, with only two sub lists, will change direction more frequently.

4.0.3 Run Time

Run time is another important metric for evaluating treemap algorithm usability. In this case, run time for the implementation of the two algorithms is compared. This is done ten (10) different times for each algorithm on a Laptop Computer with the specification such as: Intel® Core™ 2 CPU T5200, 1.60 GHz, RAM 1015MB, 32-bit Operating System. The readings obtained are presented in Figure 9.

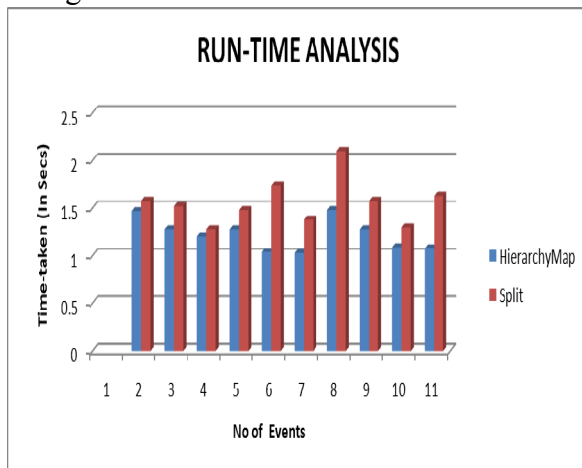


Figure 9: Column graph showing the Run-time Analysis for the two Algorithms

It is observed in Figure 9 above that HierarchyMap has a lower run time in all the events compared with Split treemap algorithm.

4.0.4 Number of thin rectangles

Another treemap efficiency metric very close to that of aspect ratio is the number of thin rectangles. The number of thin rectangles in a treemap determines the aspect ratio in the treemap. A treemap with a high number of thin rectangles has a high aspect ratio while a low number of thin rectangles has low aspect ratio. Figure 10 shows the number of thin rectangles generated by Split and HierarchyMap algorithms for different number of items displayed.

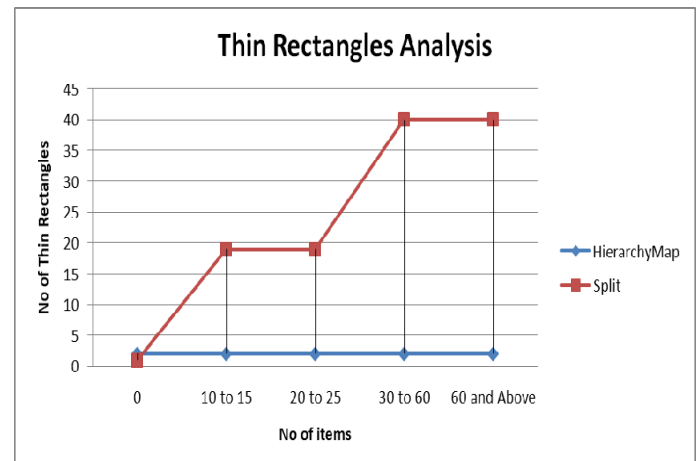


Fig. 10 Thin rectangle analysis

The thin rectangle analysis in Figure 10 shows that the number of thin rectangles generated by Split is more than the number of thin rectangles generated by HierarchyMap Treemap. Hence, Split has high aspect ratio than HierarchyMap treemap

4.0.5 Usability: HierarchyMap treemap algorithm by its implementation has been

observed to be capable of generating high volumes of hierarchical information on a 2-D space than Split treemap algorithm. It was interesting to observe that when the number of items to be displayed was more than 60, HierarchyMap treemap became more stable and did not flicker. Hence, HierarchyMap treemap algorithm is more efficient than Split algorithm in laying out hierarchical data in a 2-D space like a Computer screen.

