

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

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

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

*Indexed and technically co-sponsored by :*



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

## CALL FOR PAPERS

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

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 Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

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

**Deadline:** see web site

**Notification:** see web site

**Revision:** see web site

**Publication:** see web site

Context-aware systems

Networking technologies

Security in network, systems, and applications

Evolutionary computation

Industrial systems

Evolutionary computation

Autonomic and autonomous systems

Bio-technologies

Knowledge data systems

Mobile and distance education

Intelligent techniques, logics and systems

Knowledge processing

Information technologies

Internet and web technologies

Digital information processing

Cognitive science and knowledge

Agent-based systems

Mobility and multimedia systems

Systems performance

Networking and telecommunications

Software development and deployment

Knowledge virtualization

Systems and networks on the chip

Knowledge for global defense

Information Systems [IS]

IPv6 Today - Technology and deployment

Modeling

Software Engineering

Optimization

Complexity

Natural Language Processing

Speech Synthesis

Data Mining

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

arXiv.org    Google scholar

SCIRUS  
search engine for science

ScientificCommons

Scribd

.docstoc  
find and share professional documents

BASE  
Bielefeld Academic Search Engine

CiteSeerX beta

uni-trier.de  
Computer Science Bibliography

DOAJ  
DIRECTORY OF OPEN ACCESS JOURNALS

EBSCO HOST

ProQuest

## Editorial Message from Editorial Board

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

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

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

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

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

*ISSN 1947-5500 © IJCSIS, USA.*

*Journal Indexed by (among others):*



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



**Bibliographic Information**

ISSN: 1947-5500

Monthly publication (Regular Special Issues)

Commenced Publication since May 2009

**Editorial / Paper Submissions:**

**IJCSIS Managing Editor**

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

**Pennsylvania, USA**

**Tel: +1 412 390 5159**

# IJCSIS EDITORIAL BOARD

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

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

ISSN 1947 5500 Copyright © IJCSIS, USA.

## TABLE OF CONTENTS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Dr. Surjeet Dalal, Shilpa Kukreja*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Thanh Tung Khuat, My Hanh Le*

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

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

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

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

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

*Bhawana Mathur, Manju Kaushik*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Kesava Rao Alla, Soong Der Chen*

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

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

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

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

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

*Syed Muhammad Sajjad, Muhammad Yousaf*

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

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

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

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

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

*Jasdeep Singh Chauhan, Amanpal Singh Rayat*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*H. Hossein zadeh, E. Salehpour*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Walid Ellili, Mounir SAMET, Abdennaceur KACHOURI*

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

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

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

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

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

*Muhammad Asim, Waseem Iqbal*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Shohel Sayeed*

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

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

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

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

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

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

*Rohini, Dr. Indu Chhabra*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Saurabh Gupta, Manish Pandey*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Gamal H. Eladl, PhD*

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

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

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

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

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

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

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

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

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

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

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

*Khalaf F. Khatatneh, Samer S. Khanfar*

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

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

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

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

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

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

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

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

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

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

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

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

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

*Ruchi Lawaniya, Manish Pandey*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Keywords:* Inverse problem, Jacobi matrix, Eigenpair

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

*Santosh Kumar, Manish Pandey*

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

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

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

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

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

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

*Mahsa Roghanian, Mohammad Naderi Dehkordi*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

# A Fuzzy Based Approach for the Elicitation of Attributed Values of Goal Models

**<sup>1</sup>Sameena Naaz**

**Farhana Mariyam**

Department of Computer Science and Engineering, Faculty of Engineering and Technology  
Jamia Hamdard, New Delhi, India

<sup>1</sup>Corresponding Author

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

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

## I. INTRODUCTION

Requirements elicitation is the first sub-process of requirements engineering (RE) and it consists of group decision making approaches for the selection and prioritization of requirements [1, 2]. They visualize RE as a process that includes five sub processes namely requirements elicitation, requirements analysis, requirements modeling, requirements verification and validation, and requirements management [3]. Out of these sub-processes requirement elicitation has one of the major roles, as it has cascading effect on other sub-processes. Requirements elicitation means to identify the requirements of a product according to the need of the customers and users. Various methods are there for requirements elicitation like goal oriented requirements elicitation process [4, 5, 6, 7, 8], quality function deployment (QFD) [9, 10] etc. In goal oriented requirements elicitation process, the main objectives represented as high level are broken down into sub-goals at lower levels to construct an AND/OR graph. This graph is used to find out the non-functional and functional requirements of software.

Goals are high level objective of an organization and are the basis of goal oriented requirements engineering (GORE) [11]. In requirements elicitation process, a group of stakeholders together identify the requirements on the basis of domain knowledge product vision and business objective [3, 12, 14]. In an AND/OR graph, there are two types of decomposition of a goal into sub-goals that use AND and OR logical connectives. An AND decomposition means that a parent goal can be achieved/satisfied if all its sub-goals are achieved/satisfied. In the same manner, an OR decomposition achieves a parent goal if it's any sub-goal is achieved [15].

There are many fundamental challenges in requirement elicitation like producing well formulated requirements, managing the constant flow of requirements, requirements traceability, requirements volatility and interdependence, selecting the right process, and software release planning [16] pointed out that among various RE challenges; software release planning is a major challenge in RE. Requirement elicitation looks into the selection and prioritization of goals for the next release of software product according to the requirement of stakeholders. Software companies are facing problems in order to determine the set of goals for the next release of software product. A system may have thousands of goal/requirements proposed by stakeholders; and it is not possible to select these goals/requirements that add most value to the software product [17].

Since different stakeholders take part in requirements elicitation process, therefore, they have different preferences for the same goal because of their experience, understanding levels, and knowledge [18, 19]. Key elements in human thinking is not numbers, but linguistic terms or fuzzy set levels [20]. The values given by these stakeholders to a goal are often qualitatively described or imprecisely measured because human judgments including preferences are often vague and cannot estimate his preference with exact numerical values [21, 22]. Fuzzy logic can handle imprecise and linguistic variables in a well-defined mathematical way, for example, expressions such as: "high performance" and low cost are imprecise [19, 20, 23].

In literature, fuzzy based decision making methods have gained acceptance over the last decade in different areas of science and engineering, for example, technology transfer selection in biotechnology [9, 24], supplier selection problem [19], selection of object oriented simulation software for production system analysis [24, 25], and a non-functional requirements tradeoffs model in trustworthy software [27], software component selection [26] etc. In [1] a fuzzy based method for requirements prioritization in goal oriented requirements elicitation process has been presented. The contribution of this paper is to extend the attributed goal oriented requirements analysis (AGORA) method by applying the binary sort tree method and the fuzzy based approach for the selection and prioritization of goals in goal oriented requirements elicitation process. In this paper the proposed method has been demonstrated considering the ten requirements [1].

In software industry, requirements related problems have contributed to challenged and failed software projects. The reason for failure of these software projects is that requirements knowledge is distributed across myriads of stakeholders. Based on the review of goal oriented requirements elicitation process, it has been identified that Attributed Goal Oriented Requirements Elicitation (AGORA) method do not support how to elicit the attributed values, i.e., contribution values and preference matrices, of goals. Therefore, the objective of this paper is to propose a fuzzy based method for the elicitation of attributed values of goal models.

## II. BACKGROUND AND RELATED WORK

Several methods have been proposed in literature to select and prioritize the goals using goal models. For example, in 2015 Sadiq and Jain proposed a method for the prioritization of requirements and goals using fuzzy preference relation. In 2015 Subramanian et al. proposed a method for the quantitative reasoning of goal satisfaction in goal models. Apart

from these issues several other studies have been published. For example, in 2014 Sadiq and Jain proposed a method for the identification of stakeholders in goal oriented requirements elicitation process [1]. In 2002, Kaiya et al. [5] proposed an Attributed Goal Oriented Requirements Analysis (AGORA) for the analysis of goal. One of the limitations of AGORA is that it does not support how to select and prioritize the goal when several stakeholders participate during the requirements elicitation process.

### III. PROPOSED METHOD FOR THE ELICITATION OF ATTRIBUTED VALUES OF GOAL MODELS

In this part the method for the elicitation of attributed values of goal models has been presented. Proposed method includes the following steps:

Step 1: Identification of stakeholders

Step 2: Elicitation of attributed values

Step 3: Collection of decision maker's fuzzy assessment

Step 4: Computation of attributed values

#### *Step 1: Identifying Stakeholders:*

For prioritization of any requirement it is very important to identify its stakeholders and their goals. They may be any organization or any person who is actively involved in a project, or whose interests may be affected positively or negatively by execution of a project. Stakeholders can be internal to the organization or external. A Questionnaire may be used to analyze specific stakeholder's expectations, influences and needs.

#### *Step 2: Elicitation of Attributed Values:*

In this step the linguistic terms have been defined as triangular fuzzy numbers by means of the following compatibility functions:

$$F(\text{very good}) = (0.75, 1.0, 1.0)$$

$$F(\text{good}) = (0.5, 0.75, 1.0)$$

$$F(\text{fair}) = (0.25, 0.5, 0.75)$$

$$F(\text{poor}) = (0, 0.25, 0.5)$$

$F(\text{very poor}) = (0, 0, 0.25)$ , which are shown in Figure 1.

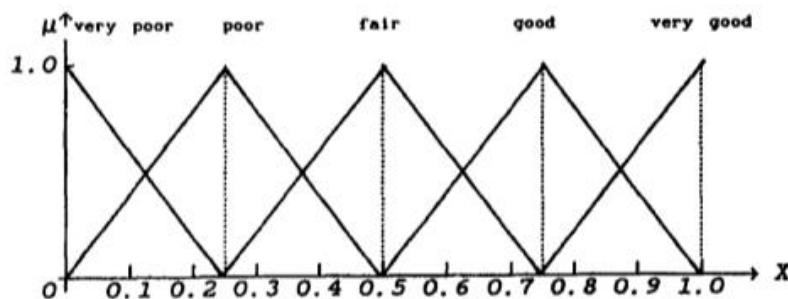


Figure 1: Triangular Fuzzy numbers for linguistic terms.

*Step 3: Collection Of Decision Maker's Fuzzy Assessment*

In this step all the decision makers express their opinion in the form of linguistic terminology such as very good, good, fair, poor, very poor and we combine the fuzzy assessment of each decision makers on every goal and its sub goals on the basis of the importance of its requirements.

*Step 4: Computation of Attributed Values*

**Step 4.1:** For the prioritization of requirements on the basis of various criteria's, the fuzzy performance rating is aggregated through all decision maker by means of extended addition and scalar multiplication to form a comprehensive performance matrix P. In this matrix the performance rating is:

$$p_{ij} = (1/n) \odot (P_{ij}^1 \oplus P_{ij}^2 \oplus \dots \oplus P_{ij}^n)$$

is a triangular fuzzy number of the form:

$$(p_{1ij}, p_{2ij}, p_{3ij}) = (\frac{1}{n} \sum_{k=1}^n p_{1kj}, \frac{1}{n} \sum_{k=1}^n p_{2kj}, \frac{1}{n} \sum_{k=1}^n p_{3kj}) \quad (1)$$

Now calculate the fuzzy weight through all decision makers by means of extended addition and scalar multiplication which gives a comprehensive weight vector W. Once the comprehensive performance and weight matrix has been obtained, the following steps are applied (Li 1999):

$$(w_{1j}, w_{2j}, w_{3j}) = (\frac{1}{n} w_{1jk}, \frac{1}{n} \sum_{k=1}^n w_{2jk}, \frac{1}{n} \sum_{k=1}^n w_{3jk}) \quad (2)$$

**Step 4.2:** Aggregate fuzzy ratings with fuzzy weights by means of extended multiplication to get a weighted, comprehensive decision matrix (D), in which

$$d_{ij} = p_{ij} \odot w_j$$

is a fuzzy number with parabolic membership functions in the form of :

$$(\lambda_{1ij}, \lambda_{2ij}, \lambda_{3ij} / d_{ij} / \Delta_{1ij}, \Delta_{2ij}, \Delta_{3ij}) \quad (3)$$

Where

$$\lambda_{1ij} = (w_{2j} - w_{1j})(p_{2ij} - p_{1ij})$$

$$\lambda_{2ij} = w_{1j} (p_{2ij} - p_{1ij}) + p_{1ij} (w_{2j} - w_{1j})$$

$$\lambda_{3ij} = w_{1j} p_{1ij}$$

$$\Delta_{1ij} = (w_{3j} - w_{2j})(p_{3ij} - p_{2ij})$$

$$\Delta_{2ij} = w_{3j} (p_{3ij} - p_{2ij}) + p_{3ij} (w_{3j} - w_{2j})$$

$$\Delta_{3ij} = w_{3j} p_{3ij} \text{ and}$$

$$d_{ij} = w_{2j} p_{2ij}$$

**Step 4.3:** Define each sub-goal/requirement as a fuzzy number  $A_i$ ,  $i=1, 2, \dots, m$  by carrying out extended addition and scalar multiplication through the following criteria:

$$A_i = 1/m \odot (d_{i1} \oplus d_{i2} \oplus \dots \oplus d_{iC}) \quad (4)$$

with parabolic membership function in the form of :

$$(\lambda_{1i}, \lambda_{2i}, \lambda_{3i} / EA_i / \Delta_{1i}, \Delta_{2i}, \Delta_{3i});$$

where

$$\lambda_{ii} = \frac{1}{m} \sum_{j=1}^m \lambda_{ij}, \quad i = 1, 2, 3;$$

$$\Delta_{ii} = \frac{1}{m} \sum_{j=1}^m \Delta_{ij}, \quad i = 1, 2, 3, \text{ and}$$

$$EA_i = \frac{1}{m} \sum_{j=1}^m d_{ij}.$$

**Step 4.4:** Define extended average (EA) by means of extended addition and scalar multiplication through all alternatives (sub-goals/ requirements).

$$EA = 1/n \odot (g_1 \oplus g_2 \oplus \dots \oplus g_n) \quad (5)$$

with parabolic membership function in the form of :

$$(\lambda_1, \lambda_2, \lambda_3 / Sum\_EA / \Delta_1, \Delta_2, \Delta_3);$$

where

$$\lambda_I = \frac{1}{n} \sum_{i=1}^n \lambda_{ii}, \quad I = 1, 2, 3;$$

$$\Delta_I = \frac{1}{n} \sum_{i=1}^n \Delta_{ii}, \quad I = 1, 2, 3, \text{ and}$$

$$Sum\_EA = \frac{1}{n} \sum_{i=1}^n EA_i$$

**Step 4.5:** Define the extended difference,  $EA_i \ominus Sum\_EA$ , for each  $A_i \in R$ , with parabolic membership function in the form of:

$$((\lambda_{1i} - \Delta_1), (\lambda_{2i} + \Delta_2), (\lambda_{3i} - \Delta_3) / EA_i - Sum\_EA / (\Delta_{1i} - \lambda_1), (-\Delta_{2i} - \lambda_2), (\Delta_{3i} - \lambda_3))$$

(6)

**Step 4.6:** Calculate ranking values of each requirements

In this step, we calculate the ranking values ( $rv_i$ ) for each requirements  $A_i$  by means of F-preference relation R:

if  $(\lambda_{3i} - \Delta_3) < 0, (\Delta_{3i} - \lambda_3) \geq 0, EA_i \geq Sum\_EA$ ; then

$$rv_i = \mu R (A_i \ominus EA, 0) = \Pi^+ / (\Pi^+ + \Pi^-); \quad (37)$$

else if  $(\lambda_{3i} - \Delta_3) \leq 0, (\Delta_{3i} - \lambda_3) > 0, EA_i \leq Sum\_EA$ ; then

$rv_i = \mu R(A_i \ominus EA, 0) = \Psi^+ / (\Psi^+ + \Psi^-)$ ;  
 else if  $(\lambda_{3i} - \Delta_3) = 0, (\Delta_{3i} - \lambda_3) = 0$ , EA<sub>i</sub> = Sum\_EA then  $rv_i = \mu R(A_i \ominus EA, 0) = 0.5$ ;  
 else if  $(\lambda_{3i} - \Delta_3) \geq 0, (\Delta_{3i} - \lambda_3) > 0$ , EA<sub>i</sub>  $\geq$  Sum\_EA then  $rv_i = \mu R(A_i \ominus EA, 0) = 1$ ;      (7)  
 else if  $(\lambda_{3i} - \Delta_3) < 0, (\Delta_{3i} - \lambda_3) \leq 0$ , EA<sub>i</sub>  $\leq$  Sum\_EA then  $rv_i = \mu R(A_i \ominus EA, 0) = 0$ .

