

**IJCSIS Vol. 13 No. 7, July 2015**  
**ISSN 1947-5500**

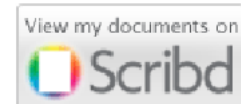
# **International Journal of Computer Science & Information Security**

**© IJCSIS PUBLICATION 2015**  
**Pennsylvania, USA**



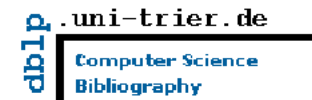
Cogprints

Google scholar



SciRate.com

CiteSeer<sup>x</sup> 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 2015 Issues

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 Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

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

**Deadline:** see web site

**Notification:** see web site

**Revision:** see web site

**Publication:** see web site

Context-aware systems  
Networking technologies  
Security in network, systems, and applications  
Evolutionary computation  
Industrial systems  
Evolutionary computation  
Autonomic and autonomous systems  
Bio-technologies  
Knowledge data systems  
Mobile and distance education  
Intelligent techniques, logics and systems  
Knowledge processing  
Information technologies  
Internet and web technologies  
Digital information processing  
Cognitive science and knowledge

Agent-based systems  
Mobility and multimedia systems  
Systems performance  
Networking and telecommunications  
Software development and deployment  
Knowledge virtualization  
Systems and networks on the chip  
Knowledge for global defense  
Information Systems [IS]  
IPv6 Today - Technology and deployment  
Modeling  
Software Engineering  
Optimization  
Complexity  
Natural Language Processing  
Speech Synthesis  
Data Mining

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



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

## Editorial Message from Managing Editor

*The International Journal of Computer Science and Information Security (IJCSIS) is a comprehensive and refereed, international publication featuring emerging research findings and industry solutions involving all aspects of computing and security. The editorial board is pleased to present the July 2015 issue. The purpose of this edition seeks to improve and expand existing research in experimental and theoretical research from both industry and academia in the broad areas of Computer Science, ICT & Security and related areas. As the editors of this issue, we are glad to see variety of articles focusing on the major topics of innovation and computer science; high performance computing; security; genetic algorithms; interdisciplinary applications & mobile technologies etc. This scholarly resource endeavors to provide international audiences with the highest quality research manuscripts and accounts of the constant evolution of information science and technology in whole. Researchers, academicians, practitioners and students will find this journal as a critical source of reference for all advanced technological applications and developments.*

*Over the last years, we have witnessed significant growth of IJCSIS in several key areas, include the expansion of scope to recruit papers from emerging areas of green & sustainable computing, cloud computing security, forensics, mobile computing and big data analytics. IJCSIS archives all publications in major academic/scientific databases and is indexed by the following International agencies and institutions: Google Scholar, CiteSeerX, Cornell's University Library, Ei Compendex, Scopus, DBLP, DOAJ, ProQuest, ArXiv, ResearchGate and EBSCO.*

*We thank and congratulate the wonderful team of publication staff members, associate editors, and reviewers for their dedicated services to select and publish extremely high quality papers for publication in IJCSIS. In particular, we would like to thank all authors for submitting their high quality papers to IJCSIS and the readers for continued support to IJCSIS by citing papers published in IJCSIS. Without their continued and unselfish commitments, IJCSIS would not have achieved its current premier status.*

*"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication."*

*For further questions please do not hesitate to contact us at [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com).*

*A complete list of journals can be found at:*

<http://sites.google.com/site/ijcsis/>

*IJCSIS Vol. 13, No. 7, July 2015 Edition*

*ISSN 1947-5500 © IJCSIS, USA.*

*Journal Indexed by (among others):*





**Bibliographic Information**

ISSN: 1947-5500

Monthly publication (Regular Special Issues)  
Commenced Publication since May 2009

**Editorial / Paper Submissions:**

**IJCSIS Managing Editor**

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

**Pennsylvania, USA**

**Tel: +1 412 390 5159**

# IJCSIS EDITORIAL BOARD

**Professor Yong Li, PhD.**

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

**Professor Ying Yang, PhD.**

Computer Science Department, Yale University, USA

**Professor Hamid Reza Naji, PhD.**

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

**Professor Elboukhari Mohamed, PhD.**

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

**Professor Mokhtar Beldjehem, PhD.**

Sainte-Anne University, Halifax, NS, Canada

**Professor Yousef Farhaoui, PhD.**

Department of Computer Science, Moulay Ismail University, Morocco

**Dr. Alex Pappachen James**

Queensland Micro-nanotechnology center, Griffith University, Australia

**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

**Dr. T. C. Manjunath**

HKBK College of Engg., Bangalore, India

**Dr. Naseer Alquraishi**

University of Wasit, Iraq

**Dr. Shimon K. Modi**

Director of Research BSPA Labs, Purdue University, USA

**Dr. Jianguo Ding**

Norwegian University of Science and Technology (NTNU), Norway

**Dr. Jorge A. Ruiz-Vanoye**

Universidad Autónoma del Estado de Morelos, Mexico

**Prof. Ning Xu**

Wuhan University of Technology, China

**Dr . Bilal Alatas**

Department of Software Engineering, Firat University, Turkey

**Dr. Ioannis V. Koskosas**

University of Western Macedonia, Greece

**Dr Venu Kuthadi**

University of Johannesburg, Johannesburg, RSA

**Dr. Kai Cong**

Intel Corporation, & Computer Science Department, Portland State University, USA

**Dr. Omar A. Alzubi**

Prince Abdullah Bin Ghazi Faculty of Information Technology  
Al-Balqa Applied University (BAU), Jordan

**Dr. Zhihan Iv**

Chinese Academy of Science, China



# TABLE OF CONTENTS

## **1. Paper 30061528: Evaluating Classifiers in Detecting 419 Scams in Bilingual Cybercriminal Communities (pp. 1-7)**

*Alex V Mbaziira, Ehab Abozinadah, James H Jones Jr  
Dept of Computer Science, George Mason University, Fairfax, VA, USA*

*Abstract* — Incidents of organized cybercrime are rising because of criminals are reaping high financial rewards while incurring low costs to commit crime. As the digital landscape broadens to accommodate more internet-enabled devices and technologies like social media, more cybercriminals who are not native English speakers are invading cyberspace to cash in on quick exploits. In this paper we evaluate the performance of three machine learning classifiers in detecting 419 scams in a bilingual Nigerian cybercriminal community. We use three popular classifiers in text processing namely: Naïve Bayes, k-nearest neighbors (IBK) and Support Vector Machines (SVM). The preliminary results on a real world dataset reveal the SVM significantly outperforms Naïve Bayes and IBK at 95% confidence level.

*Keywords-Machine Learning; Bilingual Cybercriminals; 419 Scams;*

## **2. Paper 30061525: Discrete Flower Pollination Algorithm for Resource Constrained Project Scheduling Problem (pp. 8-19)**

*Kirils Bibiks, Jian-Ping Li, Fun Hu  
Faculty of Engineering and Informatics, University of Bradford, Bradford, United Kingdom*

*Abstract* - In this paper, a new population-based and nature-inspired metaheuristic algorithm, Discrete Flower Pollination Algorithm (DFPA), is presented to solve the Resource Constrained Project Scheduling Problem (RCPSP). The DFPA is a modification of existing Flower Pollination Algorithm adapted for solving combinatorial optimization problems by changing some of the algorithm's core concepts, such as flower, global pollination, Lévy flight, local pollination. The proposed DFPA is then tested on sets of benchmark instances and its performance is compared against other existing metaheuristic algorithms. The numerical results have shown that the proposed algorithm is efficient and outperforms several other popular metaheuristic algorithms, both in terms of quality of the results and execution time. Being discrete, the proposed algorithm can be used to solve any other combinatorial optimization problems.

*Keywords- Flower Pollination Algorithm; Discrete Flower Pollination Algorithm; Combinatorial optimization; Resource Constrained Project Scheduling Problem; Evolutionary Computing.*

## **3. Paper 30061501: Cloud Computing Security: A Survey (pp. 20-28)**

*Amjad Mehmood, Muhammad Roman, M. Munir Umar, Institute of Information Technology, Kohat University of Science and Technology, Kohat  
Houbing Song, Department of Electrical and Computer Engineering, West Virginia University, USA*

*Abstract* — Cloud computing brings new possibilities for individuals and firms to utilize computing as a utility. It utilizes computing power irrelevant of user's location and devices. Thus it has become more demanding due to its performance, high computing power, cheapness, elasticity, accessibility, scalability and availability. Cloud computing offers ubiquitous operation with different security challenges. In this paper we discuss security challenges and vulnerabilities as well as limitations of current security modules. This paper will serve as a baseline guide for new researchers in this area.



*Index Terms—Cloud Computing Security, Infrastructure-as-a-Service(IAAS), Platform-as-a-Service(PAAS), Software-as-a-Service(SAAS), Private Cloud, Public Cloud, Hybrid Cloud, Trust, Vulnerabilities.*

#### **4. Paper 30061514: A Framework for Web Search Log Evaluation for Testing Information Search in Cloud (pp. 29-33)**

*Althaf Ali A, Bharathiar University, Coimbatore, Tamilnadu, India*

*Dr. R. Mahammad Shafi, Dept. Of MCA, Sree Vidyaniketan Engineering College, Tirupati, India.*

*Abstract* — To meet the rapid growth of cloud technologies, many web information provider application are developed and deployed, and these applications run in the cloud. Because of the scalability provided by the clouds, a Web application can be visited by several millions or billions of users. Therefore, the testing and evaluation of performance of these applications are becoming increasingly important. Web application usage log evaluation is one of the promising approaches to tackle the performance problem by adapting the content and structure of application to the needs of the users by taking advantage of the knowledge acquired from the analysis of the users searching activities from the web search logs. We propose a framework for web search log evaluation using classification and clustering method for effective testing information search in cloud. It also provides an information search ranking method to refine and optimizes the search evaluation process. We evaluate the proposed approach through implementing a web proxy in a server to record the user search logs and measure the retrieval precision rate for different users. A rate of 25% precision improvement is observed using different cluster testing for different users.

*Keywords-* Cloud, Web Search, Web Log, Classification, Clustering, Information Search, Testing.

#### **5. Paper 30061527: An Integrated Mobile Application for Enhancing Management of Nutrition Information: Case Study of Arusha Region (pp. 34-38)**

*Neema Mduma, Khamisi Kalegele, School of Computation and Communication Science &Engineering, Nelson Mandela African Institution of Science and Technology, Arusha, Tanzania*

*Abstract* — Based on the fact that management of nutrition information is still a problem in many developing countries including Tanzania and nutrition information is only verbally provided without emphasis, this study proposes mobile application for enhancing management of nutrition information. The paper discusses the implementation of an integrated mobile application for enhancing management of nutrition information based on literature review and interviews, which were conducted in Arusha region for the collection of key information and details required for designing the mobile application. In this application, PHP technique has been used to build the application logic and MySQL technology for developing the back-end database. Using XML and Java, we have built an application interface that provides easy interactive view.

*Keywords - Nutrition information; MySQL; XML; Java; PHP; Mobile Application.*

#### **6. Paper 30061533: Automated Vehicle Tracking by GPS Modern Technology (pp. 39-48)**

*Fahmi Ahmed, Faculty of Engineering, University Of Development Alternative (UODA), Dhaka, Bangladesh*

*A.B.M. Rocknuzzaman, Faculty of Engineering, Universität Duisburg-Essen, Duisburg, Germany*

*Syed Foysol Islam, Faculty of Engineering, University Of Development Alternative (UODA), Dhaka, Bangladesh*

*Abstract* — This Research paper represents, Automatic vehicle tracking by GPS modern system. The results obtained in a purpose-designed computer software for track Vehicle's location and other data using by modern GPS or GLONASS technology. For this system need two types of product, one is technology product and another is web system.

*Keywords - Microcontroller, IC Max232, GPS Antenna, GPRS and GSM Antenna, Satellite Communication, UART Communication, GPS, GPRS, GSM, Web Application, Google API, Open Street AP.*

## **7. Paper 30061543: Using J48 Tree for Value-Based Customer Relations Management (CRM) (pp. 49-55)**

*Marzieh mohammadi, Department of Computer Engineering, Najafabad branch, Islamic Azad University, Isfahan, Iran*

*Hamid Rastegari, Department of Computer Engineering, Najafabad branch, Islamic Azad University, Isfahan, Iran*

*Abstract* — One of the main important issues in critical to retail success is decision support methods for marketing decisions. Different data mining techniques can be suitable for targeted marketing and efficient customer segmentation. Mainly over data mining, the extraction of hidden predictive pattern from datasets organizations can recognize forecast future behaviors profitable customers, and assist firms to create proactive, knowledge-driven choices. The mechanized, future-oriented analyses is possible with data mining move outside the analyses of previous events usually provided with history-oriented tools like decision support systems. Data mining techniques response business requests that in the previous were too time consuming to follow. However, the responses to these requests create customer relationship management probable. Therefore, in this paper, a model base on the classification of J48 tree and feature selection is proposed to predict precise marketing performance. The propose model is evaluated conducted 3datasets and the results are compared with other algorithms such as Rep tree, Random tree and J48 tree. The experimental results show that the proposed model has higher precision and lower error rate in comparison of J48 tree, Rep tree and Random tree.

*Keywords*-Customer relations management (CRM); Feature Selection; Data mining; Classification; J48 tree

## **8. Paper 30111406: A Security Architecture for Software Defined Networks (SDN) (pp. 56-61)**

*Okunade Oluwasogo Adekunle, School of Science and Technology, National Open University of Nigeria, Victoria Island, Lagos, Nigeria.*

*Osunade Oluwaseyitan, Department of Computer Science, University of Ibadan, Ibadan, Nigeria.*

*Abstract* - Software defined networking is an emerging network architecture with promising future in network field. It is dynamic, manageable, cost effective, and adaptable networking where control and data plane are decoupled, and control plane is centrally located to control application and dataplanes. OpenFlow is an example of Software Defined Networking (SDN) Southbound, which provides an open standard based interface between the SDN controller and data plane to control how data packets are forwarded through the network. As a result of rapid changes in networking, network program-ability and control logic centralization capabilities introduces new fault and easily attack planes, that open doors for threats that did not exist before or harder to exploit. This paper proposed SDN architecture with some level of security control, this will provide secured SDN paradigm with machine learning white/black list, where users application can be easily test and group as malicious attack or legitimate packet.

*Keyword* - Software Defined Networking (SDN); OpenFow; Flow table; Security control; white/black list

## **9. Paper 30061502: Design and Extraction of Facial Components in Automated Face Recognition (pp. 62-65)**

*Ms. Roshani R. Patil, Mr. P. S. Mohod, Department of Computer Science and Engineering G. H. Raisoni College of Engineering Nagpur, Maharashtra, India*

*Abstract* - Face recognition presents a challenging problem in the field of image analysis and computer vision. Face recognition system should be able to automatically detect a face in an image. This involves extracts its features and then recognize it, regardless of lighting, expression, illumination, ageing, transformations (translate, rotate and scale image) and pose, which is a difficult task. This paper presents a framework for component- based face alignment and representation that demonstrates improvement in matching performance over the more common holistic approach to face alignment and representation. Active shape model (ASM) technique that has been used often for locating facial features in face images. The proposed scheme selects robust landmark points where relevant facial

features are found and assigns higher weights to their corresponding features in the face classification stage. For alignment and cropping Procrustes analysis is used. Multi-scale local binary pattern is used for matching automated face image. In MLBP per-component measurement of facial similarity and fusion of per-component similarities is used. The proposed work is more robust to changes in facial pose and improves recognition accuracy on occluded face images in forensic scenarios.

*Keywords:- Active shape model, Multi-scale local binary pattern, Procrustes analysis, holistic method.*

#### **10. Paper 30061503: Software Reliability Estimation of Component Based Software System using Fuzzy Logic (pp. 66-71)**

*Gopal Prasad Jaiswal, Ram Nivas Giri, Department of Computer Science and Engineering, RITEE Raipur, Chhattisgarh, CSVTU Bhilai, Chhattisgarh, India*

*Abstract* — Software Reliability Modeling has been one of the much-attracted research domains in Software Reliability Engineering. Software reliability means provide reusable, less complex software, to perform a set of successful operation and his function within a provided time and environment. Software designers are motivated to develop reliable, reusable and useful software. In past, Object-Oriented Programming System (OOPS) concept is to be used in purpose of reusability but they are not providing powerful to cope with the successive changing as per requirements of ongoing applications. After that Component Based Software system (CBSS) is in floor. IT is based on reusability of his component with less complexity. This paper presents a new approach to analyze the reusability, dependency, and operation profile as well as application complexity of component-based software system. Here, we apply Fuzzy Logic approach to estimate the reliability of component-based software system with the basis of reliability factor.

*Index Terms*—Component, Object-Oriented Programming System (OOPS), Component Based Software system (CBSS), Fuzzy Logic, Fuzzy Inference System (FIS), Adaptive Neuro Fuzzy Inference System (ANFIS), Reliability, Application Complexity, Component Dependency, Operation Profile, Reusability, Fuzzification, Defuzzification, Reliability Model, Rule Based Model, Path Based Model, Additive Model, etc.

#### **11. Paper 30061536: Constructing an Add-in Tool for Enterprise Architect v7.5 To Measure the Quality of Object Oriented Design (Class Diagram) (pp. 72-85)**

*Laheeb Mohammed Ibrahim, Khalil Ahmed Ibrahim, Software Engineering, Mosul University, Collage of Computer Sc. & Math., Mosul, Iraq*

*Abstract* — Software design is very important stage in software engineering since it lies in the middle of the software development life cycle and costs can be reduced if corrections or improvements made in design phase. Some of the existing CASE tools do not have the ability to correct or improve software design like EA v7.5. The present study aims to construct a CASE tool that helps software engineers in design phase by assessing or evaluating the quality of that design using object oriented design metrics, use the developed CASE tool as add-in to work inside Enterprise Architect since it has no support for design metrics. So, this paper may be considered as an evolvement of such a well-known CASE tool like the Enterprise Architect. In this paper, three tools are developed. First, is “K Design Metrics tool (KDM)” as an add-in that works inside Enterprise Architect (EA) v7.5 which is a well-known, powerful CASE (Computer Aided Software Engineering) tool. KDM tool takes the XMI (XML Metadata Interchange) document for the UML class diagram exported by EA as input, processes it, calculates and visualize metrics, provides recommendations about design naming conventions and exports metrics as XML (Extensible Markup Language) document in order to communicate with other tools namely KRS (K Reporting Service) and KDB (K Database). A Second tool is K Reporting Service (KRS) “KRS” which takes XML document generated by KDM tool as input, parses it and gives a report. The report helps the project manager or the team leader to monitor the progress and to document the metrics. Hence KRS tool is integrated with Enterprise Architect. Lastly, K Database “KDB” which takes the same XML document generated by KDM tool as input, parses it and stores metrics in the database to be used as a historical data. KDB tool is also integrated with Enterprise Architect. Two object oriented design metrics models are used, namely MOOD (Metrics for Object Oriented Design) which measures

Encapsulation, Inheritance, Polymorphism and Coupling, and MEMOOD (Maintainability Estimation Model for Object Oriented software in Design phase) which measures Understandability, Modifiability and Maintainability. Both models are validated theoretically and empirically. These measurements allow designers to access the software early in process, make changes that will reduce complexity and improve the design. All three tools were developed using C# programming language with the aid of Microsoft Visual Studio 2010 as integrated development environment under Windows 7 operating system with minimum 4 GB of RAM and Core-i3 of CPU.

*Keywords-MOOD (Metrics for Object Oriented Design); MEMOOD (Maintainability Estimation Model for Object Oriented software in Design phase); UML (Unified Modeling Language); Object Oriented software; Enterprise Architect v7.5.*

## **12. Paper 30061517: Usability Evaluation Methods and Principles for the Web (pp. 86-92)**

*Joel Mvungi, Titus Tossy, Computer Science Studies Department, Mzumbe University, Morogoro, Tanzania*

*Abstract* - In order to determine the quality of any web application in the world, Usability is the one of the most important tool that one can use. Web analysis perform several inspections on the websites and software and use usability criteria to determine some faults on the systems. Usability engineering has being important tool for the companies as well, this is due to the fact that through usability engineering companies can improve their market level by making their products and services more accessible. Know days there some web application and software products which are complex and very sophisticated, hence usability can be able to determine their success or failure. However currently usability has been among the important goal for the Web engineering research and much attention is given to usability by the industry due to recognition of the importance of adopting usability evolution methods before and after deployment. Moreover several literature has proposed several techniques and methods for evaluating web usability. And however there is no agreement yet in the software on which usability evolution method is better than the other. Extensive usability evaluation is usually not feasible for the case of web development process. In other words unusable website increases the total cost of ownership, and therefore this paper introduces principles and evaluation methods to be used during the whole application lifecycle, so as to enhance usability of web applications.

*Keywords* - *Evolution methods, Web usability, Web usability principles, Development process.*

# Evaluating Classifiers in Detecting 419 Scams in Bilingual Cybercriminal Communities

Alex V Mbaziira  
Dept of Computer Science  
George Mason University  
Fairfax, VA, USA

Ehab Abozinadah  
Dept of Computer Science  
George Mason University  
Fairfax, VA, USA

James H Jones Jr  
Dept of Electrical Computer Engineering  
George Mason University  
Fairfax, VA, USA

**Abstract**— Incidents of organized cybercrime are rising because of criminals are reaping high financial rewards while incurring low costs to commit crime. As the digital landscape broadens to accommodate more internet-enabled devices and technologies like social media, more cybercriminals who are not native English speakers are invading cyberspace to cash in on quick exploits. In this paper we evaluate the performance of three machine learning classifiers in detecting 419 scams in a bilingual Nigerian cybercriminal community. We use three popular classifiers in text processing namely: Naïve Bayes, k-nearest neighbors (IBK) and Support Vector Machines (SVM). The preliminary results on a real world dataset reveal the SVM significantly outperforms Naïve Bayes and IBK at 95% confidence level.

**Keywords**—Machine Learning; Bilingual Cybercriminals; 419 Scams;

## I. INTRODUCTION (HEADING 1)

Cybercrime has evolved from misuse and, or abuse of computer systems to sophisticated organized crime exploiting the internet. The causes of increasing incidents of cybercrime are attributed to: widespread internet access, increasing volume of internet-enabled devices and integration of social networking in computing architectures. These global internet-driven computing architectures continue to expand and build on top of existing immeasurable vulnerabilities, which provide miscreants with low barriers to commit and profit from cybercrime.

There are numerous types of cybercrime. Some research categorizes cybercrime into content-based and technology-based crime [1]. Other studies provide elaborate classification of cybercrime to include offences against confidentiality, availability and integrity of information and information technology [2]. In each category is a list of crimes that offer cybercriminals incentives and tools with capabilities to exploit computer system vulnerabilities for high financial rewards. Criminals also use the internet to obtain sophisticated tools for exploiting their victims without being detected or apprehended. Cyberspace provides criminals with capabilities for using dissociative anonymity to assume fake identifies for committing crime [3]. However, with social media, the true identities of cybercriminals can be leaked when the actor's friends in the criminal social network do not implement the same levels of privacy to hide their identities.

This study extends work in a previous paper [4] by implementing machine learning algorithms to detect 419 scams within an actual bilingual cybercriminal community. The main contribution of this paper is evaluation of the performance of machine learning algorithms in detecting 419 scams an actual bilingual cybercriminal community in a social network . We use in English as well as English and Nigerian Pidgin to evaluate the classifiers using the unigram and bigram models. We use three classifiers to detect 419 scammers within this cybercriminal community namely: Naïve Bayes, Support Vector Machines and k-Nearest Neighbor. Support Vector Machines significantly out-performed the other classifiers on datasets comprising of both English and Nigerian Pidgin unigram and bigram models at 95% confidence level. This because Nigerian Pidgin vocabulary has fewer words compared to English hence Support Vector Machines tend to work well such datasets.

The rest of the paper is organized as follows: in Section 2 we discuss related work. In Section 3 we describe the dataset and criteria for evaluating the performance of these classifiers. In Section 4 we present the results and discussion of our experimental study and in section 5 we draw our conclusions.

## II. RELATED WORK

There is a growing body of research investigating the context and impact of cybercrime due to the increasing number of incidents and numerous vectors that criminals are exploiting to profit from crime [5], [6], [7], [8]. There are numerous types of cybercrime which are categorized as content-based and technology-based crime. Content-based cybercrime includes: scams, phishing, fraud, child pornography, spamming etc., while technology-based crime includes but is not limited to hacking, code injection, espionage [1]. In this section we review existing research on content-based crime in general but scams in particular. We also define scams and bilingual cybercriminal networks in context to this paper.

### A. Nigerian Bilingual Cybercriminals and 419 Scams

This paper investigates detection of 419 scams within a bilingual community of cybercriminals. The actors comprising the community of cybercriminals that we are studying was constructed into a graph in an earlier paper using publicly

leaked emails obtained from an online data theft service [4]. These scams are known as advance-fee fraud or 419 scams [9], [10]. 419 scams originated from Nigerian in the 1970s at smaller scale but escalated in the 1980s during the oil boom as posted letters and then transitioned to email in the 1990's with commercialization of the Internet [11]. With time the origin of 419 scam cells expanded to different West African countries like Ghana, Cameroon, Ivory Coast, Benin as well other parts of the world. Although these scams usually go unreported, a 2013 report revealed that victims lost \$12.7 billion during that year to this category of cyber-criminals [11].

Cybercriminals committing 419 scams speak at least two languages hence are bilingual. For purposes of this paper we use the term *bilingual cybercriminal community* to refer an online community of criminal actors that use English and Nigerian Pidgin to exploit victims using 419 scams. This because Nigeria as well as other West African countries with 419 scammers are very diverse countries with hundreds of local dialects. However, English and Nigerian Pidgin are the most popular and widely common spoken languages spoken in West Africa.

Nigerian Pidgin is an English-based pidgin comprising words from local Nigerian dialects and English. In Nigerian pidgin, the phrases are short compared to English while the English used in Nigerian pidgin does not follow proper grammar hence is broken English like any pidgin or Creole language.

### B. Content-based Cybercrime Detection

Various research has studied detection of different types of content-based cybercrime like fraud, phishing and spam [12], [13]. Wang et al., study spam in social networks to build a social spam detection framework that filters spam across multiple social networks namely: MySpace, Twitter and WebbSpam Corpus [14]. Bosma et al., develop a social spam detection framework that uses link analysis and this implemented on a popular social network [15]. Bhat et al., propose a community-based framework and apply ensemble classifiers to detect spammers within community nodes in online social networks [16], [17]. Other studies evaluate predictive accuracy of several machine learning algorithms like Support Vector Machines, Random Forests, Naïve Bayes, Neural Networks in predicting phishing emails [18], [19]. Other research investigates the extent at which malware and spam has infiltrated online social networks [20]. However, these studies have not tackled bilingual datasets with 419 scams which are obtained from an actual cybercriminal community and evaluated performance of machine learning algorithms in detecting such scams within online cybercriminal communities. 419 scams comprise work-at-home scams, high yield investment scams, lottery scams or rewards from pay-per-click online adds.

### C. Machine Learning

In this section we review supervised machine learning algorithms for our study. In supervised machine learning, the algorithms map inputs to specific outputs using input and output data [21]. We use three classifiers namely Naïve Bayes,

Support Vector Machines and Decision Trees to detect scam in a social network of multi-lingual Nigerian cyber-criminals because these classifiers have been well studied and applied to spam and malware classification problems.

*a) Naïve Bayes:* this a popular classifier which has been applied to a variety of learning problems that are investigating scams like phishing, spamming and injected malicious hyperlinks. The algorithm implements Bayes Theorem which assumes conditional independence in feature variables of a learning set to predict statistical outcomes [22], [23].

*b) Support Vector Machines:* this another popular algorithm and that uses hyperplanes in dimensional space to address classification problems. This algorithm has been used in studying spam, fraud, malware, and phishing [24] [25].

k-Nearest Neighbors (kNN): this is also popular algorithm that uses instance-based learning to predict outcomes in learning problems. With instance-based learning, the kNN algorithm looks at the k-nearest neighbors when determining which instance to predict [26].

## III. DATASET

We use a publicly leaked set of 1036 email addresses of Nigerian cybercriminals who are using an online data theft service called *PrivateRecovery* (which was formerly called *BestRecovery*) [27]. These cybercriminals are known for committing specific scams namely: advance-fee, online dating and *Nigerian letter* scams. Facebook lookups were conducted on each email address to identify corresponding public profiles of the criminal actors and their friends. The Facebook URLs of these actors was used in a previous paper to construct large graph of 43,125 criminal nodes [4]. These Facebook accounts of these criminal actors are real because the actors post and share a lot of personal information in form of text and photographs. The average number of friends for the 150 important criminal actors is 490 while the 4966 is the maximum number of friends these actors have. For this study, we used public data from 150 criminal nodes which had a high PageRank. During data collection, we did not engage with or friend the actors through their Facebook accounts.

### A. Dataset Description

For our experimental study, we first generate two primary datasets from records which are randomly selected from the 150 Facebook accounts with high PageRank scores. Primary Dataset 1 (PD1) has English only records while Primary Dataset 2 (PD2) has half of the records in English and the other half in Nigerian Pidgin as shown in Table 1. The data in each primary dataset is labeled and then preprocessed to remove all non ASCII characters, symbols and punctuation marks except for the apostrophes, which we escaped. The data used in our classification problem is in two languages namely English and Nigerian Pidgin both of which use Latin characters hence do not use special symbols or non ASCII characters which is typical in languages like French, Spanish etc that use such characters to emphasize accents for certain words. However, in the data there was some evidence of use non ASCII characters in form on text-based emoticons expressing emotion. We do



not stem the English words in the sub-datasets but use term frequency-inverse document frequency (tf-idf) to weight the words.

From each of the primary datasets we obtained two sub-datasets of unigram words and bigram words as shown in Table I. Sub-datasets (SD) A and B contains English unigram and bigram words respectively while Sub-datasets (SD) C and D has both unigram and bigram words respectively in both English and Nigerian Pidgin.

### B. Classifier Evaluation Metrics

Our study uses binary classification to train and test text instances in the datasets as either scam or not-scam. To evaluate our classifiers we use *Recall*, *Precision* and *F1 measure* on unigram and bigram word vectors. *Recall* measures the percentage of scam messages that are detected hence this metric determines how well a classifier performs in identifying a condition. *Precision*, however, measures how many of the scam messages are detected correctly hence this is a measure of probability that a predicted outcome is the right one [28]. F1 measure is a harmonic mean of precision and recall.

Let  $x_{ns \rightarrow ns}$  be the number of *not-scam* posts classified as *not-scam*,  $x_{ns \rightarrow s}$  be the number of *not-scam* posts misclassified as *scam*,  $x_{s \rightarrow ns}$  be the number of *scam* posts misclassified as *not-scam* and  $x_{s \rightarrow s}$  be the scam posts classified as *scam*. Therefore the equations for recall, precision and F1 will be:

$$Recall = \frac{x_{s \rightarrow s}}{x_{s \rightarrow s} + x_{s \rightarrow ns}} \quad (1)$$

$$Precision = \frac{x_{s \rightarrow s}}{x_{s \rightarrow s} + x_{ns \rightarrow s}} \quad (2)$$

$$F1\ Measure = \frac{2PR}{P+R} \quad (3)$$

### C. Experimental Setup

In this section we demonstrate how we obtain results on performance of the three classifiers in WEKA[29] using the four sub-datasets. To solve our classification problem we use three classifiers namely: Naïve Bayes (NB)[30], Support Vector Machines (LibSVM), and k-Nearest Neighbor (IBK) [26]. To obtain the results, each of the random sample sub-datasets is split into training and testing tests. 80% of the data in each sub-dataset is randomly allocated for training and 20% for testing. We also used 10-fold cross validation method to improve the performance of the classifiers. Using cross validation, each of the four sub-datasets were split up into 10 sets of equal proportion. Training was done on nine sets while testing is done of one. This process was repeated 10 times to ensure independence of the elements in the sample and also to minimize biases in the outcomes.

TABLE I. DESCRIPTION OF DATASETS USED IN ANALYSIS

PD #	SD #	Language	N-Gram Words	# Words
1	A	English	Unigram	2081
1	B	English	Bigram	12070
2	C	English & Nigeria Pidgin	Unigram	1875
2	D	English & Nigeria Pidgin	Bigram	3057

TABLE II. PRECISION, RECALL, F-MEASURE, ROC CURVE AREA, PRECISION-RECALL CURVE FOR ENGLISH UNIGRAM WORDS USING SUB-DATASET A

Classifier	Precision	Recall	F-Measure	ROC Area	PRC
NB	0.915	0.911	0.911	0.964	0.96
LIBSVM	0.886	0.885	0.885	0.947	0.945
IBK	0.833	0.78	0.771	0.822	0.811

TABLE III. PRECISION, RECALL, F-MEASURE, ROC CURVE AREA AND PRECISION-RECALL CURVE FOR ENGLISH BIGRAM WORDS USING SUB-DATASET B

Classifier	Precision	Recall	F-Measure	ROC Area	PRC
NB	0.72	0.565	0.473	0.895	0.883
LIBSVM	0.673	0.656	0.648	0.742	0.734
IBK	0.695	0.515	0.371	0.644	0.659

TABLE IV. PRECISION, RECALL, F-MEASURE, ROC CURVE AREA PRECISION-RECALL CURVE FOR ENGLISH AND NIGERIAN PIDGIN UNIGRAM WORDS USING SUB-DATASET C

Classifier	Precision	Recall	F-Measure	ROC Area	PRC
NB	0.964	0.964	0.964	0.994	0.994
LIBSVM	0.962	0.962	0.962	0.993	0.994
iBK	0.851	0.79	0.781	0.915	0.921

TABLE V. PRECISION, RECALL, F-MEASURE, ROC CURVE AREA AND PRECISION-RECALL CURVE FOR ENGLISH AND NIGERIAN PIDGIN BIGRAM WORDS USING SUB-DATASET D

Classifier	Precision	Recall	F-Measure	ROC Area	PRC
NB	0.887	0.861	0.859	0.981	0.981
LIBSVM	0.898	0.895	0.895	0.94	0.928
iBK	0.844	0.796	0.789	0.901	0.909

## IV. EXPERIMENTAL RESULTS

### A. Results of Evaluation Metrics

In this section we first present the experimental results for performance of the classifiers on unigram and bigram words for the four sub-datasets. To evaluate the classifiers we use precision, recall, F-measure, ROC Curve and PR-Curve on datasets .

Table II shows the results for performance of the three classifiers using sub-dataset A of English unigrams. The results in this table reveal has precision of 0.915, recall of 0.911, f-measure of 0.911, ROC Area of 0.964 and PRC of 0.96. LibSVM has a precision of 0.866, recall of 0.885, f-measure of 0.885, ROC Area of 0.947 and PR-curve of 0.945. IBK has a precision of 0.833, recall of 0.78, f-measure of 0.771, ROC-curve of 0.822 and PR-curve of 0.811.

Table III presents results of the 3 classifiers using sub-dataset B of English bigram words. Detailed results in this table indicate that Naïve Bayes has a precision of 0.72. recall of 0.565, f-measure of 0.473, ROC Area of 0.895 and PR curve of



0.883. Comparatively, LibSVM has precision of 0.673, recall of 0.656, f-measure of 0.648, ROC Area of 0.742 and PR-Curve of 0.734. IBK has precision of 0.695, recall of 0.515, f-measure of 0.371, ROC of 0.644 and PR-curve of 0.659.

Table IV shows results for the classifier performance on sub-dataset C which contains unigram words in both English and Nigerian Pidgin. The results indicate that Naïve Bayes has a precision of 0.964, recall of 0.964, f-measure of 0.964, ROC area of 0.994 and PR-curve of 0.994. LibSVM has a precision of 0.962, recall of 0.962, f-measure of 0.962, ROC area of 0.963 and PR-curve of 0.994. IBK has a precision of 0.851, recall of 0.79, f-measure of 0.781, ROC area of 0.915 and PR-curve of 0.921.

Table V indicates results for performance of the three classifiers on sub-dataset D which contains bigrams words in both English and Nigerian Pidgin. The results in this table indicate that LibSVM has a precision of 0.898, recall of 0.895, f-measure of 0.895, ROC Area of 0.94 and PR curve of 0.928. Naïve Bayes has a precision of 0.887, recall of 0.861, f-measure of 0.859, ROC area of 0.981 and PR-curve of 0.981. Finally, IBK has precision of 0.844, recall of 0.796, f-measure of 0.789, ROC area of 0.901 and PR-curve of 0.909.

### B. Classifier Performance Evaluation

In this section we evaluate performance of the classifiers to determine the best classifier for detecting scam within this community of bilingual cybercriminals using unigram and bigram models. We evaluate LibSVM against Naïve Bayes and IBK to establish the significance of results at 95% confidence level using the four datasets. To achieve this we use a 2-tailed T-test evaluate performance metrics of LibSVM against Naïve Bayes and IBK on the four sub-datasets. To perform this test, we run the experiment five times and for each run we perform 10-fold cross validation. During each run the instances are randomized and the dataset is split into 80% training test and 20% testing set.

The performance metrics that we use to evaluate performance of our classifiers on the sub-datasets are ROC area, PR-curve and f-measure. We develop several hypotheses to test significance of the outcomes of the classifiers predicting accuracy in detecting 419 scams on datasets with English only as well as English and Nigerian Pidgin using unigram and bigram models. We use  $H_0$  to represent the null hypothesis and  $H_1$  to represent the alternate hypothesis. We compare the performance of LibSVM with Naïve Bayes and IBK on the four sub-datasets.

TABLE VI. EVALUATING LIBSVM AGAINST OTHER CLASSIFIERS USING ROC CURVE AREA AT 95% CONFIDENCE ( $\pm$  FOR STANDARD DEVIATION)

SD #	LibSVM	LibSVM vs NB	Hypothesis ( $\alpha=0.05$ )	LibSVM vs IBK	Hypothesis ( $\alpha=0.05$ )
A	0.93±0.02	0.95 ±0.01	Not Reject	0.80 ±0.02	Reject
B	0.94±0.03	0.88 ±0.03	Not Reject	0.84 ±0.12	Not Reject
C	0.99±0.00	1.00 ±0.00	Accept	0.92 ±0.02	Reject
D	0.89±0.02	0.99 ±0.00	Accept	0.94 ±0.02	Accept

We evaluate classifier performance using ROC area as below:

- $H_0$ : LibSVM's ROC area is greater than IBK's ROC Area for English unigrams while for  $H_1$  : LibSVM's ROC area is not greater that IBK's ROC Area for English unigrams. We reject the null hypothesis  $H_0$  because LibSVM's ROC area is significantly worse at 0.8 with a standard deviation of 0.02 as shown in Table VI.
- $H_0$ : LibSVM's ROC area is greater than Naïve Bayes' ROC Area for both English and Nigerian Pidgin unigrams while for  $H_1$  : LibSVM's ROC area is not greater that Naïve Bayes' ROC Area for English and Nigerian Pidgin unigrams. We accept the null hypothesis  $H_0$  because LibSVM's ROC area is significantly better at 1.00 as shown in Table VI.
- $H_0$ : LibSVM's ROC area is greater than IBK's ROC Area for English and Nigerian Pidgin unigrams while for  $H_1$  : LibSVM's ROC area is not greater that IBK's ROC Area for English and Nigerian Pidgin unigrams. We reject the null hypothesis  $H_0$  because LibSVM's ROC area for both English and Nigerian unigrams is significantly worse at 0.92 and standard deviation of 0.02 as shown in Table VI.
- $H_0$ : LibSVM's ROC area is greater than Naïve Bayes' ROC area for English and Nigerian Pidgin bigrams while for  $H_1$  : LibSVM's ROC area is not greater that Naïve Bayes' ROC area for English and Nigerian Pidgin bigrams. We accept the null hypothesis  $H_0$  because LibSVM's ROC area for both English and Nigerian bigrams is significantly better at 0.99 as shown in Table VI.

$H_0$ : LibSVM's ROC area is greater than IBK's ROC area for English and Nigerian Pidgin bigrams while for  $H_1$  : LibSVM's ROC area is not greater that IBK's ROC area for English and Nigerian Pidgin bigrams. We accept the null hypothesis  $H_0$  because LibSVM's ROC area for both English and Nigerian bigrams is significantly better at 0.94 and standard deviation of 0.02 as shown in Table VI.

Here we continue the evaluation for classifier performance using PR area as shown below:

- $H_0$ : LibSVM's PR area is greater than IBK's PR area for English unigrams while for  $H_1$  : LibSVM's PR area is not greater than IBK's PR area for English unigrams. We reject the null hypothesis  $H_0$  because LibSVM's PR area for English unigrams is significantly worse at 0.79 and standard deviation of 0.02 as shown in Table VII.
- $H_0$ : LibSVM's PR area is greater than Naïve Bayes's PR area for English bigrams while for  $H_1$ : LibSVM's PR area is not greater than Naïve Bayes' PR area for English bigrams. We reject the null hypothesis  $H_0$  because LibSVM's PR area for

English bigrams is significantly worse at 0.86 and standard deviation of 0.02 as shown in Table VII.

- $H_0$ : LibSVM's PR area is greater than Naïve Bayes's PR area for English and Nigerian Pidgin unigrams while for  $H_1$ : LibSVM's PR area is not greater than Naïve Bayes' PR area for English and Nigerian Pidgin unigrams. We accept the null hypothesis  $H_0$  because LibSVM's PR area for English and Nigerian Pidgin unigrams is significantly better at 1.00 as shown in Table VII.
- $H_0$ : LibSVM's PR area is greater than IBK's for English and Nigerian Pidgin unigrams while for  $H_1$ : LibSVM's PR area is not greater than IBK's PR area for English and Nigerian Pidgin's unigrams. We reject the null hypothesis  $H_0$  because LibSVM's PR area for English and Nigerian Pidgin unigrams is significantly worse at 0.93 and standard deviation of 0.01 as shown in Table VII.
- $H_0$ : LibSVM's PR area is greater than Naïve Bayes's PR area for English and Nigerian Pidgin bigrams while for  $H_1$ : LibSVM's PR area is not greater than Naïve Bayes' PR area for English and Nigerian Pidgin bigrams. We accept the null hypothesis  $H_0$  because LibSVM's PR area for English and Nigerian Pidgin bigrams is significantly better at 0.99 as shown in Table VII.
- $H_0$ : LibSVM's PR area is greater than IBK's PR area for English and Nigerian Pidgin bigrams while for  $H_1$ : LibSVM's PR area is not greater than IBK's PR area for English and Nigerian Pidgin bigrams. We accept the null hypothesis  $H_0$  because LibSVM's PR area for English and Nigerian Pidgin bigrams is significantly better at 0.94 and standard deviation of 0.02 as shown in Table VII.

TABLE VII. EVALUATING LIBSVM AGAINST OTHER CLASSIFIERS USING PR CURVE AREA AT 95% CONFIDENCE ( $\pm$  FOR STANDARD DEVIATION)

SD #	LibSVM	LibSVM vs NB	Hypothesis ( $\alpha=0.05$ )	LibSVM vs IBK	Hypothesis ( $\alpha=0.05$ )
A	0.93 $\pm$ 0.02	0.95 $\pm$ 0.01	Not Reject	0.79 $\pm$ 0.02	Reject
B	0.94 $\pm$ .02	0.86 $\pm$ 0.02	Reject	0.82 $\pm$ 0.10	Not Reject
C	0.99 $\pm$ 0.00	1.00 $\pm$ 0.00	Accept	0.93 $\pm$ 0.01	Reject
D	0.89 $\pm$ 0.02	0.99 $\pm$ 0.00	Accept	0.94 $\pm$ 0.02	Accept

TABLE VIII. EVALUATING LIBSVM AGAINST OTHER CLASSIFIERS USING F-MEASURE AT 95% CONFIDENCE ( $\pm$  FOR STANDARD DEVIATION)

SD #	LibSVM	LibSVM vs NB	Hypothesis ( $\alpha=0.05$ )	LibSVM vs IBK	Hypothesis ( $\alpha=0.05$ )
A	0.86 $\pm$ 0.03	0.89 $\pm$ 0.02	Not Reject	0.76 $\pm$ 0.01	Reject
B	0.80 $\pm$ 0.05	0.48 $\pm$ 0.04	Reject	0.37 $\pm$ 0.02	Reject
C	0.94 $\pm$ 0.01	0.97 $\pm$ 0.01	Accept	0.79 $\pm$ 0.03	Reject
D	0.77 $\pm$ 0.04	0.85 $\pm$ 0.03	Accept	0.79 $\pm$ 0.04	Not Reject

We conclude the evaluation for classifier performance with f-measure as below:

- $H_0$ : LibSVM's f-measure is greater than IBK's f-measure for English unigrams while for  $H_1$ : LibSVM's f-measure is not greater than IBK's f-measure for English unigrams. We reject the null hypothesis  $H_0$  because LibSVM's f-measure for English unigrams is significantly worse at 0.76 and standard deviation of 0.01 as shown in Table VIII.
- $H_0$ : LibSVM's f-measure is greater than Naïve Bayes' f-measure for English bigrams while for  $H_1$ : LibSVM's f-measure is not greater than Naïve Bayes' f-measure for English bigrams. We reject the null hypothesis  $H_0$  because LibSVM's f-measure for English bigrams is significantly worse at 0.48 and standard deviation of 0.04 as shown in Table VIII.
- $H_0$ : LibSVM's f-measure is greater than IBK's f-measure for English bigrams while for  $H_1$ : LibSVM's f-measure is not greater than IBK's f-measure for English bigrams. We reject the null hypothesis  $H_0$  because LibSVM's f-measure for English bigrams is significantly worse at 0.37 and standard deviation of 0.02 as shown in Table VIII.
- $H_0$ : LibSVM's f-measure is greater than Naïve Bayes' f-measure for English and Nigerian Pidgin unigrams while for  $H_1$ : LibSVM's f-measure is not greater than Naïve Bayes' f-measure for English and Nigerian Pidgin unigrams. We accept the null hypothesis  $H_0$  because LibSVM's f-measure for English and Nigerian Pidgin unigrams is significantly better at 0.97 and standard deviation of 0.01 as shown in Table VIII.
- $H_0$ : LibSVM's f-measure is greater than IBK's f-measure for English and Nigerian Pidgin unigrams while for  $H_1$ : LibSVM's f-measure is not greater than IBK's f-measure for English and Nigerian Pidgin unigrams. We reject the null hypothesis  $H_0$  because LibSVM's f-measure for English and Nigerian Pidgin unigrams is significantly worse at 0.79 and standard deviation of 0.03 as shown in Table VIII.
- $H_0$ : LibSVM's f-measure is greater than Naïve Bayes f-measure for English and Nigerian Pidgin bigrams while for  $H_1$ : LibSVM's f-measure is not greater than Naïve Bayes' f-measure for English and Nigerian Pidgin bigrams. We accept the null hypothesis  $H_0$  because LibSVM's f-measure for English and Nigerian Pidgin bigrams is significantly better at 0.85 and standard deviation of 0.03 as shown in Table VIII.

As shown in Tables VI, VII and VIII, 8 of the null hypotheses are accepted while 9 hypotheses are rejected and 6 hypotheses are not rejected. All the 8 null hypotheses which

are accepted reveal that LibSVM significantly outperformed other classifiers on a unigram and bigram models that comprise both English and Nigerian Pidgin words. The rejected hypotheses reveal that IBK performed significantly worse compared to LibSVM mainly on the English only unigram and bigram models as well as on the unigram model comprising Nigerian Pidgin and English words. The 6 hypotheses that are not rejected were based on unigram and bigram model for English only words.

The LibSVM out-performed other classifiers on English and Nigerian Pidgin unigram and bigram model because these sub-datasets had fewer words in their vocabulary compared to the English words. This is because Nigerian Pidgin uses a limited vocabulary of words which are selected from both English and other local Nigerian dialects

## V. CONCLUSION

This study evaluated performance of three classifiers in detecting 419 scams within a bilingual cybercriminal community. The three classifiers we used are LibSVM, Naïve Bayes and IBK. We evaluated the performance of the three classifiers using both unigram and bigram models comprising and of English words as well as both English and Nigerian Pidgin words. In both models, LibSVM outperformed Naïve Bayes and IBK. We used a 2-tailed t-test at 95% confidence to evaluate the classifiers on both the unigram and bigram models of English words as well as both English and Nigerian Pidgin words. These results motivate future work to explore the use of ensemble learning in detecting scams in bilingual criminal communities.

## ACKNOWLEDGMENT

We would like to thank Prof Damon McCoy for his initial input and allowing us to use the publicly leaked emails from which we collected data used in the earlier paper. We also want to thank all the anonymous reviewers whose comments were used to improve this paper.