V. Conclusion and Future Work

In this work, we compared the efficiency of two Ordered treemap algorithms called HierarchyMap and Split algorithms developed to represent hierarchical data on 2-D space. In comparing the two algorithms, the two algorithms were first analyzed measure their complexity. Then standard treemap algorithm metrics like aspect ratio, readability, ordering, usability, number of thin rectangles, and run time were also used as the basis of comparing them. The measure of complexity of the two algorithms shows that HierarchyMap is more efficient in laying out items on rectangular space and results of implementation using standard treemap algorithms metrics showed that HierarchyMap and Split although maintained the same level of data ordering and usability but HierarchyMap algorithm was observed to have better aspect ratio, readability, low Run-time, and less number of thin rectangles compared to Split treemap algorithm. Since aspect ratio is one of the most important properties when using treemaps on 2-D and small screens, HierarchyMap can therefore be said to be more efficient than the Split treemap algorithm. The future effort on this work is intended to improve on HierarchyMap algorithm to have better ordering and usability.

VI. References

- [1] Bruggemann-Klein, A. and D. Wood. Drawing Trees nicely with Tex. Electronic Publishing, 2(2):101–115, 1989.
- [2] B. Johnson and B. Shneiderman. Treemaps: A space-filling approach to the Visualization of Hierarchical Information Structures. *In Proc. of the 2nd International IEEE Visualization Conference*, pages 284–291, October 1991.
- [3] B. Shneiderman. Tree visualization with treemaps: A 2-D space-filling approach. *ACM Transactions on Graphics*, 11(1):92–99, September 1992.
- [4] Bruls S., M., Huizing, K., and Van Wijk, J., 2000. Squarified treemaps. *In Proceedings of the Joint Eurographics and IEEE/TCVGSymposium on Visualization (VisSym)*, 33–42.
- [5] Bederson, B., Shneiderman, B., and Wattenberg, M. 2002. Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. *ACM Transactions on Graphics* 21, 4, 833–854.
- [6] B. Engdahl, 2005. Ordered and Unordered Treemap Algorithms and Their Applications on Handheld Devices. Master's Thesis in Computer Science at the School of Computer Science and Engineering, Royal Institute of Technology year 2005.
- [7] D.E. Knuth. Fundamental algorithms. Art of computer programming. Volume 1. Addison-Wesley, Reading, MA, 1973.
- [8] D. O. Aborisade and O.J. Oyelade. HierarchyMap: A New Approach to Treemap Visualization of Hierarchical Data. *Global Journal of Computer Science and Technology*. Vol. 9 Issue 5, Online ISSN-0975-4172, Print ISSN 0975-4350. Pages 77-81. January, 2010.

[9] G.W. Furnas. Generalized fisheye views. In Proc. of ACM CHI'86, Conference on Human

Factors in computing systems, pages 16–23, 1986. Herman H, Maurer. Data Structures and Programming Techniques. Prentice- All Incorporation. 1977.

[10] J. Bingham and S. Sudarsanam. Visualising large hierarchical clusters in Hyperbolic space. Bioinformatics Chapter 16:pg. 660-661, 2000.

[11] Malin Koksal, Visualization of threaded discussions forums on hand-held devices, Masters Thesis at NADA, 2005.

Authors's Profile

Aborisade Dada Olaniyi is a PhD student and Lecturer in the Department of Computer Science, College of Natural Sciences, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria. He bagged his first degree in in B.Sc Mathematical Sciences (Computer Science option) in 2000 from University of Agriculture, Abeokuta, Ogun State, Nigeria and Msc in Computer Science of the University of Ibadan, Oyo State, Nigeria in 2007. His research interests are in the area of Human Computer Interaction (HCI) and Computer Information Security. He's a member of Microsoft Information Technology Academy (MITA) and Nigeria Computer Society (NCS).

Oyelade Olanrewaju Jelili recieved his Bachelor degree in Computer Science with Mathematics (Combined Hons) and M.Sc degree in Computer Science from Obafemi Awolowo University, Ile-Ife, Nigeria. He obtained his Ph. D in Covenant University, Ota, Nigeria. **Dr. Oyelade, O. J.** is a senior faculty member in the department of Computer and Information Sciences, Covenant University, Ota, Nigeria. His

[12] Russel Winder and Graham Roberts, Developing Java Software, John Wiley & Sons.

1998.

[13] S.K. Card, G.G. Robertson, and J.D. Mackinlay. The information visualizer, an Information workspace. In Proc. of ACM CHI'91, Conference on Human Factors in Computing Systems, pages 181–188, 1991.