Where,

$$\begin{aligned}\Pi^+ &= -[1/4(\Delta_{1i} - \lambda_1) - 1/3(\Delta_{2i} + \lambda_2) + 1/2(\Delta_{3i} - \lambda_3)] + [1/4(\lambda_{1i} - \Delta_1)(1 - Z^4) + 1/3(\lambda_{2i} + \Delta_2)(1 - Z^3) + 1/2(\lambda_{3i} - \Delta_3)(1 - Z^2)]; \\ \Pi^- &= (1/4(\lambda_{1i} - \Delta_1)Z^4 + 1/3(\lambda_{2i} + \Delta_2)Z^3 + 1/2(\lambda_{3i} - \Delta_3)Z^2)]; \\ Z &= [-(\lambda_{2i} + \Delta_2) + \sqrt{(\lambda_{2i} + \Delta_2)^2 - 4(\lambda_{1i} - \Delta_1)(\lambda_{3i} - \Delta_3)}]/[2(\lambda_{1i} - \Delta_1)]; \\ \Psi^+ &= 1/4(\Delta_{1i} - \lambda_1)P^4 + 1/3(\lambda_{2i} + \Delta_2)P^3 \\ &\quad + 1/2(\lambda_{3i} - \Delta_3)P^2; \\ \Psi^- &= -[1/4(\lambda_{1i} - \Delta_1) + 1/3(\lambda_{2i} + \Delta_2) + 1/2(\lambda_{3i} - \Delta_3)] - [1/4(\lambda_{1i} - \Delta_1)(1 - P^4) - 1/3(\lambda_{2i} + \Delta_2)(1 - P^3) + 1/2(\lambda_{3i} - \Delta_3)(1 - P^2)]; \\ P &= [(\lambda_{2i} + \Delta_2) - \sqrt{(-\lambda_{2i} - \Delta_2)^2 - 4(\lambda_{1i} - \Delta_1)(\lambda_{3i} - \Delta_3)}]/[2(\lambda_{1i} - \Delta_1)];\end{aligned}$$

Where “*sqrt*” is a square root function

**Step 4.7:** Select the goal with the maximum ranking values. These values have to be arranged in ascending order and the goal with the maximum ranking value has to be selected as the best alternative.

#### IV. EXPERIMENTAL EVALUATION

The proposed algorithm has been applied on Institute Examination System to validate the results. A numerical example is illustrated that implement the Fuzzy group decision making approach in goal oriented requirement elicitation process for requirements prioritization of an Institute Examination System (IES). IES is an integrated, efficient and easy to use system for computerizing total examination work of an institute. This system can handle large volume of data and is quiet robust. This system is used to provide the facility to submit the examination form online, conduct online examination and generate the result of the student.

#### Step 1

Here 5 stakeholders' s1, s2, s3, s4 and s5 and two decision makers DM1 and DM2 who are participating in group requirement elicitation process have been assumed.

Table1 gives the classification of requirements into functional requirements (FR) and non-functional requirements (NFR). AND decomposition is there among nfr1, nfr2 and nfr3 but there is an OR decomposition among nfr2-1, nfr2-2 and nfr2-3 given in Table 1. Table 2 represents the classification of functional requirements for IES. Table 3 gives the requirements. Table 4 represents the judgment matrix.

Table 1: Requirements classification

Requirements								
FR			NFR					
FR <sub>1</sub>	FR <sub>2</sub>	FR <sub>3</sub>	nfr <sub>1</sub>	nfr <sub>2</sub>			nfr <sub>3</sub>	
				nfr <sub>2-1</sub>	nfr <sub>2-2</sub>	nfr <sub>2-3</sub>		

Table 2: Functional requirements classification

FR															
FR <sub>1</sub>		FR <sub>2</sub>		FR <sub>3</sub>					FR <sub>4</sub>		FR <sub>5</sub>				
fr <sub>9</sub>	fr <sub>16</sub>	fr <sub>3</sub>	fr <sub>7</sub>	fr <sub>1</sub>	fr <sub>4</sub>	fr <sub>5</sub>	fr <sub>6</sub>	fr <sub>8</sub>	fr <sub>11</sub>	fr <sub>13</sub>	fr <sub>15</sub>	fr <sub>2</sub>	fr <sub>10</sub>	fr <sub>12</sub>	fr <sub>14</sub>

where

FR<sub>1</sub>: Student Examination Module.

FR<sub>2</sub>: System Administrator Module.

FR<sub>3</sub>: Policy Enforcement Module.

FR<sub>4</sub>: Faculty Module

FR<sub>5</sub>: Examination Controller Module

fr<sub>1</sub>: Document retention that is consistent with departmental policies and contemporaneous with the examination.

fr<sub>2</sub>: Provision of a list of written instructions that document a routine activity followed by the examination system.

fr<sub>3</sub>: Online conduction of examination.

fr<sub>4</sub>: Ensure the software licensing of all the software used by the system.

fr<sub>5</sub>: Provision of a tool that can be used to document technical activities.

fr<sub>6</sub>: Establishing an event that helps in improving the performance of the data access method.

fr<sub>7</sub>: Display semester result.

fr<sub>8</sub>: Investment in a user friendly data management system.

fr<sub>9</sub>: Online submission of examination fee.

fr<sub>10</sub>: Involvement of continuous and comprehensive evaluation scheme.

fr<sub>11</sub>: Creation of explicit norms regarding data use at the system and student level.

fr<sub>12</sub>: Establishment of rules for the online submission of examination form.

fr<sub>13</sub>: Generation of complete and accurate examination report for the student.

fr<sub>14</sub>: Quick upload of all the examination related activities.

fr<sub>15</sub>: Entry of internal and external marks

fr<sub>16</sub>: Online filling of the examination forms. After successful submission of the forms, system will generate hall ticket for the student with the following information related to the student:

(a) Name of the student.

- (b) Father's name.
  - (c) Roll number
  - (d) Enrollment number
  - (e) Examination name
  - (f) Subject name(s)
  - (g) Subject code(s)
- NFR: Trustworthiness
- $nfr_1$ : Security
- $nfr_2$ : Reliability
- $nfr_3$ : Performance
- $nfr_{2-1}$ : Recoverability
- $nfr_{2-2}$ : Adaptability
- $nfr_{2-3}$ : Flexibility

## Step 2

Table 3: Triangular Fuzzy numbers of Linguistics value {EL, VL, L, M, H, VH, EH} for each goal.

S. No.	Linguistics value	Triangular fuzzy number
1	VL(Very Low)	(0,0,0.25)
2	L (Low)	(0,0.25,0.5)
3	M (Middle)	(0.25,0.5,0.75)
4	H (High)	(0.5,0.75,1)
5	VH (Very high)	(0.75,1,1)

On applying Step 3 the decision maker's fuzzy assessment given in table 4 and table 5 by two decision makers  $DM_1$  and  $DM_2$  respectively is obtained.

Table 4: Fuzzy assessment of  $DM_1$

Sub Requirement	Non Functional Requirement (NFR)		
	$NFR_1$	$NFR_2$	$NFR_3$
$FR_1$	(0.2, 0.4, 0.5)	(0.3, 0.4, 0.5)	(0.4, 0.6, 0.7)
$FR_2$	(0.2, 0.3, 0.6)	(0.2, 0.5, 0.8)	(0.3, 0.4, 0.7)
$FR_3$	(0.1, 0.2, 0.2)	(0.3, 0.5, 0.7)	(0.3, 0.6, 0.7)
$FR_4$	(0.5, 0.6, 0.7)	(0.2, 0.3, 0.6)	(0.3, 0.4, 0.5)
FR	(0.1, 0.2, 0.3)	(0.2, 0.3, 0.4)	(0.2, 0.4, 0.6)

Table 5: Fuzzy Assessment of  $DM_2$

Sub Requirement	Non Functional Requirement (NFR)		
	$NFR_1$	$NFR_2$	$NFR_3$
$FR_1$	(0.5, 0.6, 0.6)	(0.2, 0.5, 0.6)	(0.5, 0.7, 0.8)
$FR_2$	(0.3, 0.4, 0.6)	(0.2, 0.2, 0.3)	(0.3, 0.4, 0.5)

$FR_3$	(0.4, 0.6, 0.6)	(0.5, 0.6, 0.7)	(0.2, 0.3, 0.3)
$FR_4$	(0.1, 0.2, 0.2)	(0.3, 0.2, 0.2)	(0.1, 0.2, 0.3)
$FR_5$	(0.2, 0.3, 0.2)	(0.3, 0.4, 0.5)	(0.2, 0.3, 0.4)

#### Step 4

Table 6: Fuzzy Importance Weight

NFR	$DM_1$	$DM_2$
$NFR_1$	(0.1, 0.2, 0.5)	(0.2, 0.3, 0.7)
$NFR_1$	(0.2, 0.4, 0.6)	(0.5, 0.6, 0.8)
$NFR_1$	(0.3, 0.4, 0.5)	(0.1, 0.1, 0.4)

Table 7: Comprehensive Performance

Sub Requirement	Non Functional Requirement (NFR)		
	$NFR_1$	$NFR_2$	$NFR_3$
$FR_1$	(0.35, 0.5, 0.55)	(0.25, 0.45, 0.55)	(0.45, 0.65, 0.75)
$FR_2$	(0.25, 0.35, 0.6)	(0.2, 0.35, 0.55)	(0.3, 0.4, 0.6)
$FR_3$	(0.25, 0.4, 0.4)	(0.4, 0.55, 0.7)	(0.25, 0.45, 0.5)
$FR_4$	(0.3, 0.4, 0.45)	(0.25, 0.25, 0.4)	(0.2, 0.3, 0.4)
$FR_5$	(0.15, 0.25, 0.25)	(0.25, 0.35, 0.45)	(0.2, 0.35, 0.5)

Table 8: Comprehensive Weight

$NFR_1$	$NFR_2$	$NFR_3$
(0.15, 0.25, 0.6)	(0.35, 0.5, 0.7)	(0.2, 0.25, 0.45)

Application of step 4.1 gives the comprehensive performance and comprehensive weight of each requirement and sub requirement shown in table 6, 7 and 8.

Table 9: Quadratic Membership Function

Sub Requirement	Non Functional Requirement (NFR)		
	$NFR_1$	$NFR_2$	$NFR_3$
$FR_1$	$0.125_{QF}$	$0.225_{QF}$	$0.1625_{QF}$
$FR_2$	$0.875_{QF}$	$0.175_{QF}$	$0.1_{QF}$
$FR_3$	$0.1_{QF}$	$0.275_{QF}$	$0.1125_{QF}$
$FR_4$	$0.1_{QF}$	$0.125_{QF}$	$0.075_{QF}$
$FR_5$	$0.0625_{QF}$	$0.175_{QF}$	$0.0875_{QF}$

#### Step 4.2

A weighted comprehensive decision matrix is given as a quadratic membership function in table 9, can be established by applying the extended multiplication of the form of:

$$\begin{aligned}
 0.125_{QF} &= 0.015, 0.0575, 0.0525 / 0.125 / 0.0175, 0.2225, 0.33 \\
 0.225_{QF} &= 0.03, 0.1075, 0.0875 / 0.225 / 0.02, 0.18, 0.385 \\
 0.1625_{QF} &= 0.01, 0.0625, 0.09 / 0.1625 / 0.02, 0.195, 0.3375 \\
 0.0875_{QF} &= 0.01, 0.04, 0.0375 / 0.0875 / 0.0875, 0.36, 0.36 \\
 0.175_{QF} &= 0.0225, 0.0825, 0.07 / 0.175 / 0.04, 0.25, 0.385 \\
 0.1_{QF} &= 0.005, 0.035, 0.06 / 0.1 / 0.04, 0.21, 0.27 \\
 0.1_{QF} &= 0.015, 0.0475, 0.0375 / 0.1 / 0, 0.014, 0.24 \\
 0.275_{QF} &= 0.0225, 0.1125, 0.14 / 0.275 / 0.03, 0.245, 0.49 \\
 0.1125_{QF} &= 0.01, 0.0525, 0.05 / 0.1125 / 0.01, 0.1225, 0.225 \\
 0.1_{QF} &= 0.01, 0.045, 0.045 / 0.1 / 0.0175, 0.1875, 0.27 \\
 0.125_{QF} &= 0, 0.375, 0.0875 / 0.125 / 0.03, 0.185, 0.28 \\
 0.075_{QF} &= 0.005, 0.03, 0.04 / 0.075 / 0.02, 0.125, 0.18 \\
 0.0625_{QF} &= 0.01, 0.03, 0.0225 / 0.0625 / 0, 0.0875, 0.15 \\
 0.175_{QF} &= 0.015, 0.0725, 0.0875 / 0.175 / 0.02, 0.16, 0.315 \\
 0.0875_{QF} &= 0.0075, 0.04, 0.04 / 0.0875 / 0.03, 0.1675, 0.225
 \end{aligned}$$

#### Step 4.3

This step computes the sub requirement mathematically to define each sub requirement as a fuzzy number.

$$\begin{aligned}
 0.1708_{QF} &= 0.0183, 0.0758, 0.0766 / 0.1708 / 0.0191, 0.1991, 0.3508 \\
 0.1208_{QF} &= 0.0125, 0.0258, 0.1675 / 0.1208 / 0.0558, 0.2733, 0.3383 \\
 0.1625_{QF} &= 0.0158, 0.0708, 0.0758 / 0.1625 / 0.0133, 0.1691, 0.3183 \\
 0.1_{QF} &= 0.005, 0.0375, 0.0575 / 0.1 / 0.0225, 0.1658, 0.2433 \\
 0.1083_{QF} &= 0.0108, 0.0475, 0.05 / 0.1083 / 0.0166, 0.1383, 0.23
 \end{aligned}$$

#### Step 4.4

On applying step 4.4 we get the following equation.

$$0.13248_{QF} = 0.01248, 0.05148, 0.08548 / 0.13248 / 0.02546, 0.3152, 0.29614$$

After applying **step 4.5** and **step 4.6** we get the ranking values of each functional requirements

$$FR_1 = 0.882$$

$$FR_2 = 0.842$$

$$FR_3 = 0.462$$

$$FR_4 = 0.304$$

FR<sub>5</sub>=0.617

#### Step 4.7

Arrange the ranking values of the Functional Requirements in ascending order and select the requirement with the maximum ranking value for the prioritization of requirement amongst them.