## REFERENCES

- [1] M. Thangiah, S. Basri, and S. Sulaiman, "A framework to detect cybercrime in the virtual environment," in *2012 International Conference on Computer Information Science (ICCIS)*, 2012, vol. 1, pp. 553–557.
- [2] ITU, "Understanding Cybercrime: A Guide For Developing Countries," International Telecommunications Union, Switzerland, Apr. 2009.
- [3] K. Jaishankar (Ed.), *Cyber Criminology: Exploring Internet Crimes and Criminal behavior*. Boca Raton, FL, USA: CRC Press, Taylor and Francis Group, 2011.
- [4] H. Sarvari, E. Abozinadah, A. Mbaziira, and D. McCoy, "Constructing and Analyzing Criminal Networks," *IEEE Secur. Priv. Workshop*, p. 8, 2014.
- [5] S. M. Nirakhi, R. V. Dharaskar, and V. M. Thakre, "Analysis of online messages for identity tracing in cybercrime investigation," in *2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec)*, 2012, pp. 300–305.
- [6] R. Bohme and T. Moore, "How do consumers react to cybercrime?," in *eCrime Researchers Summit (eCrime)*, 2012, 2012, pp. 1–12.
- [7] A. K. Kuzmin, "State and trends of Russian cybercrime in 2011," in *2012 4th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2012, pp. 933–939.
- [8] S. Afroz, V. Garg, D. Mccoy, and R. Greenstadt, "Honor among thieves: A common's analysis of cybercrime economies," in *eCrime Researchers Summit (eCRS)*, 2013, 2013, pp. 1–11.
- [9] J. O. Aransiola and S. O. Asindemade, "Understanding cybercrime perpetrators and the strategies they employ in Nigeria," *Cyberpsychology Behav. Soc. Netw.*, vol. 14, no. 12, pp. 759–763, Dec. 2011.
- [10] D. Engber, "Who Made That Nigerian Scam?," *The New York Times*, 03-Jan-2014.
- [11] 419 Unit of Ultrascan Advanced Global Investigations, "419 Advance Fee Fraud Statistics Released in July 2014," Ultrascan Advanced Global Investigations, Amsterdam, Jul. 2014.
- [12] R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in *2010 IEEE Symposium on Security and Privacy (SP)*, 2010, pp. 305–316.
- [13] P. K. Chan and S. J. Stolfo, "Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection," in *In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, 1998, pp. 164–168.
- [14] D. Wang, D. Irani, and C. Pu, "A Social-spam Detection Framework," in *Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*, New York, NY, USA, 2011, pp. 46–54.
- [15] M. Bosma, E. Meij, and W. Weerkamp, "A Framework for Unsupervised Spam Detection in Social Networking Sites," in *Proceedings of the 34th European Conference on Advances in Information Retrieval*, Berlin, Heidelberg, 2012, pp. 364–375.
- [16] S. Y. Bhat and M. Abulaish, "Community-based Features for Identifying Spammers in Online Social Networks," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, New York, NY, USA, 2013, pp. 100–107.
- [17] S. Y. Bhat, M. Abulaish, and A. A. Mirza, "Spammer Classification Using Ensemble Methods over Structural Social Network Features," in *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 2014, vol. 2, pp. 454–458.
- [18] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "A Comparison of Machine Learning Techniques for Phishing Detection," in *Proceedings of the Anti-phishing Working Groups 2Nd Annual eCrime Researchers Summit*, New York, NY, USA, 2007, pp. 60–69.
- [19] I. Fette, N. Sadeh, and A. Tomasic, "Learning to Detect Phishing Emails," in *Proceedings of the 16th International Conference on World Wide Web*, New York, NY, USA, 2007, pp. 649–656.
- [20] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting Spammers on Social Networks," in *Proceedings of the 26th Annual Computer Security Applications Conference*, New York, NY, USA, 2010, pp. 1–9.
- [21] C. Sammut and G. I. Webb, Eds., "Supervised Learning," in *Encyclopedia of Machine Learning*, Springer US, 2010, pp. 941–941.
- [22] T. Pedersen, "A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation," in *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*, Stroudsburg, PA, USA, 2000, pp. 63–69.
- [23] E. Frank, M. Hall, and B. Pfahringer, "Locally Weighted Naive Bayes," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2003, pp. 249–256.
- [24] C. Chang and C.-J. Lin, *LIBSVM: a Library for Support Vector Machines*. 2001.
- [25] C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research," *Comput. Hum. Behav.*, vol. 28, no. 3, pp. 1002–1013, May 2012.
- [26] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, Jan. 1991.
- [27] B. Krebs, "Krebs On Security: In-depth security news and investigation," *Spy Service Exposes Nigerian "Yahoo Boys"*.
- [28] H. He and E. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
- [29] E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. H. Witten, and L. Trigg, "Weka," in *Data Mining and Knowledge Discovery*

*Handbook*, O. Maimon and L. Rokach, Eds. Springer US, 2005, pp. 1305–1314.

- [30] George H. John and Pat Langley, “Estimating Continuous Distributions in Bayesian Classifiers,” in *Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo, 1995, pp. 338–345.

# Discrete Flower Pollination Algorithm for Resource Constrained Project Scheduling Problem

Kirils Bibiks, Jian-Ping Li, Fun Hu  
Faculty of Engineering and Informatics  
University of Bradford  
Bradford, United Kingdom

**Abstract-** In this paper, a new population-based and nature-inspired metaheuristic algorithm, Discrete Flower Pollination Algorithm (DFPA), is presented to solve the Resource Constrained Project Scheduling Problem (RCPSP). The DFPA is a modification of existing Flower Pollination Algorithm adapted for solving combinatorial optimization problems by changing some of the algorithm's core concepts, such as flower, global pollination, Lévy flight, local pollination. The proposed DFPA is then tested on sets of benchmark instances and its performance is compared against other existing metaheuristic algorithms. The numerical results have shown that the proposed algorithm is efficient and outperforms several other popular metaheuristic algorithms, both in terms of quality of the results and execution time. Being discrete, the proposed algorithm can be used to solve any other combinatorial optimization problems.

**Keywords-** Flower Pollination Algorithm; Discrete Flower Pollination Algorithm; Combinatorial optimization; Resource Constrained Project Scheduling Problem; Evolutionary Computing.

## I. Introduction

Resource Constrained Project Scheduling Problem (RCPSP) consists of a set of predefined tasks and resources and its main objective is to assign tasks to resources in such way, that overall project schedule is as cheap and short as possible. To make the schedule feasible, there are constraints that need to be satisfied.

Despite the simplicity of definition, RCPSP is one of the widely described combinatorial problems in the literature and has existed for at least 50 years [1]. Blazewicz et al. [1] describes RCPSP as a generalization of classical job-shop scheduling problem which belongs to the class of NP-hard optimization problems [2]. Kolisch [3] classified methods used for solving RCPCP as exact solution [4], Priority Rules-Based (PRB) [5] and metaheuristic approaches [6-8].

Exact methods guarantee to find an optimal solution if it exists. The most common exact method is the branch and bound algorithm [4, 10-11]. In the branch and bound algorithm a tree is generated, where each node represents a task. Sprecher and Drexl [12] claimed that those methods cannot be used to solve large scale problems, as the trees increase sharply with the increase of dimension sizes.

PRB methods employ one or more schemes to construct a feasible schedule. Panwalker and Iskander [5] surveyed a range of priority rules. Davis and Patterson [13] compared standard priority rules on a set of single-mode RCPSP and demonstrated that the heuristics' performance decreases when the constraints become too tight. After examining the most common priority rules, Browning [14] presented novel heuristics, based on tasks criticality and load balancing factors, which appeared to be more suitable for solving RCPSP. Lawrence and Morton [15] described priority rules by using a combination of project-, activity-, and resource-related metrics. Hildum [16] proposed priority rules that distinguish single- and multiple-priority rules approaches and outlined that a scheduler with

multiple priority rules shows better performance. Boctor [17] also had similar observations. Comparing with the exact solution methods, PBR methods can find solution in shorter time, however they cannot acquire global solution.

In the last decades the metaheuristic evolution-based computational methods have been getting a lot of attention and been used extensively to solve RCPSP. The metaheuristic methods start with initial solution and constantly improve it by successively executing operations which transform one or several solutions into others. There are many evolution-based metaheuristic methods, such Genetic Algorithm (GA), Simulated Annealing (SA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and so on.

Husbands [18] outlined the advances of GA for scheduling and illustrated the resemblance between scheduling and sequence-based problems. Davis [19] demonstrated the benefits of using a stochastic search. Hartmann [6] proposed another implementation of GA and suggested to use a GA variation where every gene composing a chromosome is a delivery rule. Mendes et al. [20] proposed to use the priority rules to represent chromosomes in a form of a list of priority values for all activities in the project. Montoya-Torres [21] used a multi-array object-oriented model to depict chromosomes. Shahsavari et al. [22] designed a GA using a three-stage process that utilizes design of experiments and response surface methodology. Alcaraz et al. [23] developed several new variations of GA for solving RCPSP, extending the representation and operator previously designed for the single-mode version of the problem.

Aarts et al. [24] described one of the first SA approaches for scheduling problems. Palmer [25] combined planning and scheduling in a digraph representation. Boctor [26] reported fairly good performances of SA approaches on Patterson problems. Nikulin and Drexler [27] used SA to solve an airport flight gate scheduling problem which was modelled as RCPSP. Bouleimen [28] proposed that the conventional SA search scheme is replaced by a new design that takes into account the specificity of the solution space of the project scheduling problems. Zamani [29] combined a SA and time-windowing process, where SA generates an activities schedule and time-windowing improves it.

PSO is another popular metaheuristic method. Zhang [30] demonstrated good performance of PSO in solving RCPSP. Anantathanvit and Munlin [31] extended the original PSO algorithm by regrouping agent particles within the appropriate radius of circle. Li [32] replaced the complicated updating equations of the traditional PSO with one GA crossover operation to make the process quicker and less resource demanding. Linyi [33] introduced an implementation of PSO with one-point crossover for RCPSP. Zhang et al. [34] developed a variation of PSO in which the activities sequence is encoded with a simple code rule by the code orderer.

One of the first suggested uses of ACO for RCPSP was made by Merkle [35]. An improved ACO approach for solving RCPSP was introduced by Luo [36]. Wang [37] embedded a project priority indicator into ACO as the heuristic function and solved the multi-project scheduling problem. Shou [38] used an ACO with two separate ant colonies employed, where forward scheduling technique is applied by first ant colony, while backward scheduling technique is applied by the second one. The modified ACO algorithm for precedence and resource-constrained scheduling problems was presented by Lo et al. [39].

More and more approaches for solving RCPSP are being proposed in the literature. Recently, a new nature-inspired metaheuristic method called Flower Pollination Algorithm (FPA) has been developed by Yang [40]. Based

on the work done in [40], the FPA has demonstrated to be a very efficient algorithm in finding global optima with high success rates. Yang [40] showed that FPA is superior to both PSO and GA in terms of efficiency and success rate. However, since the FPA was designed for solving the continuous optimization problems, in order to apply it for RCPSP, the algorithm's core logic needs to be changed. The aim of this paper is to present a modification of the original FPA called Discrete Flower Pollination Algorithm (DFPA) which was adapted for solving the combinatorial problems.

The subsequent parts of this paper are organized as follows: The mathematical formulation of the problem is outlined in Section II; The explanation of FPA is given in Section III; The modification of FPA for RCPSP is proposed in Section IV; Simulation results and performance comparison with other popular algorithms are detailed in Section V; Finally, the conclusions and plans for future work are outlined in Section VI.

## II. Mathematical Formulation of the Problem

The main objective of the RCPSP is to find optimal schedule with minimal duration by assigning a start time to each activity, with the precedence relations and the resource availabilities taken into account.

Activities are formalized by a finite set  $A=\{A_0, \dots, A_{n+1}\}$ , where  $n$  is the total amount of activities. Activities  $A_0$  and  $A_{n+1}$  are dummy activities and they represent the start and the end of the project respectively.

The duration of each activity is indicated by vector  $p=\{p_0, \dots, p_{n+1}\}$ , where the duration of activity  $A_i$  is represented as  $p_i$ . The duration of dummy activities is  $p_0 = p_{n+1} = 0$ .

The precedence relationship of one task to another is represented by  $E$ , such that  $(A_i, A_j) \in E$  means that activity  $A_j$  can only be executed after activity  $A_i$  has been completed. Precedence relationship can also be stated by the activity-on-node graph [41], in which nodes represent activities and transitions between nodes represent precedence relationships.

The resources are defined by a finite set  $R=\{R_1, R_2, \dots, R_q\}$  and the availability of each resource is represented as  $B=\{B_1, B_2, \dots, B_q\}$ . The resource  $R_k$  is called unary or non-shareable if its availability is  $B_k=1$ . If the availability of resource is  $R_k > 1$ , the resource is regarded as shareable and can be occupied by several activities.

To represent the activities' demands for resources, the notation  $\mathbf{b}$  is used. The amount of resource  $R_k$  per one time period during the execution of  $A_i$  is defined as  $b_{ik}$ .

The starting times of activities are abstracted by a schedule  $S$ , where  $S_i$  represents the start time of activity  $A_i$ .  $S_0$  is used as a reference point. It signifies the start of the project and is always assumed to be 0. The total duration of the project, or makespan of a schedule,  $S$  will be equal to the start time of the last activity  $S_{n+1}$ .

Taking into consideration all formulation presented above, the optimization problem can then be stated as finding a non-pre-emptive schedule  $S$  of minimal makespan  $S_{n+1}$  (1) subject to resource (2) and precedence (3) constraints.

$$\text{Min: } S_{n+1} \quad (1)$$

$$\text{Subject to: } \sum_{A_i \in A_i} b_{ik} \leq B_k \quad \forall R_k \in \mathcal{R} \quad \forall t \geq 0 \quad (2)$$

$$S_j - S_i \geq p_i \quad \forall (A_i, A_j) \in E \quad (3)$$

The  $A_i$  in (2) represents a set of non-dummy activities that need to be schedules and can be calculated using (4).



$$A_i = \{A_i \in A \mid S_i \leq t < S_{i+1} \mid p_i\} \quad (4)$$

### III. Flower Pollination Algorithm

Flower Pollination Algorithm (FPA) is a novel nature-inspired metaheuristic algorithm based on the flower pollination process of flowering plants, which was created by Yang in 2012 [40].

Flower pollination process is typically associated with the reproduction of flowers, when flower pollen is transferred by various pollinators, such as insects, birds, and other animals. Flower pollination can be of two types: abiotic and biotic. About 80% of flowering plants belong to biotic pollination. This means that most of pollen is transferred by pollinators, like insects or animals. The rest 20% belong to abiotic and they can pollinate without any involvement of pollinators.

Some of pollinators are very diverse and they tend to visit only specific flower species. Such flower regularity can be regarded as evolutionary advantage, as it maximizes the transfer of the flower pollen to the same plants, therefore maximizing the reproduction of the flowers which belong to the same species.

Pollination can be achieved in two ways: self-pollination and cross-pollination. Cross-pollination refers to a process when a pollination occurs from a pollen of a flower of a different plant, while self-pollination is the fertilization of one flower from the pollen of the same species flower. Cross-pollination occurs at long distances, and is done by pollinators like bees and flies, which behave accordingly to Lévy flights behavior [42], with fly distance obeying a Lévy distribution. Moreover, flower constancy can be considered as an increment step using the similarity or difference between two flowers. According to Yang and Deb [43], in some optimization problems, the search for new solution is more efficient via Lévy Flights.

From the biological point of view, the main objectives of the flower pollination are the survival of the fittest and optimal reproduction of plants.

Based on the characteristics of the flower pollination process described above, Yang established the following rules for the FPA:

- 1) Biotic and cross-pollination processes are considered as global pollination process; Pollinators in this processes behave according to Lévy flights behavior;
- 2) Abiotic and self-pollination are considered as local pollination process;
- 3) Pollinators like insects can develop flower constancy, which is equivalent to a reproduction probability that is proportional to the similarity of two flowers involved;
- 4) Switching between local and global pollinations is controlled by probability  $p \in [0, 1]$ .

With the rules outline above, the algorithm's pseudo-code can be formulated in Fig. 1.

---

```
Objective function:  $\min/\max f(x)$ ,  $x=(x_1, \dots, x_d)$ 
Initialize a population of  $n$  flowers
Find the best solution  $g$  in the population
Define a switch probability  $p$ 
Define  $\max\text{Generation}$ 

while (generation <  $\max\text{Generation}$ )
  for  $i = 1 : n$ 
    if  $\text{rand} < p$ 
      Global pollination via  $x_i^{t+1} = x_i^t + L(g - x_i^t)$ , where  $L$  obeys Levy distribution
    else
      Choose two random flowers from population
      Local pollination
    end if
    evaluate new solutions
    If new solutions are better, add them to the population
  end for
  find  $g$ 
  generation++
end while
```

---

Figure 1: Flower Pollination Algorithm pseudo-code

#### IV. Discrete Flower Pollination Algorithm for RCPSPs

In this paper, Discrete Flower Pollination Algorithm (DFPA) is proposed as a modification of the original FPA for solving combinatorial problems, such as RCPSPs. As the original FPA was designed for a continuous optimization problems, the concepts of such algorithm elements as flower, objective function, global pollination, Lévy Flights, and local pollination were changed.

##### A. Flower

In DFPA, a flower represents an individual in a population, which is presented in a form of permutation (Fig. 2), where each element is the scheduled activity and the index of the element is the order in which this activity is going to be executed. Each flower is considered as one solution. These permutations are positioned in the space according to the order of their components. The movement in the search space is accomplished by changing the order of the components and the length of step is derived from the value generate by Lévy flights. Movement can be done in three ways: small step, amount of small steps or large jump. To estimate the amount of steps and their length, the Lévy is calculated in an interval between 0 and 1, which then is used to derive the steps.

1	2	3	4	5	6	7	8	9	10	-	Scheduled Task
1	2	3	4	5	6	7	8	9	10	-	Execution order

Figure 2: Solution representation

##### B. Objective Function

Objective function represents a numeric value which associates with the solution in the search space, therefore, the quality of the solution is evaluated by the makespan of the project.

### C. Global Pollination

Changing the order of the tasks can be done in small or large steps. For a small step the swap mutation (Fig. 3) is used. With the swap mutation, the positions of two randomly selected tasks are switched respectively. To mimic a large step, the inverse mutation (Fig. 4) is used. With inverse mutation two tasks from a solution are selected randomly and all tasks in between them are swapped with places. Understandably, when swap and inverse mutations are performed, the precedence constraints must be satisfied.

1	2	3	4	5	6	7	8	9	10	-	A
1	2	3	8	5	6	7	4	9	10	-	B

Figure 3: Swap mutation example. A – Initial schedule, B – New schedule.

1	2	3	4	5	6	7	8	9	10	-	A
1	2	3	8	7	6	5	4	9	10	-	B

Figure 4: Inverse mutation example. A – Initial schedule, B – New schedule.

### D. Lévy Flights

To improve the quality of the solutions, similarly to original FPA, the Lévy Flights (5) is used to calculate the length of the step.

$$Lévy(s, \lambda) \sim s^{-\lambda}, (1 < \lambda < 3) \quad (5)$$

Equation (5) has infinite variance with an infinite mean [42] and is used to derive the step size.

To make a choice between a small step, a number of small steps and a large step, the Lévy flights, associated with the interval between 0 and 1, is calculated. The steps are determined in the following way:

- 1)  $[0, i]$  – move by one step (swap mutation);
- 2)  $[(k-1) * i, k * i]$  – move by k amount of steps;
- 3)  $[k * i, 1]$  - perform large jump (inverse mutation).

The value of  $i$  in this process is  $(1 / (n+1))$ , where  $n$  is the maximum amount of steps; and  $k$  is in  $\{2, n\}$  region. For example, if  $n = 4$ ,  $i = 0.2$ , the whole interval will be divided into the following five parts:

- Lévy in  $[0, i] = [0, 0.2]$  – one small step;
- Lévy in  $[i, i * 2] = [0.2, 0.4]$  – two small steps;
- Lévy in  $[i * 2, i * 3] = [0.4, 0.6]$  – three small steps;
- Lévy in  $[i * 3, i * 4] = [0.6, 0.8]$  – four small steps;
- Lévy in  $[i * 4, 1] = [0.8, 1]$  – large step.

### E. Local Pollination

The local pollination occurs via a crossover method, example of which is demonstrated in Fig. 5, where two randomly selected flowers from the population are combined into one. In this crossover method, a subset of tasks is

selected from the first flower and is used to create the new solution. Any missing tasks are then added to the new solution from the second flower in the same order they were found.

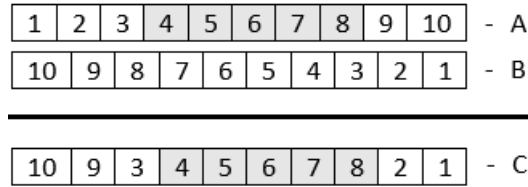


Figure 5: Local pollination Example. A – Flower 1, B – Flower 2, C – New flower.

## VI. Experimental Results

### A. Benchmark Problem

The performance and efficiency of the proposed algorithm are tested using the sets of RCPSP benchmark instances taken from the publicly available electronic library PSPLIB [44]. The PSPLIB consists of 2040 test projects with 30, 60, 90, and 120 activities, each project consisting of 4 limited resources, and each activity having a maximum of 3 successors. Due to the complexity of the RCPSP, the optimal makespan is only given for the projects with 30 activities, while optimal makespan of sets with 60 and more activities is still remains unknown. Therefore, to test the algorithm, only instances with 30 activities are considered. After all simulations are carried out, the DFPA is then compared with other recent heuristic methods which were used to solve RCPSP before, like Genetic Algorithm, Simulated Annealing, Particle Swarm Optimization, Ant Colony Optimization and Priority Rule-based scheduling.

### B. DFPA Parameter Settings Configuration

The DFPA has been implemented using Java programming language under a 64 bit Windows 8.1 operating system. All experiments were carried out on an Intel Core i7 2.4GHz laptop with 16GB of RAM.

The parameter settings (Table 1) for the DFPA were identified. Figure 6 demonstrates the impact of population sizes on the average value of all solutions found with the cases of maximum number of iterations of 25, 50 and 100, while Fig. 7 shows the effect of iterations with the same settings for the maximum number of iterations with the cases of population size of 5, 25 and 50. The experiment results, presented on Fig. 6 and Fig. 7, were received from the execution of the j3039\_3 PSPLIB instance, which has the optimal makespan of 54. Bigger population sizes and higher maximum iterations let the algorithm to find better solutions, however, this also results in higher computational time.

TABLE 1  
DFPA PARAMETER SETTINGS

Parameter	Value	Comment
$n$	20	Population size
$p$	0.8	Switch probability
$MaxGeneration$	1000	Maximum number of iteration

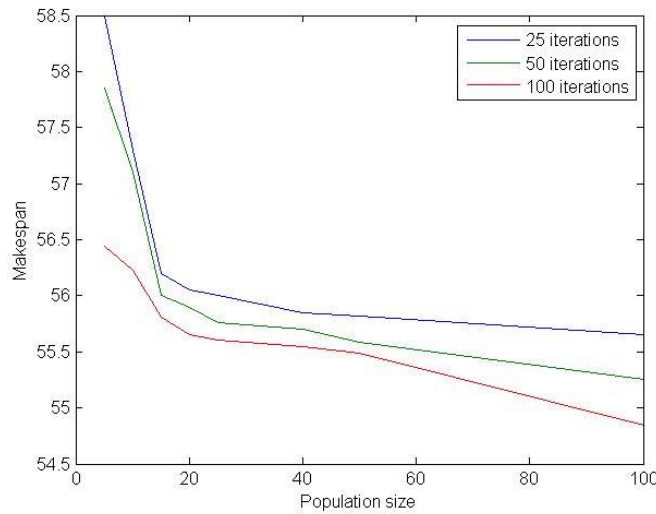


Figure 6: Dependency of average duration of best solution from population size for j3039\_3 benchmark instance set

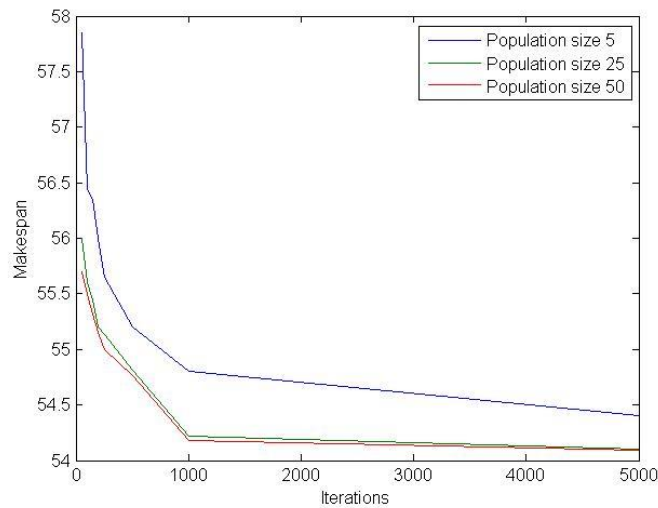


Figure 7: Dependency of average duration of best solution from maximum amount of iterations for j3039\_3 benchmark instance set

### C. Performance Evaluation

To test the algorithm in each case 100 independent runs with each benchmark instance set have been performed. The selected benchmark instances, presented in Table 2, were chosen randomly from the total amount of 480 sets. The results of the experiments are summarized in Table 2, where the first column shows the name of the instance set, the optimal makespan of the benchmark instance set taken from PSPLIB is displayed in the second column. The column “best” shows the makespan of the best solution found by the DFPA, similarly, the column “worst” shows the makespan of the worst solution. The column “average” contains the average project duration based on the 100 runs of each set. The column “Dev (%)” denotes the percentage deviation of the average solution makespan from the optimal solution makespan and is calculated using (6).

$$Dev (\%) = (solution\ makespan - optimal\ solution\ makespan) / optimal\ solution\ makespan * 100 \quad (6)$$

TABLE 2  
COMPUTATIONAL RESULTS OF DFPA SIMULATIONS

Instance name	Optimal Solution	Best	Worst	Average	Dev (%)	Time (s)
j3006_02	51	51	59	51.54	1.06	1.83
j3015_04	48	48	52	48.14	0.29	0.40
j3020_01	57	57	59	57.14	0.24	0.40
j3026_06	53	53	56	53.18	0.34	0.51
j3029_04	103	103	110	103.48	0.47	2.01
j3034_04	67	67	74	67.28	0.42	0.30
j3039_03	54	54	57	54.12	0.22	0.40
j3042_08	82	82	83	82.34	0.41	0.53
j3045_02	125	125	132	125.70	0.56	0.68
j3048_02	54	54	58	54.18	0.33	0.42
<b>Average</b>					<b>0.434</b>	

Based on the results from Table 2, it can be concluded that DFPA was capable of finding the optimal solutions for all chosen benchmark instances and the average deviation percentages from optimal solution based on 100 runs in all cases is less than 1.06%. These results, presented in Table 2, indicate that DFPA is indeed powerful algorithm and can provide adequate solutions in reasonable time.

#### D. Comparison with Other Algorithms

Lastly, in Table 3, the experimental results of the DFPA are compared with other heuristic algorithms, results of which are taken from [2, 36, 45]. The numbers 1000 and 5000 in Dev (%) column denote the maximum number of iterations and are used as a stop criterion. The algorithms presented in Table 3 were selected based on their complexity. Only original non-hybrid versions of algorithms were chosen and modified versions of metaheuristic algorithms with additional more complicated search mechanisms, e.g. radius PSO [31] or random key-based GA [20], were omitted and left out.

TABLE 3  
COMPARISON OF PERFORMANCE OF OTHER ALGORITHMS

Algorithm name	Author(s)	Dev (%)	
		1000	5000
<b>DFPA</b>	<b>This paper</b>	<b>0.434</b>	<b>0.21</b>
ACO [36]	Luo, Wang	0.39	0.22
SA [28]	Bouleimen, Lecocq	0.38	0.23
GA [6]	Hartmann	0.54	0.25
PSO [31]	Anantathanvit, Munlin	0.41	0.33
Tabu Search [46]	Baar et al.	0.86	0.44
Adaptive sampling [47]	Kolisch	0.74	0.52
Serial sampling LFT [47]	Kolisch	0.83	0.53
Serial random sampling [48]	Schrimer, Riesenber	0.71	0.59
Parallel sampling WCS [47]	Kolisch	1.40	1.28
Parallel sampling LFT [47]	Kolisch	1.40	1.29

Overall, the comparison of the performance with other algorithms can be regarded as satisfactory and it can be noted that DFPA has managed to outperform all algorithms presented in Table 3. Better performance of DFPA over other algorithms can be explained with a good balance between exploitations and exploration, intelligent use of Lévy Flights and the reduced number of parameters that need to be configured to provide the optimal performance. Another DFPA's advantage, which plays an important role in deciding which algorithm is better, is its simplicity. Being very simple, the DFPA is easy to implement, which makes it more attractive to be used in other combinatorial problems.

## VII. Conclusions

In this paper, a new metaheuristic FPA is selected and then modified for solving the combinatorial optimization problems. As the original FPA was designed for solving a continuous optimization problems, in order to adapt it for solving combinatorial problems, the concepts of such algorithm elements as flower, objective function, global pollination, Lévy Flights, and local pollination were changed. Further, the algorithm's performance has been tested on a set of PSPLIB benchmark instances, and despite being simple and relatively easy to implement, the proposed algorithm has managed to find optimal solutions in all benchmark instances and its average deviation from the optima based on 100 runs in all cases was less than 1.06%, which has validated algorithm's effectiveness. Lastly, the algorithm has been compared with other popular metaheuristic non-hybrid algorithms, like GA, SA, PSO, and ACO and the results of comparison have shown that DFPA has managed to outperform all selected algorithms in terms of average deviation percentage from the optimal solution, therefore proving its competitiveness and superiority over selected algorithms for comparison. These results indicate that despite being very simple, the DFPA is yet very powerful and efficient algorithm.

In the future, the work on improvement of DFPA will be carried on and the algorithm will be applied in solving more complicated scheduling problems. The probable areas of further application will include traveling salesman problem and knapsack problem. One of the possible areas of improvement is the better exploitation of global solution to make the chance of falling in local trap even less than it is now. Further, after this improvement is done, it will be compared with genetic algorithm with the aim of finding which algorithm finds the global solution more efficiently.

## REFERENCES

- [1] J. Blazewicz, J. Lenstra and K. Rinnooy, "Scheduling subject to resource constraints: Classification and complexity," *Discrete Applied Mathematics*, vol. 5, no. 10, pp. 11-24, 1983.
- [2] B. Chen and G. Quan, "NP-Hard Problems of Learning from Examples," in *Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, Shandong, 2008.
- [3] R. Kolisch and S. Hartmann, *Handbook on Recent Advances in Project Scheduling*, Kluwer: Dordrecht, 1999.
- [4] J. Stinson and W. Davis, "Multiple Resource-Constrained Scheduling Using Branch and Bound," *AIEE Transactions*, vol. 10, no. 3, pp. 252-259, 1978.
- [5] W. Iskander and S. Panwalker, "A survey of scheduling rules," *Operations Research*, vol. 77, no. 2, pp. 81-98, 1979.
- [6] S. Hartmann, "A competitive genetic algorithm for resource-constrained project scheduling problem," *Naval Research Logistics*, vol. 45, no. 10, p. 733-750, 1998.
- [7] H. Zhang, "Particle swarm optimization-based schemes for resource-constrained project scheduling," *Automation in Construction*, vol. 14, no. 3, p. 393-404, 2005.
- [8] W. Zhang, Y. Huang and M. Wang, "Synthetic Optimization in Project Schedule by Using Ant Colony Algorithm," in *World Congress on Intelligent Control and Automation*, Dalian, 2010.
- [9] L. Lawler and E. Wood, "Branch and Bound Methods: A Survey," *Operations*, vol. 14, no. 4, pp. 699-719, 1966.



- [10] T. Johnson, An algorithm for the resource-constrained project scheduling problem., Cambridge, MA: Doctoral Thesis, Massachusetts Institute of Technology, 1967.
- [11] H. Muller, "A method for planning the optimum input use," *The journal of economical production*, vol. 64, no. 4, pp. 135-140, 1967.
- [12] A. Sprecher and A. Drexl, Manuscripts from the institutes of Business Administration, Kiel: University of Kiel, 1996.
- [13] E. Davis and J. Patterson, "A Comparison of Heuristic and Optimum Solutions in Resource- Constrained Project Scheduling," *Management Science S2: Application Series*, vol. 21, no. 8, p. 944-955, 1975.
- [14] T. R. Browning and A. A. Yassin, "Resource-constrained multi-project scheduling: Priority rule performance revisited," *International Journal of Production Economics*, vol. 126, no. 2, p. 212-228, 2010.
- [15] S. R. Lawrence and T. E. Morton, "Resource-constrained multi-project scheduling with tardy cost: Comparing myopic, bottleneck, and resource pricing heuristics," *European Journal of Operational Research*, vol. 64, no. 2, pp. 168-187, 1993.
- [16] D. W. Hildum, Flexibility in a knowledge-based system for solving dynamic resource-constrained scheduling problems., University of Massachusetts, Amherst, MA: Ph.D. Dissertation, Computer Science Dept., 1994.
- [17] F. F. Boctor, "Heuristics for scheduling projects with resource restrictions and several resource-duration modes," *International Journal of Production Research*, vol. 31, no. 11, pp. 2547-2558, 1992.
- [18] P. Husbands, G. Gerny and M. McIlhagga, "Two applications of genetic algorithms," in *AISB workshop on evolutionary computing*, Chicago, CH, 1996.
- [19] L. Davis, "Hybrid genetic algorithms for machine learning," in *IEEE Colloquium on Machine Learning*, London, 1990.
- [20] J. J. Mendes, "A random key based genetic algorithm for the resource constrained project scheduling problem," *Computers & Operations Research*, no. 36, p. 92-109, 2009.
- [21] J. R. Montoya-Torres and E. Gutierrez-Franco, "Project scheduling with limited resources using a genetic algorithm," *International Journal of Project Management*, vol. 28, no. 6, p. 619-628, 2009.
- [22] M. Shahsavar and S. T. AkhavanNiaki, "An efficient genetic algorithm to maximize net present value of project payments under inflation and bonus-penalty policy in resource investment problem," *Advances in Engineering Software*, vol. 41, p. 1023-1030., 2010.
- [23] J. Alcaraz, C. Maroto and R. Ruiz, "Solving the Multi-Mode Resource-Constrained Project Scheduling Problem with genetic algorithms," *Journal of the Operational Research Society*, vol. 54, p. 614-626, 2003.
- [24] P. J. M. v. Laarhoven, E. H. L. Aarts and J. K. Lenstra, "Job Shop Scheduling by Simulated Annealing," *Operations Research*, vol. 40, no. 1, pp. 113-125, 1992.
- [25] G. Palmer, "A simulated annealing approach to integrated production scheduling," *Journal of Intelligent Manufacturing*, vol. 7, no. 3, pp. 163-176, 1996.
- [26] F. F. Boctor, "Resource-constrained project scheduling by simulated annealing," *International Journal of Production Research*, vol. 34, no. 8, pp. 2335-2351, 1996.
- [27] Y. Nikulin and A. Drexl, "Theoretical aspects of multicriteria flight gate scheduling: deterministic and fuzzy models," *Journal of Scheduling*, vol. 13, pp. 261-281, 2010.
- [28] K. Bouleimen and H. Lecocq, "A new efficient simulated annealing algorithm for the resource-constrained project scheduling problem and its multiple mode version," *European Journal of Operational Research*, vol. 26, no. 149, p. 268-281, 2003.
- [29] R. Zamani, "An efficient time-windowing procedure for scheduling projects under multiple resource constraints," *OR Spectrum*, vol. 26, no. 3, pp. 423-440, 2004.
- [30] H. Zhang, "Particle swarm optimization-based schemes for resource-constrained project scheduling," *Automation in Construction*, vol. 14, no. 3, p. 393-404, 2005.
- [31] M. Anantathanvit and M. Munlin, "Radius particle swarm optimization for resource constrained project scheduling problem," in *International Conference on Computer and Information Technology*, Khulna, 2014.
- [32] M. Li, Z. Yuanbiao, W. Jiang and J. Xie, "A Particle Swarm Optimization Algorithm with Crossover for Resource Constrained Project Scheduling Problem," in *IITA International Conference on Services Science, Management and Engineering*, Zhangjiajie, 2009.
- [33] D. Linyi and Y. Lin, "International Conference on Computational Intelligence and Security," in *International Conference on Computational Intelligence and Security*, Harbin, 2007.
- [34] K. Zhang, G. Zhao and J. Jiang, "Particle swarm optimization method for resource-constrained project scheduling problem," in *International Conference on Electronic Measurement & Instruments*, Beijing, 2009.
- [35] D. Merkle, H. Schmeck and M. Middendorf, "Ant colony optimization for resource-constrained project scheduling," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 4, pp. 333 - 346, 2002.
- [36] S. Luo, C. Wang and J. Wang, "Ant colony optimization for resource-constrained project scheduling with generalized precedence relations," in *International Conference on Tools with Artificial Intelligence*, Beijing, 2003.
- [37] J. Wang, S. Zhang and J. Chen, "Resource-constrained multi-project scheduling based on ant colony optimization algorithm," in *Intelligent Computing and Intelligent Systems*, Xiamen, 2010.
- [38] Y. Shou, "A Bi-directional Ant colony algorithm for resource constrained project scheduling," in *Industrial Engineering and Engineering Management*, Singapore, 2007.
- [39] S. T. Lo, R.-M. Chen, Y.-M. Huang and C.-L. Wu, "Multiprocessor system scheduling with precedence and resource constraints using an enhanced ant colony system," *Expert Systems with Applications: An International Journal*, vol. 34, no. 3, pp. 2071-2081 , 2008.
- [40] X.-S. Yang, "Flower Pollination Algorithm for Global Optimization," *Unconventional Computation and Natural Computation*, vol. 7445, pp. 240-249, 2012.
- [41] Y. Zhou and Y. Chen, "Business process assignment optimization," in *IEEE International Conference on Systems, Man and Cybernetics*, Paris, 2002.
- [42] M. F. Shlesinger, G. M. Zaslavsky and U. Frisch, "Lévy Flights and Related Topics in Physics," in *Proceedings of the International Workshop*, Nice, France, 1994.
- [43] X.-S. Yang and S. Deb, "Cuckoo Search via Lévy flights," in *World Congress on Nature & Biologically Inspired Computing*, Coimbatore, 2009.
- [44] PSPLIB, "Project Scheduling Problem LIBrary," [Online]. Available: <http://129.187.106.231/psplib/>. [Accessed 6 May 2015].
- [45] Y. Zhou and Y. Chen, "Business process assignment optimization," in *IEEE International Conference on Systems, Man and Cybernetics*, Paris, 2002.

- [46] T. Baar, P. Brucker and S. Knust, "Tabu Search Algorithms and Lower Bounds for the Resource-Constrained Project Scheduling Problem," in *Metaheuristics: Advances and Trends in Local Search Paradigms for Optimizatio*, Springer US, 1999, pp. 1-18.
- [47] R. Kolisch and S. Hartmann, "Heuristic Algorithms for the Resource-Constrained Project Scheduling Problem: Classification and Computational Analysis," *International Series in Operations Research & Management Science*, vol. 14, pp. 147-178, 1999.
- [48] A. Schirmer and S. Riesenber, "Parameterized Heuristics for Project Scheduling - Biased Random Sampling Methods," *Manuskripte aus den Instituten für Betriebswirtschaftslehre der Universitäten Kiel*, vol. 471, pp. 129-141, 1997.

# Cloud Computing Security: A Survey

Amjad Mehmood<sup>1</sup>, Muhammad Roman<sup>1</sup>, M. Munir Umar<sup>1</sup>, Houbing Song<sup>2</sup>

<sup>1</sup>*Institute of Information Technology, Kohat University of Science and Technology, Kohat*

<sup>2</sup>*Department of Electrical and Computer Engineering, West Virginia University, USA*

[amjad.mehmood@kust.edu.pk](mailto:amjad.mehmood@kust.edu.pk), [m.roman@kust.edu.pk](mailto:m.roman@kust.edu.pk), [muneerumar@hotmail.com](mailto:muneerumar@hotmail.com), [Houbing.Song@mail.wvu.edu](mailto:Houbing.Song@mail.wvu.edu)

**Abstract**—Cloud computing brings new possibilities for individuals and firms to utilize computing as a utility. It utilizes computing power irrelevant of user's location and devices. Thus it has become more demanding due to its performance, high computing power, cheapness, elasticity, accessibility, scalability and availability. Cloud computing offers ubiquitous operation with different security challenges. In this paper we discuss security challenges and vulnerabilities as well as limitations of current security modules. This paper will serve as a baseline guide for new researchers in this area.

**Index Terms**—Cloud Computing Security, Infrastructure-as-a-Service (IAAS), Platform-as-a-Service (PAAS), Software-as-a-Service (SAAS), Private Cloud, Public Cloud, Hybrid Cloud, Trust, Vulnerabilities.

## I. INTRODUCTION

Computing is becoming the need of every firm and individuals. A significant time is spent on maintenance of resources and updation of hardware and software components. It is required to keep things synchronized while trying to provide access to remote resources without the need of special devices that have the abilities to do all the processing on the local processing unit. Cloud computing provided a solution for all these queries by providing computing power as a utility for every user. Regardless of what hardware devices they are using and what are the processing capabilities of those machines, cloud computing provides it users with a high processing power as per user demands and requirements and charges per usage time. Cloud users don't need to maintain and update hardware and software resources themselves. Thus cloud computing provides a way to minimize our IT expenses, in most of the cases. Turning to cloud computing technology allows IT team to minimize the time spent on maintenance and focus on activities that have higher impact. Cloud computing integration with other technologies is much easier, giving it backward portability with the legacy systems. It is much more scalable and recoverable than ever possible as users get what they demand on their servers. It is highly customizable according to the users' requirements providing a platform where they can easily deploy their system. For application developers cloud computing provides thousands of pre-built and tested modules ready for integration in their new application. The user's data is kept on a single repository making it accessible remotely from anywhere in the world. The user gets a synchronized data from their personal

computer, mobile devices and from anywhere via internet through a browser. They can easily share their data with their friends and fellows. Still the user can control what to show and what not to while giving a maximum level of accessibility and availability. Amazon's EC2, Google AppEngine, Salesforce.com, SaaSGrid and GoGrid are some of the examples of cloud computing.

## II. DESIGN LAYERS AND TYPES OF SERVICES PROVIDED BY CLOUD COMPUTING

The services provided by cloud computing are divided into three categories, according to the level of abstraction of capabilities provided by each of these layers [2], [3]. These layers are viewed as a layered architecture in which the services of lower layer form the bases of higher layer [5].

### A. Software as a Service (SaaS)

SaaS delivers applications using web interface which is maintained and managed by the provider of the particular software application [13]. It is highly adoptable as most of the people know internet [8]. It has a lower learning curve. Most of the applications provided by SaaS are directly accessible from browser and there is no need to download or install any other software on the local machine. All the requirements of the application are managed by the vendor which includes Applications, Runtime, Data, Middleware, OS, Virtualization, Servers, Storage and Networking [20]. The user need not to be aware of backups of the data and software, update and upgrades of the software and its modules. The license of the running application is also purchased and maintained by SaaS provider and the customers are not required to purchase their own license for using the application on cloud servers. The customer is charged for the application either on monthly subscription bases or based on the total number of users accessing the application on SaaS [19]. Salesforce.Com, CRM, SugarCRM, are some of the examples of SaaS. Gmail, GMail, Microsoft Office365, LyncOnline, ExchangeOnline, Sharepoint Online are some of the applications which are running on cloud and provided as SaaS.

### B. Platform as a Service (PaaS)

PaaS is another layer of abstraction which is considered as the most complex of the three [20]. This layer is basically for the software development teams who utilize the services of cloud computing for developing new application for their customers.

Software development requires a platform for development which requires web-server, database servers and etc [14]. In order to run them on a local machine the development platform need to be setup, managed and administered by user. In case of PaaS the development, hosting, testing and deployment of applications in done very quickly and is cost-effective. The customers get an environment where they can develop and deploy software without worrying about the processing power and memory resources it requires [7]. It also eliminates the need for setting up the underlying hardware and software requirements. It also provides some pre-built software modules that can be integrated directly into the software. Provider still has to manage runtime, middleware, OS, Virtualization, Servers, Storage and Networking like in SaaS but application and data is to be handled by the user of PaaS.

A very good feature of PaaS is that the users do not need to worry about the site to get down during its maintenance. It is highly scalable the platform upgrades do not interfere with user application. The customers are charged on the bases of incoming and outgoing network traffic, CUP time per hour used by the customer, data storage size. Sometime the customers are charged on monthly bases for the type of service being provided. Usually the cost of PaaS is not predictable and a multi-dimensional pricing model is used.

GoogleAppEngine, Apprenda’s SaaSGrid and Force.com are some of the famous examples of PaaS.

C. Infrastructure as a Service (IaaS)

IaaS offers computation, storage and communication as virtualized resources. Instead of purchasing servers, software and network resources these resources are rented by the customers of cloud computing on demand and billed for these resources as per usage [6]. By paying to the IaaS providers customers are allowed to create virtual servers on their infrastructure. Unlike other two services customers of IaaS are responsible for setting and managing applications, runtime, data, OS and middleware. IaaS provides virtualization, servers, hard drives, storage and networking as a service. The users of IaaS are usually it department who save their cost by renting a fully outsourced infrastructure for which they do not need to worry about updation, upgrades and maintenance. The customers are charged based on CPU hours, gigabytes of storage and network bandwidth used by the customers if IaaS. Amazon’s EC2, GoGrid, Mosso and FlexiScale are some of the examples of IaaS [21].

	Examples	Advantages	Deliverables	Interface	Used By	Customer’s Responsibilities	Vender’s Responsibilities	Billing Method
SaaS	Salesforce.Com, CRM, SugarCRM, GMail, Microsoft Office365, LyncOnline, ExchangeOnlin, Sharepoint Online	Software maintained by vendors, Adoptable, No need to purchase software, Autobackup, Autopdate, Painless Upgrade, Accessed from anywhere	Software Application	Web Interface	End Users	To manage web client software (the browser)	Applications, Virtualization, Runtime, Middleware , Data, OS, Servers, Storage and Networking	Monthly subscription fee, Per-User monthly changes
PaaS	SaaSGrid, Google-Appengine, Froce.Com, Bungee, Heroku, Web Role, Work Role, Map Reduce, Blob, Message Queue, Service Bus, Cloudfront, Marketplace	Quick, Simple, Cost-Effective, No To Purchase Underlying Hardware And Software, Site Is Up During Maintenance, Services to deploy, test, host and maintain applications,	Framework for Developing Applications, Plug-in, Modules	Integrated Developm ent Environment (IDE)	Software Developme nt Teams	Applications and Data	Virtualization, Runtime, Middleware , OS, Servers, Storage and Networking	Incoming/O utgoing network traffic, CPU time per hour, Stored data, Monthly service changes
IaaS	EC2, GoGrid, Mosso, FlexiScale	No need to purchase computing resources, storage and network	Computer Infrastructure , Storage, Networking	Web Interface	IT Department s	Applications, Runtime, Data, OS and Middleware	Virtualization, Servers, Hard Drives, Storage, And Networking	CPU hours, gigabytes of storage, Bandwidth utilization (More Predictable)

Table 1: Cloud Computing Services Comparison Table

### III. DEPLOYMENT MODELS

There are variations in physical distribution and location of cloud computing differentiating them into three types as shown in fig. 3. Thus a cloud computing can be classified as public, private, community or hybrid [3] based on their physical location and distribution, regardless of the services they are providing to the end users.

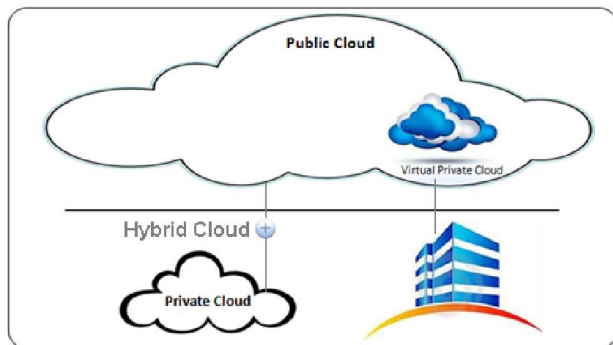


Fig. 3: The Cloud Computing Deployment Models

#### A. Public Cloud

Public cloud dynamically allocates computing resources on per-user-bases on demand via a web interface [8]. The customers of public cloud utilize the infrastructure implemented and managed by the provider. Computing power is provided as a utility and the customers can save money by metered billing approach in which they pay only for what they have used, without worrying about the management of overall system locally. It can be access from anywhere at any time from a any supporting devices like smart phone or a laptop connected to internet. Amazon and Google are two well-known providers that offer public cloud computing services [22].