[14] Wattenberg, M. 1999. Visualizing the stock market. In Extended Abstracts on Human Factors in Computing Systems (CHI), ACM Press, 188–189.

research interests are in Bioinformatics, Clustering, Fuzzy logic and Algorithms. He is a member of International Society for Computational Biology (ISCB), Africa Society for Bioinformatics and Computational Biology (ASBCB), Nigeria Society of Bioinformatics and Computational Biology (NISBCB), the Nigerian Computer Society (NCS), and Computer Professional Registration Council of Nigeria (CPN).

Obagbuwa Ibidun Christiana is a lecturer in the Department of computer science, Lagos state University Ojo, Lagos state, Nigeria. She obtained her first degree (B.Sc Computer Science) in 1997 from University of Ilorin, Ilorin, Kwara state. She proceeded to University Of PortHarcourt, Rivers state and obtain Degree of master in Computer science in 2005. She is currently working on her Doctoral degree (PhD) in Computer science. Her area of specialization include Computer security, Computational intelligence/softcomputing, Telecommunication & Networking and Databases. She is happily married with Three children. She is a member of Nigeria Computer Society (NCS), and Computer Professionals (Registration Council) of Nigeria (CPN)

Oladipupo O. O. received her Bachelor degree in Computer Science in University of Ilorin and M.Sc degree in Computer Science from Obafemi Awolowo University, Ile-Ife, Nigeria. She obtained her Ph. D in Covenant University, Ota, Nigeria. **Dr. Oladipupo, O. O.** is a senior faculty member in the department of Computer and Information Sciences, Covenant University, Ota, Nigeria. Her research interests are in Artificial Intelligence, Data Mining, and Soft Computing Technique. She is a member of Nigerian Computer Society (NCS), and Computer Professional Registration Council of Nigeria (CPN).

Itunuoluwa Ewejobi received her Bachelor's degree (First Class honours) in Computer Science and M.Sc degree in Computer Science from Covenant University, Ota, Nigeria. She is a Ph.D student in the Bio-informatics research group of the Department of Computer and Information Sciences, Covenant University, Nigeria. She is currently on a DAAD (German Academic Exchange Service) Sandwich Scholarship at the Ruprecht-Karls Universität, Heidelberg, Germany to carry out some part of her Ph.D research titled *“Transcription Factor(s)-Target Detection in the malaria parasite Plasmodium falciparum”*. Her research interests include; Artificial Intelligence, Transcriptomics and Modeling of biological systems and Algorithms.

Analysis of Network Security Policy – Based Management

Aliyu Mohammed
Universiti Teknologi Malaysia
Faculty of Electrical Engineering

Sulaiman Mohd Nor
Universiti Teknologi Malaysia
Faculty of Electrical Engineering

Muhammad Nadzir Marsono
Universiti Teknologi Malaysia
Faculty of Electrical Engineering

Abstract— Network security and management policy in information communication is the desire to maintain the integrity, validity and consistency of a system or network, its data and its immediate environmental infrastructure. Well established and secured infrastructure would help in no means making the network safe from all kinds of intrusion. Protecting all these resources is another very important concept that is needed of any computer system. Harnessing, accessing and configuring relevant security policies are very important roles to be played in safeguarding the complex network infrastructure. The paper therefore analysis some of the desired policies and assessment guidelines that should be followed by network administrators for effective and strong network management, security facilities and data optimization

Key words: Network Security; Management Policy; Intrusion; Domain Infrastructure .

1. INTRODUCTION

The research focus nowadays is on network security policy – based management as against the previous undertakings that are more on the equipments. The focus also tries to look at the new trend of malware detection, control and containment through network access control, and effective policy enforcement. The aim is to bring about policy improvement that will help refine the network requirement for proper control and so as to appropriately provide protection for the information security and its application targets.