FR<sub>1</sub> > FR<sub>2</sub> > FR<sub>5</sub> > FR<sub>3</sub> > FR<sub>4</sub>

#### V. CONCLUSION

In order to strengthen the goal oriented requirements elicitation process a method for the elicitation of attributed values in goal models has been proposed. Binary tree sort method has been used here to get the prioritized list of requirements. In this example, five functional goals and three non-functional goals have been assumed. The functional goals are: student module (FR1), teachers module (FR2), controller of examination module (FR3), administrative module (FR4), and online conduct of examination module (FR4); and three non-functional goals (NFR) or criteria are security (NFR1), reliability (NFR2), performance (NFR3). Two stakeholders' are participating as decision maker here. On the basis of the analysis, it has been identified that for institute examination systems, FR1 is the most important sub-goal and it is the best alternative. So this goal has to be considered in the first release of the software.

#### REFERENCES

1. Sadiq M and Jain S.K., "A Fuzzy Based Approach for Requirements Prioritization in Goal Oriented Requirements Elicitation Process", International Conference of Software Engineering and Knowledge Engineering (SEKE), pp. 54-58, 2013 USA.
2. Sadiq M, Shahid M, "Elicitation and prioritization of software requirements", International Journal of Recent Trends in Engineering and Technology 2(3):138–142, 2009.
3. Sadiq M and Jain SK, "An insight into Requirements Engineering Processes", International Conference of Advances in Communication, Networking and Computing (CNC), LNCSIT, Springer, PP. 313-318, 2012 India.
4. Kavaklı E, "Goal oriented requirements engineering: a unifying framework", Requirements Engineering. Springer, 2002, London.
5. Kaiya H, et al., "AGORA: Attributed goal oriented requirements analysis", Proceedings of the IEEE Joint Conference on Requirements Engineering, 2002.
6. Anwer S, Ikram N, "Goal Oriented Requirement Engineering, a Critical Study of Techniques", 12th Asia Pacific Software Engineering Conference, 2006.
7. Yu ESK, "Towards Modeling and Reasoning Support for Early and Late Phase of Requirements Engineering", IEEE International Conference on Requirements Engineering, 1997.
8. Zickert F, "Evaluation of the Goal Oriented Requirements Engineering Methods KAOS", American Conference on Information system, pp. 1–9, 2010.
9. Chan L-K, Wu M-L, "Quality Function Deployment", A Comprehensive Review of its Concepts and Methods, Qual Eng, Taylor, Francis 15(1):23–35, 2007.
10. Pai WC, "A quality Enhancing Software Function Deployment Model", Information System Management 19(3):20–24, 2006.
11. Lamsweerde AV, "Goal-Oriented Requirements Engineering, a Guided Tour", Proceedings of 5<sup>th</sup> IEEE International Symposium on Requirements Engineering, pp. 249–263, 2001.
12. Nuseibeh B, Easterbrook S, "Requirements Engineering, a Roadmap", Proceedings of the Conference on the Future of Software Engineering, 2002, New York.
13. Sadiq M and Jain SK "A Fuzzy Based Approach for the Selection in Goal Oriented Requirements Elicitation Process", International Journal of System Assurance Engineering and Management, Springer, 2014.

14. Potts C, "Invented Requirements and Imagined Customers: Requirements Engineering for off-the Shelf Software", 2<sup>nd</sup> IEEE International Symposium on Requirements Engineering, pp. 128–130, 1995.
15. Kaiya H et al, "Improving the Detection of Requirements Discordances among Stakeholders", Requirements Engineering Springer, Berlin, pp 289–303, 2005.
16. Karlsson L, Dahlstedt AG, Regnell DJN, Persson A, "Requirements Engineering Challenges in Market Driven Software Development-an Interview Study with Practitioners", Information Software Technology, 49:588–604, 2007.
17. Akker JM, Brinkkemper S, Diepen G, and Versendaal J, "Determination of the Next Release of a Software Product: An Approach Using Integer Linear Programming", Proceedings of the CAiSE, pp. 119–124, 2005.
18. Wang X-T, Xiong W, "An Integrated Linguistic-Based Group Decision Making Approach for Quality Function Deployment", Experts System Application 38:14428–14438, 2011.
19. Lee H-S, "On Fuzzy Preference Relation in Group Decision Making", International Journal of Computer Math, Taylor and Francis 82:133–140, 2005.
20. Lin HY, Hsu PY, Sheen GJ, "A Fuzzy Based Decision Making Procedure for Data Warehouse System Selection", Expert System Application-Elsevier 32:939–953, 2007.
21. Zadeh LA, "Fuzzy Sets", Infor Control 8:338–353, 1965.
22. Zadeh LA, "The Concept of a Linguistics Variable and its Application to Approximate Reasoning (I). Infor Sci 8:199–249, 1975.
23. Hwang CL, Yoon K, "Multiple Attributes Decision Making Methods and Applications", Springer, 1981, New York.
24. Chang PL, Chen YC, "A Fuzzy Multi-Criteria Decision Making Method for Technology Transfer Strategy Selection in Biotechnology", Fuzzy Sets System 63:131–139, 1994.
25. Cochran JK, Chen H, "Fuzzy Multi-Criteria Selection of Object Oriented Simulation Software for Production System Analysis", Computer Operation Res 32:153–168, 2005.
26. Yamamoto K, Saeki M, "Using Attributed Goal Graphs for Software Component Selection", An Application of Goal-Oriented Analysis to Decision Making, In: 26th International Conference on Conceptual Modeling-ER, 2007
27. Zhu MX, Luo X-X, Chen XH, Wu DD, "A Non-Functional Requirements Trade off Model in Trustworthy Software", Information Science- Elsevier 191:61–75, 2012.

# Real Time Lane Departure Warning System for Drivers

Kadir İLERİ<sup>1</sup>, Salih GÖRGÜNOĞLU<sup>2</sup>

<sup>1</sup>*Department of Electrical & Electronics Engineering, Karabuk University, Karabuk, Turkey*

<sup>2</sup>*Department of Computer Engineering, Karabuk University, Karabuk, Turkey*

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

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

## I. INTRODUCTION

Increase in the number of vehicles and traffic accidents have become a major problem all over the world. Most of these accidents occur due to driver inattention. Driver warning systems which make the driver realize a dangerous situation have been developed to avoid these kinds of accidents. The driver is warned against potential hazards which are determined by various sensor systems such as radar and camera. The system in this paper is about lane detection and lane departure warning system. The main purpose of this system is to avoid accidents that take place due to unaware lane departures. The main benefit of image based system is that single camera can get a more data than a lot of embedded sensors. Therefore, real time vision based lane departure warning system helps greatly to the reduction and prevention of accidents.

There are various methods for lane detection. An algorithm which uses the color information is more easily than others [1][2]. These algorithms use that the color of road surface is almost black and lane marks are white, yellow and red. Another method is edge-based approach [3]. However, in this method, it is difficult to determine the threshold value and classify the pixel belong to edge or not. Another approach that combines an edge distribution function and the Hough transform with linear parabolic model is also developed for lane detection and lane tracking [4]. It can detect lane boundaries successfully even in very noisy environment.

The algorithm, in this paper, uses geometric information of camera and road. First, the region of interest (ROI) which has lane marks by making use of geometry of camera configuration in input image captured from camera is determined. Thus, the processing load is reduced to a considerable degree. Then, fixed width of lane mark is supplied by using inverse perspective mapping in ROI. The image is filtered by estimating the approximate width of the lane mark. The Hough transform is applied to the lane marks and the slopes of the lane marks are found. Whether the vehicle makes lane departure or not is determined by having regard to these slopes. Fig. 1 shows the flowchart of developed lane departure warning system.

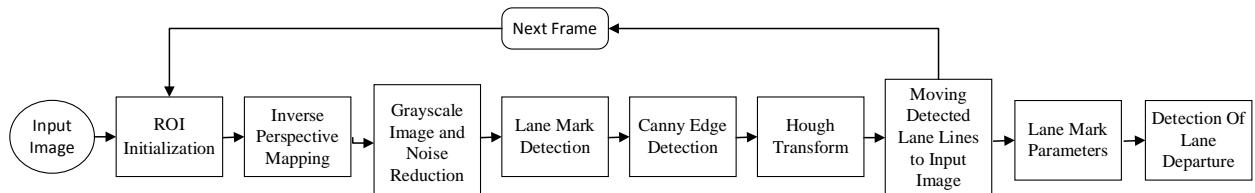


Figure 1. Flowchart of proposed lane detection algorithm.

## II. PROPOSED LANE DETECTION ALGORITHM

### A. ROI Initialization

In initializing the ROI, the area which has a high potential of being lane marks is selected. First, the ROI is defined when we capture the input image from the camera as seen in Fig. 2 by considering the height and tilt angle of camera which has placed to the vehicle. Applying lane detection algorithm to this ROI increases performance and also cleans the noise by not processing unrelated parts of road. Thus, it provides more accurate results and the memory usage is reduced at a remarkable rate.

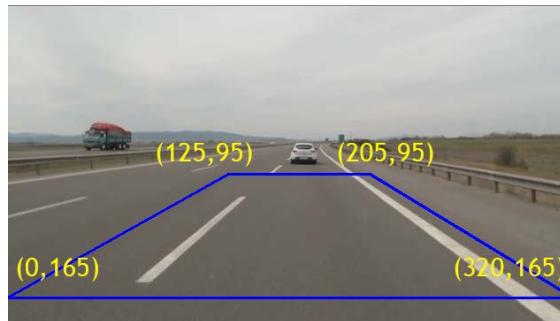


Figure 2. The definition of ROI.

### B. Inverse Perspective Mapping

There are at least two drawbacks of perspective view in input image including indefinite lane mark width and length. Eliminating such perspective effects will greatly facilitate lane detection. Converting input image to a different space will not lose the extra time much instead of applying algorithm to directly input image.

Perspective effect, a general problem in cameras, occurs match of rectangular shape in image domain with isosceles trapezoid shape field of view of the camera [5]. Various methods have been presented in the literature to reduce or eliminate perspective effects [6][7]. Inverse perspective mapping uses 3x3 homography matrix to convert points from image plane to real-world coordinates [8]. Fig. 3 shows both coordinate systems for a fixed position of camera. In Fig. 3, isosceles trapezoid shape which is on the left shows camera field in image plane and rectangular shape which is on the right shows camera field in real-world plane.

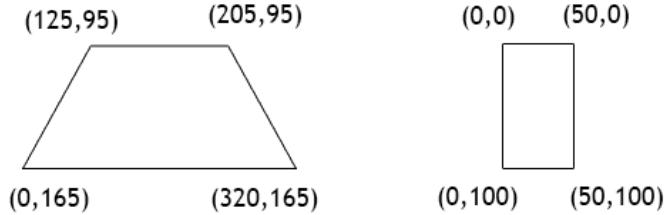


Figure 3. Image and real-world coordinate systems.

To convert the image coordinates to real-world coordinates can be performed with (1).

$$\begin{bmatrix} x^* \\ y^* \\ k \end{bmatrix} = H \begin{bmatrix} j \\ i \\ 1 \end{bmatrix} \quad (1)$$

where  $x^*$  and  $y^*$  are the unnormalized real-world coordinates,  $k$  is the normalization factor,  $H$  is known as the  $3 \times 3$  homography matrix,  $i$  and  $j$  are the row and column indices of the pixel in the image coordinate system, as given in Fig. 3, respectively.  $x^*$  and  $y^*$  should be normalized by  $k$  to obtain the real-world coordinates. A ROI and its inverse perspective mapping image are shown in Fig. 4a and Fig. 4b respectively.

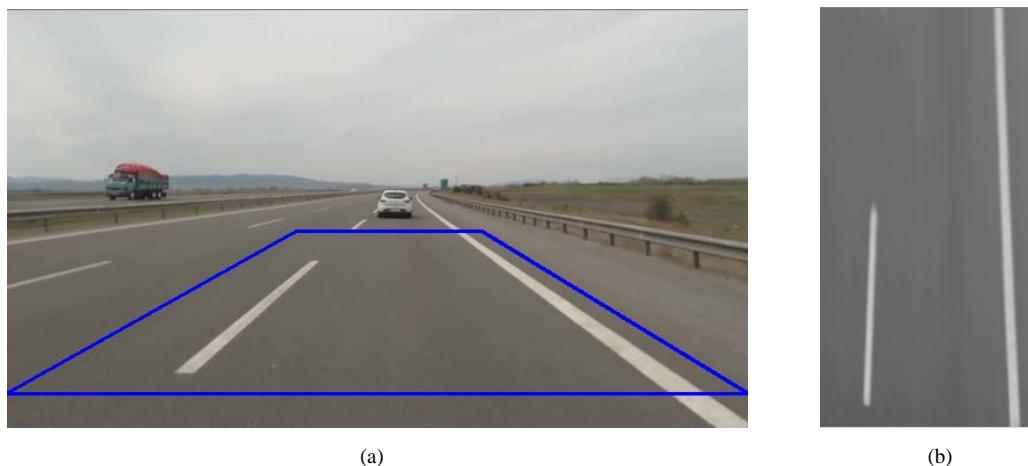


Figure 4. A representative of ROI image and its inverse perspective mapping image, a) ROI, b) inverse perspective mapping image.

### C. Grayscale Image and Noise Reduction

In this step, colorful input image captured from the camera is converted to grayscale image to speed up the method to be applied. Thus, the computational complexity of algorithm can be reduced significantly and the algorithm does less sensitive to environmental conditions [9].

Noise is a problem for computer vision as a problem for most systems. Developed algorithms should have noise tolerance or the image should be purified from noise. There are various filters such as median, mean and Gaussian filters to remove or reduce the noise. Gaussian filter is similar to the mean filter. Gaussian filter scans all image with mask you specify its matrix size, calculates the weighted average of matrix for each step of scanning, even though mean filter calculates just the average. Therefore, Gaussian filter provides a smoother blur compared to the mean filter. Thus, Gaussian filter has been preferred in our application.

$$I' = G * I \quad (2)$$

where  $I'$  is image purified from noise,  $G$  is 5x5 Gaussian filter mask and  $I$  is grayscale image.

#### D. Lane Mark Detection

Lane marks' width is narrower in upper part of the image compared to lower part because of perspective effect. In Section B, we have applied inverse perspective mapping to ROI which we determined in Section A. Thus, we have made lane marks' width equal in all part of image. Consequently, specific lane marking detector has been used for detection. The analysis is done by independently filtering each image row of intensity values, denoted as  $\{x_i\}^{Wi=1}$ , resulting in a new filtered data array  $\{y_i\}^{Wi=1}$ , defined as

$$y_i = 2x_i - (x_{i-\tau} + x_{i+\tau}) - |x_{i-\tau} - x_{i+\tau}| \quad (3)$$

where  $\tau$  is the width parameter that governs the filtering process. This filter produces high responses for positions with  $x_i$  values that are higher than their neighbors on the left and right at a distance  $\tau$ . The last term in (3) penalizes cases in which the difference between the right and left neighbors is high, so that a higher response is given to positions with similar right and left neighbors [10]. This last term makes this filter less prone to errors than other lane marking detectors presented in the literature [11][12]. Fig. 5 shows the filtered image.