#### B. Private Cloud

Private cloud is also known as “internal cloud computing”. It is implemented under the control of IT department within the corporate firewall. It is the next generation of virtualization. The complete infrastructure is under the control of IT department and must be run and managed by them. This gives them choice to skip any security implementations as the access to the cloud resources is limited as compared to public cloud. But on the other hand ROI (return on investment) is a drawback of private computing and it requires the capital expense of IT infrastructure. Examples of private cloud are VMWare vCloud and Citrix VDI [22].

#### C. Virtual Private Cloud

Elasticity is one of the main feature of cloud computing. Scaling up and down private cloud resources and services can be very much costly and cannot be achieved without user’s

interaction. On the other hand public cloud does not provide a security level of a private cloud. Combining both of them leads to another cloud known as virtual private cloud. This allows enterprise customers to connect to a public cloud services via VPN. The customer can create their own virtual private cloud and define private block of IP addresses and subnets for it. All traffic to virtual private cloud will rout through the VPN providing security of private cloud and elasticity and ROI of public cloud. Amazon’s VPC is an example of virtual private cloud which enables customers to connect to its Elastic Compute Cloud (EC2) services through a VPN [23, 24].

#### D. Community Cloud

A community cloud is the one which is shared among several organizations who are working on same project or have same aims like a mission or a target [3]. The infrastructure is shared among several organizations around the world. These organizations are from a specific community and share common concerns. It can be managed internally or by a third party. Community cloud can be hosted internally or externally. The overall infrastructure implementation and management cost is distributed among the users like in public cloud. But as the number of customers is less so the individual cost is higher than that of a public cloud.

#### E. Hybrid Cloud

Sometimes our private cloud is not enough to provide us all the services and capabilities we need. In this case we get registration with a public cloud and our private cloud is thus supplemented with a public cloud, known as hybrid cloud. This approach is termed as “cloud-bursting” [9]. A hybrid cloud combines a customer’s hardware resources with cloud computing. Software is also required to interact with the provider services. For example, Cisco’s IronPort Email security is provided as a hybrid solution. Google also provides a hybrid email archiving software known as Postini [22].

### IV. CLOUD COMPUTING KEY CHARACTERISTICS

Cloud computing being considered as one of the most promising technologies, has a number of key characteristics pointed out by US National Institute of Standards and Technology (NIST). The definition provided by NIST along with its various characteristics is now becoming a de-facto standard for cloud computing definition [14].

*On demand self service:* Cloud computing must provide an interface to manage and order services without any direct interaction with the cloud providers. This can be done via a web portal and management interface. The overall service provisioning interface must be automated and without any human interaction.

*Broad network access:* Resources are utilized over the network, usually the Internet, from anywhere in the world. This promotes the use of heterogeneous platforms like mobile

phones, PDA and laptops. The access is defined using standard access mechanisms and protocol.

*Location Independence:* The execution of job is independent of the location of processing unit, as cloud may be distributed physically in all over the world. The user is not usually concerned and does not know about the servers where the data is being executed or saved. But there are cases when the user may be required to specify the location at the higher level of abstraction due to legality issues which changes from one country to another country or from one region to another region.

*Rapid elasticity:* The resource required by a customer can be scaled up and down depending on the requirement changes by the user. The cloud computer must be elastic in service provisioning and should be able to adjust these changes without any human interaction.

*Resource Pooling:* Computing resources are realized as shared resources which are equally available to all the users. This is done by using the technique of virtualization.

*Economies of scale and cost effectiveness:* Cloud implementation is as large as required in order to take full advantage of economies of scale. The services provided by the cloud are cost effective as compared to local infrastructure. Large cloud deployment is usually located where the power is provided at lower cost and where real estate prices are low.

*Measured services:* The usage of resources and services are calculated and metered constantly. The usage reporting are communicated to the customers in order to pay-as-per-usage model of utility based computing.

Cloud computing in its very nature has the capabilities to address various limitations and issues in traditional computational architecture. However, it has introduced a number of issues which were never introduced before in traditional computing architectures. Some of them are new and solutions to these issues were never discussed before the advent of cloud computing. Others require modifications in the currently provided solutions. In the next section we discuss general issues which are there due to the core technologies used in cloud computing. We also discuss cloud specific issues which are newly introduced with advent of cloud computing.

## V. CLOUD COMPUTING SECURITY

Cloud computing customers use services provided by cloud computing. These services are hosted and maintained by cloud providers. All of the user's confidential data is saved on cloud servers. Thus the user is handing over their confidential data to the cloud providers. For cloud computing to work the customers are required to completely trust the cloud providers. We trust a system less when we do not have much control over it [15]. For example, while withdrawing money from ATM we trust more because we will get some amount at the

end of the transaction. But when we use an ATM for a deposit we do not know what will happen after we give money to the machine. Trust in cloud computing means that the providers will provide services and confidentiality to its customers as promised. In a distributed processing environment jobs are entered into the system and then it is out of the control of the user where the data is being processed. The user is not aware of legal rules followed in the region where the data is being saved or processed.

An earliest example of trust in late 70s and 80s is Trust Computer System Evaluation Criteria (TCSEC). In that trust was used to convince the customers that the system was correct and secure. The customers of cloud computing will trust the providers if they believe that provider will behave exactly as expected and promised. The characteristics of trust are credibility and consistency. We trust a system less if it does not give enough information about its expertise. Merely claiming to be "trust me" or "secure cloud" has most of the time no impact on trust by the customer. The system must be transparent. The user of the system must know, in distributed environment, where the data is being processed and where the data is being saved. A control mechanism may also help reduce the level of discomfort. This mechanism is used to manage where the data should be processed and saved physically on cloud machines.

Trust heavily depends upon the deployment model of the cloud computing infrastructure. A private cloud is more trusted as overall system is controlled by the same enterprise internally. Community clouds are less vulnerable as all the users are from same community and most of the time from same enterprise. Public cloud is used by different types of customers from different locations at the same time. You cannot trust all the users of the system. This makes public cloud more vulnerable thus less trustworthy. Service Level Agreements (SLAs) play an important role in establishing trust and most of the time is the only way to establish trust. But these might not be helpful in some cloud computing environment. This is because for most companies breach of data is irreparable and no money can recover the cost, as promised by contractual agreements. Therefore cloud trust model focuses more on preventing failure than post-failure compensation [15]. Currently claim-based access control, security assertion markup language (SAML), security token service and federated identity approach are some of the techniques which help in establishing trust. One of the very prominent solution to trust problem is the establishment of an independent security certification authority that can certify cloud services as discussed in [16].

Cloud computing is vulnerable to a variety of security attacks. According to open group risk taxonomy, "Vulnerability is the probability that an asset will be unable to resist the actions of a threat agent. Vulnerability exists when there is a difference between the force being applied by the threat agent, and an object's ability to resist that force" [14]. It measures the likelihood of an attack and the possible consequences of that attack on the system. Loss occurs when

an attacker successfully exploits vulnerability. The frequency of this loss depends upon two things. First: the frequency of the attack to exploit vulnerabilities. This depends upon the motivation of the threat agent to the system and level of access to the system. Second: the ability of the system to resist the attack for exploiting vulnerabilities [14]. Thus, computer vulnerabilities show the level of strength of a system against computer bugs and attacks.

Different articles, blogs and publications are discussing vulnerabilities of cloud computing very often. But most of them miss to differentiate between the general issues and issues specific to cloud computing. There are issues which are not there due to advent of cloud computing but were there before cloud computing gained fame. Cloud computing is based some core technologies like virtualization and the Internet. These technologies have some issues which are associated with them and directly to the cloud computing architecture. Of course, cloud computing make them severe and most of the time the proposed solution for those vulnerabilities are not helpful in case of cloud computing. Like cloud computing provides a wider and flexible access to the resources from anywhere in the world. This may increase the frequency of threat agent's attack on the system, which can help them understand the policies being applied. Now to understand what cloud specific vulnerabilities are, following are some of questions asked before deciding. A vulnerability is cloud specific if

- is present due to the very nature of core cloud computing technology,
- is due to the any of the NIST's essential cloud features as discussed above in this paper,
- came in existence only when cloud computing innovated and was found that it was difficult to control in this particular case, or
- is dominant in cloud offerings.

Now to examine each of these signs we first understand what cloud computer core technologies are.

#### A. Cloud Computing Core Technologies and their Vulnerabilities

Cloud computing is based on a number of core technologies. These core technologies Cloud computing is built heavily on certain core technologies. These are the technologies without which cloud computing cannot be fruitful and in some cases not possible at all.

*Web Services and Applications:* Web application and web services technologies are the baseline of software as a service (SaaS) and platform as a service (PaaS). SaaS is usually access as web applications by the end users. PaaS makes development process easier by exposing web services and integrating them into the user web applications. PaaS makes the development of new application easier by using pre-built services. Similarly, infrastructure as a service is typically

administered using a web interface, like managing access control of different users.

*Virtualization Offerings:* Virtualization lets users run multiple isolated virtual machines (VMs) on a single physical machine, simultaneously. It is the a core technology for providing high computing power to the customers while keeping the system elastic and keep pay-as-you-go model. Virtualization provides pooled resources to the users while giving a best utilization of the installed infrastructure. SaaS and PaaS are built on top of the virtualized infrastructure.

*Cryptography:* For a majority of the cloud computing security cryptography is the only technique used for security data on the cloud servers.

Web applications, web services, virtualization and cryptography have vulnerabilities that are either core vulnerabilities or are introduced only when these technologies are used in cloud computing. Following is a discussion about some of these vulnerabilities.

*VM Hopping:* It lets attacker on one virtual machine (VM) gain access to another VM, which is being attacked by the attacker. In this attack the attacker can modify the victim's configuration settings, monitor resource usage and delete confidential data. This may lead to harm confidentiality, integrity and availability of the user's data. The only requirements for this attack are that the attacker VM must be on the physical machine where the victim's VM is residing. And the attacker must know the victim's VM IP address. As multiple machines are running on the same physical machine simultaneously in cloud computing and they belong to different firms therefore VM hopping can be a worse in case of cloud computing. Thus we can say that VM hopping is a reasonable vulnerability in case of cloud computing. As several computers are running on the same machine, one or more VMs can become victims of this attack. VM hopping is particularly crucial to PaaS and IaaS. But SaaS can also be affected indirectly as it is also based on PaaS and IaaS. It can affect SaaS confidentiality and integrity of user's data [18].

*VM Mobility:* VM's virtual disc contents are stored as files on the physical machine. This makes VM possible to be carried away and moved from one physical machine to another. This gives mobility to VM which helps in quick deployment of the system. The mobility of VM also brings some issues into the system, like vulnerable configuration spreading. An attacker will encapsulate such vulnerable configurations into his/her own VM. When this VM is moved to any other physical server, the vulnerabilities will also be moved along with the virtual machine. This can act as man-in-the-middle attack. The guest operating system can merely loose confidential data (which is of course not a small issue itself) or can completely compromise the new guest machine. We cannot completely stop this as VM mobility makes the overall system very much flexible. Service level agreements can be helpful in minimizing the possible impacts of VM mobility.



*VM Diversity:* Securing and maintaining of virtual machines is difficult due to the wide range of operating systems. These OSs can be deployed in seconds [17]. The diversity of VM makes the maintenance and securing them a very challenging job.

*VM Denial of Service (DOS):* In virtualization all the VM share same physical resources. These resources include CPU, memory and network bandwidth. There are cases when a VM takes all the resources and denies the services for the rest of the VMs on the machine. To prevent this attack the resource allocation configuration is done prior to assigning resources to a VM. In cloud computing SLA can be very helpful to stop DOS and configuration for each customer is clearly defined.

*Session Riding/Hijacking:* Cloud computing is based on web application/services. Web applications and web services use HTTP as the carriage protocol. HTTP by design is stateless, which means that the state of the application is lost between multiple requests to the server. To manage the state of HTTP multiple state management techniques are used. Among those technique is session. There are many techniques to manage sessions, including query string, cookies and state server. But one way or the other session management are vulnerable to session riding and hijacking. As cloud computing uses web interfaces, session riding and hijacking is very much associated to cloud computing architecture.

*Cryptanalysis:* Cryptography techniques are used to secure data on cloud. There are ways to render cryptographic mechanism and algorithms.

#### B. Essential Cloud Characteristic Vulnerabilities

We have defined NIST basic characteristics of cloud computing. These vulnerabilities are mainly concerned with those characteristics of cloud computing. Below are a few examples:

*Internet Protocol:* Cloud services are provided over the network which in most cases is the Internet. Internet uses standard protocols which are not considered trusted in most cases. Thus Internet protocol vulnerabilities are relevant to cloud computing.

*Un-Authorized Access:* One of the NISTs key features was on-demand self service without human interaction. This requires a web based management interface accessible from anywhere in the world. The management interface can be accessed by unauthorized users over the Internet. In cloud this is more important than in traditional system as the management interface is accessible to more people.

*Data Recovery:* Resource pooling requires that same resources will be shared by many users at different time. A resource used by one user will be used by another user at some other time. An attacker on the same physical machine can thus recover data from memory and storage devices.

#### C. Limitations in Known Security Techniques

If cloud computing is directly affecting the currently applied security techniques, such that they are no longer helpful in cloud computing environment. Like, standard IP-based network zoning cannot be applied in cloud computing environment. IaaS provides does not allow network based vulnerability as friendly scans cannot be distinguished from attackers' scanning. In a virtualized environment network traffic means communication on real and virtual network. A virtual network is network among different VM on the same physical machine. Such issues are new with the advent of cloud computing.

Similarly, poor key management is also one of the security control issues. Cloud computing require storage, generation and management of many different kinds of keys. VM are geographically distributed and do not have a fixed physical hardware, thus some hardware security module (HSM) incorporation is difficult in case of cloud computing.

Finally, the users of cloud computing are not provided any security metrics that can be used to monitor security status of their cloud services. This is because currently there are not security metrics adapted in cloud computing. Audit, accountability and security controls are more difficult to apply until these security metrics are implemented for cloud computing.

#### D. Vulnerabilities in Cloud Offerings

Cloud computing offers some state of the art offerings in the market. If the vulnerability is found in state of the art offering it is also called cloud specific vulnerability. Weak authentication and injection attacks are two examples of such vulnerabilities.

Injection is performed by providing an input to an application such that part of it is executed as command on the server. These code lines do the attacker's desired functionality which can definitely harm the overall system. Examples are SQL injection, command injection and cross-site script injection.

Similarly, weak authentication is also a problem in cloud computing. Cloud computing provides web interfaces to its customers. Web uses username password authentication techniques most of the time which is not considered a secured authentication mechanism due to insecure user actions (choosing weak password, remember password, and so on) and one-factor authentication technique.

There are vulnerabilities in cloud computing infrastructure and platform. Cloud computing infrastructure provides basic IT resources including storage, computing resources and communication as services to the higher layers of cloud computing. These resources are usually virtual resources on top of physical resources. Cloud computing platform provides



application development and runtime environment for the services developed in one of the supported languages. The vulnerabilities involved in them are discussed below:

*i. Storage Security Risks*

Resource pooling and elasticity are playing main role in making cloud data vulnerable. As same data storage devices are being used by different users of cloud. If a user's confidential data is stored on primary memory or backup storage. When this user stops using that address space and this address space is assigned to some other user who can be an attacker. This attacker can recover this data back and get access to the confidential data of the previous customer.

Media sanitization is also harder in cloud environment using both hard and soft media sanitization. Data sanitization is used to avoid the possible outcomes from data remanence. Data remanence is the footprint of the data that remains after the data has been removed and deleted from a storage media. This may occur by a nominal file deletion operation. Media sanitization is usually done by formatting the storage media, which is not possible in case of cloud computing. Most of the organizations destroy the storage media physically, a hard sanitization. This hard sanitization is also not possible in cloud computing as the overall storage is being shared among various customers and they have valuable data stored on it.

Cryptography is used as a solution to data storage problems. But as discussed in vulnerabilities section poor key management and storage of keys is a challenge that threatens the use of cryptography in cloud computing.

*ii. Securing Communication*

Due to elasticity and resource pooling certain networking infrastructure is also shared among the customers of cloud computing. Cross-tenant attacks may occur by utilizing the shared network infrastructure resources such as domain name system, dynamic host configuration protocol and internet protocol. This usually happens in IaaS environment. Due to virtualization the network not only means real network infrastructure but also the virtual network among the VM in the same physical environment. Network based security implementation are sometime not possible to integrate in a virtual networking environment.

*iii. Identity, Authentication, Authorization and Auditing (IAAA)*

Identity management, authorization, authentication and auditing are major requirements for almost all of the services provided by cloud computing. In some cases these services can be offered as third party services but most of the time they are part of the process to utilize the services by the customers. We have already discussed weak user authentication mechanism in cloud computing. Here we mention a few more related cloud specific problems:

*Weak Credential Reset Procedure:* The process of the resetting credential details in case of forgot or loss of credential must be in accordance with cloud computing as most of the cloud computing providers manage user credential themselves.

*Denial of Account and Denial of Services:* One of the policies defined especially in username password authentication mechanism is to lock out a user in case of many wrong credential entries. This often requires some human interaction, like the user of captcha, for the next authentication verification attempt. In case of desktop client applications which are pre-configured to login at remote locations, the services will be denied until human interacts with the system.

*Authorization Checks:* Web applications and services often provide insufficient authorization checks which may lead to helping the attacker guess the next changes to get an unauthorized record. For example, in case of a record being displayed to the customer (who an attacker) by id in the query string may help the attacker guess what is the next possible record to be accessed. The authorization checks need to be applied on individual services and at any location where there is a chance of bypassing a pre-determined flow. Like, the provider thinks that the user will come to a particular page after logging into the system but the attacker directly types a URL and append a query string to it in order to get a particular record which is consider to be unauthorized.

*Customizable Authorization:* Cloud service management interfaces should provide very much customized authorization configuration platform. The users can be categorized as groups and even within same groups the privileges can be vary from user to user and from time to time. The management interface should provide a configuration panel from where each and every user should be strictly provided only what s/he needs at that time.

*Activity Logging and Monitor:* Currently there is not standard for logging and monitoring the activities performed by users of cloud computing. Log files record everything being done on the servers and it is hard to filter them out for a particular users, particular access region and so on. For auditing user actions and traces it is very much required to provide a standard mechanism for logging and monitoring the users activities performed on the cloud system.

*iv. Management Interface Security*

On demand self service requires cloud computing to provide an interface to the customers from where they can self service them by rapidly providing changing the service provisioning without any provider's intervention. This interface is a web interface and it inherits all the problems of the protocols used for web application/services. Keeping all the control at a single place makes it more attractive for the attacker and breaching the security of this one point may lead the maximum possible loss that can occur in cloud computing.

Cloud Computing Core Technologies' Issues	Essential Cloud Characteristics Vulnerabilities	Issues in Known Security Policies	Vulnerabilities in Offerings of Cloud Computing
<ul style="list-style-type: none"> <li>• Web-Services and Applications</li> <li>• Virtualization Offerings</li> <li>• Cryptography</li> <li>• VM Hopping</li> <li>• VM Diversity</li> <li>• VM Denial of Services (DOS)</li> <li>• Session Riding/Hijacking</li> <li>• Cryptanalysis</li> </ul>	<ul style="list-style-type: none"> <li>• Internet Protocol</li> <li>• Unauthorized Access</li> <li>• Data Recovery</li> </ul>	<ul style="list-style-type: none"> <li>• IP-Based Network Zoning</li> <li>• Friendly Scanning VS Attacker's Scanning</li> <li>• Key-Management</li> <li>• Security Metrics to Monitor Security</li> </ul>	<ul style="list-style-type: none"> <li>• Weak Authentication</li> <li>• Injection (SQL, Command, Cross-Site Script)</li> <li>• Storage Security Risks</li> <li>• Communication Risks</li> <li>• IAAA Issues</li> <li>• Management Interface Security</li> </ul>

**Table 2: Cloud Computing Specific Vulnerabilities**

## VI. CONCLUSION

Cloud computing is growing fast in the market as for the offerings it promises with its customers. For most of the customers cloud security has always been a main concern. The level of security cloud is providing to the customers is unknown as there are no security standards defined which focuses on cloud computing. Current security controls are not appropriate in cloud computing in many cases. Changes need to be brought in the current security policies such that they become effective in cloud computing environment. There is a great deal of interest required in adapting those security modules accordingly. Similarly, there are some new security vulnerabilities which were not introduced before cloud computing evolution. New security modules are required to be introduced which are applicable in cloud environment. These security policies can be implemented as a service of cloud computing on demand. Not all the services of cloud computing require same level of security. For example, telemedicine and e-commerce may require a high level of security but provision of public information can do well in a less secured environment. Similarly, not all the users of the same service require same level of security. For example, for business discussion the voice conversation is required to be highly secured whereas security may not be of any concern while calling a friend using the same voice service on cloud. Thus protecting at the highest level of security is not always considered a good practice as different services require different level of security when used by different types of users. Keeping highest level of security for all the services and all types of customers can be costly.

## REFERENCES

- [1] Dinh, H. T., Lee, C., Niyato, D., & Wang, P. (2013). A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless communications and mobile computing*, 13(18), 1587-1611.
- [2] Rong, C., Nguyen, S. T., & Jaatun, M. G. (2013). Beyond lightning: A survey on security challenges in cloud computing. *Computers & Electrical Engineering*, 39(1), 47-54.
- [3] Ryan, M. D. (2013). Cloud computing security: The scientific challenge, and a survey of solutions. *Journal of Systems and Software*, 86(9), 2263-2268.
- [4] Whaiduzzaman, M., Sookhak, M., Gani, A., & Buyya, R. (2014). A survey on vehicular cloud computing. *Journal of Network and Computer Applications*, 40, 325-344.
- [5] Khan, A. R., Othman, M., Madani, S. A., & Khan, S. U. (2014). A survey of mobile cloud computing application models. *Communications Surveys & Tutorials*, IEEE, 16(1), 393-413.
- [6] Raju, R., Amudhavel, J., Kannan, N., & Monisha, M. (2014, March). Interpretation and evaluation of various hybrid energy aware technologies in cloud computing environment—A detailed survey. In *Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on* (pp. 1-3). IEEE.
- [7] Appistry Inc., *Cloud Platforms vs. Cloud Infrastructure*, White Paper, 2009.
- [8] Manvi, S. S., & Shyam, G. K. (2014). Resource management for Infrastructure as a Service (IaaS) in cloud computing: A survey. *Journal of Network and Computer Applications*, 41, 424-440.
- [9] P. T. Jaeger, J. Lin, J. M. Grimes, and S. N. Simmons, Where is the cloud? Geography, economics, environment, and jurisdiction in cloud computing, *First Monday*, 14(4 5): 2009.
- [10] "Cloud Computing", [http://en.wikipedia.org/wiki/Cloud\\_computing](http://en.wikipedia.org/wiki/Cloud_computing), Retrieved 2013-10-10 21:30.
- [11] "Cloud Computing", [http://en.wikipedia.org/wiki/Cloud\\_computing](http://en.wikipedia.org/wiki/Cloud_computing), Retrieved 2013-10-10 21:30.
- [12] Nagaraju Kilari, D. R. S. (2012, September). A survey on security threats for cloud computing. In *International Journal of Engineering*

Research and Technology (Vol. 1, No. 7 (September-2012)). ESRSA Publications.

- [13] Venters, W., & Whitley, E. A. (2012). A critical review of cloud computing: researching desires and realities. *Journal of Information Technology*, 27(3), 179-197.
- [14] Bernd Grobauer, Tobias Walloschek, Elmar Stöcker, "Understanding Cloud Computing Vulnerabilities", March/ April 2011, *International Journal of Computer and Reliability Societies IEEE*.
- [15] Wei, L., Zhu, H., Cao, Z., Dong, X., Jia, W., Chen, Y., & Vasilakos, A. V. (2014). Security and privacy for storage and computation in cloud computing. *Information Sciences*, 258, 371-386.
- [16] Fernandes, D. A., Soares, L. F., Gomes, J. V., Freire, M. M., & Inácio, P. R. (2014). Security issues in cloud environments: a survey. *International Journal of Information Security*, 13(2), 113-170.
- [17] Dhar, S. (2012). From outsourcing to Cloud computing: evolution of IT services. *Management Research Review*, 35(8), 664-675.
- [18] Ryan, M. D. (2013). Cloud computing security: The scientific challenge, and a survey of solutions. *Journal of Systems and Software*, 86(9), 2263-2268.
- [19] Mehmood, A., Majeed, M. F., Noor-ur-Rehman, A. K., & Sher, T. Inter-Operation between IPV4 & IPV6 by Embedding an Agent-Oriented Technology at Cloud Computing.
- [20] Bera, S., Misra, S., & Rodrigues, J. J. (2015). Cloud computing applications for smart grid: A survey. *Parallel and Distributed Systems, IEEE Transactions on*, 26(5), 1477-1494.
- [21] Toosi, A. N., Calheiros, R. N., & Buyya, R. (2014). Interconnected cloud computing environments: Challenges, taxonomy, and survey. *ACM Computing Surveys (CSUR)*, 47(1), 7.
- [22] Khan, R., & Mehmood, A. (2013). Realization of Interoperability and Portability among Open Clouds by using Agent's Mobility and Intelligence. *International Journal of Advanced Computer Science*, 3(11).
- [23] Patel, R., & Patel, S. (2013, February). Survey on resource allocation strategies in cloud computing. In *International Journal of Engineering Research and Technology (Vol. 2, No. 2 (February-2013))*. ESRSA Publications.
- [24] Mehmood, A.; Song, H.; Lloret, J. (2014). Multi-agent based framework for secure and reliable communication among open clouds. *Network Protocols and Algorithms*. 6(4):60-76. doi:10.5296/npa.v6i4.6028.

# A Framework for Web Search Log Evaluation for Testing Information Search in Cloud

Althaf Ali A

Research Scholar  
Bharathiar University, Coimbatore  
Tamilnadu, India

Dr. R. Mahammad Shafi

Professor and Head, Dept. Of MCA,  
Sree Vidyaniketan Engineering College,  
Tirupati, India.

**Abstract** — To meet the rapid growth of cloud technologies, many web information provider application are developed and deployed, and these applications run in the cloud. Because of the scalability provided by the clouds, a Web application can be visiting by several millions or billions of users. Therefore, the testing and evaluation of performance of these applications are becoming increasingly important. Web application usage log evaluation is one of the promising approaches to tackle the performance problem by adapting the content and structure of application to the needs of the users by taking advantage of the knowledge acquired from the analysis of the users searching activities from the web search logs. We propose a framework for web search log evaluation using classification and clustering method for effective testing information search in cloud. It also provides an information search ranking method to refine and optimizes the search evaluation process. We evaluate the proposed approach through implementing a web proxy in a server to record the user search logs and measure the retrieval precision rate for different users. A rate of 25% precision improvement is observed using different cluster testing for different users.

**Keywords-** Cloud, Web Search, Web Log, Classification, Clustering, Information Search, Testing.

## I. INTRODUCTION

The dependency over cloud based web application need increase with the growth of internet and web services, at the same time extensive challenges are builds for developers to provide conveniently and better services[5][18]. In general, web applications presents the same service for different users when they request for the services, not considering about their different needs and preferences [2]. For information retrieving from cloud based web application retrieves information based on the user request input, but the return response often miss the target web service page which user looking for due to very short scope of request interpretation[19]. The response on different subtopics or meanings of a request will be mixed together in the response list, thus implying that the user may have to sift through a large number of irrelevant items to locate those of interest. However search engines are good for searching but the search results acquired might not always be helpful to the user, as search engine fail to recognize the user intention behind the request.

A typical web search engine provides similar set of results without considering the intention of the user [1][3]. Therefore, an efficient model is needed which can give accurate outputs to the user with higher accuracy.

The most challenging problem that must be met during the information retrieval process is user privacy violation. Many users are reluctant to disclose personal information either implicitly or explicitly and be reluctant to visit websites that use cookies or avoiding to disclose personal data in the registration forms [17]. In both cases, the user's anonymity is lost and all records of their actions and, in many cases without their permission, used to know. A user with cookie technology, the site also agreed to supply personal information, in addition to the disclosure of such information without the consent of the user, can be exchanged between the sites [7].

The standard approaches automatically search and interprets by user agents in a standard format that allows Web sites to express their privacy policies. Key information about the data collected by a Web site to automatically convey to the user, and a site of personal exposure can differ from the methods and consumer preferences, therefore, the process of reading privacy policies, users will be trimmed for the data is automatically flagged [6].

Many traditional methodologies [9][10][11] are proposed based on the location preferences, collaborative filtering, hybrid content-based collaborative filtering techniques which have been developed for the websites, with the support of the recommendations such as the web search[12]. However, most of these approaches are suffer from a major drawback in which users can surf websites anonymously by proxy, and their identities are hidden and difficult to get. Some of the positive development of web testing are based on user feedback or to subscribe their interest. Users of these systems is time consuming and hence the desire to use such methods are not. More recent techniques derived from data stored in Web server logs, which aims to discover interesting search patterns are based on web search logs [14]. We propose a novel web search log evaluation over distributed cloud to overcome the difficulty of web search using user web search log classification and clustering techniques.

## II. RELATED WORKS

Web search displays user interest using their relevancy information of each user information retrieval [7]. As competition in the search market is growing rapidly, some search engines introduced web search service. For example, Google Cloud Personalized Web Search allows the user to define a category of websites of interest. Some systems use a Web search for the accuracy of the information processing requirements of the user feedback or ask users to register their demographic data already, so as to provide better service [11]. Since these methods require users to take additional measures to specify your preferences manually beyond search. The approaches that are capable of implicitly recognizing user's information needs should be developed. Since the need for personalized cloud Web search is increasing, many researches have to be done to provide the relevant information by considering the users situations [16].

Lu. M.T. and Yeung [9] proposed a framework that improve the effectiveness of the commercial web applications. They designed a group of rules to facilitate the web application development. A Meta model of a generic web application structure was described in [4], and splits the websites into several components: web pages, frames, links, and forms. Based on this model Tonella et. al.[10] designed a system to automatically analyse the websites.

J. Yu et. al. [12] describes the user context of mining on the basis of calculations for interactive web search. He describes that web search network is an effective way for same request and how to achieve the requirements of the user in real-time information as a key theme in web search. Han J Kim et al. [8] described on the development of user profile based on the concept of a network for a web search. This explains the innovative methodology for development of a network user profile concept for web search over cloud [10].

F. Akhlaghian et.al [13] described a web search using ontology-based network and fuzzy theory. The proposed web search engines use an automatic network fuzzy concept. The main objective is to use the concepts of ontology to improve the design of common fuzzy network built according to user profile. C.Biancalana et al., [15] proposed a new way for web search in the Web using social tagging in the expansion of the request. Social networks and common labelling systems quickly achieve greater recognition as the most important elements for categorization and data sharing using users tag and bookmarks, so as to facilitate the distribution of information and subsequent visits.

## III. WEB SEARCH LOG EVALUATION FRAMEWORK

The purpose of the processing of web usage logs of a user to record information in the log files of the Web server is back. The user's interest with regard to the behaviour patterns of web usage data by applying statistical and data mining techniques to the web site, it is also possible correlations between pages and the user can identify a web page as shown in Figure-1.

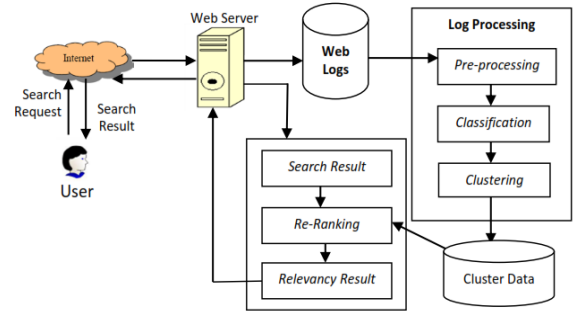


Figure -1 Web Search Log Evaluation Framework

User sends a search request to internet web server where server performs a search operation and sends the retrieved result for relevancy. Web server logs all user activity into web usage. User access log record time, request that the requested URL and the status code contains the IP address of the user. The recorded log usages construct cluster data model using Hierarchical clustered method and frequent pattern mining classification. The log usage processing is a background process which updates cluster data periodically. This approach minimizes the processing cost in real time. The clustered data are utilized for re-ranking and relevancy results.

Web access process data in manufacturing, innovation and a model for the analysis of the sample pre-processing, such as classification and clustering run in three different stages.

### A. Pre-processing

Web server records server log file for each access of Web page. Web log data, users, sessions, page views, etc. of a server log acts as a simple text file that can be processed in order to identify a user.

Each line represents a web request in the log file. When a visitor requests a Web page that contains two images, three lines will be appended to the log file in relate to the images which is included in the web page. Each line of the current page/folder in the Web site's URL with the IP of the requested file and request the user ID of the requested file having date and time stamp of the request type, location and making the user's computer's address status code, name and size which referred.

By removing noise and conflict in the log file data with pre-processing to improve the quality of data will help to normalize the data. It cleans the web log data from irrelevant entries recorded from page accesses, for example, an error, graphics, script file etc. The process parses the log file and transforms the data to normalize for ease classification and clustering.

### B. Web Search Log Classification

Classification applied in order to detect patterns using data mining methods. It helps to extract related data based on some rules. A multilevel association rule mining is applied on the proposed framework to build a pattern on pre-processed data. Multi-level association rules can mine data log efficiently with

the use of the concept of hierarchy under the support and trust framework. Overall, from the top down strategy is employed, which counts accumulate to calculate the associated item sets approach at all levels, from level 1, 2 and 3 concepts and working down the hierarchy to a more detailed conceptual levels until there you can find more related item sets.

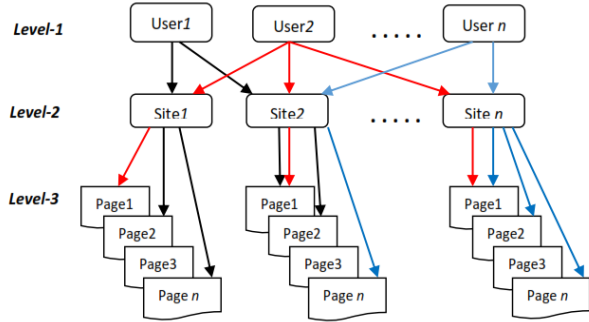


Figure -2 Multilevel Association Classifications of users log data

### C. Web Search Clustering

Clustering groups similar and dissimilar to one another on the basis of other data, the same group in the group listing process the data. A cluster of data that can be treated collectively as a group, and thus can be considered as a form of data compression. Separate classification groups is an effective tool, but it is a large set of models that characterize the sample in each group, in spite of the need for a proper collection and labelling.

The top-down manner in the framework of a hierarchical clustering method to group data into groups to implement a tree. This top-down strategy, starting with a cluster of top items, and each item that satisfies the conditions to cancel some of its own or to form a cluster, a cluster of tiny pieces and subdivides. Web search on the back so that they can monitor the cluster samples are stored in the ranking process.

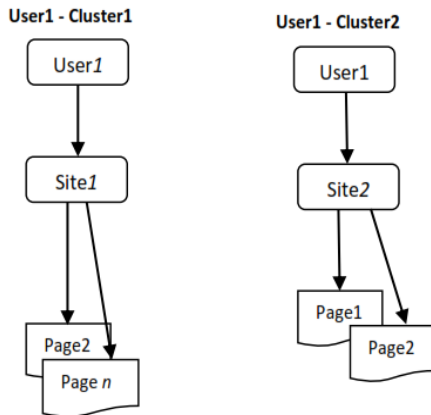


Figure-3 Hierarchical clustering of user-1 log data based on the above classification

## IV. SERVICE RE-RANKING METHOD

The framework implements the re-ranking algorithm to retrieved web search result. A search request  $Q$  requested by the user  $U$  to retrieve a search result. We assume that a set of cluster pattern data as  $C_d$  of user  $U$  has stored in cluster database of the framework. To compute the re-ranking for relevancy we compute the user web site visiting frequency as  $W_{freq}$ , link accessed frequency as  $A_{freq}$  and average value as  $P_{avg}$ , using  $W_{freq}$  and  $A_{freq}$ .

Let's assume that top 10 result set as  $S_R$  has obtained having results as  $R_1 \dots R_{10}$  and each result has link pointing a site as  $L_1 \dots L_{10}$  as on posing a request  $Q$  to a search engine.

To compute the  $W_{freq}$  of the obtained result of  $S_R$  we need to find the frequency of web site url as  $t$  accessed by user against the total distinct web sites url using clustered data pattern as,

$$W_{freq} = \frac{\sum_{i=0}^n ((t_i \in C_d) \rightarrow U)}{\text{distinct}(((t \rightarrow C_d) \neq t_i) \rightarrow U)} \quad (1)$$

To compute  $A_{freq}$  we need to find the frequency of  $t$  and link as  $L_i$  accessed for the user against the clustered data pattern of the user as,

$$A_{freq} = \frac{\sum_{i=0}^n ((t_i, L_i \in C_d) \rightarrow U)}{\text{distinct}(((t, L \rightarrow C_d) \neq t_i, L_i) \rightarrow U)} \quad (2)$$

To effective result ranking value we compute an average value as  $P_{avg}$  using  $W_{freq}$  and  $A_{freq}$  as,

$$P_{avg} = \frac{(W_{freq} + A_{freq})}{2} \times 100 \quad (3)$$

Based on the obtain  $P_{avg}$  of each result will be reorder for relevancy. The relevancy result will be send to user as search response. This approach of re-ranking will be effective as user usage log is a collection of all browsing activities. It might possible that user visit a site directly entering base URL instead of web search. The proposed approach utilized both search and direct visit log to build the relevancy result, which improvise the user web search and meet the required interest.

## V. EXPERIMENT TEST AND EVALUATION

To measure the effectiveness of proposed approach we measure the relevancy precision as  $PP_R$  of the obtain results. The measure of relevancy precision defined based on the number result re-ranked.

Let's assume a search result as  $S_R$  and a relevancy result as  $P_R$  for a given set of clustered data as  $C_D$  for a user generated. If  $P_R(r) \neq S_R(r)$  and  $P_{avg} > 0$ , where  $r$  is the result record, then we call it as relevancy precision of the result and it can be compute as,

$$PP_R = \frac{[\text{No. of result Relevant}]}{[\text{No. of Search Result}]} \times 100 \quad (4)$$



For experiment we install web proxy in a server to record the user logs. We perform a repetitive search as also direct visiting to various sites to download music files. To construct a clustered data pattern we run a infinitely interval based background java program which implements classification using multilevel association rule and clustering using hierarchical clustering method.

To evaluate the proposed approach we select 3 users and each user submits a request as “download mp3 songs” to a popular search engines as Google. We collect top 10 results from search engine and implements the re-ranking algorithm on each result to compute  $W_{freq}$ ,  $A_{freq}$  and  $P_{avg}$ . We generate the relevancy result based on the obtain  $P_{avg}$  value by result re-ranking and send to the user as search result. This process repeated for each user on submits of search request. Every search and navigation increase user web usage log, as cluster pattern generation is an infinite program run on intervals generates more relevant patterns of a user. An increase in user clustered patterns improves the precision of relevancy and user interest.

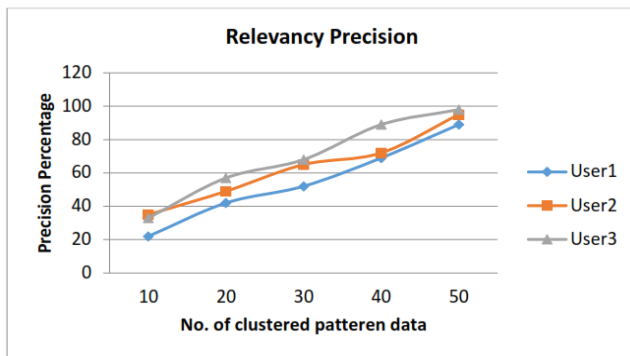


Figure-4 Relevancy Precision percentage

Figure-4 shows the relevancy precision with increase of clustered pattern data of users. An improve in precision percentage is observed with the increase of cluster pattern data, which suggests that web usage log can be useful input for user web search relevancy and the proposed approach efficiency.

## VI. CONCLUSION

Web search engines, user information need only be satisfied with a little ambiguous question. It is difficult to provide basic information retrieval and search results are customized to each user's web search. In this paper, a new classification and grouping approach using web usage for effective website search for the web relevancy is proposed, in studying how a search can be relevancy accurately identify user's web usage log data. An advance preparation of support resource required for relevancy reduces the recall cost. Experiment shows an improvisation in Relevancy Precision percentage which suggests that web usage log can be useful input for user web search relevancy and user interest and can effective model for cloud testing.

- [1]. J. Lai and B. Soh, "Personalized Web search results with profile comparisons", 3rd International Conf. on Information Technology and Applications-2005, Vol. -1, Pages. 573 - 576, 2005
- [2]. G. Candea, S. Bucur, and C. Zamfir, "Automated software testing as a service", In the 1st ACM symposium on Cloud computing (SoCC '10), 2010.
- [3]. M. Kutub, R. Prachetaa and M. Bedekar, "User Web Search Behaviour", 3rd International Conference on Emerging Trends in Engineering and Technology, Pages. 549 - 554, 2010.
- [4]. Xiaolin Xu, Hai Jin et.al., "URMG: Enhanced CBMG-Based Method for Automatically Testing Web Applications in the Cloud", Tsinghua Science and Technology, 19(1): 65-75, February 2014
- [5]. Pang Xiong Wen, Li Dong, "Quality Model for Evaluating SaaS Service", Fourth International Conference on Emerging Intelligent Data and Web Technologies, 2013
- [6]. W. Lin and Y. Liu, "A novel website structure optimization model for more effective web navigation", in Proceedings of Workshop on Knowledge Discovery and Data Mining, Los Alamitos, CA, USA, pp. 36-41, 2008.
- [7]. Fang Liu, C. Yu and Weiyi Meng, "Personalized Web search for improving retrieval effectiveness", IEEE Transactions on Knowledge and Data Engineering, Vol.-16, No.-1, Pages. 28 - 40, 2004.
- [8]. Han-joon Kim, Sungjick Lee, et.al., "Building Concept Network-Based User Profile for Personalized Web Search", 9th International Conference on Computer and Information Science (ICIS), Pages. 567 - 572, 2010.
- [9]. M. T. Lu and W. L. Yeung, "A framework for effective commercial web application development", Internet Research: Electronic Networking Applications and Policy, vol. 8, pp. 166-173, 1998.
- [10]. P. Tonella and F. Ricca, "Analysis and testing of web applications", in Proceedings of the 23rd International Conference on Software Engineering, Los Alamitos, CA, , pp. 25-34, USA, 2001.
- [11]. Xuwei Pan, Zhengcheng Wang and Xinjian Gu, "Context-Based Adaptive Personalized Web Search for Improving Information Retrieval Effectiveness", International Conference on Wireless Communications, Networking and Mobile Computing, Pages. 5427 - 5430, 2007
- [12]. Jie Yu and Fangfang Liu, "Mining user context based on interactive computing for personalized Web search", 2nd International Conference on Computer Engineering and Technology, Vol.-2, Pages. 209-214, 2010.
- [13]. F. Akhlaghian, B. Arzanian and P. Moradi, "A Personalized Search Engine Using Ontology-Based Fuzzy Concept Networks", International Conference on Data Storage and Data Engineering (DSDE), Pages. 137 - 141, 2010.
- [14]. H. Liu and V. Keselj, "Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users future requests", Data and Knowledge Engineering, Elsevier, 2007.
- [15]. C. Biancalana and A. Micarelli, "Social Tagging in Query Expansion: A New Way for Personalized Web Search", International Conference on Computational Science and Engineering (CSE '09), Vol. 4, Pages. 1060 - 1065, 2009.
- [16]. Jerry Gao, Xiaoying Bai, and Wei-Tek Tsai, "Cloud Testing- Issues, Challenges, Needs and Practice", Software Engineering : An International Journal (SEIJ), Vol. 1, No. 1, Sept-2011.
- [17]. G. Ruffo, R. Schifanella, M. Sereno, and R. Politi, "WALTy: A user behavior tailored tool for evaluating web application performance", in Proceedings of Third IEEE International Symposium on Network Computing and Applications (NCA 2004), Los Alamitos, CA, USA, 2004, pp. 77-86.
- [18]. L. M. Riungu, O. Taipale, K. Smolander, "Research Issues for Software Testing in the Cloud", 2nd IEEE International Conference on Cloud Computing Technology and Science, pp. 557-564, Nov 2010.

- [19]. T. Vengattaraman, P. Dhavachelvan, R. Baskaran, "Model of Cloud Based Application Environment for Software Testing", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 3, 2010.



# An Integrated Mobile Application for Enhancing Management of Nutrition Information in Arusha Tanzania.

Neema Mduma

School of Computation and Communication Science & Engineering  
Nelson Mandela African Institution of Science and Technology  
Arusha, Tanzania

Khamisi Kalegele

School of Computation and Communication Science & Engineering  
Nelson Mandela African Institution of Science and Technology  
Arusha, Tanzania

**Abstract** — Based on the fact that management of nutrition information is still a problem in many developing countries including Tanzania and nutrition information is only verbally provided without emphasis, this study proposes mobile application for enhancing management of nutrition information. The paper discusses the implementation of an integrated mobile application for enhancing management of nutrition information based on literature review and interviews, which were conducted in Arusha region for the collection of key information and details required for designing the mobile application. In this application, PHP technique has been used to build the application logic and MySQL technology for developing the back-end database. Using XML and Java, we have built an application interface that provides easy interactive view.

*Keywords-* Nutrition information; MySQL; XML; Java; PHP; Mobile Application.

## I. INTRODUCTION

The mobile technology has been the most fastest growing media technology used in the healthy sector in Tanzania in recent years compared to other media technologies [1]. This technology directly targets the general public through engaging users in health related activities, and thereby improving accessibility to quality health information, health services, and encouraging user behavior that involves seeking preventive health solutions [2]. The wide spread of mobile phones has led to significance increase in mobile applications for providing access to various information that are needed by the community. Mobile applications have been designed to run on mobile devices and allow users to interact with service providers. Our proposed system is in the form of an integrated mobile application, which is designed to enhance management of nutrition information.

The proposed mobile application will be integrated with the existing health centre system. The health centre system used is Open MRS. The proposed system will allow nutrition practitioners to send information and recommendations to the targeted user. In this aspect, the user will be able to access nutrition information and request any other nutrition related details or seek advice when necessary. The application will also provide reminders so as to notify the user on necessary events such as clinic visit for vitamin A supplements. In this application, nutrition tips will be generally provided and available for all users, and the users will be able to view

nutrition tips and request for new tips based on their concern and nutrition practitioners will respond to the request accordingly. In responding to the nutrition tips enquiries, nutrition practitioners' profile will be specified so as to show the validity of the tip. In this application, the researcher will be able to generate nutrition reports based on provided information and the administrator will monitor the overall activities of the system and be responsible for user approval. The application will allow user interaction whereby the authorized user will be able to view the historical recommendation and request assistance when needed. The system will be user interactive and support two way flow of nutrition information.

## II. METHODOLOGY

The requirement gathering was conducted in Arusha region. The method used in this study based on qualitative research methods such as literature review and interviews whereby casual talks were conducted for the collection of information. Through the interviews, we interacted with the nutrition practitioners together with prenatal and post-natal mothers and collected data relevant for specifying the requirements for developing the mobile application.

## III. REQUIREMENT SPECIFICATION

This study involves both functional and non-functional requirements. The functional requirements for developing this mobile application covers the issues of recommendation as set of nutritional information that are suggested by nutrition practitioners to the user based on the user's described information and nutrition tips as the set of nutritional information concerning nutrition improvements added by nutrition practitioners for the user. The functional requirements also include a reminder as notification provided to users based on necessary nutrition events and reports that are generated by researcher based of nutrition information provided. The non-functional requirements of the system cover the issues of maintainability, operability, performance and security of the system.

## IV. DESIGNING THE PROPOSED SYSTEM

In this study, the design part was illustrated on two major parts using Data Flow Diagram (DFD). Fig.1 shows administration management data flow diagram and Fig.2 shows tips and recommendation management data flow diagram.