The desire for security policy and enforcement is to maintain integrity, validity and consistency [1]. At present most of the studies in the field tend to focus towards the realization of effective business information system which emphasizes the integrity description of the users, process and its implementation [2].It has been put forward by a number of security and

integrity principles such as good factor transformation, authorized implementation as being the right basis for the establishment of security policy [3]. The validity conception of security policy and enforcement is to meet the needs of system security as presented by [4, 5],

this is an information security assessment process which generally comes through the abstraction model of information security. It identifies the impact of the assets of the user, and analyses their vulnerability, threat and risk factor. The given consistency of security policy requirements in the policy rules definition and implementation is to try and avoid possible areas of conflict, as they reflect the executable and compatible features.

This paper proposes an infrastructural enterprise for information security policy enforcement assessment model. The security domain partition and security domain policy establishment tries to analyze the characteristics of the network attributes and consider the network security policy enforcement capability to enable for effective handling of the growing number of threats and exploits.

The remainder of the paper is organized as follows. Section 2 introduces basic security model design pattern and security policy control management. Section 3 considers security policy implementation, enforcement and assessment. Section 4 concludes the paper.

11. NETWORK SECURITY POLICY MANAGEMENT

The concept of technicality and managerial approach in any system design are complementary measures that are quite necessary in considering any form of security management. The factor of insecurity is a necessary reflect on the organizational management and staffing. This becomes a serious question that needs to be handled in computer network security characteristics. Thus managers require paying special attention to all kinds of security issues. It is a novel idea for an organization to have a well planned and cohesive policy framework to enable a constructive structural enterprise based on a hierarchical security structure [8].

With a well placed security structure, the effective running and operation of the agency could be achieved. The responsibilities of the entire structure is to be able to articulate the following activities ; monitor the entire network operation and safety information; auditing at all levels of the network and conventional analysis of performance information; maintenance of safety equipment; security strategy planning, formulation and implementation; handling network security incidents and so on. Although, the security regulation of an enterprise information system is in two categories:

(i) The information security management in the Laws, regulation, enforcement view point, the rules is normally worked out by the secrecy Bureau and the Ministry of Public security. (ii) The enterprise's own system and computer network system; all tends to complement each other with the network security mechanisms.

A. Enterprise Security Mechanisms

The enterprise security mechanism is strongly based on the concept of (P2DR) Policy, Protection, Detection, and Response which is widely recognized by professionals and

entrepreneur [17]. In the overall context the model is aimed at controlling the system infrastructure under the guidance of the comprehensive use of protective equipment. Using detection and access control tools enables for understanding and accessing the state of the security system. This includes state of condition at the end- point. There is a correlation from the management point of view to understand the health condition of the end-point when they are connected to the enterprise. In this regards their access and the possible threat that they could exhibit is being looked into by the management [17, 19] as indicated on the Table 1 below.

Table 1. Access control at End- point.

Access Management		Threat Management	
Authentication	Access Rights	End point Integrity	Behavior Monitoring
Certificate / tokens	Virtual LAN	Health Checks	Flow Analysis
IEEE 802.1 x	Access control list	Antivirus Status	NBAD(Network-Based Anomaly Detection)
MAC – based	Firewall Policy	Anti spyware	IPS/ IDS
Web- Auth.	Application Software	No Trojans	Log management / SEM(Security Enforcement Management)
		Policy Checks Allowed software Required applications	

The more unique functions of P2DR are adding the time factor such as the time of Intrusion, response time to include prevention and enforcement, so that it becomes an ideal security framework [10].

B. Network Security Policy Information

Network security policy information is a model with a set of commands and rules that are used for control mechanisms relevant to the network security management policy implementation. The model has five (5) detailed policy rules which are explicit and effective. This is indicated in the Table 2 shown below.

Table 2. Classification of Policies

Policy Name	Condition	Action

Attack signature policy	Snort signature or other monitors	Alert packet, drop
Packet filtering policy	Source IP, Destination IP, Source port, Destination port, Protocol, TCP, flags, ICMP type, CMP code	Permit, Deny
Rate Limit Policy	Source IP, Destination IP, Source port, Destination port, Protocol	Transmit, Drop
Routing Control Policy	Destination IP	Drop
Alert Control Policy	Source IP, Destination IP, Source port, Destination port, Protocol, attack ID, Time Interval	Filtering, Sampling

The detection of known attacks such as the signature policy is referred to from given Snort or any other monitor rules of any current version in use. The detection policy is such that it drops packets and / or sends an alert to the server after which the attack signature policy is compared to a pattern of packets with policy rules [9], the packets filtering policy which decides to either permit or deny packets that are incoming from a firewall or router in accordance with the value of the packets header fields. Thus the rate limit policy prescribes control against excessive traffic in a router or a traffic control device; this is held as to increase the network performance.