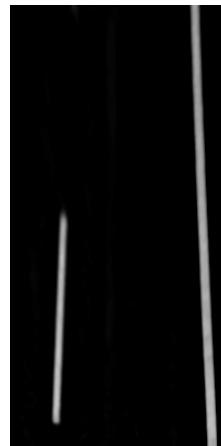


Figure 5. The filtered image.

#### E. Canny Edge Detection

Sobel, Roberts Cross and Prewitt are the most used methods in the edge detection. Prewitt and Sobel are composed of horizontal and vertical directional 3x3 masks. The major limitation of these operators cannot find the edges properly in high noisy environment. A more advanced technique is Canny edge detection method. Canny edge detection method is used in our application. This edge detection algorithm developed by John F. Canny in 1986 [13]. Nowadays, this edge detection method has become a standard and widely used in many researches. Fig. 6 shows the result image of Canny edge detection.

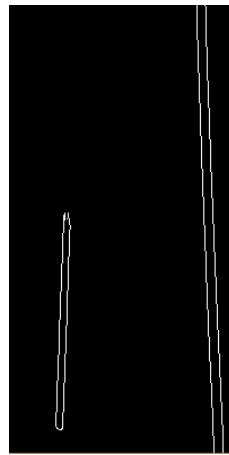


Figure 6. The resultant image of Canny edge detection.

#### F. Hough Transform

Hough transform is a technique commonly used for the detection of particular shapes in images. This technique was developed by Paul Hough in 1962 and it provides an effective approach in determining the boundaries of the object of interest. The main focus point of the section is to find lane mark parameters such as angle and length of straight lines by using Hough transform due to the fact that lane marks are almost straight.

A straight line can be described in 2D coordinate system in many different ways. For example in a Cartesian coordinate system, it is described with (4).

$$y = m*x + b \quad (4)$$

where  $m$  is the slope and  $b$  is the intercept of  $y$ -axis. However, Cartesian coordinate system cannot be used in this system in that  $m$  in (4) is infinite for vertical lines. Polar coordinate system is more suitable for Hough Transform. A straight line can be described with (5) in Polar coordinate system.

$$\rho = x*\cos(\theta) + y*\sin(\theta) \quad (5)$$

where  $\rho$  is the length of a line perpendicular to this line, starting from the origin and  $\theta$  is the orientation angle of  $\rho$  with respect to the  $x$ -axis. The conversion between these two systems can be easily done. A point in Cartesian coordinate system which corresponds to a sinusoid in Polar coordinate system is shown in Fig. 7a and Fig. 7b respectively.

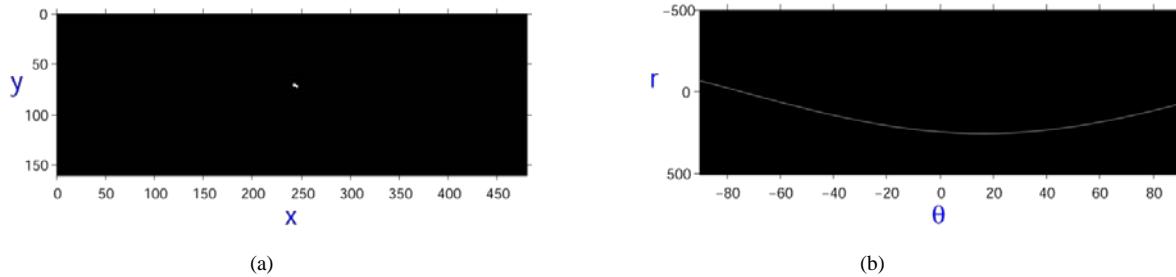


Figure 7. a) A point in Cartesian coordinate system, b) its sinusoid in Polar coordinate system [14].

All sinusoids obtained from different points in the Cartesian coordinate system intersects at the same point in Polar coordinate system as shown in Fig. 8a and Fig. 8b.

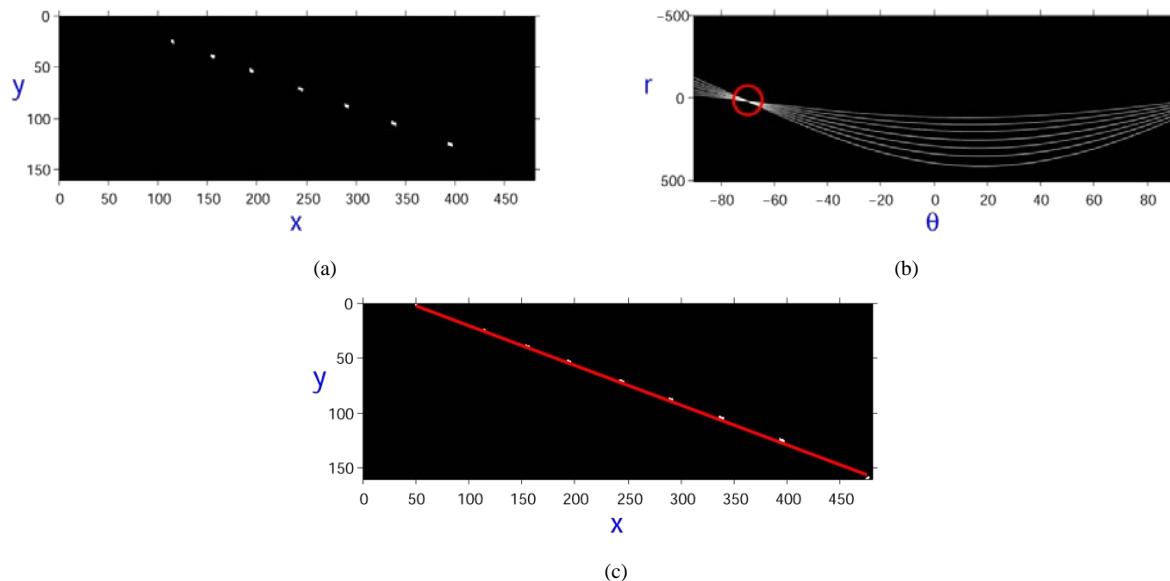


Figure 8. a) Points in Cartesian coordinate system, b) corresponding sinusoids in Polar coordinate system, c) detected line in Cartesian coordinate system [14].

Dashed, interrupted, even locally distorted lines in input image will still get detected by Hough transform. Because their undistorted parts will form the intersection points which indicate presence of lines. So if we search for local maxima in Polar space, extract these points of intersection, map them back to Cartesian space and overlay this image on the original image. Thus, we obtain the detected lines (Fig. 8c). The resultant images after Hough transformation is applied to images in our application are shown in Fig. 9.

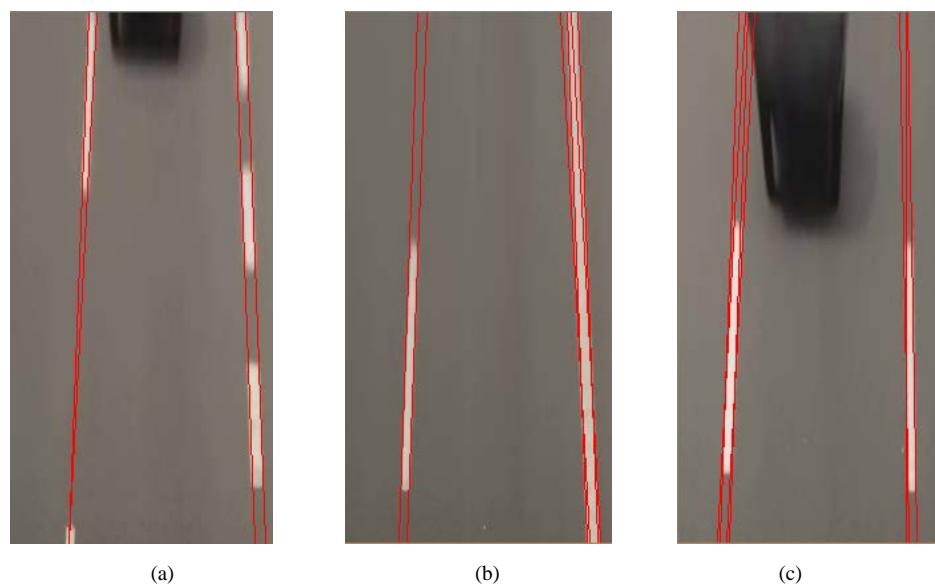


Figure 9. Images after Hough transformation.

#### G. Moving Detected Lane Lines to Input Image

Hough transform, as shown in Fig. 9, can produce multiple results for each lane mark. We need to express each lane mark with a single line by converting multiple lines obtained as a result of the Hough transform to single line for use in the next section of our application. The following steps are applied in order to accomplish this.

- Each line obtained as a result of Hough transform is determined by looking at left and right lane marks and the next steps are applied separately for each lane mark in Polar coordinate system.
- The  $\rho$  (the perpendicular length of a line from origin to this line) value of new single line is determined by taking arithmetic mean of  $\rho$  values.
- The  $\theta$  (the orientation angle of  $\rho$  with respect to the x-axis) value of new single line is determined by taking arithmetic mean of  $\theta$  values.
- $\rho$  and  $\theta$  values calculated for each lane mark is converted to Cartesian coordinates and the lines are obtained as seen in Fig. 10.

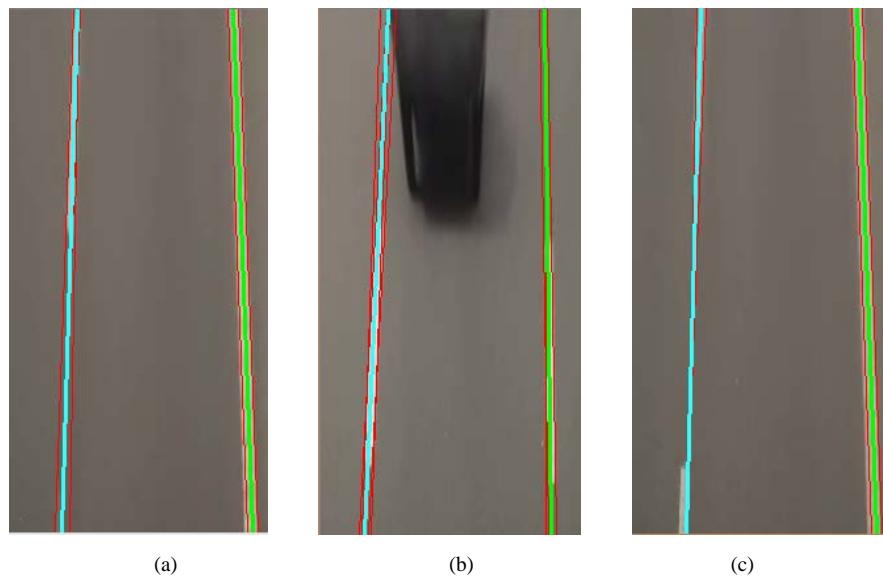


Figure 10. Images of multiple Hough transformation lines converted to single line for each lane mark.

Ratio and proportion has been used to move the lane marks we obtained in Fig. 10 to input image from inverse perspective mapping image. Input image results are shown in Fig. 11.





Figure 11. Detected lane marks in input images.

### III. LANE DEPARTURE WARNING SYSTEM

#### A. Lane Mark Parameters

While the vehicle is on a normal course, two lane marks appear as Fig. 12a. Left and right lane marks are symmetrical with respect to each other along a vertical line, and their angles with respect to horizontal axis are equal. If the vehicle is making lane departure, one lane mark will appear more vertically than the other one. If the vehicle is making left lane departure, left lane mark is more vertical than right one (Fig. 12b) or if the vehicle is making right lane departure, right lane mark is more vertical than left one (Fig. 12c). We can easily detect that the vehicle is making right or left lane departure by using this feature.

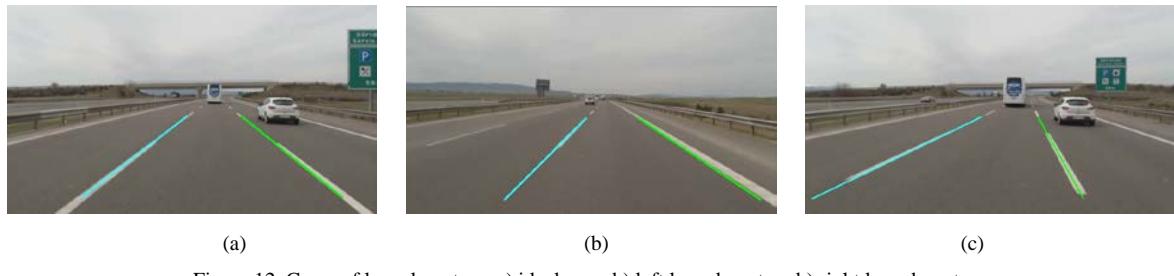


Figure 12. Cases of lane departure, a) ideal case, b) left lane departure b) right lane departure.

#### B. Detection Lane Departure

It can be decided which lane mark is more vertical by looking at their angles with respect to horizontal axis. Greater angle value with respect to x-axis will indicate more vertical line between two lane marks. Both angles are equal ( $\theta_l = \theta_r$ ) in the ideal situation as seen in Fig. 14a and there is no any lane departure. If one of the angles is greater than the other, lane departure starts at the direction of which lane mark's angle is great as seen in Fig. 13b and Fig. 13c.

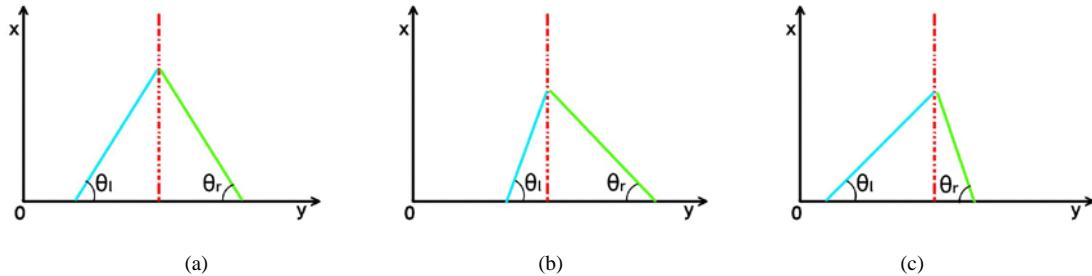


Figure 13. Graphical representation of lane departure cases, a) ideal case, b) left lane departure, c) right lane departure.