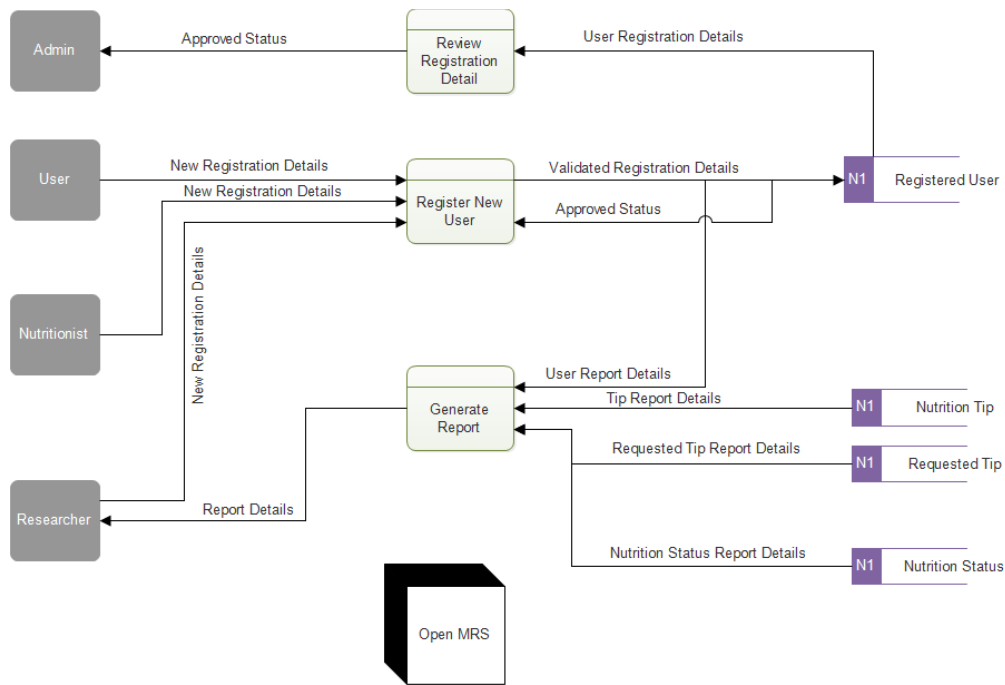


Fig. 1 Administration management data flow diagram

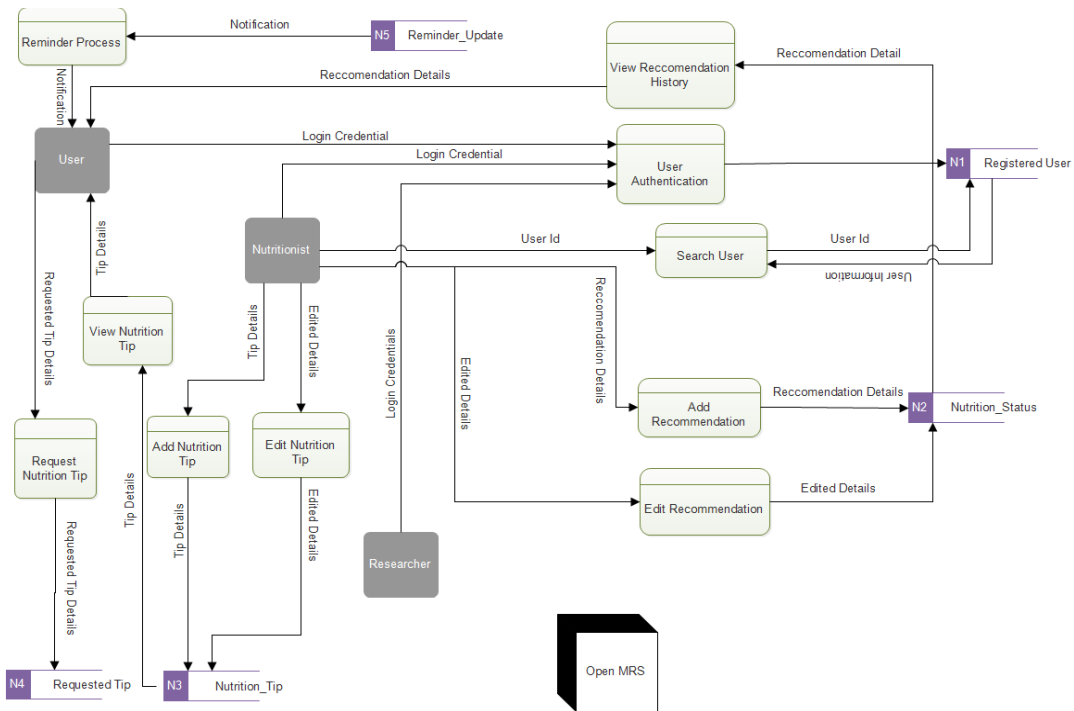


Fig. 2 Tips and Recommendation management data flow diagram

To accomplish the implementation part for the proposed application, a model adopted from SDLC has been chosen for developing a successful information system. The software development life cycle (SDLC) is a framework that defines the tasks performed at each step in the software development process. It consists of a meticulous plan that describes the processes for developing, maintaining, replacing and altering the specific software. The SDLC defines the method for software quality enhancement and the overall development process [3]. To make the complete product to deliver faster, we decided to use the Rapid Application Development (RAD) model.

#### A. Rapid Application Development

The RAD is a model designed to facilitate much faster software development and provides higher quality results compared to the traditional lifecycle; this model delivers faster and higher quality product [4]. In this study, we preferred to use RAD as it proved to be successful tool for developing our mobile application. Fig. 3 shows the Rapid Application Development model of our system.

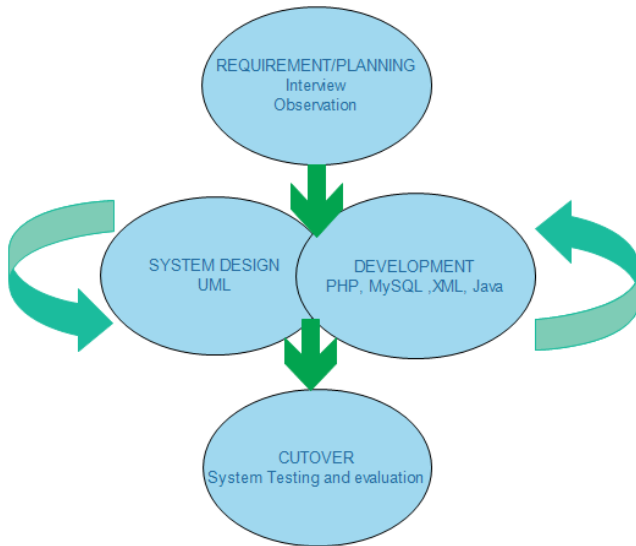


Fig. 3 The System's Rapid Application Development Model

#### B. Mobile Application

Mobile application is a type of application software that takes advantage of the mobile technology, and it can be used with any other technology apart from mobile phones [5]. The numerous functions and services offered prompt the extensive use of the mobile applications. In this paper, we use android mobile application in order to distinguish with other Unstructured Supplementary Service Data (USSD) applications that provide limited information and don't support storing of provided information. The reason is to provide two way flow of information by supporting interaction and allow access of large amount of information.

1) *Why choose Android:* Android is one of the most powerful and flexible open source platforms and its increasingly becoming popular. There are no licensing fees; this increases preference of many developers. In this study, we

preferred developing mobile application supported by this operating system with consideration of market terms. The growth of mobile devices such as mobile phones is a worldwide phenomenon with mobile phone ownership outstripping computer ownership in many countries. Also there is an increase in smart phones, which created a growth market for advanced mobile applications [6].

#### C. PHP

In developing the mobile application, we used the Hypertext Pre-processor (PHP) because this is one of the server-sided languages widely-used in software development and is an open source scripting language that we found appropriate for developing our system.

1) *Why use PHP:* PHP was preferred in this development study because, first it is simple and thus easy to learn. It efficiently runs on the server side and its codes runs faster due to the fact that it runs in its own memory space so it has a fast loading time. The PHP has tools that are open source software, and thus are freely available for use. Furthermore, it is flexible for database connectivity and it supports a wide range of databases. Additionally, the PHP can connect to a number of databases, but MySQL is the most commonly used as it can also be used at no cost [7].

In addition, PHP is compatible with almost all servers and its security features allow many functions to protect users against certain attacks. This language runs on various platforms such as Android, Windows and so many others.

#### D. MySQL

MySQL is one of the database systems that run on a server and uses the standard Structured Query Language (SQL). It is easy to use, reliable and it runs very fast. In this study, we used MySQL Database so as to enable the cost-effective delivery of reliable and high-performance application. The data in a MySQL database are stored in tables and offers a flexible programming environment [8]. Database systems are vital in computing and can be used as standalone utilities or as part of other applications.

1) *Why use MySQL:* The MySQL database server provides the ability to handle applications that are deeply embedded and offers platform flexibility; this is a MySQL stalwart feature. It allows customization so it is easy for a programmer to improve the database server by adding unique features.

MySQL has been used by many database professionals due to the unique storage-engine architecture that allows configuration of the database server remarkable end results performance in particular applications.

Apart from that, MySQL offers a variety of unique high-availability database server options ranging from high-speed master/slave replication configurations, specialized cluster servers offering instant failover, to third party vendors. So it provides high availability for programmers to rely on it.

MySQL protects data through its outstanding security features; it has powerful mechanisms, which ensures that access to the database server is possible only to authorized users and other users are limited to the client machine level.

MySQL also has granular object privilege framework, which ensures that users can only see what they are supposed to see. Another important feature is that it has powerful data encryption and decryption functions, which protects sensitive data from unauthorized users. Secure Shell (SSH) and Secure Sockets Layer (SSL) are provided to ensure safe and secure connections. It also provides backup and recovery utilities so as to allow complete logical and physical backup, and also full and point-in-time recovery.

MySQL offers full support needed for development of applications and developers can get all they required for developing information systems that are based on databases. [9].

#### E. XML

Extensible Markup Language (XML) is designed to describe data [10]. This language is used as a medium for carrying information independently from the involved software and hardware of the information system. Through the XML, you can create information formats and structured data can be shared electronically. XML data is self-describing, which means the data and its structure are embedded replacing the need for pre-building the structure for storing the data when it arrives. XML allows sharing of information in a consistent way due to its simpler format [11].

1) *Why XML:* XML has good features for storing and transmitting information, which simplifies data storage and sharing. This language is useful in accurately describing and identifying information without mistake so as to allow information to be understood [12]. Standardized description and control of particular types of document structure is possible in XML. It provides messaging systems' common syntax to facilitate information exchange between applications. In this study, we decided to use XML because it is free so we don't need to pay and it is easier to upgrade without losing data.

#### F. Java

Java is a programming language and computing platform that is designed to support many applications to work. This language is fast, secure, and reliable so as to ensure developers about performance, stability and security of the developed application [13].

1) *Why Java:* In this study, we decided to use Java because it is platform independent so applications can run on many different types of devices such as computers and even mobile phones. Java is essentially made up of objects, which are programming elements, and therefore it is object-oriented [14]. This language is very simple, so it is easier for the developer to engage it in application development.

### V. RESULTS

An integrated mobile application has been developed as a result for enhancing management of nutrition information and integration with existing health system as declared in this study. Results show the system interface that was developed by using XML and Java so as to allow user interaction with the system.

#### A. System Interface

Designing an interface is described as the process of developing a method in a system to connect and communicate so as to allow exchange of information. This acts as a channel of communication between user and application. Interface design focuses on anticipating what users might need to do and ensuring that the interface has elements that are easy to access, understand, and use to facilitate those actions [15].

1) *Interface for mobile application:* This section provides some of the developed interface for this application. First, the system administrator will register the users by approving their registration requests as no one can use the system without registration. Users will be using mobile phones to access this application. The application interface is presented in the Fig. 4 below.

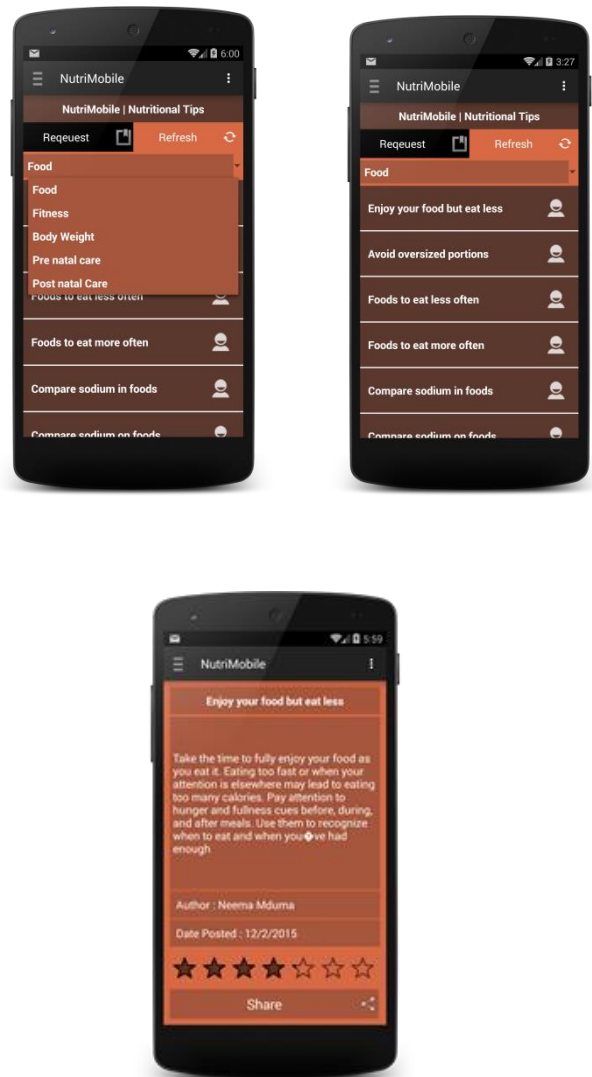


Fig. 4 System interfaces

## VI. CONCLUSION

This study was on developing an integrated mobile application for managing nutrition information in Tanzania. The system development used various methods and materials, which were determined after the design process discussed in this paper, and which culminated into development of a mobile application for management of nutrition information. Mobile phones were chosen as the tool to manage nutrition information so as to allow interaction without time and place limitations due to ownership issues. After registration, only authorized users will be able to access the information. The system administrator is the one responsible for the approval of user registration and this will provide security. All nutrition information is provided by nutrition practitioners and the system will allow sharing of that information via social networks. The user will be reminded in case of any necessary event concerning nutrition and clinic visits so as to increase efficiency. On the other hand, the user can request for nutrition information and nutrition practitioners will respond accordingly.

## REFERENCES

- [1] H. Cole-Lewis and T. Kershaw, "Text messaging as a tool for behavior change in disease prevention and management," *Epidemiologic Reviews*. 2010.
- [2] Rick van Genuchten, Wouter Haring, Daan van Kassel, and Kaoutar Yakubi, "Mobile phone use in Tanzania," *TAN2012 Mark. Res.*, 2012.
- [3] K. Schwaber, "Software Development Life Cycle (SDLC)," *Tutorials point simple easy Learn.*, 2001.
- [4] M. A. Hirschberg, "STN 2-1 Topic: Rapid Application Development (RAD): A Brief Overview," 2015. [Online]. Available: <https://www.mendeley.com/library/viewer/?fileId=6f1c4f30-d7da-50d9-67a7-2c02747d1f83>.
- [5] J. Muthee and N. Mhando, "AMDII-BBC-summary-report," *African Media Dev. Initiat.*, 2006.
- [6] R. Meier, *Professional Android 2 Application Development*. 2010.
- [7] Pitts Monica, "5 Reasons Why PHP Is a Great Programming Language | Mayecreate Design," 2015. [Online]. Available: <http://www.mayecreate.com/2013/12/5-reasons-php-great-programming-language/>. [Accessed: 16-May-2015].
- [8] Heng Christopher, "What is MySQL? What is a Database? What is SQL? (thesitewizard.com)," 2010. [Online]. Available: <http://www.thesitewizard.com/faqs/what-is-mysql-database.shtml>. [Accessed: 16-May-2015].
- [9] MySQL, "MySQL :: Top Reasons to Use MySQL," 2015. [Online]. Available: <https://www.mysql.com/why-mysql/topreasons.html>. [Accessed: 16-May-2015].
- [10] Hall Marty and Brown Larry, "Microsoft PowerPoint - XML.ppt - xml.pdf," 2001. [Online]. Available: <http://www.cs.toronto.edu/~mlou/csc309/xml.pdf>. [Accessed: 17-May-2015].
- [11] Rouse Margaret, "What is XML (Extensible Markup Language)? - Definition from WhatIs.com," 2001. [Online]. Available: <http://searchsoa.techtarget.com/definition/XML>. [Accessed: 17-May-2015].
- [12] Tutorialspoint, "xml\_tutorial.pdf," 2014. [Online]. Available: [http://www.tutorialspoint.com/xml/xml\\_tutorial.pdf](http://www.tutorialspoint.com/xml/xml_tutorial.pdf). [Accessed: 17-May-2015].
- [13] Eck David, Hobart, and Colleges William, "()- javanotes6-linked.pdf," 2014. [Online]. Available: <http://math.hws.edu/eck/cs124/downloads/javanotes6-linked.pdf>. [Accessed: 17-May-2015].
- [14] Lowe Doug and Burd Barry, "What Is Java, and Why Is It So Great? - For Dummies," 2007. [Online]. Available: <http://www.dummies.com/how-to/content/what-is-java-and-why-is-it-so-great.html>. [Accessed: 17-May-2015].
- [15] Usability, "User Interface Design Basics," May 2014.

# Automated Vehicle Tracking by GPS Modern Tecnology

Fahmi Ahmed  
Faculty of Engineering  
University Of Development  
Alternative (UODA)  
Dhaka, Bangladesh

A.B.M. Rocknuzzaman  
Faculty of Engineering  
Universität Duisburg-Essen,  
Duisburg, Germany

Syed Foysol Islam  
Faculty of Engineering  
University Of Development  
Alternative (UODA)  
Dhaka, Bangladesh

**Abstract**— This Research paper represents, Automatic vehicle tracking by GPS modern system. The results obtained in a purpose-designed computer software for track Vehicle's location and other data using by modern GPS or GLONASS technology. For this system need two types of product, one is technology product and another is web system.

**Keywords**- Microcontroller, IC Max232, GPS Antenna, GPRS and GSM Antenna, Satellite Communication, UART Communication, GPS, GPRS, GSM, Web Application, Google API, Open Street AP.

## I. INTRODUCTION

Building the Automated Vehicle Tracking by Modern Technology needs two major design units. One is Electronic side, Embedded hardware for vehicle Unit, which one sends the data to protocol of vehicle position and other status. Another one is customized software for monitor the vehicle which one receives data from sander (vehicle hardware unit) and store date for future. For this system find out the vehicle position, speed of vehicle and other status (AC, Ignition, Oil or Gas etc).

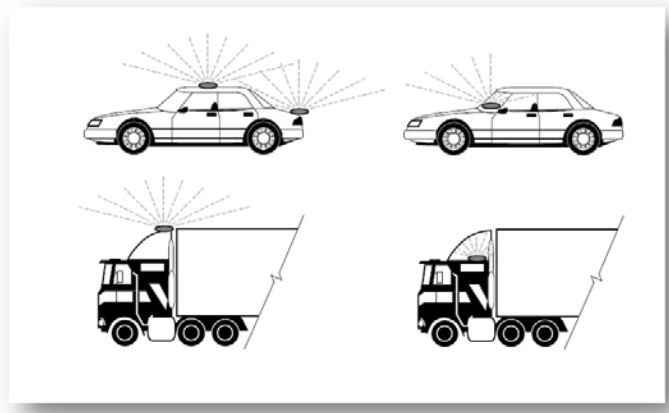


Figure 1: Vehicle Tracking by GPRS [4].

## II. HARDWARE OF VEHICLE ANALYSIS

Vehicle hardware unit have several parts. As like GPS receiver, which one is quality that must be provided to the GSM Cellular System. Another one is Central Processing Unit

which makes the status of vehicle. Another is desktop application which collects the data and sends the data to the receiver [2].

### A. GPS Reciever

GPS receiver is used to capture the current location and vehicle speed but this one is not in human understandable format. This raw data needs to be processed to convert it into useful information that can be displayed by a beacon on the map. CPU is required to process this raw data. SiRF Star III single-chip GPS receiver is used which comes integrated with SIM548C - GPS which is GSM/GPRS modem which is used for data transmission [3].

### B. Design of vehicle unit

In-Vehicle Unit is designed using OEM module Telit GM862-GPS GSM/GPRS modem and microcontroller PIC18F248 manufactured by Microchip. Figure-1 shows the block diagram of In-Vehicle Unit.

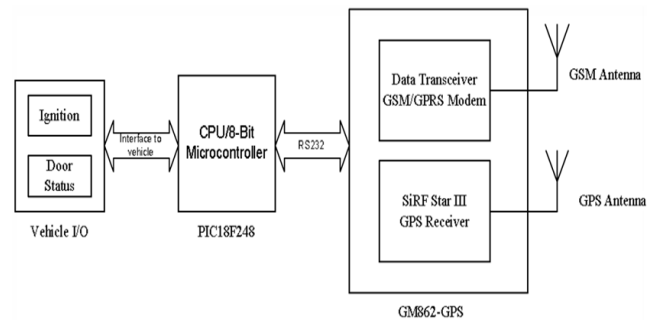


Figure 2: Design of In-Vehicle Unit [3].

GPS antenna receives signals from GPS satellites and it must face towards sky for correct computation of the current location by GPS receiver. Location data is transferred to microcontroller through serial interface. After processing of the data provided by GPS receiver, microcontroller transmits this information to remote location using GSM/GPRS modem. Microcontroller controls the operation of GSM/GPRS modem through serial interface using AT commands. External GSM antenna is required by the GSM/GPRS modem for reliable transmission and receiving of data. When modem receives any command sent by tracking server, it passes this information to microcontroller which analyses received information and

performs action accordingly (i.e. turns on/off ignition of vehicle, transmits current location, restarts GPS receiver, restarts whole system etc). Some of microcontroller I/O ports are connected to vehicle ignition on/off circuitry and door status output of vehicle. Information packet sent to server also contains status information of these I/O ports [4].

C. Vehicle Unit Software Design

Microcontroller is acting as Central Processing Unit for Vehicle unit. All operations of the In-Vehicle Unit are to be controlled by the microcontroller. Microcontroller needs instructions to operate the whole system. These instructions are provided to microcontroller by writing the software into microcontroller’s flash memory. It reads the software instruction by instruction and performs the action as required by instruction. Complete software is broken down into small modules as shown in Figure-2 [5].

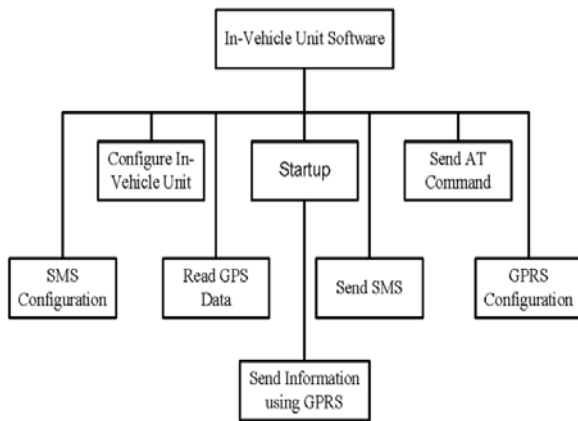


Figure 3: In-Vehicle Unit Software Design [5].

TABLE I: List of Vehicle Unit parts [1].

Parts	Model No.
Battery	
Buzzer	
Button	
Capacitor:	1 $\mu$ F
	10 $\mu$ F
	100 $\mu$ F
	1000 $\mu$ F
	1 nF
	100 nF
Communication Port:	D89 Male
	D89 Female
Connectors	
Diode:	1M1170ZS5
	1N4007
	Zener Diode
Display:	20*4 LCD
EEPROM:	FM24C64
GPS/GSM/GPRS Module:	SIM548C
GPS and GPRS Antenna	

IC:	MAX 232
Inductor:	6.8 nH
LED	
Microcontroller:	Atmega128
Opamp:	4136
SIM Card and SIM Holder	
Transistor:	2N3390
	MJE340
Voltage Regulator:	IC-7805
	IC-LM317

TABLE II: List of Simulation and Design parts [1].

Unit	Name of software
Simulation Coding:	AVR Studio
Simulation Software:	Proteous
PCB Design Software:	Proteous
	Orcad
Signal Capture and Analysis:	Logic Analyzer
Digital Oscilloscope	
Map:	Google Map

D. Desktop Application Design

1. By using C#.NET customized software has developed and tests the communication with satellite, communication with web server and check AT commands.
2. Hyper Terminal is also used for check communication. In this project we use MikroC compiler’s build in USART Terminal for checking data communication.

i) Algorithm send “AT” command and others

This subroutine is the basic routine which handles all the communication with GM82-GPS.

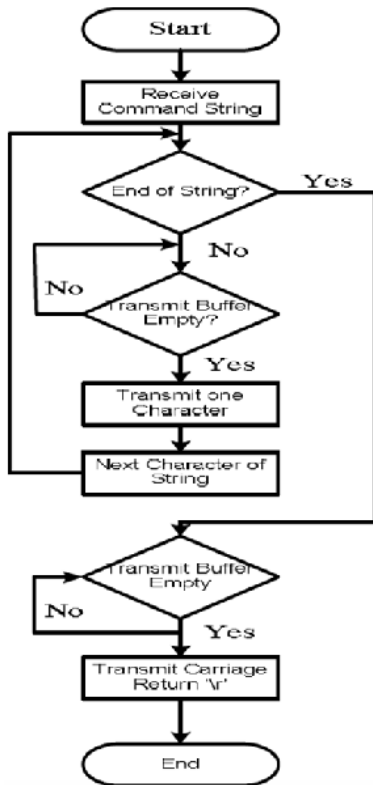


Figure 4: Flowchart of subroutine send "AT" commands [6].

This routine accepts the string containing "AT" command input in its parameters and sends this string character by character to module. GM862-GPS accepts carriage return ("\r") as a command terminating character. As this character is received it sends back the response to microcontroller. Figure-5 shows the flowchart.

As shown in Figure-6, the flow chart routine checks each character of string, if the character is not null, it will check the transmit buffer contents. If transmit buffer is empty, it will write new character into the buffer. Transmit buffer is a hardware register of UART. As soon as an 8-bit data is written into the transmit buffer, UART hardware transmits that character at the specified baud rate. Each character of command string will be sent in this way. When the null character is found, it specifies end of string and routine terminates by sending carriage return to the module. When Response received from the module will be handled in another subroutine [6].

ii) Algorithm Subroutine- Startup

Startup routine is executed only when device is powered on. It initializes all hardware of the In-Vehicle Unit and configures GM862-GPS. It performs various tests to ensure the GM862-GPS is working properly and is ready to use.

All peripherals in use need to be initialized in this step. After initializations of local peripherals, GM862-GPS needs to be tested. Microcontroller sends "AT" command to GSM module using subroutine AT Command.

Every commands sent to module are using this subroutine. If the device responds with "OK", it means microcontroller can communicate with module. If device doesn't respond after expiration of timeout routine is restarted. If problem persists definitely something in hardware is damaged. After receiving "OK" response from module various parameters of module need to be initialized. SIM presence is checked by sending command "AT+CPIN?". If device responds with "+CPIN: READY" message, SIM is ready to use. Any other response message will be considered as an error and routine will be restarted after expiration of timeout. When SIM card is ready, it is important to test whether module is connected to network or not. Network status can be tested with command "AT+CREG?". If module responds with "+CREG: 0, 1" module is connected to network and data can be sent over network. If any other response is received module keeps on checking for network status until it connects to network. Once it makes sure that module is connected to network, subroutine is terminated.

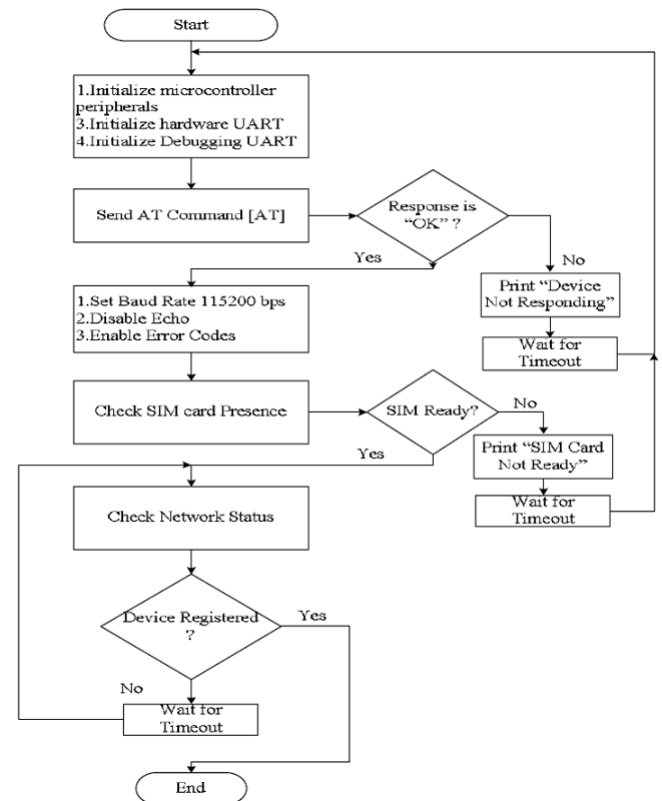


Figure 5: Flow Chart of Startup Subroutine [7].

iii) Subroutine- Read GPS Data

GPS controller is by default powered on when module is switched on. Figure 5 shows the flow chart for Read GPS Data subroutine. As shown in the flow chart subroutine first of all checks whether GPS controller is powered on? To check this "AT\$GPSP?" is sent to the module. If it responds with \$GPSP: 0 it is not powered up. If it is not already powered up; it can be switched on by sending "AT\$GPSP=1". Once GPS controller is powered



up location information can be read from it by sending "AT\$GPSACP". The module responds with a long NMEA sentence. The information of interest is latitude, longitude, speed, number of satellites used in calculating latitude and longitude. This information is extracted from the received response and saved in formatted string. This string can be later on passed to Send SMS subroutine to send it to remotely located Tracking Server [7].

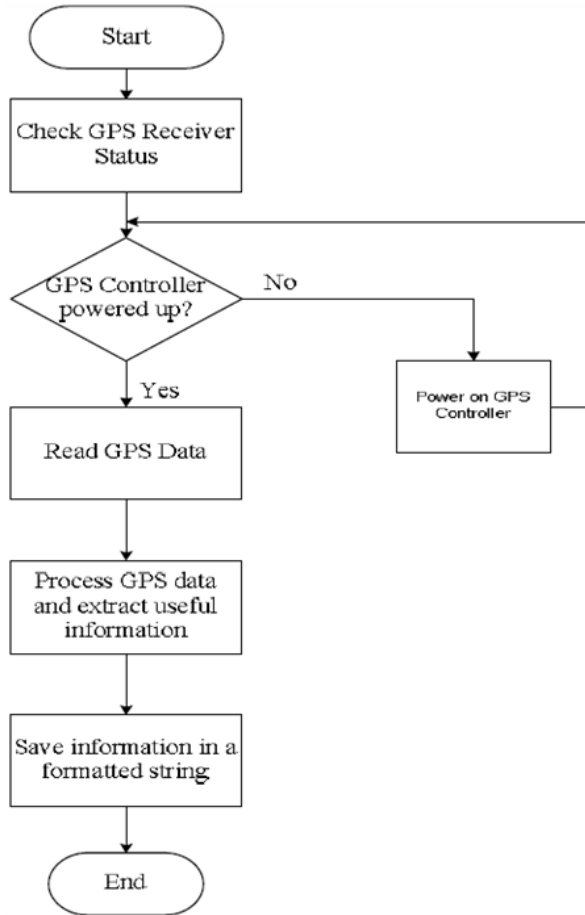


Figure 6: Flow Chart Of Subroutine Read GPS Data [8].

iv) Subroutine- Send SMS

This subroutine accepts message string as input parameter which needs to be transmitted. Subroutine adds a terminating character Ctrl-Z at the end of message string as shown in Figure-8.

Then it checks whether module is in Text SMS mode. It can be checked by sending command "AT+CMGF?" If module responds with "+CMGF: 0" it is in PDU mode. Mode can be changed to text by sending command "AT+CMGF=1". To send an SMS module requires destination phone number that is sent to module using command "AT+CMGS= da" where da represents the destination phone number. This phone number will be read from microcontroller internal memory which is stored during programming. After sending destination number module waits for prompt ">". When prompt appears message string is sent using Send AT Command subroutine. If message sent successfully, module

responds with +CMGS: where mr is message reference number. If any error occurs subroutine tries to resend the message until it is successfully sent.

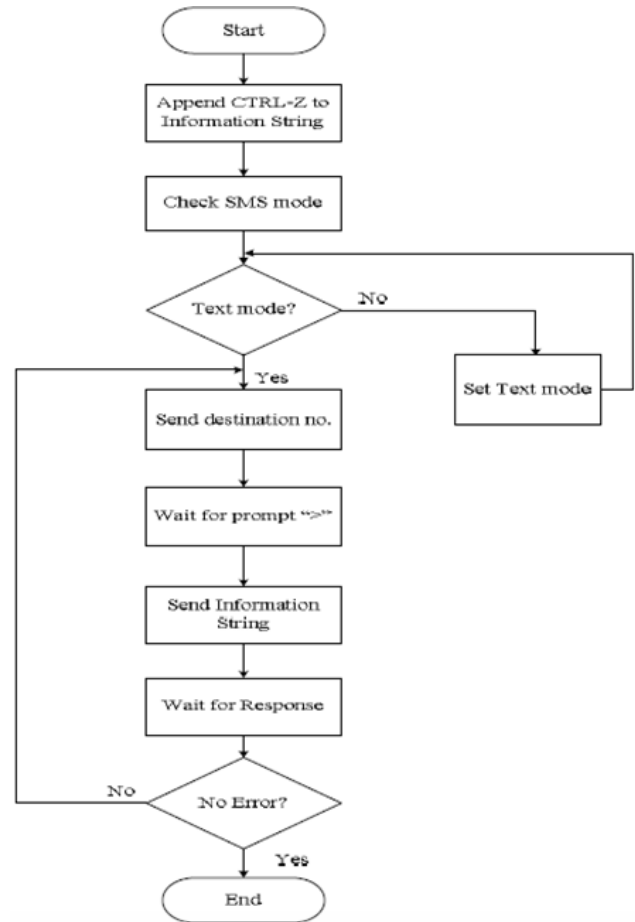


Figure 7: Flow Chart Of Subroutine Send Sms [10].

v) Subroutine- SMS configuration

SMS configuration subroutine is call after startup routine. It is basically called once after powering up the In-Vehicle unit like startup routine.

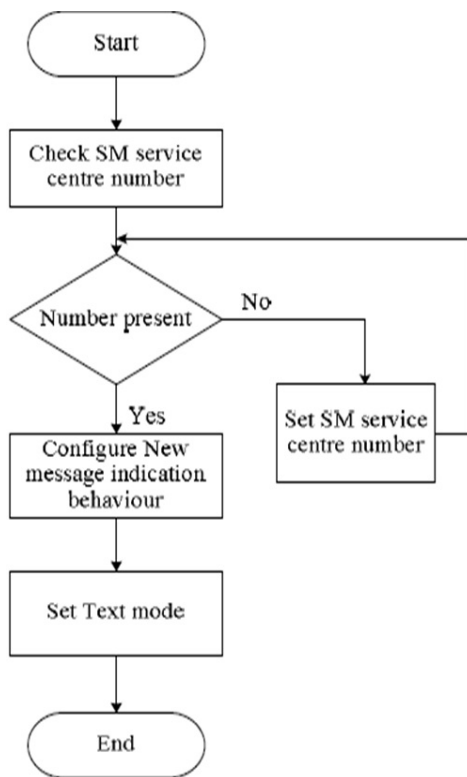


Figure 8: Flowchart of Subroutine – SMS Configuration [9].

It can be part of startup routine but it is separated because it does configuration of the module related to SMS only. Figure shows the flow chart. This subroutine checks the SMS service centre number by sending the command “AT+CSCA?” Service centre number is required because SMS is routed to destination via SMS service center. The module responds with “+CSCA: number”. If no number is present it can be saved in module by sending the command “AT+CSCA= number, type” type could be 145 if number is in international number format (i.e. it begins with +) or it could be 129 if number is in national format. When new message is received by module an unsolicited indication is generated. This indication may be sent to microcontroller, buffered if microcontroller is busy or discarded. In this case new message must be immediately sent to microcontroller or buffered if microcontroller is busy. This configuration can be done by sending command “AT+CNMI=1, 1, 0, 0, 0” when GSM modem receives a new message it will send “+CMTI: “SM”, message index no” where message index no is location of message in memory and it can be then read by sending command “AT+CMGR=message index no”. After configuring new message behavior module is set to Text mode for SMS. It can be done by sending command “AT+CMGF=1”. All configuration related to SMS is finished and subroutine terminates [8].

vi) Subroutine- Configure GPRS

When GPRS service is available, it is cost effective and more efficient to transmit vehicle information through GPRS. In order to connect to GPRS, it needs to be

configured. Figure 8 shows the steps required to configure the GMS module for GPRS data transmission. First step in configuration of GPRS is to define GPRS context. It is set of information to identify the internet entry point interface provided by the ISP. With these parameters the GPRS network identifies the ISP to be used to gain access to the internet and defines the value of IP address of the GPRS device once connected.

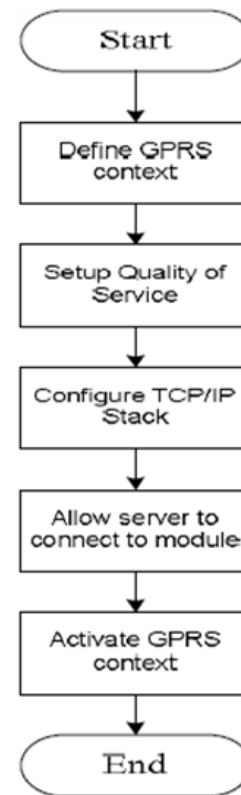


Figure 9: Flow Chart Of Subroutine Configures GPRS [10].

The command sent for defining GPRS context is AT+CGDCONT=1, “IP”, “payandgo.o2.co.uk”, ”0.0.0.0”, 0, 0. First parameters is context id, it is possible to define up to 5 contexts. Next parameter is protocol used for communication, third parameter is APN assigned by network server provider. In next step subroutine sets the parameters for Quality of service. Commands used are “AT+CGQMIN= 1,0,0,0,0,0” and “AT+CGREQ=1,0,0,3,0,0. These parameters are recommended by manufacturer of the GSM module. Along with APN network service provider also provides user name and password to connect to ISP. Next step is to set user name and password for current GPRS context. Commands used are AT#USERID=payandgo and AT#PASSW=password. Next step configures the TCP/IP stack. It basically sets the minimum packet size, data sending timeout and socket inactivity timeout. Command used for configuring TCP/IP stack is AT#SCFG=1,1,140,30,300,100. First parameter of command is connection identifier; next parameter is context

identifier for which stack is being configured. 300 is the minimum number of bytes that will be sent in one packet. Next parameters are inactivity timeout, connection timeout, and data sending timeout. Next step of the subroutine is configures the firewall settings. It allows certain computers to connect to module. In this case server IP address will be provided to firewall so that Tracking server can connect to In-Vehicle unit. Command used for firewall settings is AT#FRWL=1,"server ip", subnet mask. Server IP address will be the IP address of Tracking server and subnet mask can be provided to allow access to range of computers. Last step is activating current GPRS context. Command is AT#SGACT=1, 1. First parameter is context id to be activated and next parameter is status i.e. 1 for activation and 0 for deactivation.

vii) Subroutine-Send Information Using GPRS

When In-Vehicle unit is configured to send information using GPRS, all activities of In-Vehicle unit are controlled by this subroutine.

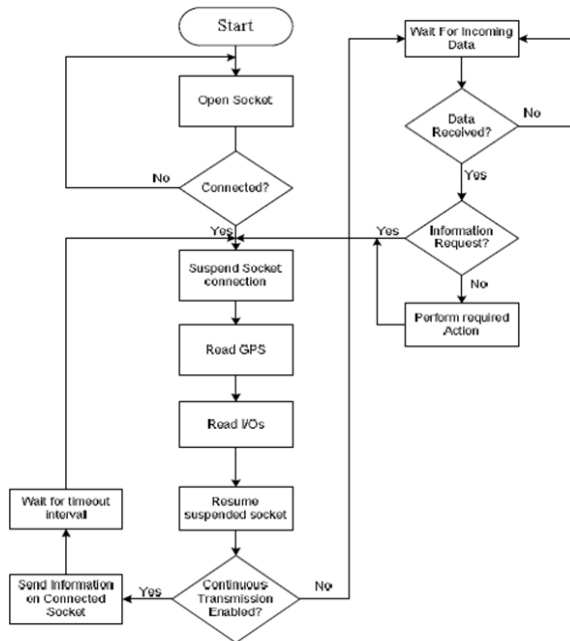


Figure 10: Flow Chart of Subroutine Send Information Using GPRS [9].

In order to send data over IP network application needs an interface to physical layer. This interface is named as socket. This subroutine starts with opening socket for currently configured TCP/IP stack. Command used to open socket for configured embedded TCP/IP stack is AT#SD=1, 1, 6534. First parameter is connection identifier of TCP/IP stack, 2nd is protocol i.e. 0 for TCP and 1 for UDP. Next two parameters are port number and IP address/host name of Tracking server respectively. If command returns the response CONNECT; connection is accepted. Data can be sent now. After getting connection, socket is suspended using escape sequence +++ to bring module in command mode. Socket remains connected while it is suspended.

When GPRS connection is alive, module can't accept AT commands and GPS data can't be read from module. Once module is in command mode this subroutine calls the routine Read GPS data which provides the information string that is to be sent to Tracking Server.

Next step is to read I/O ports of microcontroller to get vehicle's door and ignition status. Information string received from Read GPS data subroutine is appended with status of I/O ports. Socket connection is resumed and information is sent to Tracking server on this socket. If In-Vehicle unit is configured for continuous transmission of vehicle information after regular intervals, all above steps are repeated otherwise module waits for incoming requests from Tracking server. If location request is received above steps are repeated and if any other command is sent by the server according action is taken. Server can send request for vehicle shutdown, changing the data transmission from GPRS to SMS or changing the continuous transmission to polling or vice versa, restart the In-Vehicle unit. This subroutine ends only when In-Vehicle unit is restarted by Tracking server.

viii) Tracking Server

Tracking server maintains all information received from all In-Vehicle units installed in different vehicles into a central database. This database is accessible from internet to authorized users through a web interface. Authorized users can track their vehicle and view all previous information stored in database. Tracking server has a GSM/GPRS modem attached to it that receives SMS from In-Vehicle units and sends those messages to the server through serial port. Tracking server saves this information into database [9].

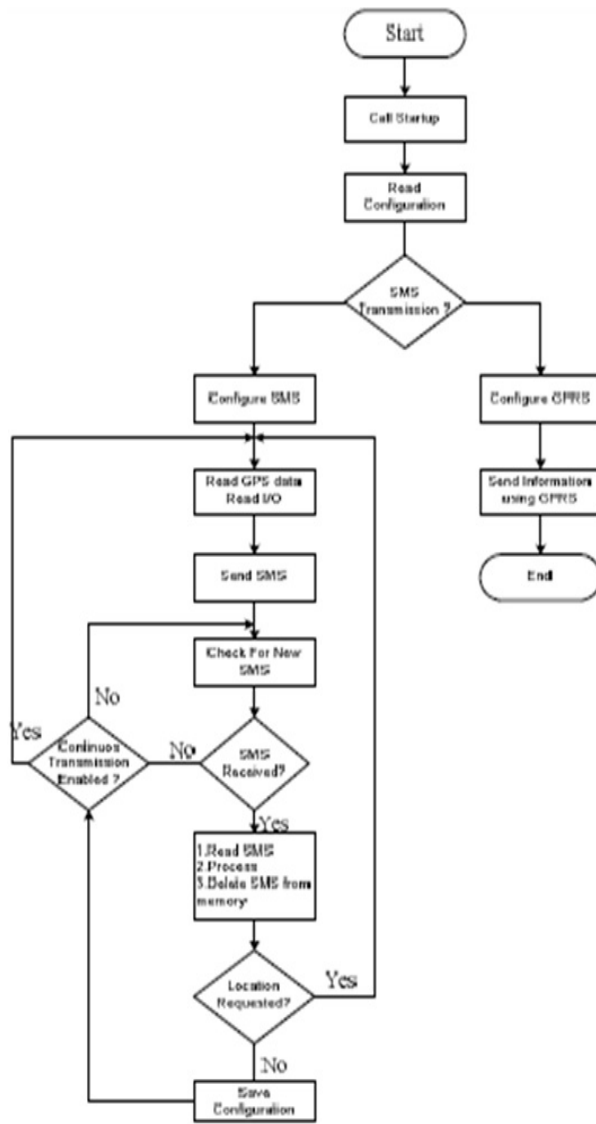


Figure 11: Flow Chart of Main Program [11].

Design of Tracking Server is partitioned into four major parts.

1. Hardware design for GSM/GPRS Modem (GM862-GPS)
2. Communication Software for GM862-GPS
3. Database
4. Web Interface

ix) *Communication Software for GPS*

GPS functions work by this procedure:

1. Configuration of GM862-GPS for sending and receiving SMS Receiving the SMS.
2. Processing received SMS and saving information into database
3. Sending SMS to in vehicle unit as required by user

4. Accepting TCP/IP connections from In-Vehicle units
5. Exchanging information with In-Vehicle units through internet.

GPS will be configured in such a way that whenever new SMS arrives, GM862-GPS will send the information about SMS to the serial port. Software will be listening at serial port; it will read the SMS from GM862-GPS memory and extract the information from SMS. After extracting the information SMS will be deleted from GM862-GPS by software and information will be written to the database [10].

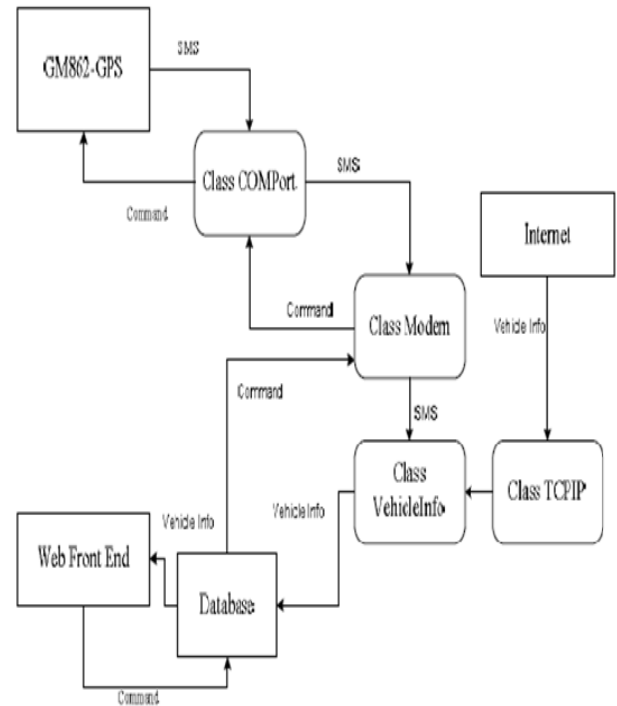


Figure 12: Data Flow of Communication Software [10].

x) *System Testing and Results*

**Testing In-Vehicle Unit (SMS Configuration):**

GPS interface board was connected to microcontroller board through a serial cable.

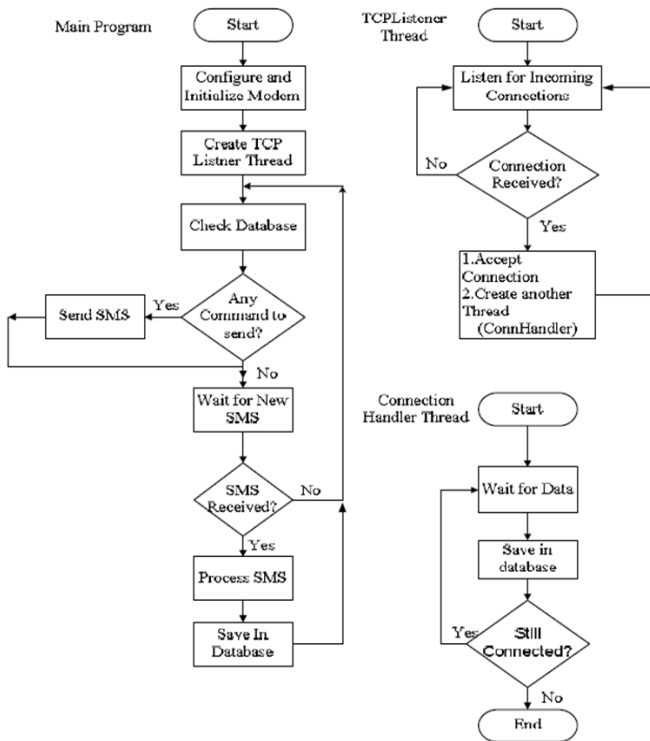


Figure 13: Flowchart Of Communication Software For Gps [10].