The characteristic of routing control policy is such that it forwards a packet to a router's bit bucket for manipulating bad traffic. In other words, the routing control policy routes unwanted packets to null. The routing control policy works only on destination addresses, since it is really part of the forwarding logic. Lastly, the alert control policy controls transmission of alerts from security devices. This policy allows a policy enhancement point (PEP) to send the sample of alerts to the filter alerts. Thus, this indicates the brief rundown of the network security policy classification of the model.

Network Security Policy Model Analysis and Implementation [12, 13] includes:-

1. Analysis: (i) Management requirements business model, Organization structure, IT management
(ii) Technical requirements-existing/planned environment, Security issues.
2. Implementation procedures-
(i) Castle defense system- critical information, physical protection, operational system hardening, information access, external access
(ii) Defense planning- threat Assessment, risk assessment, user

awareness, monitoring procedure,
attack reaction, plans recovery
programs, organization system watch.

In the context of information security, the domain will help the research and application of system security. Simply speaking, security domain is a collection of entities and resources as a subset of the environment. They share a security policy set. Therefore the security domain is a cross-system and cross-platform collection which is based on the object oriented information security, information sharing and security demand, and the security policy can be for many safety-related needs and norms. [6]

C. Domain System Analysis

The standard architecture based on policy management has been put forward by the IETF, it is to be used to control the access policy. The core aspect of the structure is the policy enforcement point (PEP) and policy decision point (PDP) which mainly take account of the circumstances operating the network management to the router on the basis of RSVP protocol. The situation is such that it does not involve the adjustment function of the quality of service technology within the system. It also does not necessarily reflect the problem of detecting and solving conflict involving the policy. Meanwhile, most of the network management functions are tailored towards avoiding issues that tend to contribute to a waste of time and network resource when policies are frequently accessed [14,6]. A model which addresses the traditional security systems that include access control and framework is conceived in fig.1. It is designed to achieve the defining factor of management policy, transmission, sharing capability and optimized implementation procedure, with the Policy Repository (PR) and management tool (PMT) in place.

The use of GUI (graphical user interface) by administration helps to define clearly the policy rules and the handling of the role – function domain issues. The role function domain manages the roles of the human and other resources. The advantage is that it makes clear role expression position in the domain structure. These bring about an improved expansibility in the system through the well spread sub-domains. When objects transfer from one domain to another, its policies will be replaced automatically by the policies of the new domain environment. Thus, there is no need for modification of the policies and managing the relationship between the policies and object manually.

Generally, the function of the policy repository accessing control differs from that of the policy control system. This is as a result of the fact that the latter belongs to the operational characteristics of COPS, SNMP and others, while the former is a LDAP format based. The abstraction layer provides the required interface between policy and the domain service application, thus any request made in the policy handling point will be channeled to the corresponding policy according to the state of each of the preceding elements in the domain. The whole process is summarized as the embodiment of applying

the policy on idiographic equipment, and transferring it into SNMP message, COPS message or CLI command and sending it to the object in the corresponding domains such as router, firewall, scanner etc.