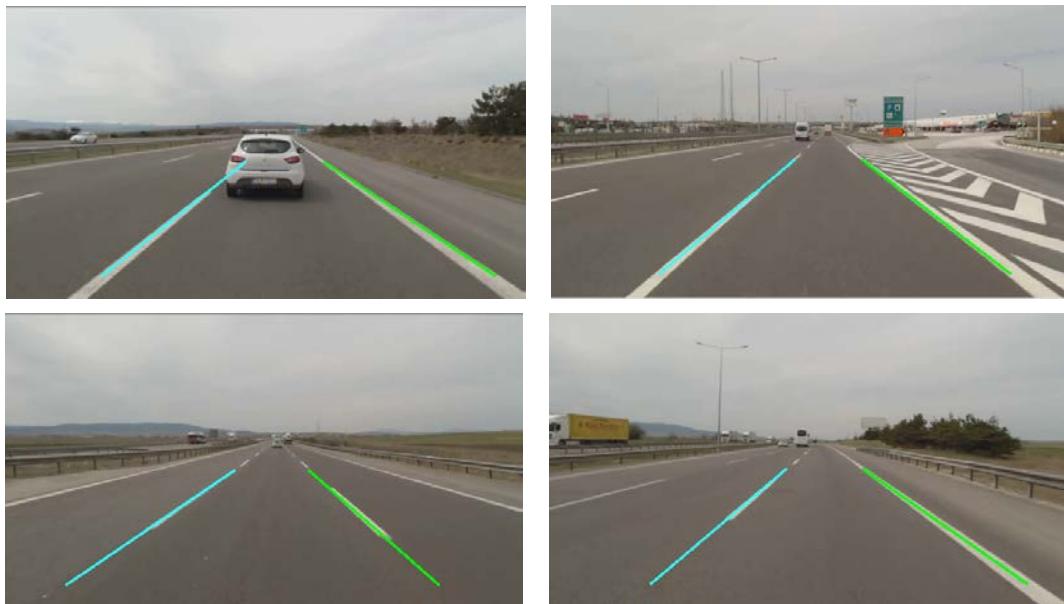
#### IV. EXPERIMENTAL RESULTS

The proposed lane departure warning algorithm has been implemented on Raspberry Pi 2 Model B ARM board (900MHz quad-core ARM Cortex-A7 processor and 1GB LPDDR2 SDRAM) and Raspberry Pi 3 Model B ARM board (1.2GHz quad-core ARM Cortex-A53 processor and 1GB LPDDR2 SDRAM) by using OpenCV library with C++ programming language. The images have been captured from camera have 320x180 image resolution. The processing time of proposed algorithm was about 20~30ms per frame on Raspberry Pi 2 Model B and 12~18ms per frame on Raspberry Pi 3 Model B.

TABLE I  
PROCESSING TIME OF EACH COMPONENT

Components of Proposed Algorithm	Average processing time (ms)	
	Raspberry Pi 2 Model B	Raspberry Pi 3 Model B
Inverse perspective mapping	8	5.7
Converting image to Grayscale	0.4	0.3
Gaussian filter	1.5	1
Lane mark detection	2	1.5
Canny edge detection	2.5	1.5
Hough transform	9	3.2
Calculating angles of lane marks and warnings	0.02	0.01
<b>Total</b>	<b>24</b>	<b>14</b>

Sample images are shown in Fig. 14.



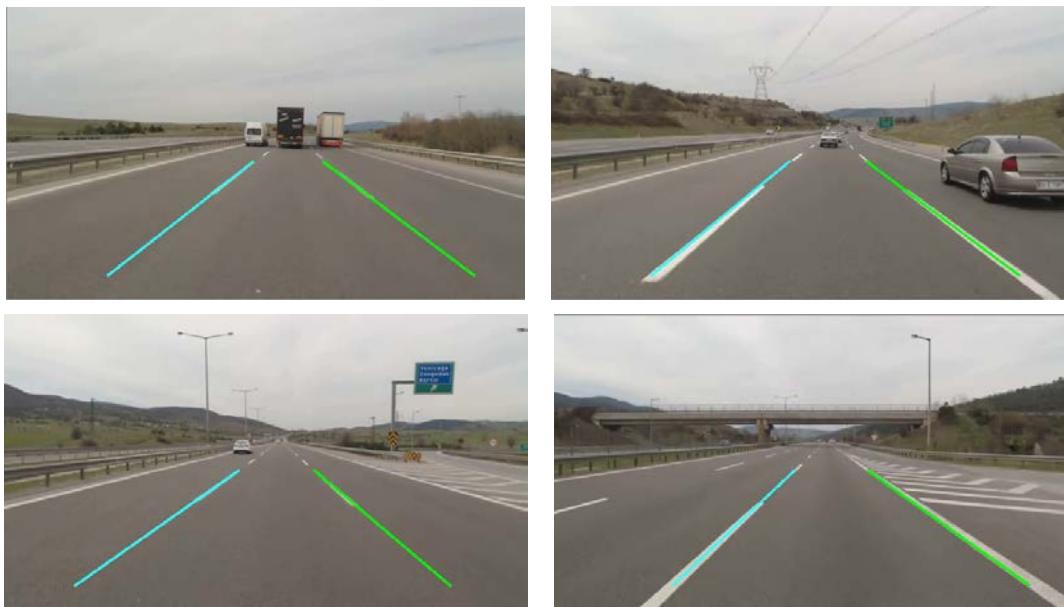


Figure 14. Sample images of the lane detection.

## V. CONCLUSION

In this paper, real time lane departure warning system is presented. The main part of this system consists of lane detection. Lane detection has been eased by avoiding disadvantage of perspective effect by using inverse perspective mapping. In addition, the proposed algorithm is more adaptive to various road environments by reason of the fact that there are no background models or any other road surface models. And this algorithm has been implemented on two ARM development boards and compared their performances. In future, this system can integrate with vehicle. For example, if driver does not signal right or left while making lane departure, the driver can be warned or steering wheel can be vibrated.

## ACKNOWLEDGMENT

This work was supported by Research Fund of Karabuk University. Project Number: KBÜ-BAP-16/1-YL-082.

## REFERENCE

- [1] C. E. Thorpe, (1993) "SCARF: A Color Vision System that Tracks Roads and Intersections," *IEEE Trans. Robot. Autom.*, vol. 9, no. 1, pp. 49–58.
- [2] Y. He, H. Wang, and B. Zhang, (2004) "Color-Based Road Detection in Urban Traffic Scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 309–318.
- [3] K. Kluge, (1994) "Extracting road curvature and orientation from image edge points without perceptual grouping into features," *Proc. Intell. Veh. '94 Symp.*, no. 3 13, pp. 109–114.
- [4] C. R. Jung and C. R. Kelber, (2004) "A robust linear-parabolic model for lane following," *Proceedings. 17th Brazilian Symp. Comput. Graph. Image Process.*, pp. 72–79.
- [5] H. A. Mallot, H. H. Bülthoff, J. J. Little, and S. Bohrer, (1991) "Inverse perspective mapping simplifies optical flow computation and obstacle detection," *Biol. Cybern.*, vol. 64, no. 3, pp. 177–185.
- [6] E. Ince, (2011) "Measuring traffic flow and classifying vehicle types: A surveillance video based approach," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 19, no. 4, pp. 607–620.
- [7] G. Y. Jiang, T. Y. Choi, S. K. Hong, J. W. Bae, and B. S. Song, (2000) "Lane and obstacle detection based

- on fast inverse perspective mapping algorithm," *IEEE Int. Conf. Syst. Man, Cybern.*, vol. 4, pp. 2969–2974.
- [8] G. Küçükyıldız and H. Ocak, (2014) "Development and optimization of a DSP-based real-time lane detection algorithm on a mobile platform," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 22, pp. 1484–1500.
- [9] A. A. Assidiq, O. O. Khalifa, M. R. Islam, and S. Khan, (2008) "Real time lane detection for autonomous vehicles," *2008 Int. Conf. Comput. Commun. Eng.*, pp. 82–88.
- [10] M. Nieto, J. Arrospide Laborda, and L. Salgado, (2011) "Road environment modeling using robust perspective analysis and recursive Bayesian segmentation," *Mach. Vis. Appl.*, vol. 22, no. 6, pp. 927–945.
- [11] M. Nieto, L. Salgado, F. Jaureguizar, and J. Cabrera, (2007) "Stabilization of Inverse Perspective Mapping Images based on Robust Vanishing Point Estimation," *2007 IEEE Intell. Veh. Symp.*, pp. 315–320.
- [12] M. Bertozzi and a Broggi, (1998) "GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection," *Image Process. IEEE Trans.*, vol. 7, no. 1, pp. 62–81.
- [13] J. Canny, (1986) "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698.
- [14] H. Jakub, (2014) "Road lane detection for Android," Bachelor Thesis, Masaryk University Faculty of Informatics, p. 42.

# Building and evaluating an adaptive user interface using a Bayesian network approach

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

ENIS, Embedded Computer System, University of Sfax, Tunisia

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

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

## 1 Introduction

The emergence of new technologies in Web and telecommunications has revolutionized the area of evaluation of the user interfaces. In fact, many research works are interested in the ergonomics and organization of information in an interface and the adaptation of this information to the user's preferences. The user's preferences and experience are at the heart of issues of presentation and adaptive interfaces. Users, then, have interests, knowledge, learning styles, and preferences. An important area of research is how the interfaces can be designed to know the characteristics of the user. In order to provide personalized information, it is necessary to monitor the user's behavior and make generalizations and predictions based on these observations. The information about the user is known as a user model [9,21]. The objective of the user model can be predicting a user's behavior to gain the knowledge of a user, or creating a database of users.

To meet these needs, current research works on user modeling is moving towards richer models in terms of uncertain knowledge. In fact, the traditional

## 2 Building and evaluating an adaptive user interface

models are insufficient to translate the human machine interface. Therefore, Bayesian Networks tend to be increasingly used to manage uncertainty in user modeling [10,24], thanks to their clear semantics and the machine learning opportunities they offer [26,11].

In this paper, the Bayesian networks approach in a web interface was used so as to infer user's preferences. Our approach used learning and inference algorithms. The obtained results proved the effectiveness of Bayesian networks for user modeling. The proposed method was then evaluated. Ten subjects were asked to perform ten tasks using both a fixed user interface and the proposed adaptive user interface. The remaining of this paper was organized as follows: the background and related work were introduced in section 2. Section three was devoted to defining the process for the adaptive user interface building. The structure of Bayesian networks model was constructed using learning algorithms in section 4. in section 5, we dealt with the update of a soft evidence distribution of our Bayesian model. An extensive discussion of the experimental results was presented in section 6. We predicted the user's preferences and we evaluated our user interface. Finally, section 7 revealed the paper conclusion and the potential future perspectives.

## 2 User modeling

User modeling has been the major topic of many studies. It seems to be an intricate issue attesting the different branches of information and human sciences related to this area such as psychology, artificial intelligence, education, human machine interface, etc. The results were mixed with undeniable successes and failures. This research field began to grow significantly in the 80's with the work in the ITS field (Intelligent Tutoring Systems)[3, 29].

Innovations in existing user models contributed to strive towards better personalization capabilities taking into account different users characteristics. An important amount of research has addressed the topic of providing personalized services for a wide variety of web applications in books [23], e-commerce [32], online research papers [22] and films [12], lately. Jie Lu and al [17], for instance, have highlighted the implementation of a hybrid fuzzy semantic recommendation approach into an intelligent recommendation system that can recommend relevant business partners to individual business users. Probabilistic models can represent the uncertainty related to the user needs and characteristics modeling so that they are well adapted to the user modeling problem. Bayesian Networks in particular provide an effective approach for manipulating probabilistic models [26]. They have been used to predict the user's preferences and objectives in several systems [14, 30]. Horvitz et al [14] have used the Bayesian networks to infer the users' objectives and needs from their interactions and from the conditional probabilities models. In the work of Horvitz, a Bayesian model was built originally by psychologists from the observation of users in situations. In another research work [15], Horvitz et al proposed a system that predicts a user's intention in an uncertain environment.

Many approaches have been tried in the field of student modeling [4]. Among these, Bayesian networks have attracted a lot of attention from theoreticians due to their sound mathematical foundations as well as their natural way of representing uncertainty using probabilities models [19]. Andes [6] is a tutorial used in e-learning in case of a problem to help with Newtonian mechanics that assesses the student's state of knowledge. Millan et al [8] have provided education practitioners with the background needed to understand Bayesian networks and used them to develop students models. Nguyen and Do [24] introduced an approach that combines the Bayesian Network and the overlay model so that it is able to infer user's knowledge from evidence gathered during the user's learning process. [20] proposed a presentation of different systems that use Bayesian Networks to select the next teaching activity to students based on their knowledge level.

Furthermore, the Bayesian network has emerged as a simple and effective approach for handling uncertainty in context awareness [7]. Korpipaa et al [18] qualified a user's context using the naive Bayes classifier. Hong et al [13] proposed a context sensitive messenger that automatically infers the user's context.

In a recent research work [16], In-Jee Song et al. have proposed a model of the ubiquitous family environment in order to implement a context adaptive user interface. They have used a Bayesian network to provide the necessary devices in each location.

The knowledge models presented above may not assign reasons for the failure in the field of knowledge. The preference models, however, provide just information about the users' preferences without taking into account the evolution of these preferences. In this paper, an approach that successfully addresses the difficult problem of user's preferences detection using probabilistic models was introduced. In particular, we constructed a Bayesian user model that determined automatically the users' preferences. Our approach takes into consideration the preferences evolution and offers a better correspondence between a user's preferences and the provided information.

Although the probability of interaction via a web application is limited by the quality of service, we were able to collect the user's preferences of our adaptive user interface and build specific models to propose thereafter the web interfaces with the most adapted content and form for each user.