Debugging serial port of In-Vehicle unit was connected to a laptop's COM port to see the debugging messages printed by microcontroller on HyperTerminal during its operation. This laptop and debugging COM port is just for debugging purposes, in real time there is no need to connect laptop to In-Vehicle unit.

After connecting the GSM antenna and GPS antenna to the In-Vehicle unit system was powered on. Following logs of microcontroller operation were captured from A custom Software.

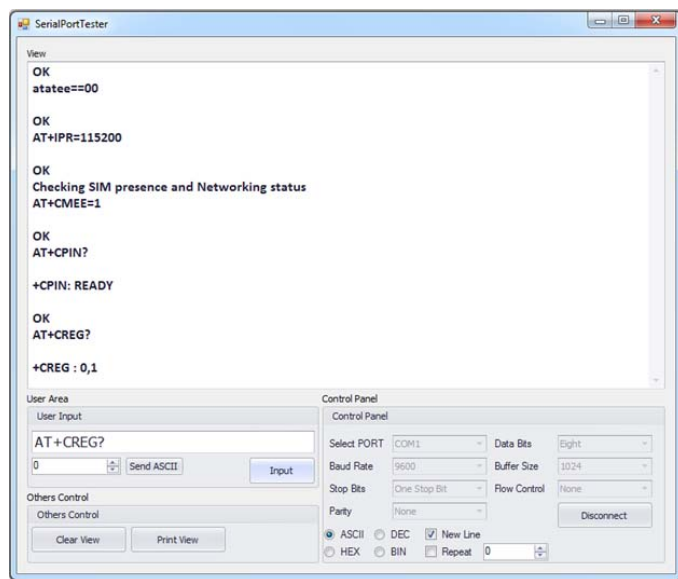


Figure 14: Serial Port Tester [11].

When In-Vehicle unit is powered on it executes Startup routine. It first reads and displays the existing configuration of the system. In next step microcontroller is configuring the GM862-GPS. It first tests the communication interface by sending “AT” command. GM862-GPS responded with “OK” message which shows that interface is working. +CPIN: READY response shows that SIM card is ready and +CREG: 0, 1 response shows that module is connected to network.

#### xi) Testing Tracking Server

In order to test server, laptop was configured to act as a server. GM862-GPS COM was connected to COM port of laptop. Apache server was run on laptop to make it act like server. MySQL DBMS was installed. After running the Communication software for GPS following results were observed.

#### xii) Web Interface Testing

Since server is setup on the local machine. Website was opened in internet explorer. After logging to the website it displayed the page as shown in Figure-15.

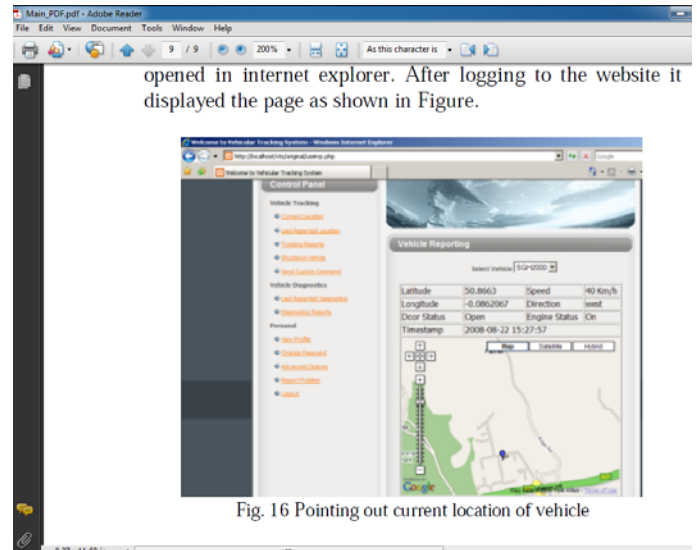


Fig. 16 Pointing out current location of vehicle

Figure 15: Pointing Out Current Location of Vehicle [11].

Upgrading this setup is very easy which makes it open to future requirements without the need of rebuilding everything from scratch, which also makes it more efficient.

Tracking vehicles in our system utilizes a wide range of new technologies and communication networks including GPRS, GSM, the Internet or the World Wide Web and GPS. All the services provided by this system had been tested in the real life. We implemented a system which is composed of a combination of a low-cost hardware unit and a user-friendly Android-based mobile application software utilized to create an on-board vehicle diagnostic system. For future work, more services could be added to the mobile application and also the graphical user interface could be improved.

### III. MICROCONTROLLER COMPILER

In this project we have used ATmega128 microcontroller which is familiar in AVR series. In this project to make a smooth instruction set we have used Win AVR Compiler.

#### A. Microcontroller Programmer

In this project we have used ATmega128 microcontroller which is familiar in AVR series. In this project to make a smooth instruction set we have used Win AVR Compiler.

#### B. Web Application

- The main part of this project is to show the live data to client. The GUI which is known as Graphical User Interface. This section has designed by using YII framework.
- *Testing Tools*

### IV. PROBLEMS AND LIMITATIONS

#### a) GPS Module

In this project we've use a simple GPS module. Though GPS module is not available in Bangladesh, so we've collect it from china. We can show the point around 10 meter. If we use higher GPS module, we can able to show 1 cm diameter range.

#### b) Microcontroller

In Bangladesh microcontroller is also not available. We've use a SMD microcontroller which is very tough to soldering manually.

#### c) PCB

In Bangladesh we don't have any automatic PCB Plant, so we made PCB locally. To make local pcb sometimes track connection will not established, it is very hard to find out the faulty point and fix it.

Due to some hardware problem, it is very tough to communication with satellite. Sometimes we cannot find the exact value of the location. In GPRS portion, we face some critical problem, because of our mobile operators' network. Whenever we are going to a garage or under some high rise building, we cannot find any GPS signal, so we cannot show the point where our vehicle exists.

Another problem is every operator do not support each features and AT commands. So it is very hard to find out the right AT Commands.

### V. CONCLUSION AND FUTURE SCOPE

Vehicle tracking system is becoming increasingly important in large cities and it is more secured than other systems. Now a days vehicle thefting is rapidly increasing, with this we can have a good control in it. The vehicle can be turned off by only with a simple SMS. This system can be used to prevent car theft by combining the device with the car alarm and also obtaining a map containing the car location if the car is thought to be stolen. Since, now a days the cost of the vehicles are increasing they will not step back to afford it. This setup can be made more interactive by adding a display to show some basic information about the vehicle and also add emergency numbers which can be used in case of emergency.

### VI. RECOMMENDATIONS

Hence, we have implement the system which provides the various facilities to the client related to the bus application like to see the all bus details such as bus route, bus timings, bus stops and also facility to the admin for adding the new routes or other bus details. The very important feature provide to the admin is finding the location and speed of the bus or client vehicle.

### VII. REFERENCES

- [1] [http://www.geocomputation.org/1998/38/gc\\_38.htm](http://www.geocomputation.org/1998/38/gc_38.htm)
- [2] <http://www.edaboard.com/thread196370.html>
- [3] [http://wctek.com/index.php?main\\_page=product\\_info&products\\_id=7](http://wctek.com/index.php?main_page=product_info&products_id=7)
- [4] [http://high-cloud.en.alibaba.com/product/453692667-209670367/SIM548C\\_Simcom\\_GSM\\_GPRS\\_GPS\\_module.html](http://high-cloud.en.alibaba.com/product/453692667-209670367/SIM548C_Simcom_GSM_GPRS_GPS_module.html)
- [5] <http://www.edaboard.com/thread232426.html>
- [6] <http://www.extpdf.com/commandes-ts-sim548c-pdf.html>
- [7] [http://www.banque-pdf.fr/fr\\_commandes-ts-sim548c.html](http://www.banque-pdf.fr/fr_commandes-ts-sim548c.html)
- [8] <http://www.8051projects.net/t21224/embedded-gsm-development/sim548c-gprs-activation.htm>
- [9] <http://www.edaboard.com/thread197640.html>
- [10] <http://factory.dhgate.com/network-communication-device/gprs/gsm-gps-module-sim508/sim548/sim548c-p39163174.html>
- [11] <http://www.openstreetmap.org/>

**AUTHORS PROFILE**

**Fahmi Ahmed**

*MEngg in Telecommunication (AIUB, Bangladesh)*  
*BSc in Electronic and Telecommunication Engg (UODA, Bangladesh)*  
*Senior Lecturer, Department of CSE and ETE*  
*University of Development Alternative (UODA)*

**A.B.M. Rocknuzzaman**

*MSc in Computer Engineering (ISE)*  
*(Universität Duisburg-Essen, Germany)*  
*BSc in Computer Science & Engineering (UODA, Bangladesh)*

**Syed Foysol Islam**

*MSc Engg in Electrical Engineering (BTH, Sweden)*  
*BSc, MSc in Computer Science (Rajshahi University, Bangladesh)*  
*Associate Professor, Department of CSE and ETE*  
*University of Development Alternative (UODA)*

# Using J48 tree for value-based customer relations management (CRM)

Marzieh mohammadi  
Department of Computer Engineering,  
Najafabad branch, Islamic Azad  
University, Isfahan, Iran

Hamid Rastegari  
Department of Computer Engineering,  
Najafabad branch, Islamic Azad  
University, Isfahan, Iran

**Abstract**— One of the main important issues in critical to retail success is decision support methods for marketing decisions. Different data mining techniques can be suitable for targeted marketing and efficient customer segmentation. Mainly over data mining, the extraction of hidden predictive pattern from datasets organizations can recognize forecast future behaviors profitable customers, and assist firms to create proactive, knowledge-driven choices. The mechanized, future-oriented analyses is possible with data mining move outside the analyses of previous events usually provided with history-oriented tools like decision support systems. Data mining techniques response business requests that in the previous were too time consuming to follow. However, the responses to these requests create customer relationship management probable. Therefore, in this paper, a model base on the classification of J48 tree and feature selection is proposed to predict precise marketing performance. The propose model is evaluated conducted 3 datasets and the results are compared with other algorithms such as Rep tree, Random tree and J48 tree. The experimental results show that the proposed model has higher precision and lower error rate in comparison of J48 tree, Rep tree and Random tree.

**Keywords**-Customer relations management (CRM); Feature Selection; Data mining; Classification; J48 tree

## I. INTRODUCTION

Exploiting large dimensions of data for superior decision making by discovering interesting patterns in the amount of data has become a key task in today's business background. Each company designs for future goals decisions to market their services and products. A lot of money and time is spent on these decisions. Leads or the prospective customers, created by these decisions come from different background and can be classified into different sets in relation to their spending power. Information about these leads is kept in the CRM database commonly [1]. Customer Relationship Management (CRM) is described about managing business dealings with the customer. CRM is analyzing, acquiring and sharing knowledge about customers [2]. CRM can be defined as the procedure of predicting customer behavior and choosing activities to effect that behavior to profit the company [3]. Customer contentment can also be developed through more effective marketing. One of the important issues in CRM is prediction and customer classification, in which a company categorizes its customers into predefined

sets with similar behavior patterns. Generally, companies make a customer prediction model to discover the prospects for a particular product. Data mining uses artificial intelligence algorithms to find useful patterns and trends from the extracted data so that it able to yield main insights including prediction models and associations can support companies understand their customer well. Analyzing and Examining data able to turn raw data into valuable information about customer's requires [4]. The forecasting of stock markets is considered as a challenging job of commercial time series prediction [5]. Classification of data is one of the main technologies in data mining. The main purpose of data classification is to create a classification model, that able to map to a specific subclass through the data in the dataset. Classification is very important to retrieve information properly, organize data rapidly. It can be suitable for the sales department if the customer data is categorized with some attributes. A decision tree is a predictive machine learning method that selects the purpose value of a new instance with different attribute values. The decision tree methods is most valuable in such classification problems. A tree is created to model the classification method, with this technique. When the tree is constructed, it is applied to every tuple of the dataset and therefore, results in a classification of that tuple. Now, the decision tree has become a significant data mining approach. It is a more common classification method approximation algorithm with machine learning [6]. Many other data mining models such as neural networks are hard to interpret. The relation between features can be features from a dataset and disregard the less essential ones. This ability to be choosy increases to their human readability and able to yield better understanding about what is the most important in a dataset. Though, since decision trees create their decision with one feature at each point in their construction, they are limited to recognizing only linear associations within a dataset. Other methods may be hard to interpret but can probably produce better results. [7]. The internal nodes of a decision tree represent the various features, the branches between the nodes show the probable values that these features can have in the observed instance, when the terminal nodes represent the final value in classification. The feature that is to be predicted is recognized as the dependent variable, since its



value depends upon, or is decided by, the values of all the other features. The other features, which help in predicting the value of the dependent variable, are identified as the independent variables in the dataset. Decision trees have proven to be effective method in controlling predicting and classification problems [8-10]. J48 decision tree is an open source Java implementation of the C4.5 decision tree algorithm in the Weka data mining tool. C4.5 tree is a method that makes a decision tree with a set of labeled input data. This algorithm was established by Ross Quinlan [11]. With the purpose of classify a new features in J48 tree, it first needs to make a decision tree with the feature values of the training dataset. Therefore, whenever it encounters a training set it detects the features which discriminates the different instances most clearly. This feature is able to present about the data instances so that it can classify the features with the highest information gain. Among the possible values of this feature, if there is any value for which there is no ambiguity, that is, for which the data instances falling contained by its class have the same value for the goal variable, then it terminate that branch and allocate to it the goal value that it has gained [12]. In this paper, a model base on J48 tree for analyzing and predicting customer behavior and choosing activities to effect that behavior to profit the company. We compare the proposed model with another data mining algorithm such as Rep tree, Random tree and J48 tree. The experimental results indicate that the proposed model has higher precision and lower error rate in comparison of J48 tree, Rep tree and Random tree. The rest of paper is organized as follows: Section 2 presents the concepts and related work. Section 3 represents the proposed model. Section 4 discusses about results and discussion. Section 5 is about conclusion.

## II. CONCEPTS AND RELATED WORK

### A. Concepts

#### 1) Rep tree

Rep Tree is a fast decision/regression tree learning algorithm which creates with information gain as the splitting criterion and prunes with decrease error pruning [13]. It just sorts values for numeric features once. Missing values are related to using C4.5's approach of using fractional instances [10].

#### 2) Random tree

The random decision tree is a simply tree creating by choosing a random feature at each node and is not pruning and is used as a control tree [13]. A random tree is a tree drawn from a group of possible trees at random with  $k$  random attribute at each node. The concept "at random" means that every tree in the group of trees has the same chance of being sampled. Random trees can be created efficiently and the combination of large groups of random trees generally causes accurate models. The models of

random tree have been widely developed in the field of data mining in the recent years [10].

### 3) Feature Selection

Feature Selection decreases dataset size by eliminating redundant/irrelevant features. It discovers minimum set of features such that resulting possibility distribution of data classes is as local as possible of creative distribution [12]. Feature selection techniques help to produce an accurate predictive model. By selecting features that will give as better or good accuracy whilst needing less data. Feature selection techniques can be used to recognize and remove irrelevant, redundant and extra features from data which do not contribute to the accuracy of a predictive model or possibly in fact reduce the accuracy of the model. Less features is required because it decreases the complexity of the model. The aim of variable choosing is three-fold: producing more cost-effective and faster predictors, improving the prediction performance of the predictors and producing a better understanding of the underlying methods that created the data [14]. There are three common classes of feature selection methods: filter methods such as (the Chi squared test, information gain and correlation coefficient scores), wrapper methods such as (a best-first search, it may stochastic such as a random hill-climbing algorithm, or it may use heuristics, like forward and backward passes to add and remove features) and embedded methods such as (regularization methods, LASSO, Elastic Net and Ridge Regression).

### B. Related Work

There are many data mining approaches suggested by various researches for different areas of businesses [15]. Many Data Mining approaches are useful for improvement of CRM's scopes. In the retention of customer, techniques such as association rules, clustering, classification and sequence discovery are utilized concentrating on customer complaint and loyalty programs which are CRM elements. In the work of [16] discusses about appropriate Data Mining tools for CRM. In [17] studies about intelligent Data Mining methods for CRM. [18] describe about building Data Mining methods for CRM. They suggested a simple approach to evaluate the profits of a data mining methods for the CRM techniques. In [19] discusses about applying Data Mining methods for better CRM. The classification of data mining approach as proposed by [20] contains two phases: in the first phase, a classifier is manufactured relating a predefined set of data classes as learning phase. In the next phase the classification is really implemented. In the work of [21] arguments classification as the way to determine the features of customers who are probable to provides a model that able to be used to forecast who they are. In this study the most significant method of data mining are decrease the fraud, to develop customer retention and acquisition is discussed. Decision trees is one of the data mining techniques are generally utilized in operations research, exactly in decision

analysis, to help recognize a model most probable to achieve a goal such as detection of spam in the work of [22]. In decision tree approach, a logical conclusion is achieved. When decision tree is used for the business decisions, the approach denotes a documented record of the inputs that are accessible. Decision trees are used to extract patterns and models to predict future behaviors or describe sequences of interrelated decisions among customer data [1, 23, 24]. In the work of [24], a proposed method based on decision tree analysis in the change recognition problem can be utilized in more structured states in which the manager has a particular research request and it also discovers the change of classification measures in a dynamically changing situation. For representing the performance of the proposed methodology a Korean Internet shopping mall instance is evaluated and practical business implications for this method are produced. In [25] to dominate the limitations of lack of information of customers of PHSS<sup>1</sup> and to construct an accurate and effective customer churn model, three experimentations (altering sample approaches for training datasets, altering sub-periods for training datasets, altering misclassification rate in churn model) are put forward to increase the prediction performance of churn model with using decision tree that is used generally, some optimum parameters (random sample approach for trainset, sub-period time being 10 days, misclassification rate being 1:5) of models are discovered with three research experimentations.

### III. THE PROPOSED MODEL

Data mining techniques help and development of CRM by producing the complete framework, that covers all scopes [15]. Fig 1 shows the general flowchart of using data mining technique in CRM.

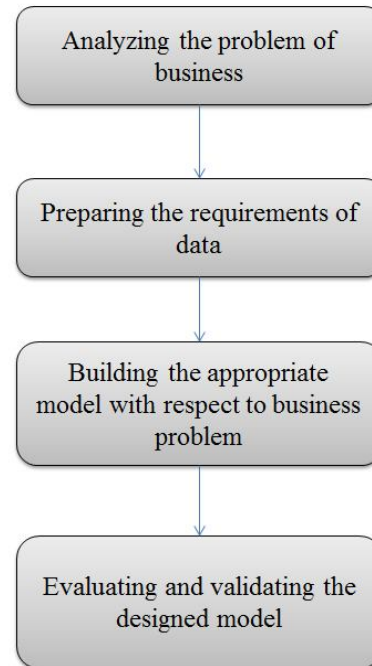


Figure 1: the general flowchart of using data mining technique in CRM

The framework contains 4 stages: analyzing the problem, preparing the requirement of data, building the proper model and evaluated the designed model. The analytical process of the data mining technique benefits to realize the hidden models and patterns which aid the organizations in decision making. Slowly the concentrate turned to other concepts of data mining approaches in CRM such as data preparations, model building and evaluation of models. Data preparation is basic for the improvement of CRM as the data for comes from different sources. Therefore, missing data, outlier and other essential work is applied on data preparation stage. Building the appropriate model is the next stage of the Data mining approaches in CRM, which constructs the different models base on the data given in the data preparation stage. The last phase is the validating and evaluation of the model, so that the suitable results in the custom of valuable patterns can be drawn from the models manufactured. Due to the flowchart described above, a model base on J48 decision tree is proposed that can forecast with a very good accuracy in scope of data mining approaches in CRM. There are 3 phase in the proposed model that includes: preprocessing, feature selection with ranker, classifier with J48 tree. Fig 2 represents the proposed model:

---

<sup>1</sup> Personal Handyphone System Service

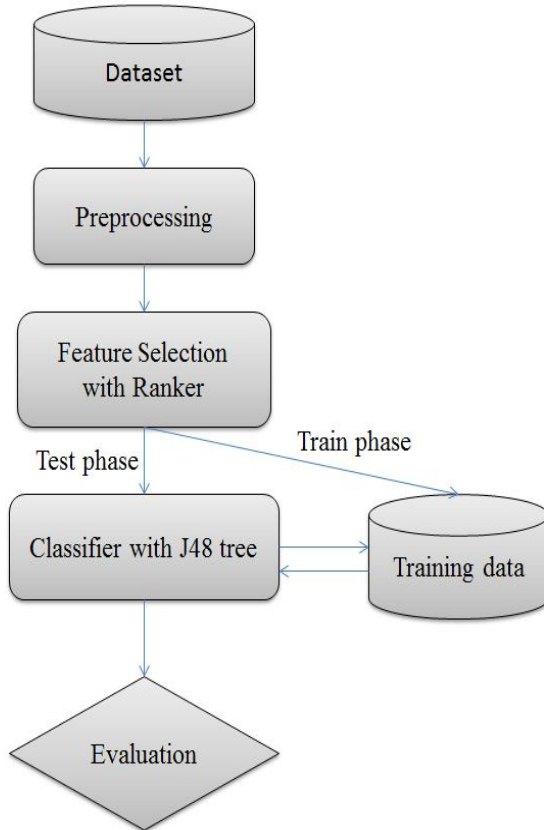


Figure 2: the proposed model

### A. Preprocessing phase

Data pre-processing is an essential stage in the data mining process. Data filtering and preparation steps can take substantial amount of processing time. Data pre-processing contains normalization, cleaning, transformation [26]. Normalization is scaling method or a mapping method or a preprocessing phase [20]. It can be useful for the prediction purpose a lot [27]. There are so many ways to predict but all can be different with each other a lot. Therefore to keep the large difference of prediction the Normalization approach is necessary to create them closer. Nevertheless there are some existing normalization techniques such as Min-Max, Z-score & Decimal Scaling. In the proposed model Z-score algorithm is used for normalizing in preprocessing phase. Therefore the unstructured data can be normalized with z-score parameter, as formulae 1:

$$V_i' = \frac{V_i - \bar{E}}{\text{std}(E)} \quad (1)$$

$V_i'$  is Z-score normalized one values.  $V_i$  is value of the row  $E$  of  $i$ th column  $n$  this method, suppose there are five rows namely  $X$ ,  $Y$ ,  $V$ ,  $U$  and  $Z$  with various columns or variables that are 'n' in each row.

$$\text{Std}(E) = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (v_i - \bar{E})^2} \quad (2)$$

$$\bar{E} = \frac{1}{n} \sum_{i=1}^n v_i \text{ or mean value}$$

Thus in each row above Z-score method can be used to estimate the normalized ones. If suppose some row having all the values are identical, therefore the standard deviation of that row is equivalent to zero and then all values for that row are equal to zero.

### B. Feature selection with ranker phase

Ranking is a general and universal method for constructing otherwise disorganized groups of objects by computing a rank for each object with the value of one or more of its features. This method allows, for instance, prioritizing tasks or evaluating the performance of produces relative to each other. Although the visualization of a ranking itself is straightforward, its interpretation is not, since the rank of an object shows only a summary of a possibly complicated association between its features and those of the other objects. It is common that different rankings be existent which essential to be compared and evaluated to achieve insight into how various heterogeneous features affect the rankings. Advanced visual examination applications are necessary to make this process efficient[28]. These algorithms try to directly optimize the value of one of the above evaluation measures, averaged over all queries in the training data. This is difficult because most evaluation measures are not continuous functions with respect to ranking model's parameters, and so continuous approximations or bounds on evaluation measures have to be used. Rank features use in conjunction with feature evaluators such as ReliefF, GainRatio and Entropy. In the ranking method, features are ranked with some measures and those that are above a definite threshold are selected. A general algorithm can be measured for such method where it just requires to decide which one if the best ranking measures to be used. It can reach a ranking with the best features from the point of view of the classification. This outcome approves with what is common knowledge in data mining techniques, which states related to training and test phases [29]. The following algorithm is about ranking features selection.

```

Main Algorithm()
Input: E training(N instances, M attributes)
Output: E reduced(N instances, K attributes)
For each attribute  $A_i \in 1..M$ 
    Sort-Method(E,i)
    
```

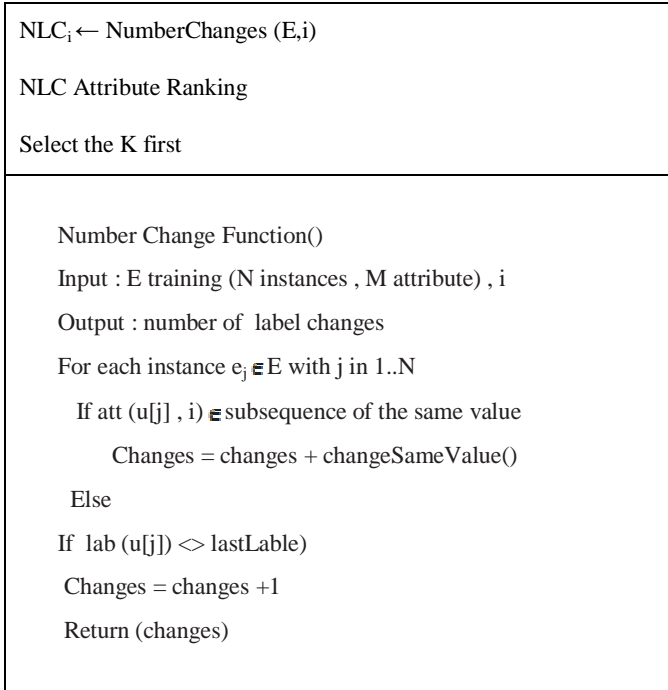


Figure 3: Feature Ranking Algorithm [29]

The algorithm is very fast and simple, see Figure 3. It has the ability to run with discrete and continuous variables as well as with datasets which have two or several classes. In the ascending-order-task for each feature, one of the sort algorithm can be used. When ordered with a feature, it can be count the label changes through the sorted projected sequence. After applying the sort algorithm, it might have repeated values with the various or same class. For this aim, the algorithm firstly sorts with value then in case of equivalence, it will seek the worst of the all probable cases Changes Same Value function. The next implementation of the algorithm may find another state, with a various number of label changes. The answer to this problem contains of finding the worst case. The heuristic is applied to gain the maximum number of label alters within the interval containing repeated values. In this case, the Changes the same value approach would create the output. This can be achieved with low cost. It can be inferred counting the elements of the class. Changes same value saves the relative frequency for every class. It is probable to be affirm that:

$$\begin{aligned} \text{If } rfi > (ne/2) \text{ then } ((ne - rfi) * 2) & \quad (3) \\ \text{else } ne - 1 & \end{aligned}$$

rfi: relative frequency for every class, with i in [1,...,k] classes.

ne: the elements' number within the interval.

Ranking algorithms create a ranked list, base on the evaluation measure applied. The process require an external parameter to take the subset from features made by the first features of the aforesaid list. This parameter creates various outcomes with different datasets. Thus, with the purpose of

establish the number of features in each situation, the range of value of the ranked lists is put between [0,1], i.e. the punctuation of the first feature of the list will be 1, and the last feature 0. Thus, the features is selected over the parameter called Reduction Factor (RF). An special analyzed on each dataset is not realized [29].

### C. Classifier with J48 tree phase

The j48 decision tree is used as the most important decision tree algorithm, since it is an implementation of the generally of used C4.5. C4.5 tree is an extension of Quinlan's earlier ID3 algorithm [11]. The decision trees produced by C4.5 can be utilized for classification and for this aim, C4.5 tree is referred to as a statistical classifier [12]. Basic Steps of J48 tree algorithm:

- a) In state the instances is possessed by the same class the tree shows a leaf so the leaf is related to label with the same class.
- b) The possible information is calculated for each feature, set by a test on the feature. Then the gain information is calculated which would consequence from a test on the feature. This process utilizes the "Entropy" which is a criterion of the data disorder. The Entropy is calculated with formulae 4:

$$\text{Entropy}(\bar{y}) = -\sum_{j=1}^n \frac{|y_j|}{|\bar{y}|} \log\left(\frac{|y_j|}{|\bar{y}|}\right) \quad (4)$$

$$\text{Entropy}(j|\bar{y}) = \frac{|y_j|}{|\bar{y}|} \log\left(\frac{|y_j|}{|\bar{y}|}\right)$$

And Gain formulae is

$$\text{Gain}(\bar{y}, j) = \text{Entropy}(\bar{y}) - \text{Entropy}(j|\bar{y}) \quad (5)$$

- c) Then the best feature is found on the basis of the present selection measure and that feature selected for branching [30].

## IV. RESULTS AND DISCUSSION

To validate the proposed model, experimental evaluation is conducted thorough over 3 dataset. We utilized the 10 folds cross validation and precision. The results are presented on 3 datasets Bank-Data.csv<sup>2</sup>, Car.arff<sup>3</sup> and Bank-full.csv<sup>4</sup>. The proposed model is compared with Rep tree,

<sup>2</sup> <http://facweb.cs.depaul.edu/mobasher/classes/ect584/WEKA/preprocess.html>

<sup>3</sup> <http://repository.seasr.org/Datasets/UCI/arff/>

<sup>4</sup> <http://mlr.cs.umass.edu/ml/machine-learning-databases/00222/>

Random tree and J48 tree. The proposed model is implemented by java Net Beans. The jar files from WEKA are imported into source code. The experiments are implemented on a system with Core i7 CPU and 4G MB of RAM. The precision and error rate formulas calculated with equation 6 and 7:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (6)$$

$$\text{Error rate} = (\text{FN} + \text{FP}) / \text{N} \quad (7)$$

TP means the numbers of features in the classifier are correctly identified. The FP concept is the numbers of features in the classifier are incorrectly identified. FN denotes the numbers of features in the classifier are incorrectly rejected and N is the number of features.

Table 1 represents the datasets which use in the experiment:

TABLE I. DATASET IN USE

Dataset	Number of Instances	Number of Features
Bank-Data.arff	41118	21
Bank-full.arff	41189	20
Car.arff	1727	7

Table2 represents the comparison of precision between REP tree, Random tree, J48 tree and proposed model.

TABLE II. THE COMPARISON OF PRECISION WITH REP TREE, RANDOM TREE, J48 TREE AND PROPOSED MODEL

dataset	J48 tree	Rep tree	Random-Tree	The Proposed model
Bank-full.arff	90.7	90.5	88.4	93.6
Bank-Data.arff	89.9	72.2	60.6	92.1
Car.arff	92.4	88	82.7	94.6

As it can be seen in table 2, the precision of proposed model is higher than the other algorithm. J48 tree is more precise than Rep tree and Random tree and Rep tree has the precise than Random tree because random tree is not pruning.

Table $\bar{r}$  shows the comparison of precision between REP tree, Random tree, J48 tree and proposed model.

TABLE III. THE COMPARISON OF ERROR RATE WITH REP TREE, RANDOM TREE, J48 TREE AND PROPOSED MODEL

dataset	J48 tree	REP tree	Random-Tree	The Proposed model
Bank-full.arff	0.258	0.252	0.332	0.096
Bank-Data.arff	0.305	0.398	0.469	0.272
Car.arff	0.171	0.198	0.241	0.0378

As it can be seen in table 3, the error rate of proposed model reduced in the experiment and this denotes an increase in the precision of this model. One of the reasons is for applying normalizing in the preprocessing phase and the next one is for using feature selection and removing the redundant and irrelevant features. After the proposed model, the error rate of J48 tree is lower than Random tree and Rep tree. Also, the error rate of Rep tree is lower than Random tree.

## V. CONCLUSION

Nowadays, many of companies spent a lot of money and time on the decisions for marketing their services and products and decision making by discovering interesting patterns in the amount of data has become a key task in today's business background. Generally, companies make a customer prediction model to discovery the prospects for a particular product. Data mining uses the algorithms to find useful patterns and trends from the extracted data so that it able to yield main insights including prediction models and associations can support companies understand their customer well. Analyzing and Examining data able to turn raw data into valuable information about customer's requires. Classifying and features selection are two main techniques of data mining. In this paper, a model base on the classification of J48 tree and feature selection with Ranker is proposed to predict precise marketing performance. The propose model is evaluated conducted 3datasets and the results are compared with other algorithms such as Rep tree, Random tree and J48 tree. The experimental results show that the proposed model has higher precision and lower error rate in comparison of J48 tree, Rep tree and Random tree.

## REFERENCES

- [1] Berry, M.J. and G.S. Linoff, *Data mining techniques: for marketing, sales, and customer relationship management*. 2004: John Wiley & Sons.
- [2] Jutla, D., J. Craig, and P. Bodorik. *Enabling and measuring electronic customer relationship management readiness*. in *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*. 2001. IEEE.
- [3] Jenkins, D., *Customer relationship management and the data warehouse*. Call Center Solutions, 1999. **18**: p. 88-92.
- [4] Cho, Y.H., J.K. Kim, and S.H. Kim, *A personalized recommender system based on web usage mining and decision tree induction*. Expert systems with Applications, 2002. **23**(3): p. 329-342.
- [5] Kannan, K.S., et al. *Financial stock market forecast using data mining techniques*. in *Proceedings of the International Multiconference of Engineers and computer scientists*. 2010.

- [6] Ngai, E.W., L. Xiu, and D.C. Chau, *Application of data mining techniques in customer relationship management: A literature review and classification*. Expert systems with applications, 2009. **36**(2): p. 2592-2602.
- [7] Gagne, D.J., A. McGovern, and J. Brotzge, *Classification of convective areas using decision trees*. Journal of Atmospheric and Oceanic Technology, 2009. **26**(7): p. 1341-1353.
- [8] McGlynn, T., et al., *Automated classification of ROSAT sources using heterogeneous multiwavelength source catalogs*. The Astrophysical Journal, 2004. **616**(2): p. 1284.
- [9] Zhang, Y. and Y. Zhao, *Automated clustering algorithms for classification of astronomical objects*. Astronomy & Astrophysics, 2004. **422**(3): p. 1113-1121.
- [10] Zhao, Y. and Y. Zhang, *Comparison of decision tree methods for finding active objects*. Advances in Space Research, 2008. **41**(12): p. 1955-1959.
- [11] Quinlan, J.R., *Induction of decision trees*. Machine learning, 1986. **1**(1): p. 81-106.
- [12] Gholap, J., *Performance tuning of J48 Algorithm for prediction of soil fertility*. arXiv preprint arXiv:1208.3943, 2012.
- [13] Witten, I.H. and E. Frank, *Data Mining: Practical machine learning tools and techniques*. 2005: Morgan Kaufmann.
- [14] Hall, M.A., *Correlation-based feature selection for machine learning*. 1999, The University of Waikato.
- [15] Ranjan, J. and V. Bhatnagar, *A review of data mining tools in customer relationship management*. Journal of Knowledge Management Practice, 2008. **9**(1).
- [16] Ranjan, J. and V. Bhatnagar, *Data mining tools: a CRM perspective*. International Journal of Electronic Customer Relationship Management, 2008. **2**(4): p. 315-331.
- [17] Wong, K.W., et al. *Intelligent data mining and personalisation for customer relationship management*. in *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*. 2004. IEEE.
- [18] Berson, A. and S.J. Smith, *Building data mining applications for CRM*. 2002: McGraw-Hill, Inc.
- [19] Ranjan, J. and V. Bhatnagar, *Critical success factors for implementing CRM using data mining*. Journal of Knowledge Management Practice, 2008. **9**(3).
- [20] Han, J., M. Kamber, and J. Pei, *Data mining: concepts and techniques: concepts and techniques*. 2011: Elsevier.
- [21] Ahmed, S.R. *Applications of data mining in retail business*. in *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on*. 2004. IEEE.
- [22] Nadimi-Shahraki, M.-H., Z.S. Torabi, and A. Nabiollahi, *Using J48 Tree Partitioning for scalable SVM in Spam Detection*. Computer and Information Science, 2015. **8**(2): p. p37.
- [23] Chen, Y.-L., C.-L. Hsu, and S.-C. Chou, *Constructing a multi-valued and multi-labeled decision tree*. Expert Systems with Applications, 2003. **25**(2): p. 199-209.
- [24] Kim, J.K., et al., *Detecting the change of customer behavior based on decision tree analysis*. Expert Systems, 2005. **22**(4): p. 193-205.
- [25] Bin, L., S. Peiji, and L. Juan. *Customer churn prediction based on the decision tree in personal handyphone system service*. in *Service Systems and Service Management, 2007 International Conference on*. 2007. IEEE.
- [26] Kotsiantis, S., D. Kanellopoulos, and P. Pintelas, *Data preprocessing for supervised learning*. International Journal of Computer Science, 2006. **1**(2): p. 111-117.
- [27] Patro, S., et al., *Technical Analysis on Financial Forecasting*. arXiv preprint arXiv:1503.03011, 2015.
- [28] Gratzl, S., et al., *Lineup: Visual analysis of multi-attribute rankings*. Visualization and Computer Graphics, IEEE Transactions on, 2013. **19**(12): p. 2277-2286.
- [29] Ruiz, R., J.C. Riquelme, and J.S. Aguilar-Ruiz. *Fast feature ranking algorithm*. in *Knowledge-Based Intelligent Information and Engineering Systems*. 2003. Springer.
- [30] Korting, T.S., *C4. 5 algorithm and multivariate decision trees*. Image Processing Division, National Institute for Space Research-INPE Sao Jose dos Campos-SP, Brazil, 2006.



# A Security Architecture for Software Defined Networks (SDN)

Okunade Oluwasogo Adekunle  
School of Science and Technology,  
National Open University of Nigeria,  
Victoria Island, Lagos, Nigeria.

Osunade Oluwaseyitan  
Department of Computer Science,  
University of Ibadan,  
Ibadan, Nigeria.

*Abstract-Software defined networking is an emerging network architecture with promising future in network field. It is dynamic, manageable, cost effective, and adaptable networking where control and data plane are decoupled, and control plane is centrally located to control application and dataplanes. OpenFlow is an example of Software Defined Networking (SDN) Southbound, which provides an open standard based interface between the SDN controller and data plane to control how data packets are forwarded through the network. As a result of rapid changes in networking, network program-ability and control logic centralization capabilities introduces new fault and easily attack planes, that open doors for threats that did not exist before or harder to exploit. This paper proposed SDN architecture with some level of security control, this will provide secured SDN paradigm with machine learning white/black list, where users application can be easily test and group as malicious attack or legitimate packet.*

**Keyword-Software Defined Networking (SDN); OpenFlow; Flow table; Security control; white/black list**

## I. INTRODUCTION

Despite the fact that Internet has led to the creation of digital globalization; traditional IP networks are complex and very hard to manage especially in the area of network configuration, according to the predefined policies and to reconfigure it to response to faults, loads and changes. The basic concept of software defined networking (SDN) is to separates the network control (brains) and forwarding (muscle) planes to make it easier to optimize. The most common protocol used in SDN networks is to facilitate the communication between the Controller and switches/routers (called the Southbound Application Programme Interface {API}) that is currently OpenFlow; although, we have some other protocols. OpenFlow is an open standard of a communication protocol that enables the control plane to interact with the forwarding plane. People often point to OpenFlow as being synonymous with SDN, but it is only a single element in the overall SDN architecture.

Figure 1, shows a traditional network of five devices with each comprising of a control plane that provides information used to build a forwarding table, application and forwarding table used to make decision on where to send frames or packets entering the device.

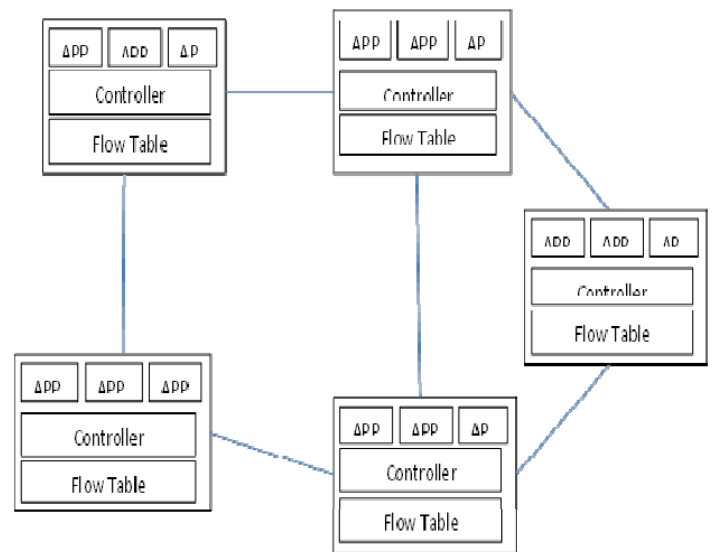


Figure 1: Traditional Network with application, distributed control on network devices

In traditional networks, routers and other network devices encompass both data and control function making it difficult to adjust the network infrastructure and operation rather than the predefined policies regardless of faults, loads and changes that may later occurs. The control plane is an element of a router or switch that determines how one individual device within a network interacts with its neighbours. Examples of control plane protocols are; routing protocols, such as Open Shortest Path First (OSPF), Border Gateway Protocol (BGP), and Spanning Tree Protocol (STP). These protocols determine the optimal port or interface to forward packets (that is, the data plane). While the control plane protocols scale very well, and provide a high

level of network resiliency. They pose limitations. For example, routing protocols may only be able to determine the best path through a network based on static metrics such as interface bandwidth or hop count. Likewise, control plane protocols do not typically have any visibility into the applications running over the network, or how the network may be affecting application performance. Data plane functionality includes features such as quality of service (QoS), encryption, Network Address Translation (NAT), and access control lists (ACLs). These features directly affect how a packet is forwarded, including being dropped. However, many of these features are static in nature and determined by the fixed configuration of the network device. There is typically no mechanism to modify the configuration of these features based on the dynamic conditions of the network or its applications. Finally, configuration of these features is typically done on a device-by-device basis, greatly limiting the scalability of applying the required functionality. While SDN abstracts this concept and places the control plane functions on SDN controller, where this controller can be a server running SDN software see Figure 3 where business requirements changes.

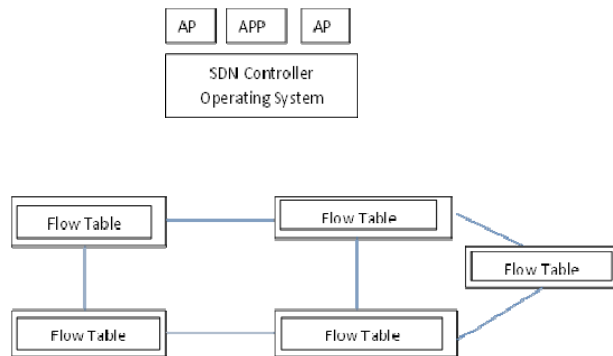


Figure 2: Software Defined Network (SDN) with decoupled Control and Application

By using an API, your controller can implement network commands to multiple devices without the need to learn the command line syntax of multiple vendor products. These are few of the benefits seen with SDN. The control plane is responsible for configuration of the node and programming the paths that will be used for data flows. Once these paths have been determined, they are pushed down to the data plane. Data forwarding at the hardware level is based on this control information. Once the flow management (forwarding policy) has been defined, the only way to make an adjustment to the policy is via changes to the configuration of the devices. The change in the

location and intensity of flows over time requires a flexible approach for successful network resource management. The numbers of handheld devices like smartphones, tablets, and notebooks have greatly increase the pressure on enterprise resources. Network resources change rapidly and management of Quality of Service (QoS) security become challenging [1]. In a security and dependability perspective, one of the key ingredients to guarantee a highly robust system is fault and intrusion tolerance [2]. According to [3] Networks are expected to operate without disruption, even in the presence of device or link failures. However, Network programmability and control logic centralization capabilities introduces new fault and attack planes, which open the doors for new threats that did not exist before or were harder to exploit [2]. OpenFlow (OF) paradigm embraces third party development efforts, and therefore suffers from potential trust issue on OF applications (apps). The abuse of such trust could lead to various types of attacks impacting the entire network [4]. This can be seen as attractive honeypots for malicious users and major concern for less prepared network operators.

The ability to control the network by means of software (always subject to bugs and a score of other vulnerabilities) and centralization of the network intelligence in the controller(s) can make anyone with unlawful access to the servers (impersonation) potentially control the entire network unlawfully. The question now is; how can the Software-Defined Network be protected from malicious attack? Since potential security vulnerabilities exist across the SDN platform. At the controller-application level, questions have been raised around authentication and authorization mechanisms to enable multiple organizations to access network resources while providing the appropriate protection of these resources (IETF Network Working Group). However, with multiple controllers communicating with a single node or multiple control processing communicating with a single, centralized controller, authorization and access control becomes more complex, potential for unauthorized access increases and could lead to manipulation of the node configuration and/or traffic through the node for malicious intent [5]. The remainder of this paper is organized as follows; Section 2 is the literature review, Section 3 introduces the framework for the Securing Software Defined Networks (SDN) from Malicious Attacks, Section 4 describes the result derived from the given framework in Section 3. Finally, important conclusion is discussed in Section 5.



## II. LITERATURE REVIEW

Software-Defined Network (SDN) create an environment where all switches and routers take their traffic forwarding clues from a centralized management controller. SDN has the following three layers/plane;

1. **Application Plane/Layer:** Control layer implement logic for flow control
2. **Control Plane/Layer:** This runs applications to control network flows
3. **Data Plane/Infrastructure Layer:** this is a Dataplane consists of the Network switch or router

The application layer contains network applications that introduces new network features, such as security and manage-ability, forwarding schemes or assist the control layer in the network configuration [6]. The application layer can receive an abstracted and global view of the network from the controllers and use that information to provide appropriate guidance to the control layer.

The interface between the application layer and the control layer is referred to as the northbound interface. This is the interface through which the SDN Application layer communicates with the Control Layer to expose the program-ability of the network [6]. SDN controller manages the forwarding state of the switches in the SDN, this management is done through a vendor neutral API that allows the controller to address a wide variety of operator requirements without changing any of the lower level aspects of the network, including topology. With the decoupling of the control and data planes, SDN enables applications to deal with a single abstracted network device without concern for the details of how the device operates. Network applications see a single API to the controller. Thus it is possible to quickly create and deploy new applications to orchestrate network traffic flow to meet specific enterprise requirement for performance or security using API. Examples of north bounds interface are FML, Procera, Frenetic, RESTful and so on.

The OpenFlow protocol provides an interface that allows control software to program switches in the network, this is called southbound. Southbound is a protocol of OpenFlow which separates the control plane from the data plane to enable centralized and fine grained control of network flows. Examples of Southbound are OpenFlow, ForCES, PCEPNetConf, IRS and so on. OpenFlow is an example of Software Defined Networking (SDN), which provides an open, standards based interface to control how data packets are forwarded through the network, Controller

communicates with a physical or virtual switch data plane through protocol that conveys the instructions to the data plane on how to forward data.

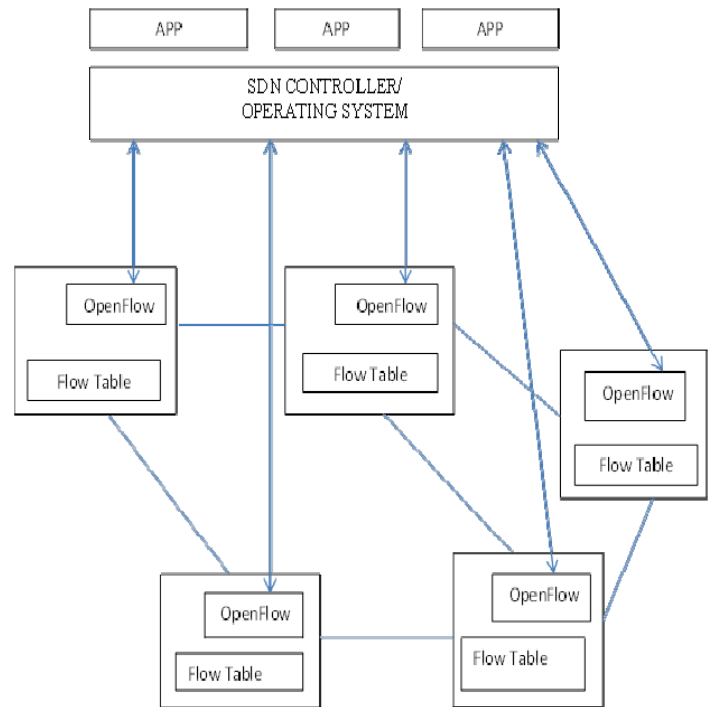


Figure 3: OpenFlow Architecture

This is a Software-Defined Network (SDN) package that enables networks to be software controlled, and used to dynamically change the network configuration. It is the most common example of southbound interface, which is standardized by the Open Networking Foundation (ONF). OpenFlow is a protocol that describes the interaction of one or more control servers with OpenFlow compliant switches. An OpenFlow controller installs flow table entries in switches, so that these switches can forward traffic according to entries. OpenFlow switches depend on configuration by controllers [6]. OpenFlow allows network switches to be configured using programmable interfaces, monitored/inspect network traffic and routing of packets [7]. OpenFlow protocol specifies the interactions between the control plane running in the controller and the infrastructure; it is a foundational element for building SDN solutions. OpenFlow framework is an embodiment of the SDN concept, framework for the implementations of Software Defined Networking (SDN) paradigm that enable communication between the controller and the switches uses a standardized OpenFlow protocol. In an OpenFlow environment, flow tables are used by devices rather than routing or MAC address table.