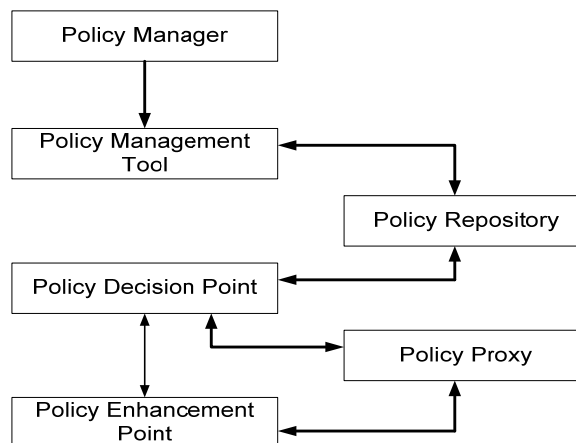


Fig. 1. Policy – Based Network Model

111. RESOURCES MANAGEMENT

In classification according to the domain, the target of the security policy is the role. The role is the property that distributes to the target resources, and has the abilities of abstract function expression. This means that the resources must support the function of role from semantic aspect. After the startup of the policy service, it can choose the role for the service, that is, choose the resources from the security equipment to provide service based on the matching of domain, role and function.

A. Risk Assessment

The critical component of the implementation of any information security framework is the performance of an appropriately-scoped risk assessment. Ideally this should be an iterative process involving input from several functions within the organization; often such a risk assessment, if performed at all, falls into the responsibility of the IT function, but in order to address all issues surrounding the organization the working group should include representatives of several functions. These include: IT, who know what information is stored and in what format; business lines, as the data owners, who can define which data is required for daily transactions and can define the sensitivity and confidentiality of each dataset and application; and legal and compliance, which can provide technical input into the regulatory and compliance frameworks within which the organization operates and explain external requirements over restricting access to systems and data [15].

The risk assessment should be regularly updated in order to address changes in the technical, regulatory and operating environments of the organization. A frequent failing in the compliance process is that reasonable conclusions have been drawn at a particular moment in time but that the area has not subsequently been readdressed and changes in the internal

environment, such as the implementation of new systems as a result of changing business requirements, or in the external environment, such as a modification in the data protection legislation in the territory, have not been taken into account

B. Management Constraints.

Traditional network management approaches lack the flexibility to configure/reconfigure the network elements according to network requirements unless it is accomplished manually. PBNM is promising network management paradigm to make administrative tasks easy and less complex. However there are certain constraints implied by the home network requirements, i.e lack of Standards- there is no standardized approach for management of heterogeneous home networks[16]: lack of Simplified Techniques-techniques and tools play a great role in network management but unfortunately there are not many simple techniques and tools available for managing home networks: lack of expertise-usually lack of technical skills and the level of expertise of typical Home area Networks (HAN) where users in the domain of network management makes it more complex because traditional approaches require high level skills and domain knowledge. Static Configurations- static configurations of network resources make network management static as well, which presents lack of adaptability in the network with the change in network requirements.

IV. CONCLUSION.

The implementation and optimization of reliable information security measures is a subject that can require a great deal of expertise, energy and resources to perform properly. No one-size-fits-all framework will be appropriate for all organizations, but by following a set of reasonable standard principles in a structured way, many organizations are able to define and meet their basic requirements in this respect.

V. REFERENCE

[1] Chenghua. T, Shuping, Y, and Zhongjie. C, "A Network Security Policy Model and realization Mechanism", *LNCIS 4318, Springer, Beijing, China*, 2006.
[2] Kittichote, R, and Hu, Liang, "Network Security Infrastructure Management" *IEEE*, 2009.
[3] Sourour, M, Adel, B, and Tarek, Abbas, "A Security Policy and Network Cartography based Intrusion Detection and Prevention Systems", *Journal of Information Assurance and Security* 4, 2009.
[4] Jiayi. Y, Anjia. M and Zhizhong. G, "Vulnerability Assessment of Cyber Security in Power Industry", *Power Systems Conference and Exposition, IEEE press, Atlanta*, 2006.
[5] W. Mees, "Risk management in coalition networks", *Information Assurance and Security*, *IEEE press, Manchester, UK*, 2007.