### 3 Bayesian Rules

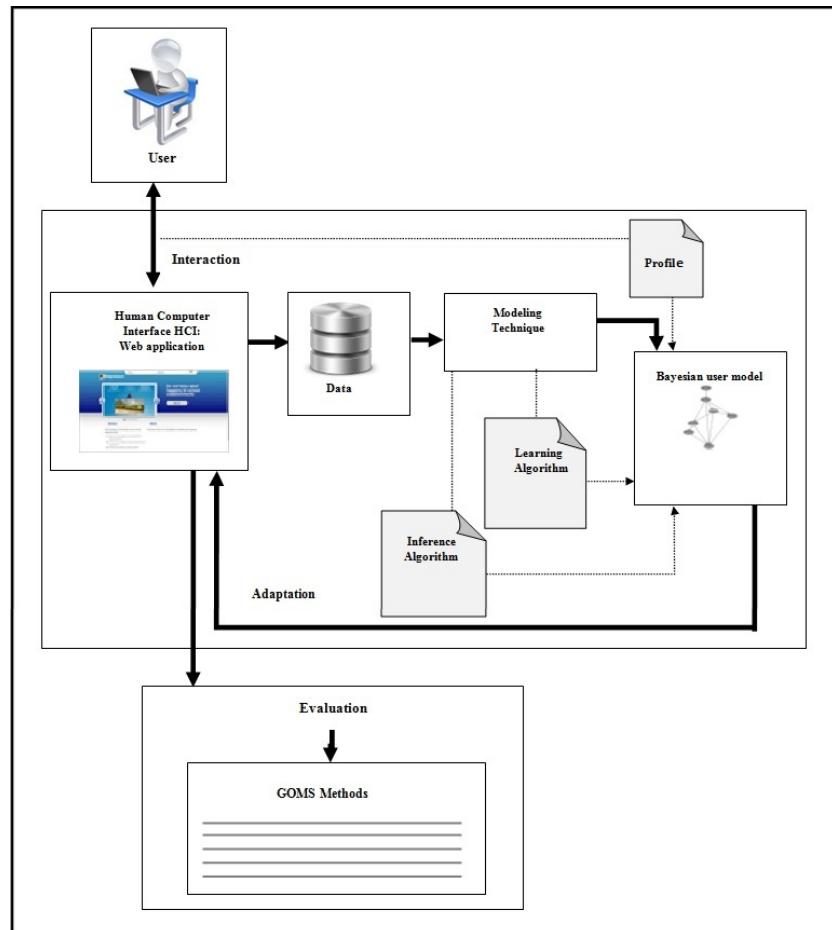
The Bayesian networks are probabilistic graphical models that consist of a qualitative part and a quantitative one. The qualitative is the structure of the network: a directed acyclic graph where nodes correspond to variables and arcs represent influences between the variables. Concerning the quantitative part, it provides the conditional probability tables that make up the network settings. The Bayesian Networks are powerful tools that take advantage of algorithms for inference and associated learning. The inference is based on the Bayes theorem to spread the knowledge within the network. The learning applies both a network structure and parameters that can be obtained from complete or incom-

#### 4 Building and evaluating an adaptive user interface

plete data. More precisely, a Bayesian Network is a set consisting of a directed acyclic graph and  $n$  random variables ( $X_1, X_2, \dots, X_n$ ) such that it is a bijection between the set of vertices graph and the set of random variables and that:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i)) \text{ where } pa(X_i) \text{ is the set of parents of } X_i \text{ in the graph.}$$

#### 4 The adaptive user interface building process



**Fig. 1.** The software architecture.

In order to collect the usage of a user's experience in an initial model, the data of our web interface was used. This web interface is a social study net-

work. Based upon a given experience, the user's interaction is used to update the parameters set of its model. Depending on the currently observed situation within the Bayesian model, the learning and inference algorithms were used to refine the Bayesian user model. The general software architecture is displayed in Figure 1. The Human Machine Interface (HMI) denotes the web interface with which the user interacts.

Thus, the modeling technique manages the overall process of adaptation using the data from the web interface so that it is able to determine which adaptations would be appropriate to the user. The Bayesian user model preserves the system knowledge about the preferences and the interactions of the end user.

Once the adaptive user interface was built, it was evaluated using the GOMS method which is one of the most widely accepted analytical modeling concepts in the Human Computer Interaction community.

## 5 Construction of bayesian variables: learning structure

We tried to determine the key variables in a user model to generate a user adapted model for a Web interface.

Therefore, we tried to pool and back up items for the user in a profile not as some variables but as a Bayesian Network model.

We can add the implementing rules to present these terms. The user preferences may therefore involve:

- The choice of services,
- The choice of graphic components not to be shown,
- The choice of publicity to display,
- The presentation of information preferences,
- The desired details level.

Assuming that our adaptive interface can infer the user's preferences, for example, from observations on a user's age, focus, gender and profession;

The idea is to assume that each user can be described by a profile. We start by defining some nodes of random variables that represent the profile. For example, we represent the node focus, with a marginal probability vector corresponding to the probability that the user has a center of interest:

$$P(\text{focus} = f_1), \dots, P(\text{focus} = f_n).$$

Next, we define the set of possible documents  $d_1; \dots; d_k$ . This can be modeled by defining another random variable called document. The choice of the documents to display depends on the user's focus, which makes us add an arc of causality between the two nodes (figure 2).

The Bayesian network could be constructed using an expert knowledge even though there was no data to train. After a sufficient user data was obtained, it was possible to train the Bayesian network. Generally, training Bayesian networks from data consists of two parts. The first is to train the structure and the other part is to train the parameter set. To train the structure can be solved by

6 Building and evaluating an adaptive user interface



**Fig. 2.** Representation of two nodes.

looking for the network structure which best matches the given data. To find the best structure with a score metric, learning algorithms can be used.

At the beginning, our database consists of 60 users, which means that it is not very large. As a result, for the test phase, we chose to use the cross validation process. The cross validation test was used to test the predictive power of Bayesian models using a single database.

Using the cross validation test, we also compared the different learning algorithms namely the K2 algorithm, the simulated annealing algorithm (SA), the genetic algorithm, the Tree Augmented Naive Bayes algorithm (TAN) and the Hill Climbing algorithm. The structure of our Bayesian user model is developed with Weka [33].

The results are shown in Table 1 below. The predictive power of the network depends on the type of algorithm used: It seems to be good with the simulated annealing algorithm (about 94 % correct values), less good with the K2 algorithm (about 91 % correct values) and only about 89 % and 87 % with the Hill Climbing and both of genetic and TAN algorithm, respectively.

Learning algorithm	percentage of correct instances
genetic algorithm	87.23 %
TAN	87.23 %
Hill Climbing	89.36 %
K2 algorithm	91.48 %
SA algorithm	93.61 %

**Table 1.** A comparison between the learning algorithms.

Hence, we used the simulated annealing algorithm to train the structure of our model. This algorithm (SA) is a global optimization technique which crucially allows the current solution to move to less optimal states based on a probability function, preventing a local optimal from restricting the algorithm. A simulated annealing algorithm is presented in Algorithm 1. The Simulated annealing algorithm makes analogies between the network contained in the structure ( $E$ ,

---

**Algorithm 1** A pseudocode for simulated annealing algorithm to train the structure of Bayesian model.

---

```

procedure SIMULATED ANNEALING()
    D // a Dataset
     $B_{current}$  // generate initial network
     $T_0, T_{final}$  :integer // set initial temperature
    N // N attempt for T given
    t=1
    repeat
        repeat
            //generate new network  $B_{new}$  from  $B_{current}$ 
            //evaluate  $E(B_{current}, D), E(B_{new}, D)$ 
            if  $E(B_{new}, D) \leq E(B_{current}, D)$  then
                 $B_{current} = B_{new}$ 
            else
                 $X = \frac{E(B_{new}, D) - E(B_{current}, D)}{T(t)}$ 
                //pick a number p chosen uniformly randomly in the range [0,1]
                if ( $e^{-x} > p$ ) then
                     $B_{current} = B_{new}$ 
                end if
            end if
            N=N-1
        until N=0
         $T_{t+1} = \tau(T_t)$ 
        t=t+1
    until  $T_{t+1} \geq T_{final}$ 
     $B_{final} = B_{current}$  //final network
end procedure

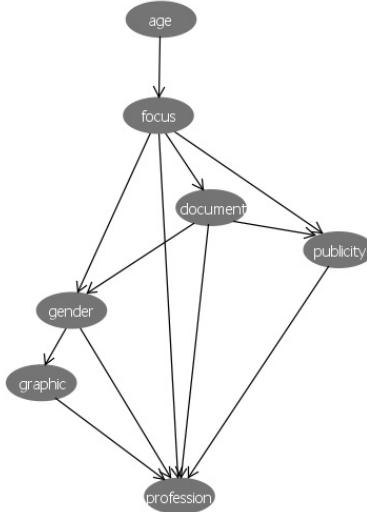
```

---

the function to be minimized), the configuration of the network structure (B) (E is a function of B) and temperature (T, an annealing schedule controlling the convergence of the algorithm to a final configuration).

Figure 3 shows the Bayesian network structure. The variables that build the Bayesian network are selected from what we saw interesting for the user. The variables represent the following states: "the member's profession" (profession), "the gender of the member" (gender), "the preferences of a member concerning the shown documents" (documents), and "the preferences of a member concerning the graphic components shown" (graphic), etc.

The previously collected data and the structure of the Bayesian model are used for the inference algorithm.

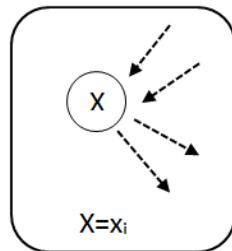


**Fig. 3.** The bayesian network structure.

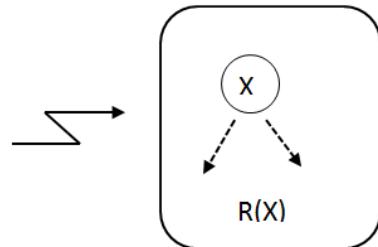
## 6 Bayesian inference: Soft evidence in our Bayesian model

Inference in the Bayesian framework entails the computation of the probability distribution over all variables given the available evidence (or set of observations).

Evidence is a collection of findings on variables. A finding may be soft or hard. A hard finding specifies which value a variable  $X$  is in. This hard finding represents the instantiation of  $X$  to the value  $x_i$  and it is characterized by  $P(X = x_i | e) = 1$  (figure 4)[1]. A soft finding is specified by a probability distribution  $R(X)$ (figure 5)[1,31].



**Fig. 4.** Hard finding on  $X$ .



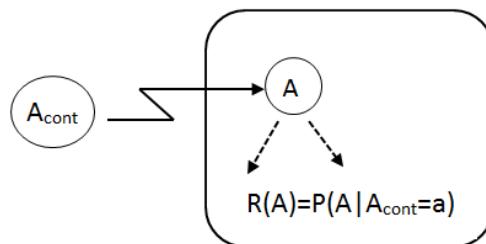
**Fig. 5.** Soft finding on X.

The term "Soft evidence" was suggested by Valtorta [31] and it is given as a distribution  $R(X)$ . This sort of evidence can be seen in many works [1, 25, 27, 31]. For example, we can know the distribution of a variable but we cannot observe its precise state. In addition, it is sometimes important to know the distribution of a variable at any given time.

### 6.1 Soft evidence coming from the fuzzy discretization

In our work, the basic context including profession, age, focus and gender could be considered as evidence to calculate the preference of each user.

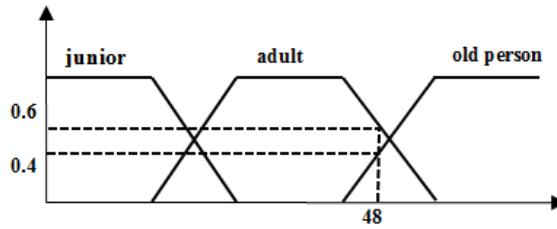
The variable age is a continuous variable discretized in the Bayesian network. A is a variable representing the age of a person with a value in  $D_A = \{junior, adult, oldperson\}$ .



**Fig. 6.** Soft finding on A: the observation of a continuous variable A is coming from a fuzzy discretization.

In Figure 6, the information comes from the observation of a continuous variable. The soft finding on A of the Bayesian network is specified by a probability distribution  $R(A)$  obtained from a fuzzy discretization of this continuous variable.

For example, the information "he is 48 years old" can be specified by a soft finding on age, by adding the knowledge of the fuzzy membership function



**Fig. 7.** The membership function of the variable "age".

linking the real age to the values of  $D_A$  (Figure 7). The result is for example:  $R(A) = (0, 0.6, 0.4)$ . This probabilistic finding is a soft finding because this distribution cannot be modified by other information. The variable P is a variable representing the profession of a person with value in  $D_P\{\text{student, teacher}\}$ . The soft finding P is specified by a probability distribution R (P) obtained from a fuzzy discretization of this variable.

## 6.2 Propagating soft evidence

Several algorithms for Bayesian network reasoning with uncertain evidences have been proposed. Most of them can be seen as combining the techniques of the Bayesian network reasoning, Jeffrey's rule, Pearl's virtual evidence method and IPFP algorithm (the iterative proportional fitting procedure). The IPFP algorithm can be applied for small Bayesian networks, because the probability distribution table should be changed at each iteration.

Here we presented the details of the  $BN - IPFP1$  algorithm [1,27]. This is the algorithm that was implemented in our adaptive interface. The advantage of the  $BN - IPFP1$  algorithm is that it can operate with any inference algorithm.  $(G, P)$  is a Bayesian network and X is all the variables associated with the nodes of the graph G. The  $BN - IPFP1$  algorithm is used to propagate a set of m soft evidences variables,  $\{Y_j, j = 1, \dots, m\} \subset X$ , specified by  $R(Y_1), \dots, R(Y_m)$  such as  $R(Y_j) \ll P(Y_j)$ . We used the following procedure: (see algorithm 2.)

## 6.3 Bayesian inference

Bayesian Networks can be used to reason about the situation of models when created. Inference algorithms are responsible for computing posterior probability distributions for all variables. In our work, we implemented the Junction Tree algorithm [28]. This is a general algorithmic framework, which provides an understanding of the general concepts that underlie inference. A variable with a higher probability is supposed to have more functions for generating an adaptive user interface than a variable with a lower probability. Thus, the Bayesian user

---

**Algorithm 2** Pseudocode for  $BN - IPFP1$  algorithm.

---

```

procedure BN-IPFP-1()
     $Q_0(X) = P(X)$ 
     $k = 1;$ 
    k: index of iterations
    repeat
         $j = 1 + (k - 1)modm; // j: index of current evidence$ 
         $l = 1 + (k - 1)/m; // l: index of iterations by evidence$ 

        Construct virtual evidence with likelihoodratio
         $L_{jl}(Y^j) = \frac{R(y_{(1)}^j)}{Q_{K-1}(y_{(1)}^j)} : \dots : \frac{R(y_{(j_x)}^j)}{Q_{K-1}(y_{(j_x)}^j)}$ 
        // where  $y_{(1)}, \dots, y_{(j_x)}$  are states configurations de  $Y^j$ 
        //Combining the previous likelihood reports  $Y^j$ 
         $L_k(Y^j) = \prod_l L_l^j(Y^j)$ 
        // Obtain  $Q_k(X)$  by updating  $Q_{k-1}(X)$  with  $L_{jl}(Y^j)$  using Pearl's virtual
        evidence method
         $k = k + 1;$ 
    until (convergence)
end procedure

```

---

model is ready to be used to predict the user's preferences (the document to show, the publicity,...). Graphics, documents and adverts are changed because the user's preferences increased.

Our web interface is adaptive in the sense that it offers suggestions to the user to choose their preferences. In the case of a new member who did not specify his preferences, the Bayesian user model must be used to determine the default values as assigned.

## 7 Experiments

To validate our proposal, several experiments were performed. Before discussing the results, we first described the experimental setup. Comparing the proposed adaptive user interface with the conventional fixed user interface, we confirmed that the adaptive interface had a superior performance.

### 7.1 Experimental setup

The web interface, which is a social study network, allows exchanging documents and courses between students and teachers, publishing courses and finding conferences, master topics and thesis topics, etc. Table 2 shows the attributes making up each object in the Bayesian model. The database was built by extracting information from our web site.