Switches implement policy using efficient packet processing hardware: this is a secure channel that connects the switch to a remote control process (called the controller), allowing commands and packets to be sent between a controller and the switch using The OpenFlow Protocol [8] in [9]. An OpenFlow network consists of a distributed collection of switches managed by a program running on a logically centralized controller, each switch has a flow table that stores a list of rules for processing packets and, each rule consists of a pattern (matching on packet header fields) and actions (such as forwarding, dropping, flooding, or modifying the packets, or sending them to the controller). OpenFlow Protocol provides an open and standard way for a controller to communicate with a switch [9].

Controller machine manages a collection of programmable switches, defines the forwarding policy for the network and configures the switches through an open and standard (south bound) interface. A controller associates packets with their senders by managing all the bindings between names and addresses, it essentially takes over DNS, DHCP and authenticates all users when they join and keeping track of which switch port (or access point) they are connected to [9]. The controller derive the desired forwarding data in software, send OpenFlow messages to update the forwarding table in the device and the messages can add, update or delete entries in the forwarding table. Controller drives a level of network convergence; consider changing the entire configuration on your network to support new network path every 10 minutes.

The SDN Controller defines the data flows that occur in the SDN data plane: each flow through the network must first get permission from the controller, which verifies that the communication is permissible by the network policy. If the controller allows a flow, it computes a route for the flow to take and adds an entry for that flow in each of the switches along the path. With all complex functions subsumed by the controller, switches simply manage flow tables whose entries can be populated only by the controller. A controller accomplishes this network programming via software and it is in this software that SDN's promise of flexibility comes in. The controller is a platform on which software is run, as well as being a communication gateway that software can communicate through. Most controller architectures are modular, allowing the controller to communicate with different kinds of devices using different methods as required.

The SDN architecture is remarkably flexible: it can operate with different types of switches and at different protocol layers. SDN controllers and switches can be implemented for Ethernet switches (Layer 2), Internet routers (Layer 3), transport (Layer 4) switching, or application layer switching and routing. SDN relies on the common functions found on networking devices, which essentially involve forwarding packets based on some form of flow definition. It encapsulates and forwards the first packet of a flow to an SDN controller, enabling the controller to decide whether the flow should be added to the switch flow table. Switch forward incoming packets out; the appropriate port based on the flow table in which the flow table may include priority information dictated by the controller. Switch can drop packets on a particular flow temporarily or permanently as dictated by the controller.

SDN controller communicates with OpenFlow compatible switches using the OpenFlow protocol, running over the Secure Sockets Layer (SSL). Each switch connects to other OpenFlow switches and possibly to end-user devices that are the sources and destinations of packet flows. Within each switch, a series of tables typically implemented in hardware or firmware are used to manage the flows of packets through the switch.

Flow table tells switch how to process each data flow by associating an action with each flow table entry. Flow table consist of flow rules that guide the controller on action to be perform on a given particular packet. OpenFlow enabled device has an internal flowtable and a standardized interface to add and remove flow entries remotely [10]. Flow table is the basic building block of the logical switch architecture, each packet that enters a switch passes through one or more flow tables. Each flow table contains entries consisting of six components; Match Fields, Priority, Counters, Instructions, Timeouts and Cookie.

SDN switches are controlled by a Network Operating System (NOS) that collects information using the API and manipulates their forwarding plane, providing an abstract model of the network topology to the SDN controller hosting the applications. The controller can therefore exploit complete knowledge of the network to optimize flow management and support service user requirements of scalability and flexibility. [7] Propose a novel network system architecture that protects network devices from intra-LAN attacks by dynamically isolating infected devices using OpenFlow on

detection. [5] has proposed an extension to the OpenFlow data plane called connection migration, which dramatically reduces the amount of data to-control-plane interactions that arise during the Inherent communication bottleneck that arises between the data plane and the control plane, which an adversary could exploit by mounting a control plane saturation attack that may disrupts network operations.

### III. METHODOLOGY

To address the previously stated problem, we present SDN architecture with some level of security control. This will provide secured SDN paradigm, where control plane will check for the authentication of users' application through the API to confirm some security measure using the inbuilt white and black list for legitimacy confirmation of the users' application who are requesting to make use of control plane. If the application is from the black list it will be discarded but permitted if from the white list. A machine learning tools will be used to update SDN architecture black /white list using the packet flow movement on the table flow for updating. Figure 4 is the proposed SDN architecture with the extension of control plane to contain some level of security control that will interact with the proposed extended data plane flow table. Figure 5 contains extension of black/white list of the applications; the extension will communicate with the control plane security control to supply the application security status to the flow table rule through the controller. This supplied security status will be used by the controller to decide particular action(s) to be taken on application requesting rule.

### IV. RESULT AND DISCUSSION

This research work comes up with a secured Software Defined Networking (SDN) Architecture (figure 6) that identified the malicious source, and therefore prevents unauthorized access to the network by blocking packet from insecure source and automatically update it white/black list. For every incoming packet(s), it compares and checks through its white/black list to identify the packet source and update the list using machine learning..

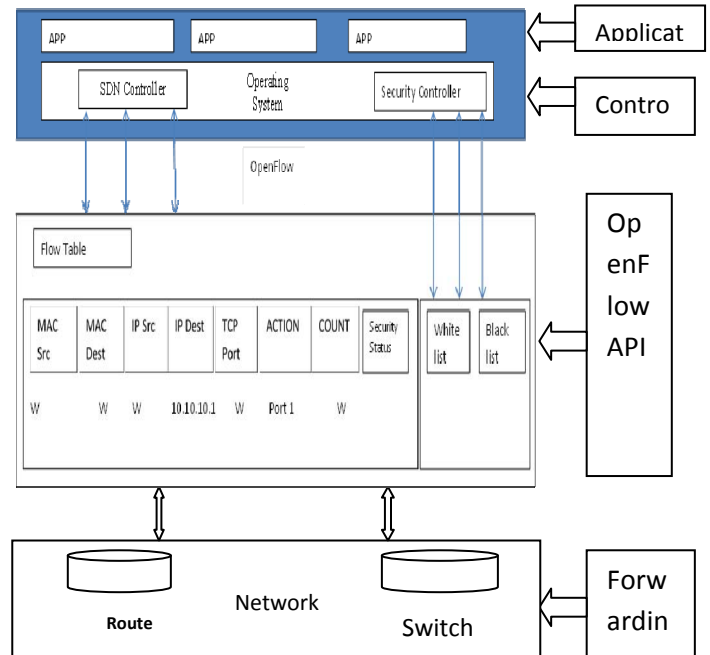


Figure 4: Proposed SDN Security Architecture

With extension of control plane and flow table with security features, a secured SDN architecture is designed, see figure 6 where user's application can be easily tested, and permit if not malicious attack will be discard when suspected as such. The SDN can be easily prevented from malicious attack, and made secured with some level of security control implementation into SDN Architecture that make it a secured SDN Architecture, this will help to prevent malicious attack by blocking packages from insecure source/networks. There is an extension of control plane called security control, this will interact with the extended aspect of flow table consist of white and black list supplying it resolution based on the security status of every incoming packets to the secured SDN controller that will then place its security status on the flow table rule extension through the security controller. This will be used for decision making on the action to be perform on the said packet that is packet security status.

### V. CONCLUSION

Despite the flexibility and successful contribution of Software Defined Networking (SDN) to network society, deployment of a secure SDN environment has become challenging. A security architecture is a Software Defined Networking (SDN) security control system, that prevents malicious attacks from having access to SDN environment. The secure architecture will help to promote and encourage the openness of the SDN and

prevent against its security challenges. The system (Security architecture for Software Defined Networking (SDN)) check security status of incoming packets populated by white/black list security controller to determine action(s) to be taken on the arrived packed either to be permitted for transaction or discarded.

- [10] J. Naus, D. Erickson, G. A. Covington, G. Appenzeller and N. McKeown, "Implementing an OpenFlow Switch on the NetFPGA platform." ANCS '08, San Jose, CA, USA. ACM 978-1-60558-346-4/08/0011

## REFERENCES

- [1] H. R.Muhammad, C. S. Shyamala, N. Ali and R. Bill, "A Comparison of Software Defined Network (SDN) Implementation." 2nd International Workshop on Survivable and V. Robust Optical Networks (IWSRON).Published by Elsevier B. Procedia Computer Science 32, 2014, pg1050–1055. Available online at [www.sciencedirect.com](http://www.sciencedirect.com)
- [2] K. Diego, M. V. Fernando, Ramos and V. Paulo, "Towards Secure and Dependable Software-Defined Networks." HotSDN'13, Hong Kong, China. ACM 978-1-4503-2178-5/13/08
- [3] R. Mark, C. Marco, G. Arjun, and F. Nate, "FatTire: Declarative Fault Tolerance for Software-Defined Networks" HotSDN'13, August 16, 2013, Hong Kong, China. ACM 978-1-4503-2178-5/13/08
- [4] W. Xitao, C. Yan and H. Chengchen, "Towards a Secure Controller Platform for OpenFlow Applications." HotSDN'13, Hong Kong, China. ACM 978-1-4503-2178-5/13/08.
- [5] S. Seungwon, Y. Vinod, P. Phillip, and G. Guofei, "AVANT-GUARD: Scalable and Vigilant Switch FlowManagement in Software-Defined Networks." CCS'13, Berlin, Germany.ACM 978-1-4503-2477-9/13/11. <http://dx.doi.org/10.1145/2508859.2516684>.
- [6] B. Wolfgang and M. Michael, "Software-Defined Networking Using OpenFlow: Protocols, Applications and Architectural Design Choices." Future Internet 2014, 6, pg.302-336; doi:10.3390/fi6020302 ISSN 1999-5903. [www.mdpi.com/journal/futureinternet](http://www.mdpi.com/journal/futureinternet).
- [7] J. Yutaka., H. Hung-Hsuan and K. Kyoji, "Dynamic Isolation of Network Devices Using OpenFlow for Keeping LAN Secure from Intra-LAN Attack." 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013. Published by Elsevier B.V. Selection and peer-review under responsibility of KES International doi: 10.1016/j.procs.2013.09.163. ScienceDirect.22, pg. 810–819 1877-0509. Available online at [www.sciencedirect.com](http://www.sciencedirect.com)
- [8] K. Eddie, M. Robert, C. Benjie, J. John and M. F. Kaashoek, "The Click modular router." ACM Transactions on Computer Systems 18(3), 2000, pg. 263-297.
- [9] N McKeown, T. Anderson and H. Balakrishnan, "OpenFlow: Enabling Innovation in Campus Networks." ACM SIGCOMM Computer Communication Review, 38(2), 2008, pg. 69-74

# Design and Extraction of Facial Components in Automated Face Recognition

Ms. Roshani R.Patil

Department of Computer Science and Engineering  
G. H. Raisoni College of Engineering  
Nagpur, Maharashtra, India

Mr.P.S.Mohod

Department of Computer Science and Engineering  
G. H. Raisoni College of Engineering  
Nagpur, Maharashtra, India

## Abstract:

Face recognition presents a challenging problem in the field of image analysis and computer vision. Face recognition system should be able to automatically detect a face in an image. This involves extracts its features and then recognize it, regardless of lighting, expression, illumination, ageing, transformations (translate, rotate and scale image) and pose, which is a difficult task. This paper presents a framework for component-based face alignment and representation that demonstrates improvement in matching performance over the more common holistic approach to face alignment and representation. Active shape model (ASM) technique that has been used often for locating facial features in face images. The proposed scheme selects robust landmark points where relevant facial features are found and assigns higher weights to their corresponding features in the face classification stage. For alignment and cropping Procrustes analysis is used. Multi-scale local binary pattern is used for matching automated face image. In MLBP per-component measurement of facial similarity and fusion of per-component similarities is used. The proposed work is more robust to changes in facial pose and improves recognition accuracy on occluded face images in forensic scenarios.

**Keywords:**-Active shape model, Multi-scale local binary pattern, Procrustes analysis, holistic method.

## I. INTRODUCTION

Face recognition has been a rapidly growing research area due to an increasing demand for biometric-based security applications. Varying factors such as cosmetics, illumination, and face disguise can hinder face recognition performance. Such varying faces are called as automated faces. Several researchers proposed different automated face recognition algorithms that perform well with unconstrained face images[4]. Recently, the face recognition algorithms based on local descriptors such as Gabor filters, SURF, SIFT, and histograms Local Binary Patterns (LBP) provide more robust performance against occlusions, different facial expressions, and pose variations than the holistic approaches. Appearance based or pixel based representation i.e. representations that extract features per specific facial components is the best technique used for automated face recognition. Using facial components that are precisely extracted through automatically detected facial landmarks, it demonstrates that descriptors computed from the individually aligned components result in higher recognition accuracies than descriptors extracted using the more common approach of dense sampling from globally aligned faces. The strong evidence of component processing in human face perception, and the lack of mature

component-based methods in automated face recognition research; a more thorough investigation of the role of component-based processing in automated face recognition is warranted [1].

## II. REVIEW OF LITERATURE

### 1. Three approaches for face recognition

The detail review of different face recognition approaches has been given by V.V. Starovoitov, D.I Samal, D.V. Briliuk. Three approaches for face recognition:

#### A. Feature base approach

The local features like nose, eyes are segmented and it can be used as input data in face detection in this approach. It is the easier task as only three parameters are used.

#### B. Holistic approach

The whole face taken as input in the face detection system to perform face recognition. It is more complicated approach as compared to above approach.

#### C. Hybrid approach

Hybrid approach is combination of feature based and holistic approach. Both local and whole face is used as the input to face detection system.

The computational cost is high, as a large set of randomly generated local deformations must be tested. Elastic bunch graph matching is used to overcome above drawback. Here bunch of jets i.e. instead of actual landmark location, information related to landmark is used. This give more accurate result.

### 2. Face recognition using local binary patterns

The detail review of face recognition by Local Binary Pattern (LBP) has been proposed by Jo Chang-Jo Chang-yeon. LBP features have worked efficiently in various applications i.e. is for texture classification and segmentation, image retrieval and surface inspection. The original LBP operator labels the pixels of an image by thresholding the 3-by-3 neighbourhood of each pixel with the centre pixel value and considering the result as a binary number. Figure shows an example of LBP calculation.

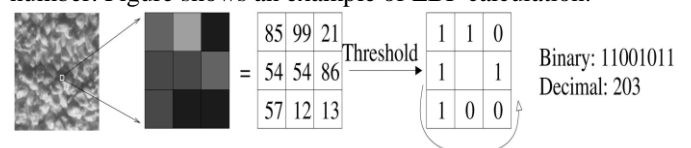


Fig2.1: Computing LBP value at each pixel.

The LBP operator has been extended to take different sizes of neighbour. In general, the operator LBP  $P, R$  refers to a neighbourhood size of  $P$  equally spaced pixels on a circle of radius  $R$  that form a circularly symmetric neighbourhood set. LBP  $P, R$  produces  $2^P$  different output values, corresponding to the  $2^P$  different binary patterns that can be formed by the  $P$  pixels in the neighbour set. It has been shown that certain bins contain more information than others. Hence, it is possible to use only a subset of the  $2^P$  LBPs to describe the textured images. Fundamental patterns with a small number of bitwise transitions from 0 to 1 and vice versa are considered. For example, 00000000 and 11111111 contain 0 transitions while 00000110 and 01111110 contain 2 transitions and so on. Concatenating patterns which have more than 2 transitions into a single bin yields an LBP descriptor [3].

The proposed work is computationally heavy to work on mobile applications. Also LBP requires more time for face recognition as compared to other latest techniques [3].

### 3. Automatic local Gabor features extraction for face recognition

The detail description of automated face recognition has been illustrated by Ben Jemaa Yousra and Sana Khanfir. It is a very important stage to detect face before face recognition. To identify a person, it is necessary to localize his face in the image. It includes following steps:-

When Gabor filters are applied to each pixel of the image, the dimension of the filtered vector are very large they are proportional to the image dimension. It leads to expensive computation and storage cost. To remove such problem and make the algorithm strong, Gabor features are obtained ten extracted fiducial points [4].

### 4. Automatic Face Recognition using Principal Component Analysis with DCT

The detail description of automatic face recognition with different techniques has been proposed by Miss.Renke Pradnya Sunil. It has proved to do instrumental work in this field of face recognition.

The increase in the number of signatures will increase the recognition rate, however, the recognition rate saturates after a certain amount of increases. Hence, it is better to use robust image pre-processing systems, such as geometric alignment of important facial feature points (eyes, mouth, and nose) and intensity normalization which increases the recognition rate and at the same time decreases the number of signatures representing images in the PCA space[5].

### 5. Enhancing the Performance of Active Shape Models in Face Recognition Applications

The detail review of active shape model has been given by Carlos A. R. Behaine and Jacob Scharcanski. It has proved to be instrumental in the field of face recognition through active shape model. Active shape model (ASM) is an adaptive shape matching technique that has been used to locate the facial feature of an image.

As structural constraints given by the face, ASM model-based detection can handle small variations in pose and expression. ASMs are sensitive only to the initial placement of landmarks prior to the iterative updating of model parameters. And insensitive if this initial placement is not closely aligned to the true landmark locations and then the ASM may converge on an inaccurate set of landmarks [6].

## III PROPOSED WORK

The detail review of component based representation has been proposed by Kathryn Bonnen, Brendan F. Klare. This work has been instrumental in identifying the key domains of research in image processing particular to recognition of automated faces.

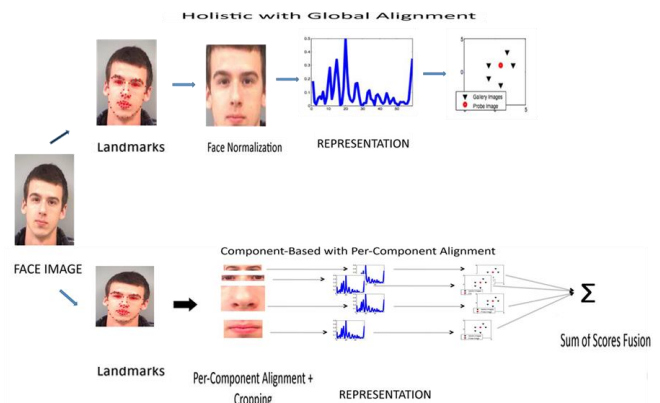


Fig.3.1 overview of comparison between holistic and component based approach

The above diagram describes outline of the per-component alignment performed to yield the proposed component-based representations. This work demonstrates the value of representing faces in a per-component manner. When compared to a globally aligned holistic representation, and other representations found in the literature, the component-based representation offers strong accuracy improvements in a number of face recognition scenarios [1].

It mainly describes component-based representations i.e. representations that extract features per specific facial components. It involves following steps:-

#### 1. Landmark Detection

For aligning the facial components is to extract a predefined set of 76 anthropometric landmarks. A subset of these anthropometric landmarks provides a general outline of the component for each component of given image. Active shape model is mainly used for landmark extraction. But ASM is sensitive if there is the small variation in pose and orientation and insensitive large variation. To overcome this problem PittPatt's Face Recognition SDK. In this first automatically detected the centre of the two eyes, and the centre of the nose. Because these three landmarks are also present in the ASM, initialized the ASM landmarks by (i) solving the affine transformation from these three ASM points to the corresponding PittPatt detected points, and (ii) applying this transformation to the set of 76 ASM



landmarks (representing the mean face in the model). The result of this step is an initial placement of facial landmarks that is well suited to correctly converge on the proper locations [1].

## 2. Alignment and Cropping

It gives the rigid transformation which minimizes the mean squared error between two ordered sets of coordinates. It reduces the variation in translation, scale, and rotation, which allows for a more accurate similarity measure between facial components after performing Procrustes analysis on each component in each face image, the rotation, translation and scaling parameters, is obtained. They are used to rigidly align the parameters. Cropping is done by creating a bounding box around the aligned landmarks. The bounding box is obtained by first performing the horizontal cropping boundaries from the minimum and maximum values. The vertical cropping boundaries are determined based on a ratio of the crop width. To improve the subsequent descriptor extraction A small pixel border around each set of landmarks is used. The same method is then later applied for per aligned and cropped component to get more accurate results [1].

## 3. Representation

Multi-scale local Binary Patterns (MLBP) is used for representation of facial components. It is the combination of local binary pattern. Each facial component is divided into regions of  $d \times d$  pixels  $m$  overlapping by pixels where  $m < d$ . For each region, a histogram of LBP values is obtained from comparisons at each pixel. The LBP value is calculated at each pixel is computed by comparisons selected pixel with the surrounding pixel at a radius of length which gives the gray value at each of the surrounding pixels. This creates a histogram of dimensionality, which further reduced by mapping LBP values without "uniform patterns" to the same value where a uniform pattern is an LBP binary string which produces 2 or fewer bitwise transitions. The MLBP representation concatenates two or more LBP descriptors [1].

## 1. Component-Based Discriminant Analysis

The RS-LDA approach includes the following steps for training.

- 1) the feature space is randomly sampled into subspaces, with each subspace sampling a fraction  $s$  ( $0 < s < 1$ ).
- 2) For each of the random  $k$  sample spaces, principal components analysis is performed in order to retain percent of the variance.
- 3) LDA subspaces are learned from each of the PCA representations.
- 4) From these trained subspaces, images then sampled into each of the  $k$  random feature subspaces, projected into the corresponding PCA and LDA subspaces
- 5) For each of the subspace vectors are combined into a final feature vector.

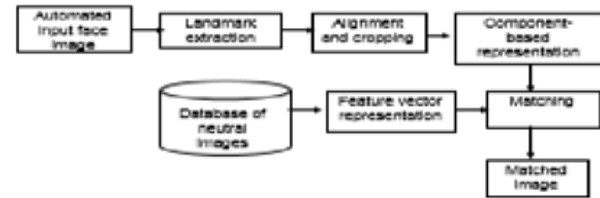


Fig 3.2 block diagram of proposed meth of propose method

This is the block diagram of proposed method. It consist of automated face image which is given as input as probe image. Landmarks are extracted from that which helps to extract components. Later on extracted components per-component alignment and cropping is performed. Each extracted component is represented in form of histogram which is obtained through multi-scale local binary pattern. With the help of histogram vector is obtained for probe image. Later on remaining images feature vector extraction is performed using random sampling linear discriminant analysis (RSLDA). The feature vector is obtained for gallery images. Matching done by cosine similarity measure. The image with minimum distance is obtained as the output or matched image which is matched with most of the components in the probe image. So, it give more accurate results as compared to holistic approach.

## IV IMPLEMENTATION AND EXPERIMENTAL RESULTS

It consists of 4 modules:

1. Landmark extraction:-It is the process of extracting the predefined set of landmarks such as eyes, nose, eyebrows and mouth which provide the general outline for component. Active shape method is used which handles variation in occluded faces.

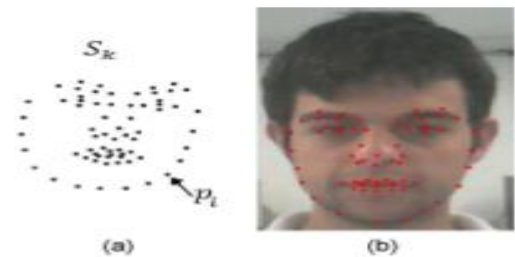


Fig 4.1 point distribution model with landmarks extracted

2. Per-component alignment and cropping:-From the outline obtained due to landmark extraction each component is cropped and aligned properly.

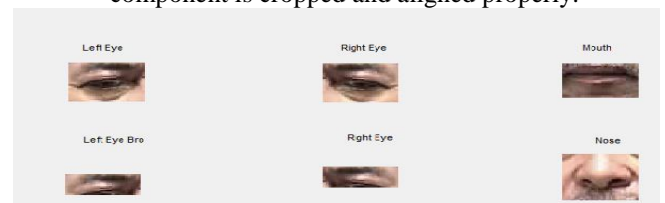


Fig 4.2 extracted components after performing per-component alignment and cropping

3. Representation of each extracted component:- Extracted component is represented through multi-scale local binary pattern which will give the

histogram. MLBP is obtained from LBP values and histogram

LBP	10011110	LBP	11001100110	LBP	11101100011101
Decimal of LBP	158	Decimal of LBP	1638	Decimal of LBP	15133
Type of LBP	Non Uniform	Type of LBP	Non Uniform	Type of LBP	Non Uniform
No of 0's	3	No of 0's	5	No of 0's	5
No of 1's	5	No of 1's	6	No of 1's	9
All ones	NO	All ones	NO	All ones	NO
All Zeros	NO	All Zeros	NO	All Zeros	NO

Fig 4.3 LBP pattern obtained with different radii

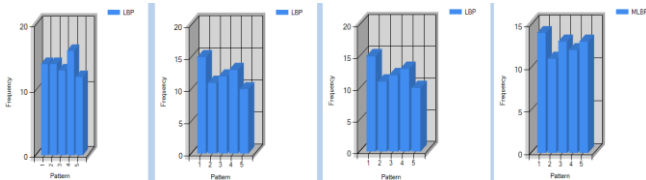


Fig 4.4 LBP histogram with  $r=1,2,3$  and their combined histogram

4. RSLDA on remaining images: - Random sampling linear discriminant analysis is applied on remaining images. The result of which is used for matching with the result of MLBP.

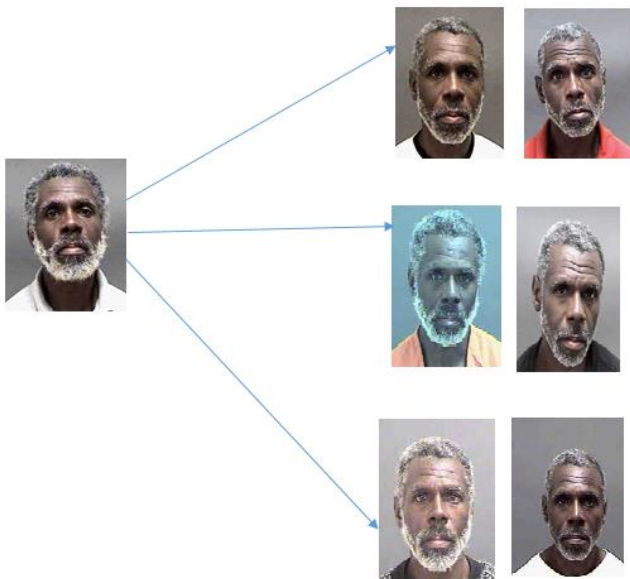


Fig 4.5 output obtained after face recognition under different illumination, pose, facial expression and age

#### IV CONCLUSION

The main objective is to demonstrate the potential of different face recognition such as face recognition by geometric approach, elastic bunch graph matching, neural network, local binary pattern, automatic local gabor features extraction, principal component analysis with discrete cosine transform, active shape model and multi-scale local binary pattern. The difficulties in extracting individual facial components prevented the effective use of component-based approaches in automatic face recognition. A viable future research topic is a dedicated study on how to best tailor learning-based methods to component-based representations which improves face recognition accuracy [1]. From the

different face recognition component based approach has proven to be more efficient. Providing the extension to component based approach such for proposed work to be used for make-up faces, water images or instead of neutral image if automated image is present in training image the system should work efficiently.

#### References

1. Kathryn Bonnen, Brendan F. Klare and Anil K. Jain "Component-Based Representation in Automated Face Recognition" IEEE Transactions on Information Forensics And Security, Vol. 8, No. 1, January 2013 239
2. V.V. Starovoitov, D.I Samal1, D.V. Briiliuk "Three Approaches for Face Recognition" The 6-th International Conference on Pattern Recognition and Image Analysis October 21-26, 2002, Velikiy Novgorod, Russia, pp. 707-711
3. Timo Ahonen, Abdenour Hadid, and Matti Pietikainen "Face Description with Local Binary Patterns: Application to Face Recognition" IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 28, No. 12, December 2006
4. Yousra BEN JEMAA and Sana KHANFIR "Automatic local Gabor features extraction for face recognition" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 3, No. 1, 2009
5. Miss.Renke Pradnya Sunil "Automatic face recognition using principal component analysis and DCT".Journal of Electronicsl and Communication Engineering (IOSR-JECE) ISSN: 2278-2834-, ISBN: 2278-8735, PP: 01-07, 2011
6. Carlos A. R. Behaine and Jacob Scharcanski "Enhancing the Performance of Active Shape Models in Face Recognition Applications" IEEE Transactions On Instrumentation And Measurement, Vol. 61, No. 8, August 2012.
7. Rabia Jafri and Hamid R. Arabnia "A Survey of Face Recognition Techniques" Journal of Information Processing Systems, Vol.5, No.2, June 2009.
8. Yousra Ben Jemma, Sana Khanfir "Automatic local Gabor features extraction for face recognition" International Journal of Computer Science and Information Security, 2009
9. C. Geng and X. Jiang, "Face recognition based on the multi-scale local image structures," Pattern Recognit., vol. 44, pp. 2565–2575, 2011
10. Yuchi Huang, Qingshan Liu, Dimitris N. Metaxas "A Component-Based Framework for Generalized Face Alignment" IEEE transactions on systems, man, and cybernetics vol.34 May 2011.
11. Divyarajsinh N. Parmar, Brijesh B. Mehta "Face Recognition Methods & Applications" Int.J.Computer Technology & Applications, Vol 4 (1),84-86 jan-feb 2013.
12. Bradshaw and G. Wallace, "Models for the processing and identification of faces," Perception Psychophys., vol. 9, pp. 443–448, 1971.
13. J. Huang, P. Yuen, W.-S. Chen, and J. Lai, "Component-based LDA method for face recognition with one training sample," in Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures, Oct. 2003, pp. 120–126
14. J. Richler, O. Cheung, and I. Gauthier, "Holistic processing predicts "Psychol.Sci.,vol.22,no.4,pp.464



# Software Reliability Estimation of Component Based Software System using Fuzzy Logic

Gopal Prasad Jaiswal

(Research Scholar)

Department of Computer Science and Engineering  
RITEE Raipur, Chhattisgarh  
CSVТУ Bhilai, Chhattisgarh, India

Ram Nivas Giri

(Asst. Professor)

Department of Computer Science and Engineering  
RITEE Raipur, Chhattisgarh  
CSVТУ Bhilai, Chhattisgarh, India

**Abstract**—Software Reliability Modeling has been one of the much-attracted research domains in Software Reliability Engineering. Software reliability means provide reusable, less complex software, to perform a set of successful operation and his function within a provided time and environment. Software designers are motivated to develop reliable, reusable and useful software. In past, Object-Oriented Programming System (OOPS) concept is to be used in purpose of reusability but they are not providing powerful to cope with the successive changing as per requirements of ongoing applications. After that Component Based Software system (CBSS) is in floor. IT is based on reusability of his component with less complexity. This paper presents a new approach to analyze the reusability, dependency, and operation profile as well as application complexity of component-based software system. Here, we apply Fuzzy Logic approach to estimate the reliability of component-based software system with the basis of reliability factor.

**Index Terms**—Component, Object-Oriented Programming System (OOPS), Component Based Software system (CBSS), Fuzzy Logic, Fuzzy Inference System (FIS), Adaptive Neuro Fuzzy Inference System (ANFIS), Reliability, Application Complexity, Component Dependency, Operation Profile, Reusability, Fuzzification, Defuzzification, Reliability Model, Rule Based Model, Path Based Model, Additive Model, etc.

## I. INTRODUCTION

Software reliability is defined as the probability of failure – free software operation for a specified period of time in a specified environment. The reliability of a software product is usually defined to be “the probability of execution without failure for some specified interval of natural units or time” [1]. Software reliability is a feather of any software. Software reliability is depends on performance of successful operations and function as well as less complexity, maintainability, portability, flexibility and so on. Basically we can say that software reliability is a feather of the software that to be depend on another feather of the software. Hence, we cannot simply define it. In a binary form we can say that if software is correct and failure-free then its reliability is 1 else 0. Reliability is still predict probabilistically as

Software Reliability = [1-probability of failure]

Software reliability is mostly depending on reusability of the software because reliability of software is directly proportional to its reusability. For this purpose many year ago object-oriented programming system (OOPS) concept is appear for software development. But he was not successful as per requirement. After that another concept is appear in development floor that is Component Based Software System (CBSS)

Component Based Software System (CBSS) is a paradigm that aims at constructing and designing systems using a pre-defined set of software components explicitly created for reuse. Component based software development is most promising approach for software development today. This approach is based on the idea that software systems can be developed by selecting appropriate off-the-shelf components and then assembling them with well-defined software architecture [2]. This new software development approach is very different from the traditional approach in which software systems can only be implemented from scratch.

This paper presents soft computing techniques for reliability estimation for the component based software system. Here we will use fuzzy logic for estimating the reliability of the software. Fuzzy logic provides logical capabilities as well as learning capabilities for decision making. Logically decision that is Fuzzy Inference System (FIS) based on fuzzy rule and learning capability based on training for decision making that is Adaptive Neuro Fuzzy Inference System (ANFIS). In this paper we will use both type of facilities are adopted with different number of membership function for estimation component based software system and analysis that which one is provide better reliability for both the models.

Rest of the paper is sorted out as follows Region-2 related research work Region-3 proposed framework. Region-4 proposed methodology for CBS reliability. In Region-5, experiments, observation and result analysis of different

approaches has been calculated. Paper is concluded with a summary and the description for future work in Region-6.

## II. RELATED RESEARCH

In this software reliability estimation many number of models proposed for estimating CBSS reliability. We can summaries these approaches into three types [3]:

- Architecture Based Reliability Models
- Mathematical Model for Estimating CBSS Reliability
- Soft Computing techniques for estimating CBSS reliability

**Architecture Based Reliability Models:** Shooman, 1976 "Structural models for software reliability prediction", here consider the possible execution paths for estimating the reliability of an application. A sequence of components along different paths is obtained by either algorithmic or experimental testing [4]. Cheung, 1980 "A user oriented software reliability model" user-oriented software reliability figure of merit is defined to measure the reliability of a software system with respect to a user environment. The reliability of a system is expressed as a function of the reliabilities of its components and the user profile, Means that the current behavior of a component is independent of its previous behavior. These models consider transfer among components to be Markov behavior, which means that the current behavior of a component is independent of its previous behavior. These models can be represented in two ways, namely, as composite models or as hierarchical models [5]. Popostojanova and Trivedi, 2001; Cai et al., 2003; Gokhle, 2007 "Architecture based approach to reliability assessment of software systems" architecture-based reliability models such as state-based and path-based models and find out CBSS reliability depends not only on the architecture but also on the operational profile for the input[6]. Yacoub, S., Cukic, B., and Ammar, H., "Scenario based reliability analysis approach for component based systems" in 2004 propose an approach to reliability analysis called scenario based reliability analysis. This approach introduces component dependency graphs (CDGs) which can be extended for complex distributed systems. This approach is based on scenarios which can be captured with sequence diagrams, which means that the approach can be automated [7].

**Mathematical Model for Estimating CBSS Reliability:** Dong, W., Huang, N., Ming, Y., 2008 "Reliability analysis of component-based software based on relationships of components" a new model for estimating CBSS reliability in which various complex component relationships are analyzed. The Markov model is used to solve these complicated relationships, which have a large impact on a system's reliability. The results were used to develop a new tool to

calculate software application reliability [8]. Huang, N., Wang, D., Jia, X., 2008 "An algebra-based reliability prediction approach for composite web services" proposed a technique based on algebra which provides a framework for describing the syntax and predicting the reliability of a CBSS. If operational profiles have been changed, the loop times of iteration will be changed [9]. Goswami V., Acharya, Y.B., 2009 "Method for reliability estimation of COTS components based software systems" proposed an approach to CBSS reliability analysis which takes the component usage ratio, which is the time allotted for a component's execution out of the application's overall execution time, into consideration. This approach can be used in real-time applications [10]. Seth, K., Sharma, A., Seth, A., 2010 "Minimum spanning tree-based approach for reliability estimation of COTS based software applications" an algebra-based reliability prediction approach (Huang, N., Wang, D., Jia, X., 2008.) is to be used [11].

**Soft Computing techniques for estimating CBSS reliability:** Dimov, Aleksandar, Sasikumar, and Punnekkat, "Fuzzy reliability model for component-based software systems" in 2010 a fuzzy reliability model for Component Based Software System (CBSSs), based on fuzzy logic and probability theory. A mathematical fuzzy logic model was based on necessity and possibility is proposed to predict the reliability of a CBSS. This model does not require component failure data because it is based on uncertainty. However, a mechanism is necessary to model the propagation of failure between components and failure behavior [12]. Lo, J., 2010 "Early software reliability prediction based on support vector machines with genetic algorithms" proposed a software reliability estimation model based on an SVM and a GA. This model specifies that recent failure data alone are sufficient for estimating software reliability. Reliability estimation area for the SVM is determined by the GA. This model is less dependent on failure data than are other models [13]. Hsu, C., Huang, C., 2011 "An adaptive reliability analysis using path testing for complex component based software systems" proposed an adaptive approach for testing path reliability estimation for complex CBSSs. Path reliability estimation: these use sequence, branch, and loop structures. The proposed path reliability can be used to estimate the reliability of the overall application [14]. Tyagi, K., Sharma, A., 2012 "A rule-based approach for estimating the reliability of component-based systems" proposed an approach based on fuzzy logic for estimating CBSS reliability. In this approach, four critical factors were identified for estimating the reliability of a CBSS. They are used to design an FIS for the estimation [15]. Kirti Tyagi, Arun Sharma 2014, "An adaptive neuro fuzzy model for estimating the reliability of component-based software systems" propose a model for estimating CBSS reliability, known as soft computing model or an adaptive neuro fuzzy inference system (ANFIS), that is based on these two basic elements FIS and ANFIS, Here, we analysis its performance with that of a plain FIS (fuzzy inference system) for different

data sets. This is a hybrid method that requires less computational time than traditional approaches and the previously proposed FIS approach. [3]

### III. PROPOSED FRAMEWORK

In region-2 research work to read various models that to be proposed reliability estimation model and conclude that all the models have their own restriction to estimate the reliability of the Component Based Software System (CBSS). We have proposed an soft computing model But still soft computing model have various techniques are available. Some soft computing techniques are listed below:

- Fuzzy Inference System (FIS)
- Artificial Neural networks (NN) and Adaptive Neuro Fuzzy Inference System (ANFIS)
- Support Vector Machines (SVM)
- Probabilistic Reasoning (PR) or Probabilistic Logic (PL)
- Evolutionary Computation (EC)
- Evolutionary Algorithms (EA)
- K-Nearest Neighbor (K-NN)
- Genetic Algorithms (GA)
- Chaos Theory (CT)
- Hybrid Model

Our proposed soft computing model is based on fuzzy logic that to be overcome previously researched restriction and estimates the nearest reliability of the Component Based Software System (CBSS).

We are using fuzzy logic for software reliability estimation. Fuzzy logic is basically if-then rules syntactically. They will provide logical capabilities as well as learning capabilities for decision making. Logical decision that is Fuzzy Inference System (FIS) and learning capability based decision making that is Adaptive Neuro Fuzzy Inference System (ANFIS). In this paper we will use both type of facility for estimation component based software system. Here, we will explain both the soft computing technique one by one:

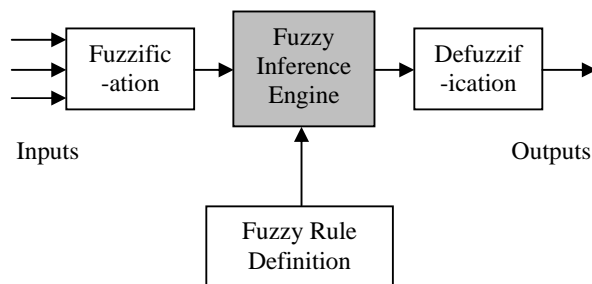


Fig. 1 Framework of Fuzzy Inference System

**Fuzzy Inference System:** A Fuzzy Inference System (FIS) is a way of mapping an input space to an output space using fuzzy logic. FIS framework is displayed at fig. 1. FIS uses a collection of fuzzy membership functions and rules, instead of binary logic, to reason about data. The rules in FIS (sometimes may be called as fuzzy expert system) are fuzzy production rules of the form [25] [26]:

if M then N, where M and N are fuzzy statements.

For example, in a fuzzy rule

if A is low and B is high then C is medium.

Here A is low; B is high; C is medium are fuzzy statements; X and Y are input variables; Z is an output variable, low, high, and medium are fuzzy sets.

**Adaptive Neuro Fuzzy Inference System:** An adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS) is a kind of artificial neural network that is based on Takagi–Sugeno fuzzy inference system. It was developed in the early 1990s [16] [17]. Since it integrates both neural networks and fuzzy logic rules, it has potential to grab the benefits of both in a single paradigm. This inference system is a set of fuzzy IF–THEN rules that have learning capability to approximate nonlinear functions [18]. Hence, ANFIS is considered to be a universal estimator [19]. Below figure-2 is basic ANFIS structure for two input variable with two membership function for each input variable [25].

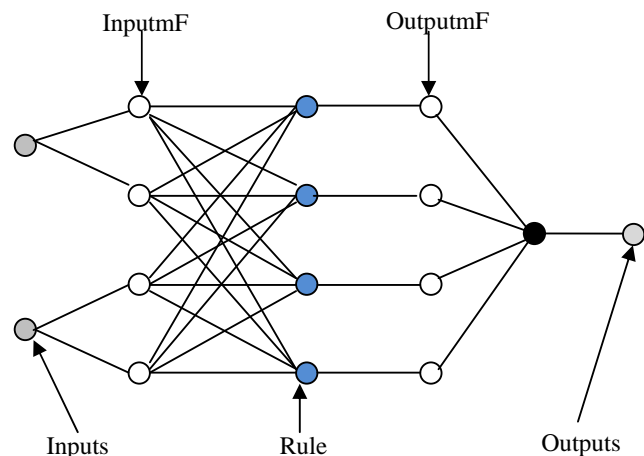


Fig. 2 Structure of Adaptive Neuro Fuzzy Inference System

### IV. PROPOSED METHODOLOGY

In this paper we will use soft computing techniques for software reliability estimation of Component Based Software System (CBSS). It paper is based on fuzzy logic based computing technique, and we are use FIS and ANFIS. This

both the model is to performed in to some input variables. There so we will use some software feather for the calculation of the software reliability. Those feathers are listed below:

**Reusability:** Reusability means how to use any component in multiple times without any failure or any other restriction called software reusability. The reliability of a component is directly proportional to its reusability. Component reusability is calculated on the basis of components feathers [3] [20] [21] [22] [23] [24].

$$\text{Component Reliability} \propto \text{Reusability}$$

Reusability of the any software will be based on attributes, sub-attributes and there selected metrics. Here we are discussed about reusability attributes or Evolutionary model [20] that is reusability of the software is depending upon various attributes. This attributes are listed below:

- Understandability
- Portability
- Maintainability
- Variability
- Flexibility

According to software Evolutionary mode,

$$\text{Reusability of Package} = [0.2 * \text{Understandability} + 0.2 * \text{Portability} + 0.2 * \text{Maintainability} + 0.2 * \text{Variability} + 0.2 * \text{Flexibility}]$$

According to Reusability attribute model reusability of any package is calculating as follows:

**Operation Profile:** Operation profile means how much number of operations was performed successfully. It will be directly proportional to its reliability [3] [15].

$$\text{Component Reliability} \propto \text{Operation Profile}$$

**Component Dependency:** Component dependency is feather of software. It gives information about how much component is dependent on another component [3] [15].

$$\text{Component dependency} \propto (1 / \text{reliability})$$

**Application Complexity:** Application complexity is feather of any software that gives information about complexity of the software. Application complexity is directly proportional to number of component [3] [15].

$$\text{Application Complexity} \propto (1 / \text{reliability})$$

After the calculating these above software feathers, we are applying FIS and ANFIS fuzzy soft computing technique in these calculated feathers (ex.-reusability, operation profile, and

component dependency and application complexity) for reliability estimation of the Component Based Software System (CBSE). Figure-3 described flow chart of our proposed model that to be given below:

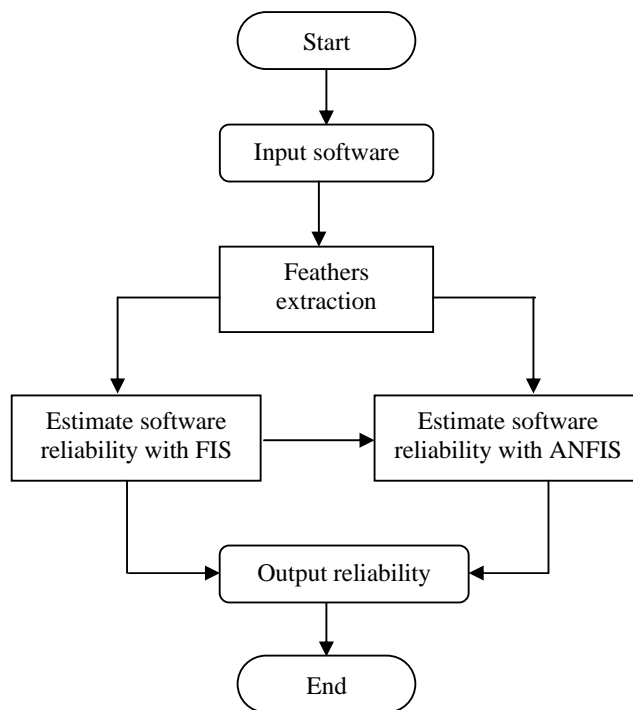


Fig. 3 Flow chart of proposed methodology

## V. EXPERIMENTS, OBSERVATIONS AND RESULT ANALYSIS

In this part, we are applying our methodology in between number of freeware software. We collected software data from [www.sourceforge.net](http://www.sourceforge.net). Here we will use software data as a Jasmin and pBeans. Both the software are various versions are available in the [www.sourceforge.net](http://www.sourceforge.net). After collecting the software data sets we are calculate the above described feather (ex.-reusability, operation profile, and component dependency and application complexity) for the estimation of software reliability.

After this we are applying our model that is FIS and ANFIS:

**Fuzzy Inference System model:** we are using describes feathers as a input data set and calculated software reliability with three and five membership function separately. In FIS with three membership function total 81rules defined for fuzzy inference engine and calculate software reliability. Similarly for five membership functions total 625 rules are defined for fuzzy inference engine and calculate software reliability with

Table-I Software reliability analysis of FIS and ANFIS					
Inputs features				Output Reliability	
Application Complexity	Operation Profile	component Dependency	Reusability	FIS model	ANFIS model
0.703988662	0.140793109	0.851358641	0.775565915	0.327908851	0.331058397
0.704845012	0.140797131	0.851868132	0.776380979	0.327926837	0.331245331
0.443147251	0.387896017	0.81008991	0.562640442	0.325130262	0.323375651
0.450605782	0.405385044	0.832507492	0.562405104	0.325	0.425774082
0.557010478	0.475951475	0.883116883	0.646119443	0.572071945	0.56692347

basis of three membership function and five membership function separately.

**Adaptive Neuro Fuzzy Inference System model:** we are using describes features as a input data set and give the respective output data or target for learning capability because ANFIS is supervised learner. ANFIS is applied for software reliability with three and five membership function separately. In FIS with three membership function total 81rules are generated automatically for learning capability of inference engine, after that give the software reliability as per input data. Similarly for five membership function total 625 rules are generated automatically for learning capability of inference engine, after that give the software reliability as per input software data.

The software reliability analysis of FIS and ANFIS is to be listed in above Table-I.