[6] J. Ns, et al, "Research for Information and communication Information Security Control Framework", *Information and communication*, vol. 29, No. 9, Sept. 2004.
[7] Seon-gyoung, Sohn, Jinoh, Jung-Chan Na, "Design of Network Security Policy information Model for Policy-based Network Management",
[8] Chenghua Tang et al, "Assessment of Network Policy Based on Security Capability", *International Conference on COMPUTER Science and Software Engineering, IEEE*, 2008
[9] Fuqian Shi, Hongbiao Xu and Haining. Wang, "A Representative Management Model of Network Security in Enterprise Information", *International Conference on Information, Innovation and Industrial Engineering*, IEEE, 2008
[10] T.S. Sobh. "Wired and Wireless Intrusion Detection System: Classification good characteristics and state-of-the-art?" *Computer Standards and Interfaces*, 28 (6), March, 2006
[11] Igor Kotenko, Vitaly Bogdanov, "Proactive Monitoring of Security Policy Accomplishment in Computer Networks", *IEEE International workshop on intelligent data acquisition and Advanced Computing System Tech. and Appl. Rome, Italy*, 2009
[12] SJ Ngobeni and M M Grobler, "Information Security Policies for Government Organizations", *the minimum Criteria, Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa*, 2009
[13] Blueprint: Enterprise Security Policy Design, "2003 Resolutions enterprises"
[14] Gu Yue- Sheng, Zhang Bao-jian, Yu Zhou, "Wireless Network Security Policy-based on Integrated Vulnerability anagement" *International Conference on Networking and Digital Society, IEEE*, 2009
[15] David Simms, "Information Security Optimization: From Theory to Practice" 2009. *Price water house coopers SA, Lausanne, Switzerland*.
[16] Annie Ibrahim Rana and Micheal O Foghlu, "New Role of Policy-based Management in Home Area Networks- Concepts, constraints and challenges", *TSSG, WIT, Ireland*, 2009

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 LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Mrs 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
Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Mr. P. Vasant, University Technology Petronas, Malaysia
Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Mr. Praveen Ranjan Srivastava, BITS PILANI, India
Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Mr. 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. Riktresh 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
Mr. Tirthankar Gayen, IIT Kharagpur, India
Ms. Huei-Ru Tseng, National Chiao Tung University, Taiwan

Prof. Ning Xu, Wuhan University of Technology, China
Mr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Mr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan
Prof. Syed S. Rizvi, University of Bridgeport, USA
Mr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India
Mr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal
Mr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P
Dr. Poonam Garg, Institute of Management Technology, India
Mr. S. Mehta, Inha University, Korea
Mr. Dilip Kumar S.M, University Visvesvaraya College of Engineering (UVCE), 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
Mr. Saqib Saeed, University of Siegen, Germany
Mr. 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
Mrs.J.Komala Lakshmi, SNR Sons College, Computer Science, India
Mr. 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, Faculty of Computer Science and Information Technology, University of
Malaya, Malaysia
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India
Mr. M. Azath, Anna University, India
Mr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh
Mr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia
Dr Suresh Jain, Professor (on leave), Institute of Engineering & Technology, Devi Ahilya University, Indore
(MP) India,
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia
Mr. Hanumanthappa. J. University of Mysore, India
Mr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)
Mr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria
Mr. Santosh K. Pandey, Department of Information Technology, 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 Proudhyogiki 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, M S Ramaiah Institute of Technology(M.S.R.I.T), Affiliated to Visweswaraiah
Technological University, Bangalore, India
M. Prabu, Adhiyamaan College of Engineering/Anna University, India
Mr. Swakkhar Shatabda, Department of Computer Science and Engineering, 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, College of Computer Science and Engineering, 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 Ahamed 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 Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India
Dr. Mana Mohammed, University of Tlemcen, Algeria
Prof. Jatinder Singh, Universal Institutiion 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 Singhanian 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 College of 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 Singhanian 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. Mahammad 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 Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh
Mrs Fazeela Tunnisa, 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 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 Praakash 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, SICSIR, 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, Lakshmanagarh, 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 Shafigh, 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 Rafrastara, 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. Dharendra Mishra, SVKM's NMIMS University, India
Prof. Shapoor 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 Golsorkhatabar 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 Institute of Engineering And Technology 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, NMIET, 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 Abdulllah Alsafi, 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 Abdullallah Alsafi, 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, Amaljiyothi 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. Sathyapraksh 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

CALL FOR PAPERS

International Journal of Computer Science and Information Security

IJCSIS 2013

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. 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 2013

ISSN 1947 5500

<http://sites.google.com/site/ijcsis/>