12 Building and evaluating an adaptive user interface

Variables in data base		attributes
Age		junior, adult, old person
Focus		computerScience, mechanic, electric, economicManagement, medicine, mathematics, physicsChemistry, physicalTraining, artsHumanities
Gender		Female, male
publicity		computerScience, mechanic, electric, economicManagement, medicine, mathematics, physicsChemistry, physicalTraining, artsHumanities
Graphic document		blue, green, yellow, orange computerScience, mechanic, electric, economicManagement, medicine, mathematics, physicsChemistry, physicalTraining, artsHumanities
Profession		student, teacher

**Table 2.** The variables in data base.

## 7.2 Evaluation of the adaptive user interface

A formal GOMS model approach was applied to the evaluation of our user interface for a specialized web application. This approach was able to identify significant usability problems embedded in the procedures by which users interact with the interface. The GOMS model concept, originally introduced by Card, Moran and Newell [4], is one of the most widely accepted analytical modeling concepts in the Human Computer Interaction community. The GOMS model describes the procedural knowledge required by a user to interact regularly with a software interface in terms of involved actions, the user's goals, the methods required to accomplish the user's goals, and the rules for selecting among competing methods. To evaluate the usability of the proposed interface, we assumed some situations that could happen in our web interface. Ten subjects (five males and five females) participated in this evaluation. Thus, using GOMS, we evaluated the generated adaptive interface based on the required time to complete the predefined tasks in those situations. The users and our profiles are presented in table 3 and the situations and detailed tasks are presented in table 4.

Using these situations and tasks, we evaluated the fixed user interface and the adaptive user interface using the required time to conduct the same tasks. Figures 8 and 9 show examples of the adaptive user interface when changing the functions; the layouts were changed according to the user's preferences. Figure 7 presents the fixed user interface given a user 1 (age = junior, profession= student, gender=female, focus=computer science). In the adaptive user interface (figure 9), the graphic, the documents and the publicity are changed according to the user's preferences.

In this figure, for example, the interface color of this user changed from blue to orange and the document and publicity are concerned with the computer science field.

No. user	age	gender	profession	focus
1	junior	female	student	computerScience
2	adult	female	teacher	electric
3	junior	female	student	electric
4	junior	female	teacher	computerScience
5	adult	female	teacher	mathematics
6	junior	male	student	computerScience
7	old person	male	teacher	mechanic
8	junior	male	student	economicManagement
9	junior	male	student	medicine
10	junior	male	student	computerScience

**Table 3.** Contextual information for ten users.

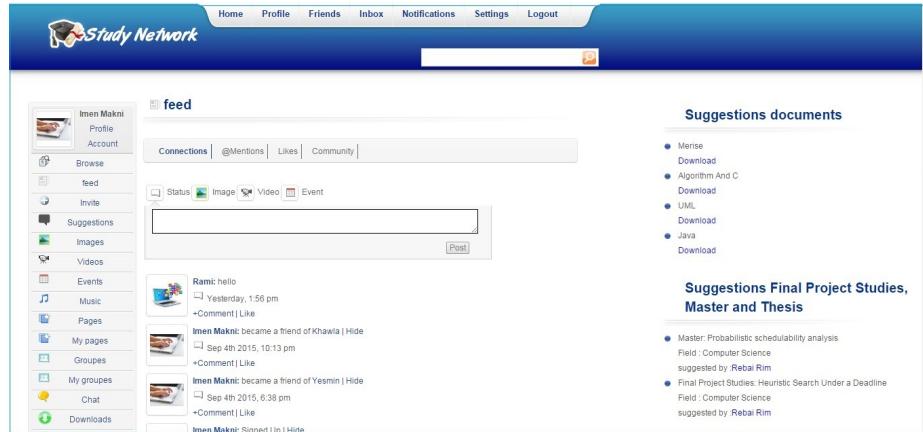
No	situation	Detailed task for situation
1	download a course	click the icon document → search among the courses offered what is appropriate → download the course
2	change the color of the graphical user interface	click the icon my account → change the color of the graphical user interface
3	find a proposal for thesis	click the icon suggestion → search among the thesis topics proposed what is appropriate
4	find a proposal for laboratory	click the icon suggestion → search among the laboratories proposed what is appropriate
5	send a message	click the icon message → send a message
6	find pages or groups	click the icon pages or groups → give the specialty of study → search among the pages or the groups offered what is appropriate
7	add a course in the website	click the icon document → add a course by specialty
8	find a conference	click the icon suggestion → search among the conferences proposed what is appropriate
9	add a new video	click the icon video → share a new video
10	add a new event	click the icon event → share a new event

**Table 4.** Ten situations and detailed tasks for usability test.

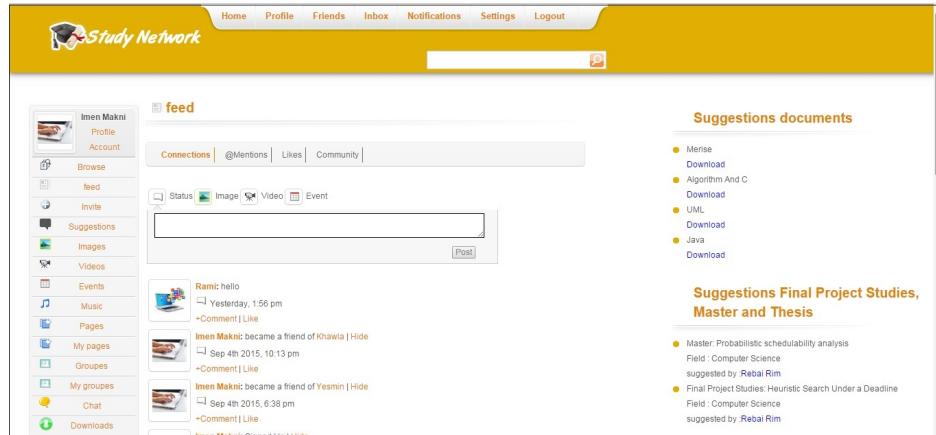
Then, we compared the required time for ten users to perform the tasks in the situations provided in Table 4 using the fixed user interface and the adaptive user interface. The results are presented in figure 10. Figure 11 shows the reduction rate of the time required when using the proposed method with the adaptive user interface. Then, when using an adaptive interface, we achieved 31.07% reduction rate in the time required compared with a fixed user interface.

Then, we carried out a survey by questionnaire for each situation to measure the attributes correctness and the user's satisfaction. The same ten users participated in the survey.

14 Building and evaluating an adaptive user interface

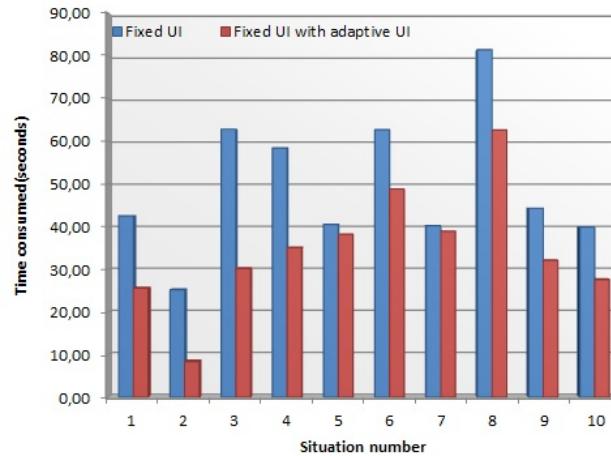


**Fig. 8.** A fixed user interface.

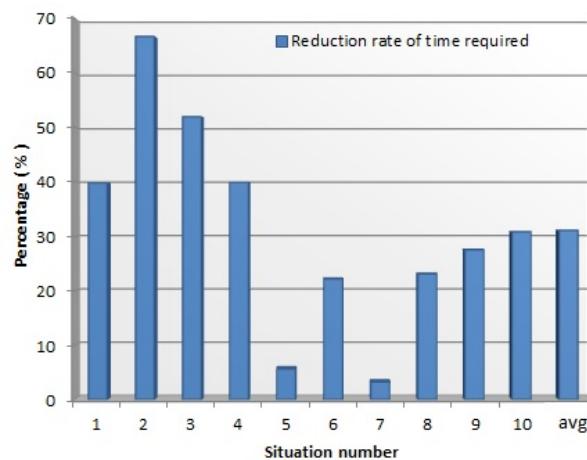


**Fig. 9.** A user interface including the adaptive interface generated by the proposed method (bayesian networks).

- The attributes correctness: The Bayesian inference was conducted to predict the user's preference (publicity, graphic, etc.). As we said, an attribute with a higher probability is supposed to have more functions to generate a user interface than an attribute with a lower probability. We compared the result of each situation achieved in the adaptive user interface with the result in the survey and we measured the percentage of correct attributes with high probability for each situation. As shown in table 5, the best results were achieved where the high probability attributes were correctly structured with about 70%



**Fig. 10.** Time to perform the tasks in ten situations.



**Fig. 11.** Reduction rate of the time required to complete the tasks in ten situations.

- The user's satisfaction: The ten users give a score to the difference between the fixed UI and the adaptive UI created by the proposed approach for the ten situations. For this test, the users were shown the two interfaces, and then we requested the users for a more appropriate personal evaluation of the two interfaces, which closely depends on their preferences satisfaction, over a 10-degree scale in each situation. We can observe a relatively good satisfaction with the adaptive interface since we got a mean score of 7.6 (table 6).

16 Building and evaluating an adaptive user interface

No.situation	1	2	3	4	5	6	7	8	9	10	avg
Attributes correctness	78%	87%	81%	76%	49%	54%	68%	59%	69%	78%	70%

**Table 5.** The percentage of attributes correctness.

No.situation	1	2	3	4	5	6	7	8	9	10	avg
Score of users	9.2	7.1	8.9	6.4	7.9	6.8	5.2	8.7	9.5	6.3	7.6

**Table 6.** The user's satisfaction.

## 8 Conclusion

In this paper, we presented our research and we proposed an adaptive user interface using the Bayesian network approach. We first proposed a comparison between the learning algorithms in order to train the Bayesian network structure and we implemented the bayesian inference in our web interface. We also validated our approach by evaluating the Web interface. Comparing to the fixed user interface, we achieved a 31% reduction rate in the time required to complete the tasks when using the adaptive user interface. We also found that the proposed interface is more satisfactory than the fixed one. Then, we needed to improve the model with a better training. A log of the users' activities was used to specify the Bayesian network. In order to improve the system evaluation we will include some qualitative measures on the user experience and satisfaction from a verbal user feedback.

As a perspective, we propose to take advantage of the Cloud Computing since it is a new paradigm which tries to provide a personalized service relative to end users.

## References

1. Ben Mrad, A. , Delcroix, V., Piechowiak, S., Leicester, P., Abid, M. : An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence*, 1-23 (2015).
2. Brown A. L. : Learning, remembering and understanding. In J. H. Flavell et E. M. Markham (dirs.), Carmichael's Manual of Child Psychology. New York :Wiley (1982).
3. Brusilovsky, P. , Eklund, L. , Schwarz, E. : Web-based Education for all: A tool for development adaptive courseware. *Seventh International World Wide Web Conference. Computer Networks and ISDN Systems*. 30, 291–300 (1998).
4. Brusilovsky, P., Milln, E. : User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), *The Adaptive Web*. vol. 4321 of lecture notes in computer science, Berlin, Heidelberg: Springer. 3-53 (2007).
5. Card, S., Moran, T. P., Newell, A. : *The psychology of human computer interaction*. Lawrence Erlbaum Associates (1983).

6. Conati, C. , Gertner, A. , Vanlehn, K. : Using Bayesian networks to manage uncertainty in student modeling. *Journal of User Modeling and User-Adapted Interaction*. volume 12 (4), 371–417 (2002).
7. Dey, A. K. : Understanding and using context. *Personal and Ubiquitous Computing*, 5, 2024 (2001).
8. Millà, E. , Loboda, T. , Prez de la Cruz, J. L. : Bayesian networks for student model engineering. *International Journal of Computers and Education*, 55, 1663-1683 (2010).
9. Fischer, G. : User modeling in human-computer interaction. *Journal of User Modeling and User- Adapted Interaction*, 11, 65-86 (2001).
10. Fink, J. , Kobsa, A. : User modeling for personalized city tours. *Artificial Intelligence Review*. 18, 33–74 (2002).
11. Gonzales, C. , Wuillemin, P.H. :Rseaux baysiens en modlisation d'utilisateurs. *Sciences et Techniques Educatives*. Volume 5, n2, 173 (1998).
12. Good, N., J.B.Schafer, J.Konstan, A.Borchers, B.Sarwar, J.Herlocker, J.Riedl : Combining collaborative filtering with personal agents for better recommendations. In Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Conference on Innovative Applications of Artificial Intelligence, American Association for Artificial Intelligence, Orlando, FL, 439-446 (1999).
13. Hong, J.H., Yang, S.I., Cho, S.B. : ConaMSN: A context-aware messenger using dynamic Bayesian networks with wearable sensors. *Expert Systems with Applications*, 37, 46804686 (2010).
14. Horvitz, E. , Breese, J. , Heckerman, D. , Hovel, D. , Rommelse, K. : The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. *Proceedings of UAI '98: the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, Wisconsin, USA. (1988).
15. Horvitz, E. , Kadie, C. M. , Paek, T., Hovel, D. : Models of attention in computing and communications: From principles to applications. *Communications of the ACM*, 46(3), 5259 (2003).
16. In-Jee, S. , Sung-Bae, C. : Bayesian and behavior networks for context-adaptive user interface in a ubiquitous home environment. *International Journal of Expert Systems with Applications*, 40, 1827-1838 (2013).
17. Jie, L. ,Qusai, S. ,Yisi, X. ,Qing, L. , Guangquan, Z. : A Web-Based Personalized Business Partner Recommendation System using Fuzzy Semantic Techniques. *International Journal of Computational Intelligence*, Volume 29, n1, 37-69(2013)
18. Korpiapaa, P., Koskinen, M., Peltola, J., Makela, S. M., Seppanen, T. : Bayesian approach to sensor-based context awareness. *Personal and Ubiquitous Computing*, 7(2), 113124 (2003).
19. Liu, C. I. : Using bayesian networks for student modeling. In R. M. Viccari, P. Augustin-Jaques, R. Verdin (Eds.), *Agent-based tutoring systems by cognitive and affective modeling*, 97-113 (2008).
20. Mayo, M. , Mitrovic, A. : Optimising ITS behaviour with Bayesian networks and decision theory. *International Journal of Artificial Intelligence in Education*. 12, 124–153 (2001).
21. McTear, M.F. : User modelling for adaptive computer systems: a survey of recent developments. *Artificial Intelligence Review*. 7, 157–184 (1993).
22. Middleton, S., N.Shadbolt, D.DE Roure : Ontological user profiling in recommender systems. *ACM Transactions on Information Systems*, 22, 54-88 (2004).
23. Mooney, R.J.,L.Roy. : Content-based book recommending using learning for text categorization. In Proceedings of the Fifth ACM Conference on Digital Libraries, San Antonio, TX, ACM, pp. 195204 (2000).