## VI. CONCLUSION AND FUTURE SCOPE

We are estimate the reliability of component based software system (CBSS). CBSS reliability is to be estimated by the FIS and ANFIS with two different number of membership function. After compression of the output reliability values for different input sets, than we are analysis that FIS and ANFIS model is provide better result for five membership function as compare three membership function. Here, CBSS reliability estimation performed based on only four factors that is Reusability, Operational profile, Component dependency and Application complexity. But CBSS reliability affected by more other factor like Fault density, Software quality, Together with functionality, Usability, Availability, Performance, Serviceability, Capability, Install ability and Maintainability. So the addition of this factor is left for future work.

## REFERENCES

- [1] Musa, John. Software Reliability Engineering, New York, NY, McGraw-Hill, 1998.
- [2] G. Pour, "Component-Based Software Development Approach: New Opportunities and Challenges," Proceedings Technology of Object-Oriented Languages, 1998. TOOLS 26, pp. 375-383.
- [3] Tyagi, K., Sharma, A., 2014, "An adaptive neuro fuzzy model for estimating the reliability of component-based software system", applied Computing and informatics 10, 38-51.
- [4] Shooman, M., 1976. Structural models for software reliability prediction, proceeding of the Second International Conference on Software Engineering, San Francisco, CA, pp. 268-280.
- [5] Cheung, R.C., 1980. A user oriented software reliability model. IEEE Trans. Softw. Eng. 6 (2), 118-125.
- [6] Popostojanova, K.G., Trivedi, K.S., 2001. Architecture based approach to reliability assessment of software systems. Perform. Eval. J. 45 (2), 179-204.
- [7] Yacoub, S., Cukic, B., Ammar, H., 2004. Scenario based reliability analysis approach for component based systems. IEEE Trans. Reliab. 53 (4), 465-480.
- [8] Dong, W., Huang, N., Ming, Y., 2008. Reliability analysis of component-based software based on relationships of components, IEEE Conference on Web Services, pp. 814-815.
- [9] Huang, N., Wang, D., Jia, X., 2008. FAST ABSTRACT: an algebra-based reliability prediction approach for composite web services, 19th International Symposium on Software Reliability Engineering, pp. 285-286.
- [10] Goswami V., Acharya, Y.B., 2009. Method for reliability estimation of COTS components based software systems, International Symposium on Software Reliability Engineering.
- [11] Seth, K., Sharma, A., Seth, A., 2010. Minimum spanning tree-based approach for reliability estimation of COTSbased software applications. IUP J. Comput. Sci. 4 (4), 13-21.
- [12] Dimov, Aleksandar, Sasikumar, Punnekkat, 2010. Fuzzy reliability model for component-based software systems, 36th EUROMICRO Conference on Software Engineering and Advanced Applications, pp. 39-46.
- [13] Lo, J., 2010. Early software reliability prediction based on support vector machines with genetic algorithms, Fifth IEEE Conference on Industrial Electronics and Applications, pp. 2221-2226.
- [14] Hsu, C., Huang, C., 2011. An adaptive reliability analysis using path testing for complex component based software systems. IEEE Trans. Reliab. 60 (1), 158-170.
- [15] Tyagi, K., Sharma, A., 2012. A rule-based approach for estimating the reliability of component-based systems. Adv. Eng. Softw. 54, 24-29.
- [16] Jang, Jyh-Shing R (1991). Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm (PDF).

- Proceedings of the 9th National Conference on Artificial Intelligence, Anaheim, CA, USA, July 14–19 2. pp. 762–767.
- [17] Jang, J.-S.R. (1993). "ANFIS: adaptive-network-based fuzzy inference system". *IEEE Transactions on Systems, Man and Cybernetics* 23 (3). doi:10.1109/21.256541.
- [18] Abraham, A. (2005), "Adaptation of Fuzzy Inference System Using Neural Learning", in Nedjah, Nadia; de Macedo Mourelle, Luiza, *Fuzzy Systems Engineering: Theory and Practice*, Studies in Fuzziness and Soft Computing 181, Germany: Springer Verlag, pp. 53–83, doi:10.1007/11339366\_3
- [19] Jang, Sun, Mizutani (1997) – *Neuro-Fuzzy and Soft Computing* – Prentice Hall, pp 335–368, ISBN 0-13-261066-3
- [20] Fazal-e-Amin, Ahmad Kamil Mahmood, Alan Oxley 2012, "An Evolutionary Study of Reusability in Open Source Software", *International Conference on Computer & Information Science (ICCIS)* 967-972.
- [21] Fazal-E-Amin, et al., "A Review of Software Component Reusability Assessment Approaches," *Research Journal of Information Technology*, vol. 3, pp. 1-10, 2011.
- [22] Fazal-E-Amin, et al., "Using open source components in software product lines - An exploratory study," in *IEEE Conference on Open Systems (ICOS) 2011*, 2011, pp. 332-337.
- [6] Fazal-E-Amin. et al., "A mixed method study to identify factors affecting software reusability in reuse intensive development," in *National Postgraduate Conference (NPC)*, 2011, pp. 1-6.
- [23] Capretz, et al., "COTS-based software product line development," *International Journal of Web Information Systems*, vol. 4, pp. 165 - 180, 2008.
- [24] Ahmed, et al., "A Model of Open Source Software-Based Product Line Development," in *Computer Software and Applications*. 2008. *COMPSAC '08*. 32nd Annual IEEE International, 2008, pp. 1215 - 1220.
- [25] Mehdi Neshat, Ali Adeli, Azra masoumi and Mehdi sargolzae, 2011, "A Comparative Study on ANFIS and Fuzzy Expert System Models for Concrete Mix Design", *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 3, No. 2, May 2011 ISSN (Online): 1694-0814, 196-210.
- [26] R.Chinnaiyan and Dr.S.Somasundaram, 2008, "RELIABILITY ASSESSMENT OF COMPONENT BASED SOFTWARE SYSTEMS USING TEST SUITE - A REVIEW", *Journal of Computer Applications*, Vol – 1, No.4, Oct – Dec 2008, 34-37.

# Constructing an Add-in Tool for Enterprise Architect v7.5 To Measure the Quality of Object Oriented Design (Class Diagram)

Laheeb Mohammed Ibrahim

Software Engineering  
Mosul University, Collage of Computer Sc. & Math.  
Mosul , Iraq

Khalil Ahmed Ibrahim

Software Engineering  
Mosul University, Collage of Computer Sc. & Math.  
Mosul , Iraq

**Abstract**— Software design is very important stage in software engineering since it lies in the middle of the software development life cycle and costs can be reduced if corrections or improvements made in design phase. Some of the existing CASE tools do not have the ability to correct or improve software design like EA v7.5.

The present study aims to construct a CASE tool that helps software engineers in design phase by assessing or evaluating the quality of that design using object oriented design metrics, use the developed CASE tool as add-in to work inside Enterprise Architect since it has no support for design metrics. So, this paper may be considered as an evolution of such a well-known CASE tool like the Enterprise Architect

In this paper, three tools are developed. First, is “K Design Metrics tool (KDM)” as an add-in that works inside Enterprise Architect (EA) v7.5 which is a well-known, powerful CASE (Computer Aided Software Engineering) tool. KDM tool takes the XMI (XML Metadata Interchange) document for the UML class diagram exported by EA as input, processes it, calculates and visualize metrics, provides recommendations about design naming conventions and exports metrics as XML (Extensible Markup Language) document in order to communicate with other tools namely KRS (K Reporting Service) and KDB (K Database).

A Second tool is K Reporting Service (KRS) “KRS” which takes XML document generated by KDM tool as input, parses it and gives a report. The report helps the project manager or the team leader to monitor the progress and to document the metrics. Hence KRS tool is integrated with Enterprise Architect.

Lastly, K Database “KDB” which takes the same XML document generated by KDM tool as input, parses it and stores metrics in the database to be used as a historical data. KDB tool is also integrated with Enterprise Architect.

Two object oriented design metrics models are used, namely MOOD (Metrics for Object Oriented Design) which measures Encapsulation, Inheritance, Polymorphism and Coupling, and MEMOOD (Maintainability Estimation Model for Object Oriented software in Design phase) which measures Understandability, Modifiability and Maintainability. Both models are validated theoretically and empirically. These measurements allow designers to access the software early in process, make changes that will reduce complexity and improve the design.

All three tools were developed using C# programming language with the aid of Microsoft Visual Studio 2010 as integrated

development environment under Windows 7 operating system with minimum 4 GB of RAM and Core-i3 of CPU.

**Keywords**—MOOD (Metrics for Object Oriented Design); MEMOOD (Maintainability Estimation Model for Object Oriented software in Design phase); UML (Unified Modeling Language); Object Oriented software; Enterprise Architect v7.5.

## I. INTRODUCTION

Software design (object oriented) is the stage in the software engineering process where the executable software system is developed. So, it plays a pivotal role in software development since it determines the structure of the software solution. Once the design has been implemented, it is difficult and expensive to change. Therefore, high design quality is vital for reducing software cost [23] [3] [17][34]

Quality assurance plays an important role in monitoring software process in the form of umbrella activities (Umbrella activities are applied throughout a software project and help a software team manage and control progress, quality, change, and risk [15]) and in the form of measurement or metrics. Without measurements (or metrics), it is impossible to detect problems early in the software process, before they get out of hand. Metrics therefore can evaluate the process and serve as an early warning system for potential problems [20].

Many object oriented design metrics have been proposed specifically for the purpose of assessing the design of a software system such as MOOD (Metrics for Object Oriented Design), CK (Chidamber and Kemrer), Lorenz and Kids metrics [12]. Some of these metrics (or models) are supported by CASE tools due to their importance in evaluating or assessing the design of the software system.

Enterprise Architect (EA) is a well-known CASE tool that is used in over 130 countries for designing and constructing software systems. EA differentiates from other tools in that it supports a comprehensive UML modeling, have a built-in requirements management, test management, extensive project



management support, Code engineering, and .. Etc. But it does not support metrics on software design [25].

As been mentioned earlier that software design is a very important stage in software engineering since it lies in the middle of the software development life cycle and costs can be reduced if corrections or improvements made in design phase. Some of the existing CASE tools do not have the ability to correct or improve software design like EA v7.5.

The present paper aims to construct a CASE tool that helps software engineers in design phase by assessing or evaluating the quality of that design using object oriented design metrics, using two metrics models namely MOOD (Metrics for Object Oriented Design) which measures Encapsulation, Inheritance, Polymorphism and Coupling, and MEMOOD (Maintainability Estimation Model for Object Oriented Systems in Design phase) which measures understandability, modifiability and maintainability, and using the developed CASE tool as add-in to work inside EA since it has no support for design metrics. So, this paper may be considered as an evolvement of such a well-known CASE tool like the EA v7.5

## II. RELATED WORK

Many researchers have worked on object oriented design by means of quality assurance. Some of them propose tools that calculate metrics, other have made surveys about quality models. Following a brief explanation about their works:

Paterson, T et al. (2002) demonstrated the potential for deriving a suite of object-oriented design metrics by the XSLT (Extensible Style Sheet Language Transformation) processing of XMI representations of UML class diagram models. They propose a tool that extracts metrics like number of classes [13].

Girgis, M.R et al. (2009) proposed a tool that automates the computation of the important metrics that are applicable to the UML class diagrams. The tool collects information by parsing the XMI format of the class diagram, and then uses the data to calculate the metrics like CK, MOOD [6].

Poornima, U.S (2011) stated that quality metrics are helpful for the designers in measuring solution architecture for better products. By understanding the solution domain of object oriented systems and measuring the quality of the design using metrics yields to future enhancements [14].

Mago, J. et al. (2012) proposed a model based on fuzzy logic which serves as an integrated means to provide an interpretation of the object oriented design metrics and also surveyed MOOD metrics with other metrics [11].

Rani, T. et al. (2012) proposed a tool (SD-Metrics) that measures the complexity of a class diagram using class metrics from XMI files from Argo UML [16].

Sharma, A.K. et al. (2012) reviewed quality metrics suites namely, MOOD, CK and Lorenz & Kidd, selected some metrics and discarded others based on the definition and capability of the metrics [22].

Hilera, J.R. et al. (2012) made a web service for calculating the metrics of UML class diagrams from XMI document. They stated that as UML becomes a standard format for specifying

classes, it is useful to have a web service that quickly runs metrics on the diagram and gives developers feedback on the class quality [7].

Jassim F. et al. (2013), the main goal was to predict factors of MOOD metrics for object oriented design using a statistical approach. They also used a linear regression model to find out the relationship between factors of MOOD and their influence on object oriented software measurements [10].

Ahmed S.H. et al. (2013), proposed a hybrid metrics suite for evaluating the design of object oriented software early in UML design phase. A metrics extraction tool was developed which operated on UML design models and corresponded XMI files to assure independency results [2].

All studies state that design metrics are important to access the software design early in process and make changes that will improve the design. None of the above mentioned studies fully automate MOOD metrics. In this paper, all MOOD metrics were fully automated and another model (MEMOOD) is used as an add-in inside EA. None of the above studies integrates or improves an existing CASE tool.

## III. SOFTWARE ENGINEERING AND QUALITY ASSURANCE

According to [9], *Software Engineering* can be defined as the “application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software; that is, the application of engineering to software”. Building an information system using the Software Development Life Cycle ( SDLC ) follows a similar set of phases see Fig. 1, requirements phase, design phase, implementation phase, test phase, installation/checkout phase, and operation/maintenance phase [22][26][35].



Figure 1. Software Development Life Cycle (SDLC)

Quality must be defined and measured if improvement is to be achieved. Yet, a major problem in quality engineering and management is that the term ‘Quality’ is ambiguous, so it is commonly misunderstood. The confusion may be attributed to several reasons. First, quality is not a single idea, but rather a multidimensional concept. Second, for any concept there are levels of abstraction; when people talk about quality, one party could be referring to it in its wide sense, whereas another might be referring to its specific meaning. Third, the term quality is a part of the daily language; the popular and professional uses of it may be very different [32]. So, just to be clear, *Quality* can be defined as [9]. The degree to which a system, component, or process meets a customer or user’s needs or expectations.

A key element of any engineering process is measurement. Using measures allows for better understanding of the



attributes of the models that will be created and assessing the quality of the engineered products or systems to be built.

**Measure** is defined as a quantitative indication of the extent, amount, dimension, capacity, or size of some attribute of a product or process whereas **Measurement** is the act of determining a measure [15]. **Metric** is “a quantitative measure of the degree to which a system, component, or process possesses a given attribute [9]. When a single data point has been collected (e.g., the number of errors uncovered within a single software component), a measure has been established. Measurement occurs as the result of the collection of one or more data points. Software metric relates the individual measures in some way (e.g., the average number of errors found per review) [15].

#### IV. ENTERPRISE ARCHITECT (EA)

**EA** is a CASE tool for designing and constructing software systems, for business process modeling, and for more generalized modeling purposes [30][28]. EA was developed by Sparx Systems © and it covers all aspects of the software development cycle from requirements gathering, through analysis, model design, testing, change control and maintenance to implementation, with full traceability (identifies the way a given process has been, or is to be, developed in a system). [30].

EA has proven to be highly popular across a wide range of industries and is used by thousands of companies worldwide, from large, well known, multinational organizations to smaller independent companies and consultants [24]. Sparx Systems © software is used in the development of many kinds of applications and systems in a wide range of industries, including aerospace, banking, web development, engineering, finance, medicine, military, research, academia, transport, retail, utilities (such as gas and electricity) and electrical engineering. It is also used effectively for UML and enterprise architecture training in many prominent colleges, training companies and universities around the world [24] [29]. For all those reasons mentioned earlier in addition to the powerful description of UML class diagrams as XMI, this paper tends to use EA as a platform for the proposed tools to work with.

EA is a great UML CASE tool, but we can make it even better by adding and extending new functionality in the form of an add-in. To fully understand the steps necessary to get the add-in running, we should first understand how EA's add-in architecture works[33]. When EA starts up, it will read the registry key [HKEY \_ CURRENT \_ USER \ Software \ Sparx Systems \ EAAddins].

Each of the keys in this location represents an add-in for EA to load. The (default) value of the key contains the name of the assembly and the name of the add-in class separated by a dot. EA then asks Windows for the location of the assembly (An assembly is a file that is automatically generated by the .NET compiler upon successful compilation of every .NET application. It can be either a DLL or an executable file), which is stored on the COM codebase entries in the registry,

and it will use the public operations defined in the add-in class[33].

#### V. ANALYSIS AND DESIGN OF KDM, KRS AND KDB TOOLS

This section explains in detail the proposed tools from the analysis and design point of view. These tools are named KDM, KRS and KDB Tool, which helps the software engineer in the design phase of the software life cycle. For modeling the proposed tools, The following CASE tools (Edraw Max, Microsoft Visio and EA) are used.

Before start analyzing the proposed tools in detail, it is needed to describes them in a general way by showing how the final user of the proposed tools like a software engineer, project manager or programmer will use them. KDM tool is used to calculate the metrics for the OOD and considered being the main tool, while KRS tool can help with the documentation of the results, and finally KDB tool can help by storing the metrics in the database.

##### A. K-Design Metrics (KDM) Tool

EA does not support any tool that measures the class diagram. So, in this paper KDM tool was developed to work from inside the EA as add-in to help software engineer understanding the design of the software better by scrutinizing the class diagram of that software by means of design metrics. In addition, the KDM tool (add-in) can be deployed to work on other machines not just on the machine where it is developed, so that other software engineers can use it. The proposed KDM tool accepts XMI 1.1 for the UML 1.3 generated by the EA v7.5 as **input** and calculates metrics for that design.

The **output** of KDM tool is the value of metrics and 3-dimension pie chart which visualizes the value of each metric. It also gives statistics about that design and produces XML document. See Fig 2 which shows the input and output of the KDM tool.

##### 1. KDM Tool in SDLC

KDM tool operates on UML class diagram either in the analysis phase (high-level design) or in the design phase (low-level or detailed design). KDM tool is classified as Upper CASE Tool (front-end) since it works in the upper level of the SDLC.

##### 2. How KDM Tool Works

KDM tool imports XMI document which then will be fed into the XMI parser. The parser will extract the required information from XMI document and pass it to the metric module which contains the MOOD model and MEMOOD model which in turn calculates the metrics for that design. KDM tool draws 3D pie chart, gives recommendations about design naming conventions, and gives design statistics, also exports XML document.

XMI is a way of saving UML diagrams as XML so it contains huge data that describes the UML diagram (in this paper the class diagram) in detail such as the name of each class, its attributes, operations, relationships, style, etc.

XMI document is stored either as XML or XMI extension which means that the information is represented or structured as **tags**.

XMI document has a large set of tags. Some are important but others are not, such as the style of each class, date of creation, etc. The tags used to calculate the metrics in this paper are listed in table I with their description.

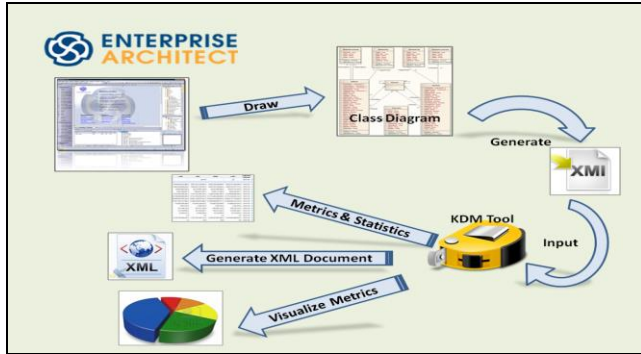


Figure 2. Input and output for KDM tool (All inside EA)

TABLE I. XMI TAGS USED IN THIS PAPER

Tag	Description
<UML:Classes>	This tag is used to represent the class element. <b>UML:</b> is a namespace (Namespace provides a means to distinguish one XML vocabulary from another, which enables to create richer documents by combining multiple vocabularies into one document type [8]) which stands for "omg.org/UML1.3"
<UML:Attribute>	This tag is used to represent the attribute of the class
<UML:Operation>	This tag is used to represent the methods of the class
<UML:TaggedValue>	Tagged Values are a way of adding additional information to an element
<UML:Generalization>	This tag is used to represent inheritance relationship and it has two tagged values: "ea_sourceName" which represents source class (sub class) that inherits from target class (super class). "ea_targetName" which represents target class (super class) in which subclass inherits from it.
<UML:Association>	This tag is used to represent association, aggregation and composition. We can tell the difference between them by their tagged values. It has two tagged values for the source and the target classes.

• **XMI Parser**

XMI parser is used to extract data from XMI document, especially those tags listed in table I. Two important programming technologies were used in building XMI parser. They are: LINQ (Language Integrated Query) -to-XML and Lambda expressions. XMI parser will store all values of tags in lists like a list which contains the names for all classes, operations for each class, source classes and target classes for generalization relationship, ...etc. In order to find the name for the classes, attributes or operations in the XMI document, the following algorithm can be used:

**Algorithm:**

- Step 1: Read XMI document and load it into XDocument object
- Step 2: Determine the tag = "Class"
- Step 3: Repeat for each tag
  - Step 3-1: Extract the value of the *name* attribute of the tag
  - Step 3-2: Save the name in the class list
  - Step 3-3: If not finish reading all tags, go to Step 3
- Step 4: Display the class list

"Class list" will contain the name of each class in the XMI document; this list is the basis for all other methods in the XMI parser, because in order to find the name of each method in some class, it is needed to know the class name first (to which class they belong). In addition, to find attributes or methods names, the same algorithm can be used except for the tag which can be either as "Attribute" or "Operation".

In case of inheritance relationship, when it is needed to find the source classes (sub classes) and target classes in generalization relationship, the following algorithm can be used:

**Algorithm:**

- Step 1: Read XMI document and load it into XDocument object
- Step 2: Determine the relation = "Generalization"
- Step 3: Repeat for each tag
  - Step 3-1: Extract the value of the ea\_source tag attribute of the relation tag
  - Step 3-2: Save the name in a source list
  - Step 3-3: Extract the value of the ea\_target tag attribute of the relation tag
  - Step 3-4: Save the name in a target list
  - Step 3-5: If not finish reading all tags, go to Step 3
- Step 4: Display the list

"Source list" and "target list" contain the subtype classes and supertype classes in XMI document. By knowing the source and target classes in the generalization relationship, this will help calculating metrics like **MIF** or **AIF** which are related to inheritance concept. Thus, to find the source and the target list of another relation, only tag relation will change. Now consider the following Fig. 3 which represents a simple class diagram for aircraft types.

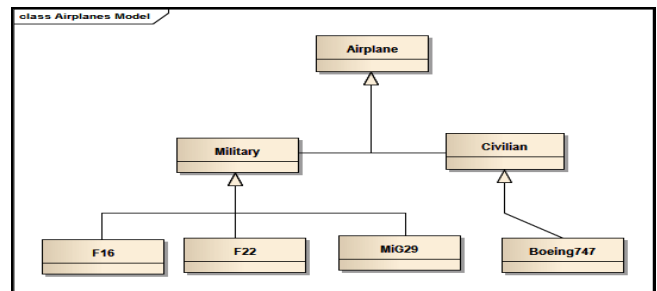


Figure 3. Simple class diagram for aircraft classification

According to the above algorithms, "class list" will contain the names for all classes, see table II, "Source list" and

“target list” for generalization relationship can be seen in table III. After collecting all required information, it is time to calculate the metrics for that design.

TABLE II. SAMPLE OF CLASS LIST

Classes Name
Airplane
Military
Civilian
Boeing747
MiG29
F16
F22

TABLE III SAMPLE OF SOURCE AND TARGET LISTS

Source Classes	Target Classes
Military	Airplane
Civilian	Airplane
Boeing747	Civilian
F16	Military
F22	Military
MiG29	Military

### 3. Suggested Algorithms For Mood Model

The person who sets the MOOD metrics was Fernando B. Abreu [1]. MOOD refers to a structural model of the object oriented paradigm like *encapsulation* as (Method Hiding Factor (MHF) and Attribute Hiding Factor (AHF)), *inheritance* as (Method Inheritance Factor (MIF) and Attribute Inheritance Factor (AHF)), *polymorphism* as (Polymorphism Factor (POF)), and *message passing* as (Coupling Factor (CF)). Each of the metrics was expressed to measure where the numerator defined the actual use of any one of the feature for a particular design. In MOOD model, there are two main features, namely methods and attributes [3].

Attributes are used to represent the status of object in the system and methods are used to maintain or modify several kinds of status of the objects [21][3].

MOOD metrics are designed to meet a particular set of criteria. They were also proposed by the MOOD project team. MOOD model in detail that will help to explain how to calculate each equation of MOOD model.

#### Algorithm for MHF Metric

MHF metric is used to measure encapsulation for the class diagram, actually for the invisibilities of methods for that class.

##### Algorithm:

- Step 1: Import and verify XMI document (verification means it is XMI document)
- Step 2: Parse XMI document
- Step 3: Define a list for each access modifier of the methods
- Step 4: Repeat for each class

Step 4-1: Store methods for each class; where the public methods are stored in the public list, private methods in private list and protected methods (if existed) in the protected list

Step 4-2: Go to step4

Step 5: Calculate MHF equation

$$MHF = \frac{\sum_{i=1}^{TC} [\sum_{m=1}^{Ma(Ci)} (1 - V(Mmi))]}{\sum_{i=1}^{TC} Ma(Ci)} \dots\dots (1)[19]$$

Where: TC = total number of classes. Summation occurs over i=1 to TC. Ci = class with index i (current class). Md (Ci) = the number of methods defined in class Ci. V (Mmi) = Visibility value of a member (method or attribute), i.e. a value between 0-1 where public members = 1, private members = 0, and semi-public (e.g. protected) members are calculated as the number of classes that can access the member / total classes in the system (if working with different packages at the same time then the protected member is calculated. Otherwise it is considered the same as public in which it is equal to 1).

Step 6: Display MHF for the design

#### Algorithm for AHF Metric

AHF metric is used to measure encapsulation for the class diagram, actually for the invisibilities of the attributes for that class.

##### Algorithm:

- Step 1: Import and verify XMI document
- Step 2: Parse XMI document
- Step 3: Define a list for each access modifier of the attributes
- Step 4: Repeat for each class

Step 4-1: Store attributes for each class; where public attributes are stored in the public list, private attributes in private list and protected attributes (if existed) in the protected list

Step 4-2: Go to step4

Step 5: Calculate AHF equation

$$AHF = \frac{\sum_{i=1}^{TC} [\sum_{m=1}^{Ad(Ci)} (1 - V(Ami))]}{\sum_{i=1}^{TC} Ad(Ci)} \dots\dots\dots (2)[19]$$

Where: TC = total number of classes. Summation occurs over i=1 to TC. Ci = class with index i (current class). Ad (Ci) = the number of attributes defined in class Ci . V(Ami) is the same as V(Mmi) except it is for the attribute not for the method

Step 6: Display AHF for the design

#### A Suggested Algorithm for Finding the Root of Generalization or Aggregation Relationship

Sometimes a number of either a generalization hierarchy or aggregation hierarchy exists. This means that there are a number of roots in the design. In order to find the root of either of them, the following algorithm is suggested.

**Algorithm:**

- Step 1: Import and verify XMI document
- Step 2: Parse XMI document
- Step 3: Determine the type of the relationship
- Step 4: Define lists for root classes, subclasses, and super classes.
- Step 5: Repeat for each class in the list of super classes
  - Step 5-1: If any class is not in the list of source (subclasses), it means that the class does not inherit from other classes, so it is a root, add it to root list
  - Step 5-2: Go to Step 5
- Step 6: If some class is repeated more than once, then delete it.
- Step 7: Display root list

**Algorithm for MIF Metric**

MIF metric is used to measure the inheritance of the class diagram, which is the ratio of the inherited methods in it.

**Algorithm:**

- Step 1: Import and verify XMI document
- Step 2: Parse XMI document
- Step 3: Define a list of source classes (subclasses) and another list of the target classes (super classes).
- Step 4: Find the root of the generalization relationship
- Step 5: Repeat for each class in source and target lists
  - Step 5-1: Store inherited methods in a list called inherited list
  - Step 5-2: Go to 5
- Step 6: Calculate the equation of MIF
 
$$MIF = \frac{\sum_{i=1}^{TC} Mi(Ci)}{\sum_{i=1}^{TC} Ma(Ci)} \dots\dots\dots (3) [19]$$

Where: TC = total number of classes. Summation occurs over i=1 to TC. Ci = class with index i (current class). Mi is the number of inherited methods in Ci  
Ma is the number of available methods defined in Ci  
Md is the number of declared methods and not inherited in Ci  
Ma = Md + Mi of class Ci
- Step 7: Display MIF for the design

**Algorithm for AIF Metric**

AIF metric is used to measure the inheritance of the class diagram, which is the ratio of the inherited attributes in it.

**Algorithm:**

- Step 1: Import and verify XMI document
- Step 2: Parse XMI document
- Step 3: Define a list for source classes (subclasses) and another list for the target classes (super classes).
- Step 4: Find the root of the generalization relationship
- Step 5: Repeat for each class in source and target lists
  - Step 5-1: Store inherited attributes in a list called inherited list
  - Step 5-2: Go to Step 5
- Step 6: Calculate the equation of AIF
 
$$AIF = \frac{\sum_{i=1}^{TC} Ai(Ci)}{\sum_{i=1}^{TC} Aa(Ci)} \dots\dots\dots (4) [19]$$

Where:TC = total number of classes. Summation occurs over i=1 to TC. Ci = class with index i (current

class). Ai is the number of inherited Attributes in Class Ci. Aa is the number of available attributes defined in class Ci. Ad is the number of attributes declared in the class Ci. Aa = Ad + Ai of class Ci.

- Step 7: Display AIF for the design

**Algorithm for POF**

POF measures the polymorphism of the class diagrams. This metric calculates the ratio of the polymorphic methods (degree of overriding in class diagram).

**Algorithm:**

- Step 1: Import and verify XMI
- Step 2: Parse XMI document
- Step 3: Calculate the source and target classes
- Step 4: NC = total number of classes
- Step 5: Repeat for each class while < NC
  - Step 5-1: Find the descendant classes for each class in the target list
  - Step 5-2: Find the new (declared) method for each class and put them in a list
  - Step 5-3: Find the overridden methods and put them in a list
  - Step 5-4: Go to Step 5
- Step 6: Calculate POF

$$POF = \frac{\sum_{i=1}^{TC} Mo(Ci)}{\sum_{i=1}^{TC} [Mn(Ci) + DC(Ci)]} \dots\dots\dots (5) [19]$$

Where: TC = total number of classes. Summation occurs over i=1 to TC. Ci = class with index i (current class). Mo(Ci) is the overridden methods for Class Ci. Mn(Ci) is the new methods defined in Class Ci. DC(Ci) is the descendant counts (number of subclasses) for Class Ci .

- Step 7: Display POF for the design

**Algorithm for CF**

CF is used to measure the coupling of the class diagram when one class calls a method of another class, then they are coupled.

**Algorithm:**

- Step 1: Import and verify XMI
- Step 2: Parse XMI document
- Step 3: Find the source and target classes of the association relationship
- Step 4: Concatenate the target list with the source list, remove duplication and put it into a new list called c list
- Step 5: Repeat for each class in c list
  - Step 5-1: If a class has any relationship but not generalization then put it into a list
  - Step 5-2: Go to Step 5
- Step 6: Apply CF equation

$$CF = \frac{\sum_{i=1}^{TC} [\sum_{j=1}^{TC} is\_client(Ci,Cj)]}{TC^2 - TC} \dots\dots\dots (6) [19]$$

Here: TC = total number of classes. Summation occurs over i=1 to TC. Ci = class with index i (current class). is\_client(Ci,Cj)=1 if Ci contains at least one non inheritance reference to a method or attribute of a class and Cj=0 otherwise.

Step 7: Display CF

Now, after all algorithms about MOOD model are explained, consider the following example which illustrates all algorithms above, see Fig. 4. Now consider the table IV which represents the class diagram as numbers.

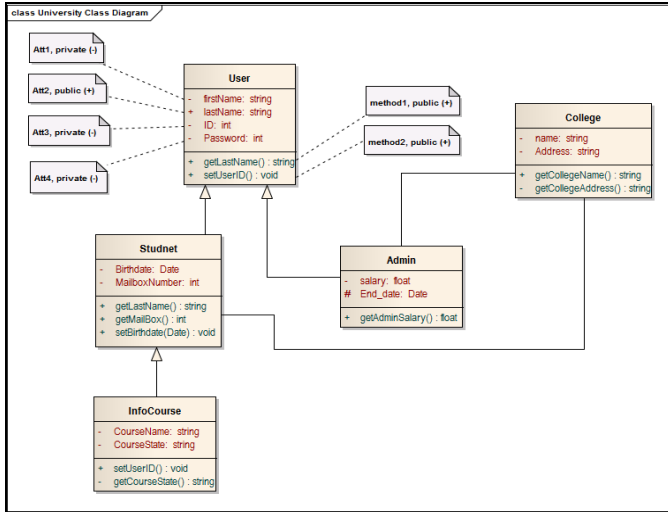


Figure 4. Simple class diagram for some of university entities

TABLE IV CLASS DIAGRAM ANALYSIS

Class	Att.	Method	+ Att.	- Att.	#Att.	+M.	- M.	#M.
User	4	2	1	3	0	2	0	0
Student	2	3	0	2	0	3	0	0
Admin	2	1	0	1	1	1	0	0
InfoCourse	2	2	0	2	0	1	1	0
College	2	2	0	2	0	1	1	0

Where: Att. is an abbreviation for attribute. M. is an abbreviation for method. + prefix means public modifier. - prefix means private modifier. # prefix means protected modifier.

Fig. 4 shows a simple class diagram with 5 classes, 12 attributes and 10 methods, to calculate the metrics according to the table above using metrics equations.

1.Encapsulation (Eq. (1) and Eq. (2))

$$MHF = \frac{0+1+0+0+1}{2+2+1+3+2} = \frac{2}{10} = 20\%$$

$$AHF = \frac{3+2+1+2+2}{4+2+2+2+2} = \frac{10}{12} = 83.33\%$$

2. Inheritance (Eq. (3) and Eq. (4))

$$MIF = \frac{0+2+2+5}{2+5+3+7} = \frac{9}{17} = 52.94\%$$

$$AIF = \frac{0+4+4+6}{4+6+6+8} = \frac{14}{24} = 58.33\%$$

3.Polymorphism (Eq. (5))

$$POF = \frac{0+1+0+1}{4+2+0+0} = \frac{2}{6} = 33.33\%$$

4.Coupling (Eq. (6))

$$CF = \frac{3}{25-5} = \frac{3}{20} = 15\%$$

It is concluded from table V, that AHF, MIF, AIF are within the limit while MHF, POF, CF are not within the standard limit. So, a correction or a review of the design is needed.

TABLE V Standard intervals for mood model [1]

Metrics	Minimum Value	Maximum Value
MHF	12.7%	21.8%
AHF	75.2%	100%
MIF	66.4%	78.5%
AIF	52.7%	66.3%
POF	2.7%	9.6%
CF	4.0%	11.2%

4. Suggested Algorithms for MEMOOD model

The ever changing world makes maintainability a strong quality requirement for the majority of software systems. The maintainability measurement during the development phases of object oriented system estimates the maintenance effort. It also evaluates the likelihood that the software product will be easy to maintain. Despite the fact that software maintenance is an expensive and challenging task, it is not properly managed and often ignored. One reason for this poor management is the lack of proven measures for software maintainability [18].

• Algorithm for Maintainability

**Maintainability** is defined as “the ease with which a software system or component can be modified to correct faults, improve performance or other attributes, or adapt to a changed environment” [9]. As class diagrams play a key role in the design phase of object-oriented software, early estimation of their maintainability may help designers to incorporate required enhancements and corrections in order to improve their maintainability and consequently the maintainability of the final software to be delivered in future. Two quality attributes of class diagram, namely understandability and modifiability are focused to estimate their maintainability [18].

Maintainability means how easy it is for software engineer to maintain the design by means of adapting, correcting or improving the design. In order to calculate the maintainability, understandability and modifiability are used along with a number of constants to form the equation (see Eq.( 7)).

Algorithm:

- Step 1: Import XMI
- Step 2: Parse XMI document
- Step 3: Calculate understandability and modifiability
- Step 4: Apply maintainability model

$$Maintainability = -0.126 + 0.645 * Understandability + 0.502 * Modifiability \dots \dots (7)$$

Step 5: Display maintainability



• **Algorithm for Understandability**

**Understandability** means how much the software engineer understands the design that he is working on or how easy to comprehend it. In order to calculate the understandability of the design it is needed first to find two metrics, named NC and NGenH (see table VI). These two metrics along with some constant numbers are used to calculate the understandability of the design.

**Algorithm:**

Step 1: Import XMI

Step 2: Parse XMI document

Step 3: Calculate NC and NGenH for the design

Step 4: Apply understandability

$$Understandability = 1.166 + 0.256 * NC - 0.394 * NGenH \dots (8)$$

Where: NC is the total number of classes. NGenH is the number of generalization hierarchies.

Step 5: Save the value of understandability

Step 6: Display understandability

• **Algorithm for Modifiability**

**Modifiability** means the ability of software engineer to modify the design without affecting it. In order to calculate the modifiability of the design it is needed first to find five metrics, named NC, Ngen, NgenH, NaggH and MaxDIT (see table VI). These five metrics along with some constant numbers are used to calculate the modifiability of the design.

**Algorithm:**

Step 1: Import XMI

Step 2: Parse XMI document

Step 3: Calculate NC, Ngen, NgenH, NaggH and MaxDIT for the design

Step 4: Apply modifiability equation

$$Modifiability = 0.629 + 0.471 * NC - 0.173 * NGen - 0.616 * NAggH - 0.696 * NGenH + 0.396 * MaxDIT \dots (9)[18]$$

Where: NC is the total number of classes. NGen is the number of generalization relationship (inheritance relationship between super class and sub class). NAggH is the number of aggregation relationship hierarchies. NGenH is the number of generalization hierarchies in the design. MaxDIT is the maximum depth of the inheritance in the design

Step 5: Save the value of modifiability

Step 6: Display modifiability

Metrics in table VI have been selected for quantifying understandability and modifiability of class diagram. It had already been empirically validated that these metrics are correlated with understandability and modifiability of class diagram [27][4].

In order to calculate the Maintainability Estimation model see Fig.5 [18], Both the Understandability and the Modifiability of the design are used. *Understandability* in our context means the extent of users (software engineer or

programmer) capability with different backgrounds to understand the software design. Understandability of the design can be calculated as in Eq. (10)[18].

$$Understandability = 1.166 + 0.256 * NC - 0.394 * NGenH \dots (10)$$

TABLE VI SIZA AND CTRUCTURAL COMPLEXITY METRICS FOR UML CLASS DIAGRAM

Metric Name	Metrics Definition
Number of classes (NC)	The total number of classes
Number of attributes (NA)	The total number of attributes
Number of methods (NM)	The total number of methods
Number of associations (NAssoc)	The total number of associations
Number of aggregation (NAgg)	The total number of aggregation relationships within a class diagram (each whole-part pair in an aggregation relationship)
Number of dependencies (NDep)	The total number of dependency relationships
Number of generalizations (NGen)	The total number of generalization relationships within a class diagram (each parent-child pair in a generalization relationship)
Number of aggregations Hierarchies (NAggH)	The total number of aggregation hierarchies (whole-part structures) within a class diagram
Number of generalizations Hierarchies (NGenH)	The total number of generalization hierarchies within a class diagram
Maximum depth of inheritance (MaxDIT)	It is the maximum of the DIT (Depth of Inheritance Tree) values obtained for each class of the class diagram. The DIT value for a class within a generalization hierarchy is the longest path from the class to the root of the hierarchy
Maximum aggregation hierarchy (MaxHAgg)	It is the maximum of the HAgg values obtained for each class of the class diagram. The HAgg value for a class within an aggregation hierarchy is the longest path from the class to the Leaves.

*Modifiability* in our context is the capability to modify the design without affecting the overall system. See equation (11)[18].

$$Modifiability = 0.629 + 0.471 * NC - 0.173 * NGen - 0.616 * NAggH - 0.696 * NGenH + 0.396 * MaxDIT \dots (11)$$

After calculating understandability and modifiability quality attributes it is possible now to find the maintainability of software design. See equation (12)[18].

$$Maintainability = -0.126 + 0.645 * Understadability + 0.502 * Modifiability \dots (12)$$

The values of understandability, modifiability and maintainability are of immediate use in the software development process. These values may help software designers to review the design and take appropriate corrective measures, early in the development life cycle, in order to control or at least reduce future maintenance cost [18].

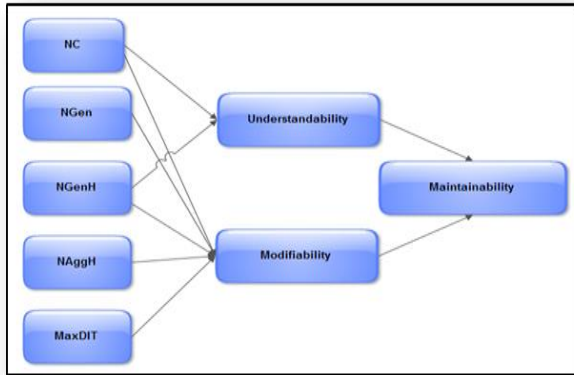


Figure 5. Maintainability Estimation Model (MEMOOD) [18]

Now go back to Fig.4, the following table VII can be deduced.

TABLE VII METRICS USED TO CALCULATE MEMOOD MODEL, ALSO SEE TABLE VI

NC	Ngen	NgenH	NaggH	MaxDit
5	3	1	0	2

Where

- NC is the total number of classes = 5.
- Ngen is the number of generalization relationships = 3.
- NgenH is the number of generalization hierarchy =1, since there is only one generalization tree.
- MaxDit is the maximum number of depth of inheritance tree =2, since the User class is in level 0 of the generalization hierarchy, Student and Admin classes are in level 1, and InfoCourse class is in level 2.
- NaggH is the number of the aggregation hierarchy = 0, since there is no aggregation hierarchy.

1. **Understandability (Eq. 10)**

$$\text{Understandability} = 1.166 + 0.256*5 - 0.394*1 = 2.05$$

2. **Modifiability (Eq. 11)**

$$\text{Modifiability} = 0.629 + 0.471*5 - 0.173*3 - 0.616*0 - 0.696*1 + 0.396*2 = 2.56$$

3. **Maintainability (Eq. 12)**

$$\text{Maintainability} = -0.126 + 0.645* 2.05 + 0.502*2.56 = 2.48$$

5. XML

XML is a standard technology that is concerned with the description and structuring of data by means of **tags** that are similar to HTML ones. XML can be used almost in every application especially in the web. See Fig. 6 which represents a sample of XML. It can be seen from the figure above that XML is used to describe a book; its title, author, price, etc. XML sometimes is used as intermediate data that flow between applications and these applications passes these XML between each other, so XML can be used as a bridge between various applications. Going back to Fig. 2, it can be seen that XML was used as **output** from KDM tool which is the main tool, KRS and KDB are developed to support it. So, how can these tools communicate between each other? The answer is by using XML as a bridge between them. XML parser was built for that XML in which it will be understood and used properly. A specific structure of XML is proposed in this paper (see table VIII).

TABLE VIII XML STRUCTURE OF THE KDM XML OUTPUT

Tag	Description
<Metrics>	This tag is used as root for a number of metrics
<Metric Id="" />	This tag is used as an identifier for the metrics
<DesignerName>	This tag is used to describe the designer name
<ModelName>	This tag is used to describe the model name
<MHF>	This tag is used to describe the MHF metric
<AHF>	This tag is used to describe the AHF metric
<MIF>	This tag is used to describe the MIF metric
<AIF>	This tag is used to describe the AIF metric
<CF>	This tag is used to describe the CF metric
<POF>	This tag is used to describe the POF metric
<Understandability>	This tag is used to describe the understandability
<Modifiability>	This tag is used to describe the modifiability
<Maintainability>	This tag is used to describe the maintainability

A sample of XML document can be seen in Fig. 7 which is also the XML output from KDM tool.

```

C:\Users\Senior Furlala\Desktop\XML TryItOut.C...
- <Book>
  <title>Night Fall</title>
  <author>Demille, Nelson</author>
  <publisher>Warner</publisher>
  <price>$26.95</price>
  <contentType>Fiction</contentType>
  <isbn>0446576638</isbn>
</Book>

```

Figure 6. XML sample

```

<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
- <Metrics>
  <Metric Id="10" />
  <DesignerName>Khali Ahmed</DesignerName>
  <ModelName>My Model</ModelName>
  <MHF>7.14</MHF>
  <AHF>84.21</AHF>
  <MIF>63.04</MIF>
  <AIF>55.56</AIF>
  <CF>11.11</CF>
  <POF>9.52</POF>
  <Understadability>3.08</Understadability>
  <Modifiability>3.83</Modifiability>
  <Maintainability>3.78</Maintainability>
</Metrics>

```

Figure 7. XML output from KDM tool

## 6. KDM Tool Sequence Diagram

KDM tool sequence of operations starts after importing XMI document and ends with exporting XML, see Fig. 8

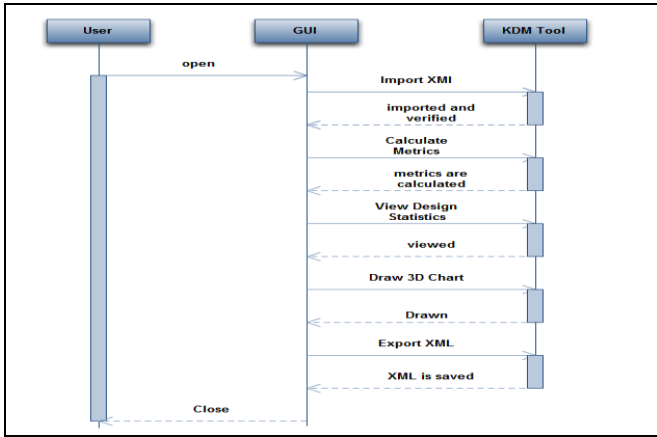


Figure 8. KDM tool sequence diagram

### B. KRS (K Reporting Service) Tool

KRS tool is a reporting tool that is integrated with EA. The purpose of this tool is to document metrics as a report for a project manager or maybe for the team leader. It is said earlier that KRS tool supports KDM tool and they communicate by exchanging XML. The **input** for KRS tool is the XML output of KDM tool. So, KRS has a parser for the XML generated by KDM tool. The **output** of KRS tool is a crystal report which contains metrics and two graphs, see Fig.9.

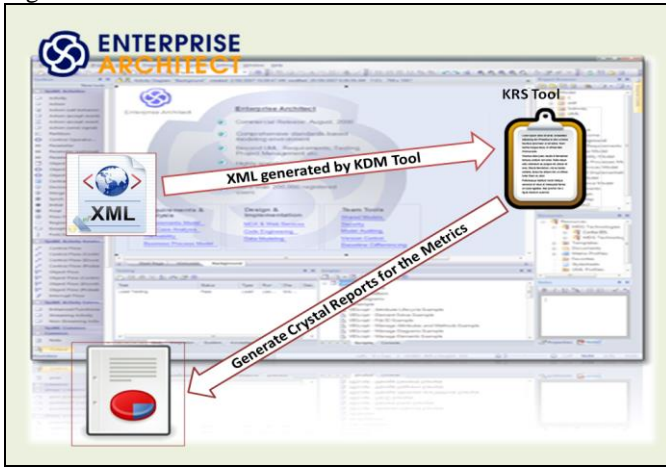


Figure 9. Input and output for KRS Tool

### 1. How KRS Tool Works

Before discussing how it works, it is needed to know where KRS tool works, and in which phase it supports in SDLC. KDM tool is an Upper CASE tool and since KRS works with the documentation of metrics in the same phase, it is deduced that KRS tool is also an Upper CASE tool. KRS tool accepts XML document that is generated by KDM tool as input, see Fig. 7. Then XML document proceeds to XML parser which extracts the information and prepares it to be fed into the report generator and produces a crystal report of the design metrics. See Fig. 10.

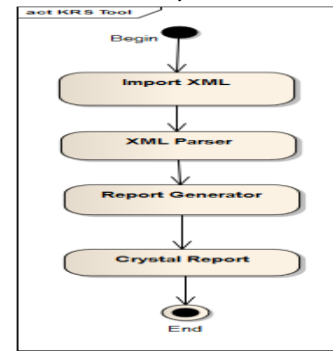


Figure 10. KRS Tool Workflow

### 2. XML Parser

XML parser extracts the value of each tag listed in table VIII from the XML document which is depicted in Fig. 11.

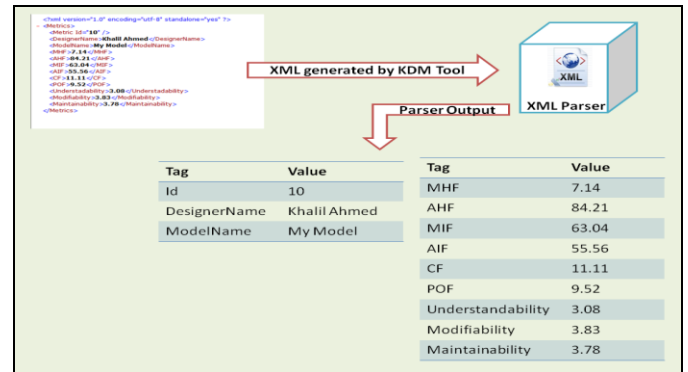


Figure 11. XML Parser

### Algorithm:

- Step 1: Import and verify XML
- Step 2: Extract model name and designer name from XML document and put them in a list.
- Step 3: Extract all metrics and put them in a list
- Step 4: End

### C. KDB (K Database) Tool

KDB tool is a database tool that is integrated with EA; KDB tool is used to store metrics in a database; may be for checking the metrics against another system (design) which has similar requirements or used as a historical data. KDB tool has the same XML parser of KRS tool. KDB tool accepts the same XML which is generated by KDM tool as **input**, and stores the numeric (double data type) value of metrics in the database

#### 1. How KDB Tool Works

Any tool that supports any phase in SDLC is considered a CASE tool, otherwise it is not. Since KDB tool stores metrics which are software engineering information and supports KDM tool, so, as a result, it is a CASE tool and can be considered an Upper CASE tool. KDB tool takes XML document which is generated by KDM tool, parses it, and formats metrics in a way that can be stored in the database.



VI. TESTING THE PROPOSED TOOLS

A case study has been taken from [5] and modified so that all metrics can be calculated. The case study is about a Student Registration System at university. By using this system, students have access to the information of the available courses, and they can also register in the system, it is managed by a special user who is allowed to modify the required courses in the catalogue. This system was modeled using EA v7.5. See Fig. 12 which represents the class diagram for the system. This class diagram is exported from EA as XMI which will be the input for KDM tool, and by pressing on Metrics button the metrics are calculated. See Fig. 13.

It can be seen that class names group box is filled with all classes from the class diagram. MOOD and MEMOOD values are calculated. Design statistics can be seen in Fig. 14.

Form the statistics above, it can be seen that there are 9 classes, 19 attributes, 28 methods with 9 relations, one aggregation hierarchy, one generalization hierarchy and the maximum depth of inheritance is 2. Required information group box is used to export XML document that contains the metrics along with the model name, designer name and model id. This document is used as input for KRS and KDB tool. Visualization of metrics for example can be seen in Fig. 15 where red color means that the design is needed to be reviewed according to metric value, green color means that the metric value is within the allowed range and no review is needed.

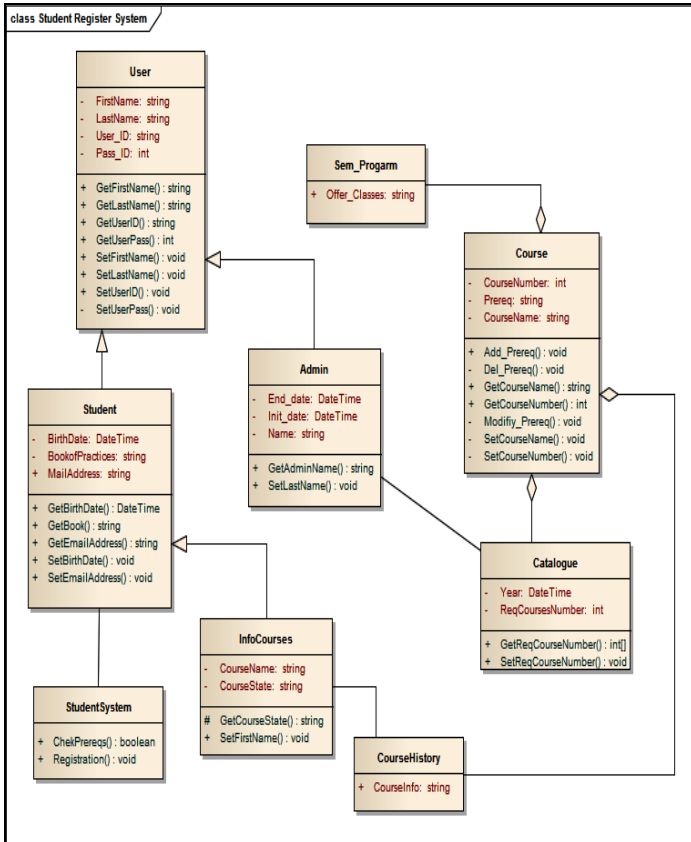


Figure 12. Student Registration System Class Diagram



Figure 13. Metrics for Student Registrations System.

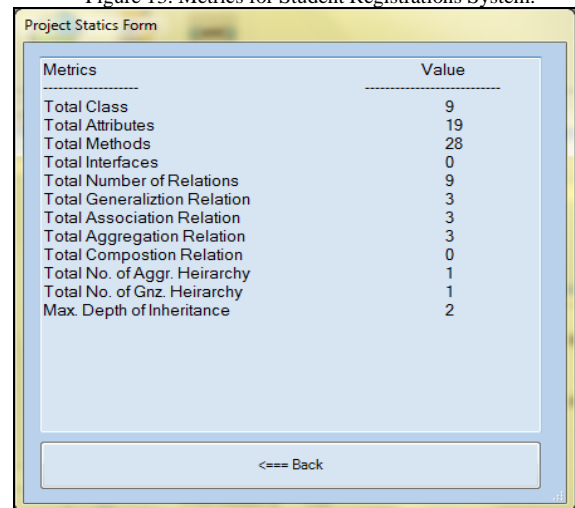


Figure 14. Design Statistics for the Class Diagram.

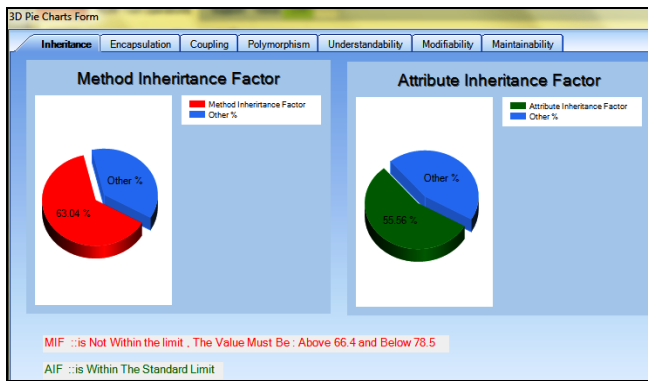


Figure 15. MIF and AIF Metrics 3D-Pie Chart.

KDM tool supports JAVA design naming conventions of the design. See Fig. 16.

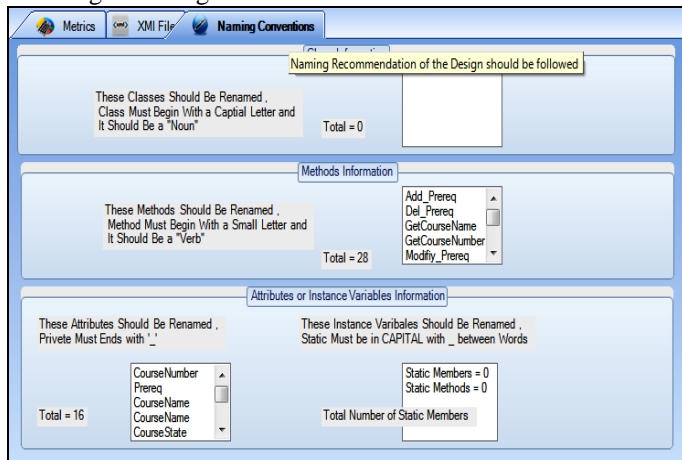


Figure 16. Naming Conventions.

By opening KRS tool and importing the XML document generated by KDM tool, the output is a crystal report. See Fig. 17.



Figure 17. A Crystal Report for the Metrics of the System.

By opening KDB tool and importing XML document generated by KDM tool, XML parser will extract metrics and KDB tool will load them into the text boxes and into the XML

tab . By pressing on View All Data button, a new form will open and it will contain the metrics that are stored in the database. See Fig 18.

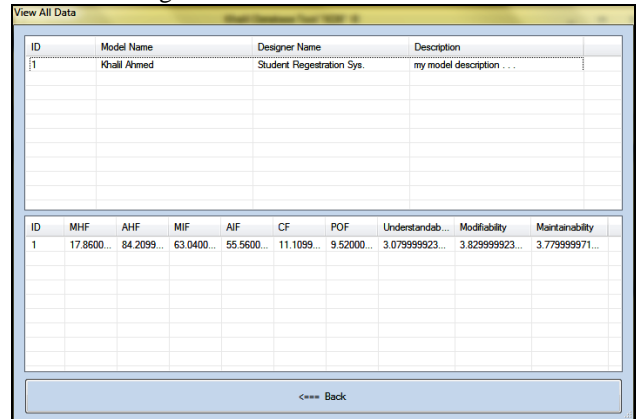


Figure 18. View All Data Form.

## VII. DISCUSSION OF TESTING RESULTS

KDM tool has succeeded in calculating MOOD and MEMOOD metrics and it gives 100% correct results, because the metrics are calculated by hand and have the same values of KDM tool. From the results of KDM tool, table IX can be deduced.

TABLE IX Metrics discussion

Metric	Recommendation	Value	Within the limit	Outside the limit
MHF	No recommendation is needed	17.86	✓	-
AHF	No recommendation is needed	84.21	✓	-
MIF	<b>It is recommended that the number of inherited methods in the design should be reduced</b>	<b>63.04</b>	-	✓
AIF	No recommendation is needed	55.56	✓	-
CF	No recommendation is needed	11.11	✓	-
POF	No recommendation is needed	9.52	✓	-
Understandability	No recommendation is needed	3.08	✓	-
Modifiability	No recommendation is needed	3.83	✓	-
Maintainability	No recommendation is needed	3.78	✓	-

From the table above, it can be concluded that the design is fine and a review must be taken for MIF value.

### A. Evaluation of the Proposed Tools

In this paper questionnaire as in [31] has been conducted by a twenty person who are considered as users of the proposed tools (programmers and software engineers). The samples were taken from people within the field (Computer Science and Software Engineering). The questionnaire divided into four sections namely:

1. Evaluating the tools generally,
2. Evaluating KDM Tool,
3. Evaluating KRS Tool, and
4. Evaluating KDB Tool.

Using SPSS program to get the results, see the table below.

TABLE X QUESTIONNAIRE RESULTS

Tool Name	Questionnaire Result
Evaluating the tools generally	94.4
Evaluating KDM Tool	93.8
Evaluating KRS Tool	96.6
Evaluating KDB Tool	90.7

### VIII. CONCLUSION

Through the building and testing of KDM, KRS, and KDB tools, conclusions were that; KDM tool accepts XMI or XML documents generated by EA since EA export UML diagram as .XML or .XMI extension. Documentation of metrics do helps project managers or team leaders to monitor the progress by using KRS tool. Storage of metrics can help designers to compare the metrics of some system with others. So, it can be used as a historical data by using KDB tool. MOOD model help to identify problems of the design by means of metrics that uses the OO concepts which allow software engineers to early access software design and yet improve it. MEMOOD model calculates understandability, modifiability, and maintainability of the design which are vital to know early in design phase. Without XMI, no UML diagram can be described. XML can be used as a bridge between tools or as intermediate data. Generics in C# (Lists) are really important due to their dynamic allocation. When EA does not support database or reports as add-in, integration must be used.

Future works can be summarized as the follows: Developing an add-in for ArgoUML and StarUML to calculate metrics since they also do not support metrics for the design. Evolving KDM tool to take not just XMI or XML as input but also the source code of Java, C# and C++.

### REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529-551, April 1955. (references)

[2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[1] Abreu, F.B., (1995), "Design Metrics for Object-Oriented Software Systems", ECOOP'95 Quantitative Methods Workshop in Aarhus, August

[2] Ahmed, S.H., Soliman, T.H.A., and Sewisy, A.A., (2013), "A Hybrid Metrics Suite for Evaluating Object-Oriented Design", *International Journal of Software Engineering*, Vol.6, Issue-1, pp: 65-82, ISSN: 16876954

[3] Al-Zobaidy, L. M. and Ibrahim, K. A., (2012), "Existing Object Oriented Design Metrics a Study and a Comparison", published in the 5th Scientific Conference in Information Technology (CCIT), Computer Science and Mathematics, Dec 19-20, University of Mosul, Iraq

[4] Calero, C., Genero, M., and Piattini, M., (2002), "Empirical Validation of Class Diagram Metrics", *Journal of Empirical Software Engineering*, Proceedings of the International Symposium on Empirical Software Engineering (ISESE'02), pp:195-203, ISBN: 07695179, IEEE

[5] Genero, M., Piattini, M., and Calero, C., (2002), "An Empirical Study to Validate Metrics for Class Diagrams", Department of Computer Science, University of Castilla-La Mancha, Spain

[6] Girgis, M.R., Mahmoud, T.M., and Nour, R.R., (2009), "UML Class Diagram Metrics Tool", *International Conference on Computer Engineering & Systems*, pp: 423-428, IEEE

[7] Hilera, J.R., and Fernández, L., (2012), "A Web Service for Calculating the Metrics of UML Class Diagrams", *Dobb's Special Report June*, www.drdbobs.com, The world of software development

[8] Hunter, D., Rafter, J., Fawcett, J., Vlist, E., Ayers, D., Duckett, J., Watt, A., and McKinnon, L., (2007), "Beginning XML", 4th Edition, Wiley Publishing, ISBN:978-0-470-11487-2

[9] IEEE Standard Glossary of Software Engineering Terminology, (1990), IEEE-Std 610.12, ISBN 1-55937-067

[10] Jassim, F., and Altaani, F., (2013), "Statistical Approach for Predicting Factors of Mood Method for Object Oriented", *Uiversity of Irbid, Irbid, Jordan*, arXiv preprint arXiv:1302.5454.

[11] Mago, J., and Kaur, P., (2012), "Analysis of Quality of the Design of the Object Oriented Software using Fuzzy Logic", *International Journal of Computer Applications (IJCA)*, Vol. iRAFIT Issue: 3, pp: 21-25

[12] Mehra, S., and Maini, R., (2011), "A METRIC FRAMEWORK FOR ANALYSIS OF OOD", *Journal of Global Research in Computer Science*, Vol. 2, No-7, pp: 67-70, ISSN :2229-371X

[13] Paterson, T., Russell, C., and Dewar, R., (2002), "Object-oriented software design metrics from XMI", School of Mathematical and Computer Science, Heriot-Watt University, Scotland

[14] Poornima, U.S., (2011), "A Quantitative Measure for Object Oriented Design Approach for Large-Scale Systems", *International Journal on Computer Science and Engineering (IJCSSE)*, Vol. 3 No-9, pp: 3237-3242, ISSN : 0975-3397

[15] Pressman, R., (2011), "Software Engineering: A Practitioner's Approach" 7th Edition, McGraw-Hill, ISBN :978-0-07-337597-7

- [16] Rani, T., Sanyal, M., and Garg, S., (2012) , "Measuring Software Design Class Metrics:- A Tool Approach" , International Journal of Engineering Research & Technology (IJERT), Vol.1,Issue-7, ISSN: 2278-0181
- [17] Reißing, R., (2001), "Assessing the Quality of Object-Oriented Designs", OOPSLA, Institute of Computer Science, University of Stuttgart , Germany
- [18] Rizvi, S.W.A., Khan, R.A., (2010), "Maintainability Estimation Model for Object-Oriented Software in Design Phase (MEMOOD)" Journal of Computing, Vol. 2, Issue-4, ISSN: 2151-9617
- [19] Sastry, J.S.V.R.S. , Ramesh, K.V., and Padmaja, M. , (2011), "Measuring Object-Oriented Systems Based On The Experimental Analysis Of The Complexity Metrics" , International Journal of Engineering Science and Technology (IJEST) , Vol. 3 , No. 5 , pp: 3726-3731, ISSN : 0975-5462
- [20] Schach, S.R., (2011) ,"Object-Oriented and Classical Software Engineering" , 8th edition, McGraw Hill , ISBN: 978-0-07-337618-9
- [21] Sharma, A., and Dubey, S.K., (2012), "Comparison of Software Quality Metrics for Oriented System", International Journal of Computer Science & Management Studies (IJCSMS), Vol. 12, June 2012 , ISSN:2231 –5268
- [22] Sharma, A.K., Kalia, A. , and Singh, H. , (2012) , " Metrics Identification for Measuring Object Oriented Software Quality" ,International Journal of Soft Computing and Engineering (IJSC), Vol. 2, Issue-5, November 2012 , ISSN: 2231-2307
- [23] Sommerville, I. , (2011) , "Software Engineering" , 9th Edition , Addison Wesley, ISBN-13: 978-0-13-703515-1
- [24] Sparx Systems, (2009) , "Enterprise Architect User Guide v7.5 " , Enterprise Architect Software , <http://www.sparxsystems.com/products/ea/7.5/>
- [25] Sparx Systems, (2013) , "Enterprise Architect 10 Reviewer’s Guide" , Sparx Systems Community , <http://www.sparxsystems.com/resources/whitepapers/>
- [26] Summers, B.L., (2011), " Software Engineering Reviews and Audits", CRC Press , ISBN:978-1-4398-5145-6
- [27] Web page ,Class diagram examples, 2012, <http://www.programsformca.com/2012/03/uml-diagrams-for-hospital-mgmt-system>
- [28] Web page, [http://www.sparxsystems.com/enterprise\\_architect\\_user\\_guide/9.3/index.html](http://www.sparxsystems.com/enterprise_architect_user_guide/9.3/index.html)
- [29] Web page, <http://www.sparxsystems.com/enterprise-architect/index.html>
- [30] Website, <http://www.sparxsystems.com/>
- [31] Al-Hadidi, K. A., Constructing a Tool for Measuring a Quality of Object Oriented Design to Enterprise Architect v7.5, MSc Thesis , Software Engineering, Mosul University, Collage of Computer sc. And Mathematic Mosul , Iraq.
- [32] Kan, S.H., (2002) , "Metrics and Models in Software Quality Engineering" , 2nd Edition , Addison Wesley , ISBN:0-201-72915-6
- [33] Bellekens, G. , (2011) , "Tutorial: Create your first C# Enterprise Architect add-in in 10 minutes" , Sparx System Community.
- [34] Al-Zobaidy, L. M. and Ibrahim, K. A. , (2012) , "Existing Object Oriented Design Metrics a Study and a Comparison" , published in the 5th Scientific Conference in Information Technology (CCIT), Computer Science and Mathematics ,Dec 19-20, University of Mosul , Iraq
- [35] Ibrahim, K. A. , (2013), " Constructing a Tool for Measuring a Quality of Object Oriented Design to Enterprise Architect v7.5" , Master Thesis, College of Computer Sciences and Mathematics ,University of Mosul

AUTHORS PROFILE

**First Author :Dr. laheeb M. Alzubaidy**, have BSc. In 1987, MSc. In 1992 And PhD in 2002, in computer Sc. From Dept. of computer Sc, university of Mosul, Iraq. Associative professor in 2003, Head of Dept of Computer Sc. In 2003, visiting lecturer in Isra private university in 2004, head of Dept of Software Engineerinh in 2007, Visiting lecturer in USM university , NAV6 center in 2009, interested research fields are in Artificial Intelligent technique, network security, image processing , pattern recognition, software eng.

# Usability Evaluation Methods and Principles for the Web

Joel Mvungi  
Computer Science Studies Department  
Mzumbe University  
Morogoro, Tanzania

Titus Tossy  
Computer Science Studies Department  
Mzumbe University  
Morogoro, Tanzania

**Abstract**-In order to determine the quality of any web application in the world, Usability is the one of the most important tool that one can use. Web analysis perform several inspections on the websites and software and use usability criteria to determine some faults on the systems. Usability engineering has being important tool for the companies as well, this is due to the fact that through usability engineering companies can improve their market level by making their products and services more accessible. Know days there some web application and software products which are complex and very sophisticated, hence usability can be able to determine their success or failure. However currently usability has been among the important goal for the Web engineering research and much attention is given to usability by the industry due to recognition of the importance of adopting usability evolution methods before and after deployment. Moreover several literature has proposed several techniques and methods for evaluating web usability. And however there is no agreement yet in the software on which usability evolution method is better than the other. Extensive usability evaluation is usually not feasible for the case of web development process. In other words unusable website increases the total cost of ownership, and therefore this paper introduces principles and evaluation methods to be used during the whole application lifecycle, so as to enhance usability of web applications.

**Keywords**-*Evolution methods, Web usability, Web usability principles, Development process.*

## I. INTRODUCTION

Web development is a complex and challenging process that has to deal with number of heterogeneous interacting components(Murugesan,2008).However the construction of Web applications has changed in some discipline, but there still lack of proper engineering approach for developing web systems, and the whole development process is still unengineered(Ahmad et al.,2005).Due to the challenges emerge in developing of more usable Web applications, this has led to the rise of several techniques, methods, and tools of which address usability issues. However much knowledge exist on how to develop usable web applications, but many of the applications still don't meet most of customer usability expectation (Offutt [29]).On top of that many company know days have decline as result of not taking to account web usability issues (Becker and Mottay [5]).Hence therefore there

is need of identifying those usability evaluation methods (UEMs) which have been successfully applied to the Web development.

However web-based applications have influenced several domains, which provide access to information and services by variety of users showing different characteristics and backgrounds. Most user visits websites, and also return back to the previously accessed sites, if they some easily useful information, organized in a way that facilitates access and navigation and presented according to a well-structured layout. Therefore we can say the acceptability of Web applications by users strictly rely on their usability.

Most of literature has reported that most work on web applications has been done, on making them more powerful but relatively little has been done to ensure the quality of those applications. Some of important factors for quality of web application are reliability, availability, usability and security. It's estimated that 90% of web sites provide inadequate usability. The ISO/ISEC 9126-1 standards mentions six principle categories of quality characteristics. Which are functionality, reliability, usability, efficiency, maintainability and portability.

Therefore we can say Web usability is a core component of web quality. Without a good usability features, the web quality will always be a question mark.

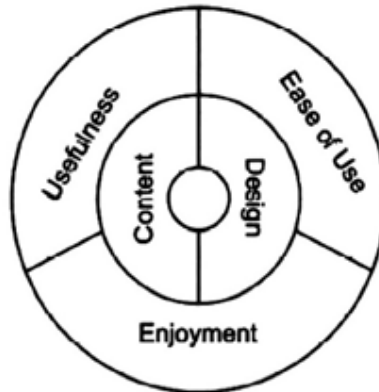
## II.RELATED WORK ON UIMS FOR WEB APPLICATIONS

Bray introduced an attempt trying to measure the web in 1996, he tried to answer the question such as the size of the web, its connectivity and visibility of its sites. Moreover Stolz et al. (2005) came with new technique to access the success of information driven websites that merged user behavior, content of site and structure while utilizing user feedback.

Dominic and Jati(2010) evaluated the usability and quality of Malaysian University websites based on factors like load time, frequency of updates, accessibility errors, and broken links using the following tools: Website optimization, Check link validator,HTML validator and accessibility testing software. Moreover from Treiblmaier and Pinterest's (2010) point of view, you can describe a website based on two main criteria: "What is presented?" and "How is it presented".

All academic efforts for developing UIMs for Web applications there is still room for improvement. (Rivero and Conte, 2012a) identified that emerging of UIMs for Web applications should be able to: Find usability problems in initial stage of its development, Aid in both identification and solution of usability problems. There is an important shortage of standard criteria for comparison, therefore UEMs cannot be evaluated. Several studies have been done to see which measure has been so common and majority of the study used thoroughness measure (ratio between the number of real usability problems found and the total real usability problems).

However Palmer (2002) emphasized on the importance of metrics in helping organization generate more effective and successful websites. Another survey by Hong (2007) on Korea organizations also found that a key enabler of website success measurement is website metrics. These metrics play two important roles: Determining if website perform to the expectation of the users and the business running the site, and also identify website design problems.



Source: Treiblmaier and Pinterest, 2010

### III. DEFINING THE TERM USABILITY

Usability is the term that is generally described as factor of system quality, it defines the quality of systems and products from human point of view who use the systems (Andrian and Emilio, 2003). However the term usability was derived from the term 'user friendly'. However the concept of usability is somehow complex to define this is due to the fact that it is used in many different context such as performance, execution time, and user satisfaction as well as easy of learning. The concept is also applied in areas like consumer electronic products and communication. Also may refer to efficient design of mechanical object such as door locks. In other words usability means those people whom use products such as software application, can learn it quickly and use it easily to accomplish their tasks (Azeem and Kamran, 2008). Usability enables employee to concentrate more on their work rather than on the tool the use to perform the tasks.

A usable product may refer to a product which:

- Is efficient to use
- Is easy to learn
- Provide quick recovery from errors
- Is easy to remember
- Is visually pleasing
- Is enjoyable to use

Moreover there several definition of usability which vary according to the model they are based on ISO standards defined usability as "the extent to which a product can be used

by specified users to achieve specified goals within effectiveness, efficiency and satisfaction in specify context of use". Where effectiveness means accuracy and completeness with which user archive specified goals, efficiency means the resource expended in relation to the accuracy and completeness with which user achieve goals and satisfaction is described as the comfort and acceptability of use. Where by usability problem refer to the aspect that make application ineffective, inefficient and difficult to learn and use.

Nielsen's defined usability as Learnability: the ease of learning the functionality and behavior of the system, Efficiency: the level of attainable productivity once the user has learne. Memmoability: the ease of remembering the system functionality so that the casual user can return to the system after a period of non-use without needing to learn again to use it. Few errors: the capability of the system to feature a low error rate, to support users making few errors during the use of the system and in case they make errors to help them to easy recover. User's satisfaction: the measure in which the user finds system pleasant to use.

### IV. WEB USABILITY CRITERIA

General usability principles are achieved through usability criteria (Tayana and Jobson, 2010), the criteria provide guidelines to designers in restricting the space for design alternatives hence prevent designers from developing products that are not usable.

There are three important dimensions that any web developer has to focus on i.e. hypertext, data and presentation design each dimension consists of number of criteria this part there will be explanations on the mentioned dimensions which represent great impact on usability of any Web application.

The criteria could be discussed as follows:

#### *A. Content Visibility*

Refer to the understanding of information structure offered by the application, and get oriented with the hypertext, user must be able to identify main conceptual classes of the contents of the application.

#### **Concepts of identification of Core Information**

The content visibility can be supported by content design, where by main classes of content are identified and hence structured (Azeem and Kamran, 2000). The identification of information entity modeling concept could provide a way in full filling the requirement.

The content will help in centering Data design, and gradually evolve by detailing their structure in terms of elementary components, and adding further auxiliary contents for accessing and browsing them.

#### **Hypertext Modularity**

The design of hypertext must be able to support users to perceive where core concepts are located, therefore:

- The hypertext can be organized on areas i.e. modularization constructs, where you group pages with homogeneous contents.
- However areas must be defined as global landmarks accessible through links grouped in global navigation bars that are displayed in any page of the application interface.
- For each area, the most representative pages can be defined as local landmarks, reachable through local navigation bars displayed in any page within an area.

However learnability and memorability could be enhanced by the use of hierarchical landmarks within pages. Landmarks indeed provide intuitive mechanisms for highlighting the available contents and the location within the hypertext where are placed.

#### *B. Ease Of Content Access*

After users have identified main classes of content the application deals with, they have to be provided with facilities for accessing the specific content items they are interested in.





Figure 2. User login to the system



Figure 3. User home page

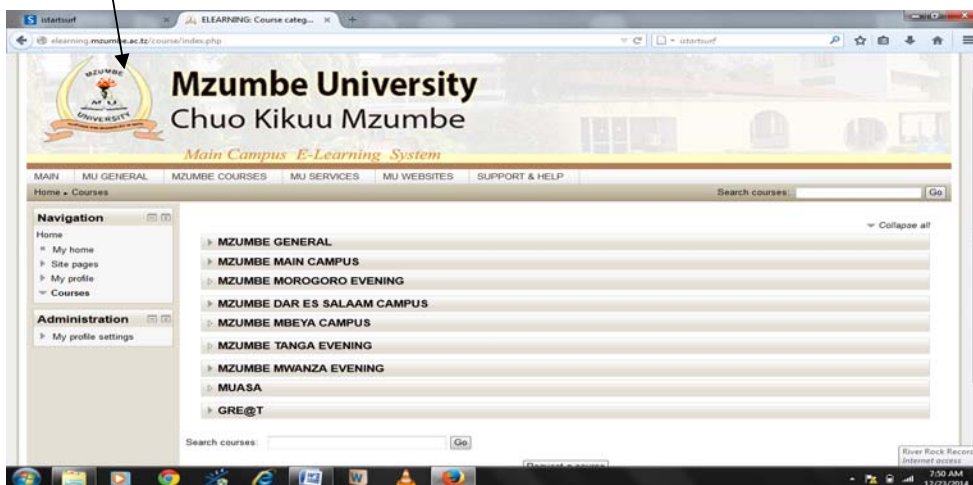


Figure 4. User specifics page

Source: [www.elearning.mzumbe.ac.tz](http://www.elearning.mzumbe.ac.tz)

## Identification of Access Information Concepts

As discussed the design of access paths for retrieving core content items can be facilitated if designers augment the application contents with access concepts corresponding to classification criteria or contexts over core concepts, enabling users to progressively move from broader to narrower categories, until they locate the specific core concept of interest.

## Navigation Access and Search-Based Access

In order to facilitate the access to specific instance of core concepts, access concepts defined at data level should be used to construct navigational access mechanism that typically consist of multi-level indexes (Alan and Gregory, 2004), which possibly distributed on several pages, bridging of pages with high visibility, such as the Home Page or the entry page of each area, to pages devoted to the publication of core concepts.

Moreover navigational access is very often complemented with direct access, especially in large Web applications i.e. keyword-based search mechanisms, which enable bypass navigation and rapidly reach the desired information object. Also direct access mechanism are essential in interfaces (such as those of mobile devices) that are unable to support multiple navigation steps. In traditional hypertext interfaces they enhance orientation when users get lost by moving along navigational access mechanisms.

### C. Ease of Content Browsing

Usually the auxiliary contents related to each single core concept must be easily identified by users, as well as the available interconnections among different core concepts.

## Core Concepts Structuring and Interconnection

The user understanding of content structuring and of the semantic interconnection defined among different content classes, enhance the ease of use and learnability of the web application (Luis and Tavana, 2013). And therefore when the identified core concepts represent a structured and complex concept, it is recommended to expand them via top-down design into a composite data structure.

Moreover the semantic interconnection among core concepts must be established for reproducing a knowledge network through which users can easily move for exploring the information contents.

## V. EVALUATION METHODS

Evaluation methods are mainly aimed on assess the application functionality, to verify the effect of its interface on user, also to identify any specific problem with the application such as aspects which show unexpected effects when used in their intended context (Azeem and Kamran, 2008). Also evaluating Web applications in particular consists of verifying if the application design allows users to easily retrieve and browse content, and invoke available service and application they need. This therefore implies not only having appropriate contents and service available into the application but also making them easily reachable by users through appropriate hypertexts.

However the development of a Web system is a continuous process with an interactive life cycle of analysis. Design, implementation and testing (Murugesan 2008). However what we need really is a different focus on evaluation methods and a new categorization system according to the purpose and platforms as Web and Website evaluation methods according to Stolz et al. & Hasans work.

- Website evaluation methods (WSEMs) could be:
  - i) User-based usability evaluation methods
  - ii) Evaluator-based usability evaluation methods
  - iii) Automatic website evaluation tools ie Bobby, Lift
- Web evaluation methods (WEMs) could be:
  - i) Web analytics tools ie Google analytics
  - ii) Link analysis methods ie Page Rank

### A. Website Evaluation (WSEMs)

Limited number of website can be measured manually or automatically by the WEMs measure, based on some criteria so as to achieve quality website. However the manual testing can be done as well but output of such evaluation consists of list of usability problems and recommendation to improve the tested website. Some of other evaluation methods are:

#### 1) User-based Usability Evaluation Methods

Process of design for usability, user testing and redesign is called User centered Design (Former and Bosch, 2004; Nielsen, 1993). The term "usability evaluation" refers to the entire test, planning and conducting the evaluation and presenting the results. However the main goal of usability evaluation is to measure the usability of the system and identify usability problems that can lead to user confusion, errors or dissatisfaction (Larusdottir, 2009). The user evaluation approach consists of set of methods that employs representative user to execute some tasks on specific system. The user performance and satisfaction with the interface are then recorded. And the most useful method in this category is user testing.

**User Testing**, when users use the system they normally work towards accomplishing specific goals in their minds (Stone et al., 2005). A goal is an abstract end result indicating what is to be achieved, and can be attained in numerous ways. Also each goal breaks into a task specifying what a person has to do, and then each task decomposes into an individual step that needs to be undertaken. User should be able to do basic tasks correctly and quickly. In order to select tasks the examiner begins by exploring all the tasks within the website then narrowing them down to those that are the most important to users. Moreover a good task is the one that discovers a usability problem or one that reveals and error that is difficult to recover from. Next step is how to present selected task to the participants and one way to do this is by using a "scenario" in which the task is embedded in a realistic story.

However it is important to test users individually and let them solve the problems on their own. Actually the purpose of usability study is to test the system and not the users, and this

aspect must be explicitly explained to tested users (Nielsen, 1993; Stone et al., 2005). Metrics can be collected from user testing; time for users to learn a specific function speed of task performance, type and rate of users' errors, user retention of commands over time and user satisfaction.

### 2) Evaluator-based Usability Evaluation Methods

Evaluator or Experts inspects the interface and assess system usability using interface guidelines, design standards, users' tasks, or their own knowledge, depending on the method to find possible user problems (Larusdottir, 2009). Moreover the inspectors can be usability specialists or designers and engineers with special expertise (Matera et al., 2006). In this category, there are many inspection methods such as cognitive walkthrough, guideline reviews, standard inspection and heuristic evaluation (Hasan, 2009).

However heuristic Evaluation is the most efficient usability method, because it is a special variable when time and resource are scarce. There is a number of evaluators who assess the application and judge whether it conforms to a list of usability principles, namely 'heuristics' (Hasan, 2009). During the heuristic evaluation process each evaluator goes individually into a system interface at least twice, and the output obtained from such evaluation is a list of usability problems with reference to the list of violated heuristics. However by principle heuristic evaluation can be conducted by only one evaluator, whom can find about 35% of total usability problems (Nielsen, 1993). But Matera et al. (2009) believed that better results are obtained by having five evaluators and certainly not fewer than three for reasonable results.

### 3) Automatic Website Evaluation Tools

Automatic tools are software that automates the collection of interface usage data and identify potential Web problems. First study was conducted by Ivory and Chevalier (2002), who concluded that more research was needed to validate the embedded guidelines and to make the tool usable. And therefore Web professionals cannot rely on them alone to improve websites. Brajnik (2004b) mentioned several kinds of Web-testing tools that can be used: accessibility tools such as Bobby, usability tools such as LIFT, performance tools such as TOPAZ, Security tools such as Web CPO, and classifying website tools such as Web Tango. He also stated that the adoption of tools is still limited due to the absence of established methods for comparing them and also suggested that the effectiveness of automatic tools be itself evaluated.

### B. Web Evaluation Methods (WEMs)

The method studies the web as a whole by calculating statistics about the detail use of the site and providing Web-traffic data, visibility, connectivity, ranking and overall impact of a site on the Web.

#### 1) Web Analytics tools

Web analytics have been defined by the Web Analytics Association as "the measurement, collection, analysis and reporting of Internet data for the purpose of understanding and reporting Web usage" (Fang, 2007). However these tools automatically calculate statistics about the detail use of site

helping. By origin, Web analytics is a business tool that started with some webmasters inserting counters on their home pages to monitor Web traffic. However most Web analytics studies target e-commerce, the method can be applied to any website (Prom, 2007). The two well-known Web analytics tools are Google Analytics and Alexa.

**Google Analytics**, Google purchased a Web analytics company called Urchin software in 2005 and released Google Analytics to the public in 2006 (Fang, 2007; Hasan et al., 2009). However the service is free for up to five million page views per month per account. Once signed for Google Analytics, Google offers users code that must be inserted into each web page to be tracked. Visual data results are displayed with a wealth of information as to where visitors come from, what pages they visited, how long they stayed on each page, how deep into a site they navigated, etc. (Fang, 2007).

**Alexa**, refers to a website metrics system owned by the Amazon Company that provides a downloadable toolbar for Internet Explorer users. Calculates traffic rank by analyzing the Web usage of Alexa toolbar users for three months or more as a combined measure of page views and reach (the number of visitors to the site). Although this information is useful, Alexa ranking is biased towards MS Windows and Internet Explorer users (Scowen, 2007).

## VI. SELECTION OF APPROPRIATE EVALUATION METHOD(S)

The evaluation of Indian Banking website navigability performed by Kaur and Dani found that Alexa and Google PageRank do not have significant correlations with navigability metrics, indicating that popularity and importance are not good indicators of website navigability; therefore the traffic data the back-links of websites are not meaningful measures of site navigation assessment. Moreover Cho and Adams (2005) added that PageRank is not a metric of page quality; Further, Hong (2007) stated that most organizations use Web metrics to determine site traffic or popular content but seldom used them to improve navigation. Jalal et al. (2010) and Noruzi (2006) concluded that the Webometric method is an imperfect tool to measure the quality of website and that it reflects unreliable results in most cases.

The findings of these five studies hence support the argument that WEMs, such as the Web analytics tools and the link analysis methods, do not discover navigation problems accurately nor do they measure website quality. Further, it seems that WEMs are complementary approaches since they do not definitely discover usability problems of a site, rather they just indicate their probability. However Link analysis method does not discover navigation problems accurately nor do they measure website quality. It seems that WEMs are complementary approaches since they do not definitely discover usability problems of a site, rather they indicate their probability.

In other words even though usability testing demonstrates how real users interact with a website and the exact problem they face, it's not enough to measure the success of the site or describe the interaction of a large number of users with it (Hasan, 2009). Therefore this highlights the weakness that WEMs such as a user evaluator or automatic evaluation methods

cannot provide traffic data, Web ranking of site or its online visibility among others.

Therefore, the choice of appropriate evaluation method depend greatly on the purpose of the evaluation itself. If the intention is to redesign the website and wanted to discover most of its potential usability problems, then the best evaluation methods are user testing and expert evaluation, while an automatic tool or Google analytics is useful complement in this situation. If the goal of evaluation is to redesign a website then WEM is the best approach, while WEMs are not useful enough in this circumstance.

## VII.CONCLUSION

In conclusion, in order to address the challenges of developing complex Web systems, “Web engineering” is an emerging discipline for the implementation of engineering principles to promote high quality websites that attract visitors (Andrina and Viado, 2000). Web measuring has become a valuable area of ongoing research, but unfortunately the field is not yet mature; Web evaluation method are so many on literatures but they lack studies that classify, compare and determine the appropriate evaluation methods.

However some previous studies confused the term “Web evaluation method” with “Website evaluation methods” since they did not distinguish between diverse platforms of assessment methods and also did not address the purposes behind such evaluation. For example some of the study evaluate the web in terms of the ranking and connectivity of the sites, others assess specific websites to discover there usability problems.

Lastly the purpose of Web evaluation is to determine the appropriate methods to be used. If the purpose is to redesign the website, then the scope of evaluation is WSEM, and therefore as stated by the literature the best evaluation methods are user testing and expert evaluation, while automatic and Web analytics tools (complementary) could

provide a first insight into the status of the website. Similarly, if Web ranking and traffic statistics are of interest, then the scope of evaluation is WEMs; thus the best way is to use a Web analytics tool such as Alexa.

## REFERENCES

- [1]Luis Rivero\* and Tavana Conte (2013): Using an empirical study to evaluate the feasibility of a new usability inspection technique for paper based prototypes of web applications: Journal of Software Engineering research and development.
- [2]Adrian Fernandez\*, Emilio Insfrana, Silvia Abrahao: Usability Evaluation Methods for the Web: ISSI Research Group, Department of Information System and Computation, Universidad Politecnica de Valencia, Camino de Vera, s/n, 4022, Valencia, and Spain.
- [3]Dalal Ibrahim Zahran, Hana Abdullah Al-Nuaim, and Malcom John Rutter: A comparative approach to web evaluation and website evaluation methods: International Journal of Public Information Systems, vol 2014:1.
- [4]Azeem Umar, Kamran Khan Tatari: Appropriate Web Usability Evaluation Method during Product Development: Master Thesis Software Engineering Thesis no: MSE-2008-03 Feb 2008.
- [5]Alan Dix Janet Finlay, Gregory D Abowd Russel Beale (2004): Human-computer-Interaction Third
- [6]Tayana Contel, Jobson Massollar1, Emilia Mendes2, Guilherme H. Travassosl (2010): Usability Evaluation Based on Web Design Perspectives, First International Symposium on Empirical Software Engineering and Measurement.
- [7]Andrina Grank,Viado Glavink and Slavomir stankov:Usability Evaluation Methodology for web-based Educational Systems.Faculty of natural science.mathematics and Education,University Split.

## 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  
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain  
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India  
Dr Li Fang, Nanyang Technological University, Singapore  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia  
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India  
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand  
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India  
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia  
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India  
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India  
Dr. P. Vasant, University Technology Petronas, Malaysia  
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea  
Dr. Praveen Ranjan Srivastava, BITS PILANI, India  
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong  
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia  
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan  
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria  
Dr. 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  
Dr. Tirthankar Gayen, IIT Kharagpur, India  
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan

Prof. Ning Xu, Wuhan University of Technology, China  
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen  
& Universiti Teknologi Malaysia, Malaysia.  
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India  
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan  
Prof. Syed S. Rizvi, University of Bridgeport, USA  
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan  
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India  
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal  
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P  
Dr. Poonam Garg, Institute of Management Technology, India  
Dr. S. Mehta, Inha University, Korea  
Dr. Dilip Kumar S.M, 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  
Dr. Saqib Saeed, University of Siegen, Germany  
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]  
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt  
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India  
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India  
Dr. Muhammad Sohail, KUST, Pakistan  
Dr. Manjaiah D.H, Mangalore University, India  
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India  
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada  
Dr. Deepak Laxmi Narasimha, Faculty of Computer Science and Information Technology, University of  
Malaya, Malaysia  
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India  
Dr. M. Azath, Anna University, India  
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh  
Dr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia  
Dr Suresh Jain, Professor (on leave), Institute of Engineering & Technology, Devi Ahilya University, Indore  
(MP) India,  
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia  
Dr. Hanumanthappa. J. University of Mysore, India  
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)  
Dr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria  
Dr. Santosh K. Pandey, 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 Munawwarrah, 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, SICSAR, 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, Lakshmandarh, 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 Singhanian 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  
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India  
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India  
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India  
Assist. Prof M. Anand Kumar, Karpagam University, Coimbatore, India  
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex  
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India  
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India  
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat  
Mr. Sivakumar, Codework solutions, India  
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran  
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA  
Mr. Varadala Sridhar, Varadhman College Engineering College, Affiliated To JNTU, Hyderabad  
Assist. Prof. Manoj Dhawan, SVITS, Indore  
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India  
Dr. S. Santhi, SCSVMV University, India  
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran  
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh  
Mr. Sandeep Reddivari, Mississippi State University, USA  
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal  
Dr. Hazra Imran, Athabasca University, Canada  
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India  
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India  
Ms. Jaspreet Kaur, Distance Education LPU, India  
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman  
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India  
Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India  
Mr. Khaldi Amine, Badji Mokhtar University, Algeria  
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran  
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India  
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India  
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia  
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India  
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India  
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India  
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany

Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India  
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India  
Dr. Nadir Bouchama, CERIST Research Center, Algeria  
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India  
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco  
Dr. S. Malathi, Panimalar Engineering College, Chennai, India  
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India  
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India  
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan  
Dr. G. Rasitha Banu, Vel's University, Chennai  
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai  
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India  
Ms. U. Sinthuja, PSG college of arts & science, India  
Dr. Ehsan Saradar Torshizi, Urmia University, Iran  
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India  
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India  
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim  
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt  
Dr. Nishant Gupta, University of Jammu, India  
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India  
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India  
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus  
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Dr. Rahul Malik, Cisco Systems, USA  
Dr. S. C. Lingareddy, ALPHA College of Engineering, India  
Assistant Prof. Mohammed Shuaib, Interl University, Lucknow, India  
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India  
Dr. T. Thambidurai, Sun Univercell, Singapore  
Prof. Anandkumar Telang, BKIT, India  
Assistant Prof. R. Poorvadevi, SCSVMV University, India  
Dr Uttam Mande, Gitam University, India  
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India  
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India  
Dr. Mohammed Zuber, AISECT University, India  
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia  
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India  
Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India  
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India  
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq  
Dr. Urmila Shrawankar, G H Raison College of Engineering, Nagpur (MS), India  
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India

Dr. Mukesh Negi, Tech Mahindra, India  
Dr. Anuj Kumar Singh, Amity University Gurgaon, India  
Dr. Babar Shah, Gyeongsang National University, South Korea  
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India  
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India  
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India  
Assistant Prof. Ankit Garg, Amity University, Haryana, India  
Assistant Prof. Rajashe Karappa, SDMCET, Karnataka, India  
Assistant Prof. Varun Jasuja, GNIT, India  
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India  
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India  
Dr. Faouzi Hidoussi, UHL Batna, Algeria  
Dr. Naseer Ali Husieen, Wasit University, Iraq  
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai  
Dr. Ahmed Farouk Metwaly, K L University  
Mr. Mohammed Noaman Murad, Cihan University, Iraq  
Dr. Suxing Liu, Arkansas State University, USA  
Dr. M. Gomathi, Velalar College of Engineering and Technology, India  
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia  
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India  
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India  
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran  
Dr. Thiyagu Nagaraj, University-INOUE, India  
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe  
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India  
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India  
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India  
Dr. Shenshen Liang, University of California, Santa Cruz, US  
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia  
Mr. Snehasis Banerjee, Tata Consultancy Services, India  
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania  
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia  
Dr. Ying Yang, Computer Science Department, Yale University, USA  
Dr. Vinay Shukla, Institute Of Technology & Management, India  
Dr. Liviu Octavian Maftciu-Scai, West University of Timisoara, Romania  
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq  
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India  
Dr. Timothy Powers, University of Hertfordshire, UK  
Dr. S. Prasath, Bharathiar University, Erode, India  
Dr. Ritu Shrivastava, SIRTS Bhopal, India  
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India  
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania



Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India  
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India  
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India  
Assistant Prof. Mallikarjun C Sarsamba Bheemna Khandre Institute Technology, Bhalki, India  
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India  
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India  
Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq  
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco  
Dr. Parul Verma, Amity University, India  
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco  
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India  
Assistant Prof.. G. Selvavinayagam, SNS College of Technology, Coimbatore, India  
Assistant Prof. Madhavi Dhingra, Amity University, MP, India  
Professor Kartheesan Log, Anna University, Chennai  
Professor Vasudeva Acharya, Shri Madhwa vadiraja Institute of Technology, India  
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia  
Assistant Prof., Mahendra Singh Meena, Amity University Haryana  
Assistant Professor Manjeet Kaur, Amity University Haryana  
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt  
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia  
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India  
Assistant Prof. Dharmendra Choudhary, Tripura University, India  
Assistant Prof. Deepika Vodnala, SR Engineering College, India  
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA  
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India  
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan  
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India  
Assistant Prof. Chirag Modi, NIT Goa  
Dr. R. Ramkumar, Nandha Arts And Science College, India  
Dr. Priyadharshini Vydhialingam, Harathiar University, India  
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka  
Dr. Vikas Thada, AMITY University, Pachgaon  
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore  
Dr. Shaheera Rashwan, Informatics Research Institute  
Dr. S. Preetha Gunasekar, Bharathiyar University, India  
Asst Professor Sameer Dev Sharma, Uttaranchal University, Dehradun  
Dr. S. Preetha Gunasekar, Bharathiyar University, India  
Dr. Zhihan Lv, Chinese Academy of Science, China

# CALL FOR PAPERS

## International Journal of Computer Science and Information Security

IJCSIS 2015

ISSN: 1947-5500

<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

### *Track A: Security*

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity  
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

### ***Track B: Computer Science***

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com). Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



**© IJCSIS PUBLICATION 2015**

**ISSN 1947 5500**

**<http://sites.google.com/site/ijcsis/>**