- 18 Building and evaluating an adaptive user interface
24. Nguyen, L. , Do, P. : Combination of Bayesian Network and Overlay Model in User Modeling. ICCS (2), pp.5–14 (2009).
  25. Pan, R., Peng, Y., Ding, Z.: Belief update in Bayesian networks using uncertain evidence. In: ICTAI. pp. 441–444 (2006)
  26. Pearl, J. : Probabilistic reasoning in intelligent systems – Networks of plausible inference. San Mateo, CA: Morgan Kaufman. (1988).
  27. Peng, Y., Zhang, S., Pan, R.: Bayesian network reasoning with uncertain evidences. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 18(5), 539–564 (2010)
  28. Rebai, R., Maalej, M., and Mahfoudhi, Adel. Bayesian networks for user modeling : Predicting the user's preferences. International Conference on Hybrid Intelligent Systems, HIS'13, 144-148. 2013
  29. Rich, E. : Stereotypes and User Modeling. chapitre 2, pp. 35–51. Springer Verlag. (1989).
  30. Sbattella, L. , Tedesco, R. : Profiling and tutoring users in Virtual Campus. Proceedings of the 5th International Conference on Information Technology Based Higher Education and Training (ITHET '04), Istanbul, Turkey. (2004).
  31. Valtorta, M., Kim, Y.G., Vomlel, J.: Soft evidential update for probabilistic multiagent systems. International Journal of Approximate Reasoning 29(1), 71–106 (2002)
  32. WEI, K. , Huang, J. , Fu, S. : A survey of e-commerce recommender systems. In Proceedings of the International Conference on Service Systems and Service Management, Chengdu, China, pp.1-5 (2007).
  33. Weka, <http://www.cs.waikato.ac.nz/ml/weka/>

# A new approach in Steganography of digital images using saliency map algorithm and Pixel Neighbors

Meysam Ghasemi<sup>1,2</sup>, Mehran Emadi<sup>1,3\*</sup>

<sup>1</sup>Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran.

<sup>2</sup>[mr.meysamghasemi@hotmail.com](mailto:mr.meysamghasemi@hotmail.com)

<sup>3</sup>Faculty of Electrical Engineering, Mobarakeh Branch, Islamic Azad University, Mobarakeh, Iran.

[m.emadi@mau.ac.ir](mailto:m.emadi@mau.ac.ir)

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

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

## I. INTRODUCTION

Steganography is considered as the art and science of hiding information within a media agent that requires the slightest change in the appearance of the medium to be exchanged safely between source and destination. The aim is to cover up the relationship through putting the information in a carrier medium with the lowest media file quality loss. Accuracy for preventing the discovery of the hidden message is of great significance in the process of steganography. Steganography is a branch of science called hidden communication [1]. Importance of steganography in a medium (carrier file) as a method in storing information is in transferring them between two points with high security. Due to the rapid growth of science, technology, and security issues, hiding information using new methods with high security is essential. To this end, different methods have been proposed for steganography, each with their own strengths and weaknesses. In this study, the aim is to increase security in steganography in images and the most valid methods have been used to create an algorithm with minimum error and maximum security [2]. Historical records depict that Steganography has a long history and that human beings used various secret methods to exchange confidential information. Some of these methods will be briefly expressed here. In Ancient Greece, people were used as a secret message carrier. First, they shaved the messenger's head and tattoo secret message on their scalp; after the growth of hair, they were sent to the recipient and the recipient shaved the person's hair successively in order to access the message. The drawbacks of this method may be to the long process of writing and receiving the message and the long permanence of the message on the person's head as well. Another way of the ancient Greeks was writing the message on a tablet and then covering it with wax; the recipient simply had to push the wax away to access the message. In ancient Rome and also during World War II, things such as milk, vinegar, and urine were consumed as invisible ink to insert secret information between the lines of a letters. In this regard, heating the paper to some extent could reveal the message [4].

---

\* Corresponding author

L. Bourdev Et al. (2011) revealed that specifications are practical for diagnosing new objects that have not been seen before and that characteristic-oriented methods are perfectly extended to object classifications. They presented an object-diagnosis design in which characteristic classifications were taught utilizing chosen specifications (for each characteristic, one classifier) and the object's classification was determined through single-characteristic classifications. This method did not achieve good results in terms of accuracy and is applied on color images [11]. Yang and Mori (2010) presented an educated hidden model discriminatively for connecting object classes modeling and their visual specifications. In this method, the specifications of an object are deemed as a latent variable and correlations of their specifications are then obtained. This method also had weaknesses, one of the most important of which was that the images were examined as latent variables and it sometimes leads to wrong results in some of the images. In this paper the focus is on color images [2]. Farhadi et al (2012) stated an approach for localizing component-based specifications, such as the head, or a wheel. They taught detectors for secondary sectors and categories and used classifiers output to present to an object; for example, detectors "feet", "head" and "dog" are willing to vote to a "quadruped" object. In their work, the specifications used for the voting should be semantic and be well-excessed to similar images. The applied method is a proper one for diagnosing specifications in images and color images have been used in the investigation [1]. Ferrari et al (2008) diagnosed and localized low-level specifications in a productive context. They could diagnose color and texture of objects with high accuracy, in which their database was color images [4]. Tsagkatakis and Savakis (2012) use SRC (Sparse Representation Classifier) instead of SVM. They use SRC to predict the characteristic and examining its benefits for across-category generalization, and could gain accuracy above 90% in it. They mainly conducted their study on color images [7]. Joo et al. (2013), present a component-based method for identifying human with a single photo of the person's body. They suggest learning a rich photo bank from body organs with less supervision. They do this through decomposing the image network to windows that overlap (in various scales and with repeated refining of local appearance forms). They were able to obtain the best results ever for the diagnosis and working on the organs of the body [8]. Kumar et al (2009) suggested using facial specifications for diagnostic of the face. They educated the classifier for two types of specifications: visual appearance, such as "white" or "obese", and also similar expressions or areas of the face. These specifications were later used for face diagnosis. Images they processed were colored images of the human body and the results were satisfactory and impressive up to that time but later other researchers were not able to improve their technique and achieve better results [9].

## II.SYSTEM MODEL AND ASSUMPTIONS

The performance of the suggested method initiates with the entrance of the carrier received image. Receiving images is an important stage in content-based steganography system. Also, quality and clarity of the images entered to the operating system is a determining factor in decreasing errors. If images proper for the needs of the system are used, they do not need to be tailored, and this will increase speed and accuracy.

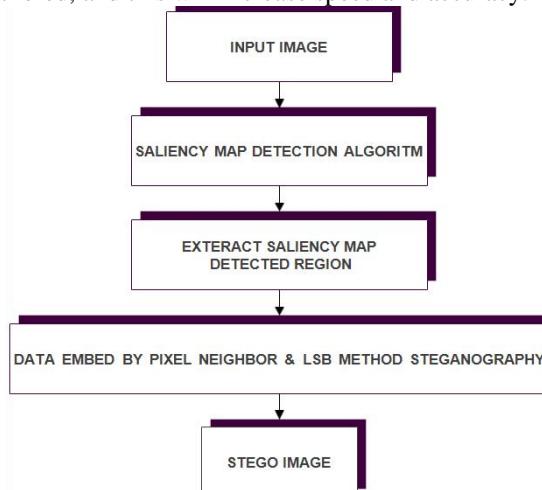


Figure 1. Proposed method flowchart

#### A. Special section in picture carrier detection

In this paper, a new method is applied to diagnose prominence. In contrast to the previous diagnosis of objects prominences that were based on a spotlight theory, in the proposed approach an attention theory based on objects is followed and the concept of object is used directly in prominence criteria.

Main objects are used as units of analysis, with a prominent element in it defined as an attached area of the image, which can turn into a possible object or a part of the object when focused on. As the object-based attention theory suggests we started with dividing a complex object into main objects and then the prominence for each main part will be examined. The most prominent main object will be considered as a prominent object.

Then the bass contour method is used for a detailed drawing of the selected object area and finally the exact area of the emphasized object will be obtained. In the method used in special section diagnosis (Prominence) in the images in the steganography system design, it is attempted to distinguish two types of views about the object prominence.

1: First, an object is prominent, if its pixels in terms of contrast and color range differ from the pixel neighbors, which is called enclosed in the center.

2: Secondly, an object is prominent, if it contains different or unique specifications, which is called merged prominence.

These two types of object prominences have supplement specification, in addition, combining these two, in the proposed algorithm, will create a new level in prominent object diagnosis and will function quite differently from conventional methods in the field.

#### B. separating Special area of the image from the original image (image segmentation)

Main objects are extracted through dividing an image into consistent vision areas that are possible candidates for prominent objects. To do so, hierarchical segmentation was used to find proper candidates for the object location that was based on two key ideas.

1 The objects can be of any size and scale. Therefore, a hierarchical segmentation strategy is used and all parts of the hierarchy are considered.

2 In order to explain different aspects of the image and its conditions, results of additional segmentation will be combined.

This strategy works for focusing the object. To create merged prominence maximizing information approach for detailed quantitative assessment of the object. In fact, image locations that are apart from other parts of an image should be prominent.

In the majority of cases examined, prominence is defined based on the maximum sampling of information. They calculate Shannon personal information according to the local image content possibility in a piece, with regard to the content of the rest of the image. Pieces with unexpected content include more information and are thereafter more prominent. Color histograms are used to reflect the image content.

Here a set of described main objects are created. Next, enclosed in center and merged prominence of each object is estimated. The enclosed in center prominence displays how much a part is different from its surrounding. Merged prominence shows if there are any unique details in that part.

These two evaluations complement each other. Thus, both are combined through adding scores of each prominence to gain the final value. The part with the highest final value is chosen as the most prominent object (MPO).

In fact, in the algorithm proposed, a framework for diagnosing the prominent object, inspired by the object-based visual focus theory, is suggested. A main object is considered as a unit of focus and it is noted that the concept of an object should be considered when evaluating object prominences.

In addition, two types of prominence are considered: enclosed in center prominence evaluates the differences an object has with its surroundings, and merged prominence examines the number of unique specifications in an object. In the following figure the algorithm proposed for diagnosing and separating the part proposed in the paper.

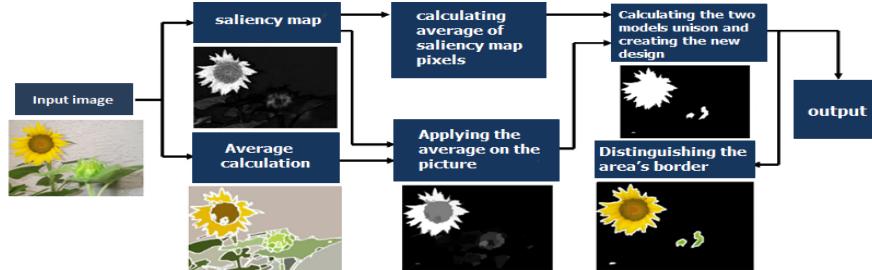


Figure 2. Steps of applying prominence diagnosis combined algorithm in the image

As shown in the figure, this process includes six main steps. All the standards of prominence diagnosis are observed in the algorithm and in this level even the area related to the border points of the prominent area can be extracted.

### C. Embedding information in the picture carrier after removing the prominent object from the image

Before embedding the information, the part selected as the prominent part in the image under process is ignored. In the proposed algorithm to embed information in the output file combined structure is used as analysis and information placement in RGB format.

In general, the system follows the following steps of information embedding. This diagram shows the simplified image of the proposed algorithm.

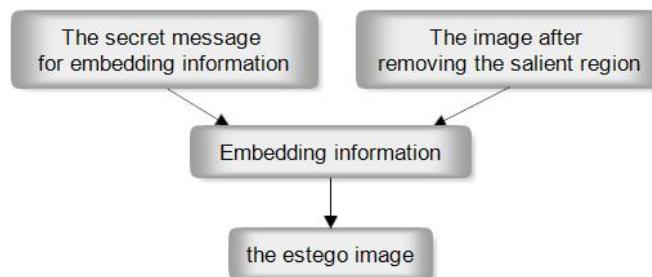


Figure 3. Diagram of embedding information in the picture after removing the prominent area form it

The purpose of embedding information in RGB is increasing the volume of storing in the three B, R, G bands. Although changes in the stego image are not obvious, a disadvantage of this algorithm is that through processing the image, for example, resizing, smoothing the image, compression, etc., some of the hidden information might be damaged and it may be irreversible.

The main levels of the proposed algorithm for diagnosing prominence in the image and steganography are:

1. Signing raw image under process to the system, at this stage, after applying a second category smoothing filter (Wiener filter can also be used) the image will move to the next step.
2. Specifying the image prominence using hybrid method (areas communal).
3. Calculating the average color pixels of each part of the image to determine the main parts of the image under process. Averaging operators are divided into two main categories.

A.Definitive and irreversible methods

B.Temporary and indefinite methods (this method is also called incidental or optimization method)

Filters available in the first category depend on the averaging the radius and do not require information about the image structure and noises in the data. In contrast, incidental operators are used for making the best decision in separating the prominent part of the image and noises.

4. Calculating the average using the prominent map gained from the second stage. Here abundant pixels (ineffective pixels) can be deleted and calculations can be done in smaller intervals that would increase the speed and accuracy of the system.

5. Applying the result of averaging on the output of the prominence map shown in the image. In this stage, a common point is captured between pixels of the area shared in the output of averaged situation of the image on the diagnosed map. Thus, the unshared parts between the two images are deleted.

6. In this stage the area output in stage 4 and algorithm output in stage 5 are subscribed, and the result is vividly shown in the figure. Through this, the complete model of prominence in the image is extracted, whose result will be used in the following stages.

7. In the last stage, the border of the prominent area can be determined through combining the sub space of prominence and the discovered prominent area. Here, the output can be used to draw the border of shared areas using snake algorithm.

In the algorithm proposed active cantor method is associatively used and the attempt has been to determine the most effective parts of the image, and area separation lines be drawn using active cantor algorithm. This selected area is ignored when embedding data.

8. The steganography information is received from the user and using pixel neighbors and LSB the information is hidden in worthless bits of the image resulted from the previous stage, and the image with hidden information is obtained.

### III. RESULT AND DISCUSSION

To conduct the tests and examination of results, 6 images have been selected as sample images.

In the images, we tried to follow these guidelines so that the test results encounter less problems and non-systematic errors:

- 1 Pictures had the same resolution.
- 2 Only one object in each image will be selected as prominence.
- 3 Bright and dark images are in the test for a better evaluation of the results.

#### A. Computing capacity, MSE And PSNR in the test sample images

According to the system performance and calculations for comparing results and drawing a comparison table, the six selected images were given to the system as the prototypes and the results are shown in the table below. The calculations are carried out in the same circumstances. It should be noted that general specifications of each image are given in the table to get a more precise comparison measure for the proposed algorithm results. The information includes embedded volume-bit, PSNR and MSE, compared through 6 ways.



Figure 4. Comparison charts with 32,000 bits of data

#### IV. CONCLUSION

In this paper, the proposed algorithm of diagnosing objects in the image and also steganography methods were examined. First, specifications, weaknesses, and the strengths of each specification, as well as, steganography methods were examined and each method's problems were pointed out. Applying the above stages and adding a new filter to embedded data (prominence diagnosis) which leads to a different way of embedding data using carrier file (carrier picture), is an innovation in steganography, leading to a higher security of the proposed algorithm. Ultimately, the operation and the accuracy of the proposed algorithm were tested and appropriate results were obtained. This method can be used as an efficient algorithm in this field.

#### REFERENCES

- [1] A.Farhadi, I.Endres, and D.Hoiem, "Attribute-Centric Recognition for Cross-category Generalization.", CVPR 2010.
- [2] W.Yang, and G.Mori. "A discriminative latent model of object classes and attributes." Computer Vision-ECCV 2010. Springer Berlin Heidelberg, 2010. 155-168.
- [3] W. Jiayun and LK Bin, "A color grouping method for detection of object regions based on local saliency," in 12th International Conference on Control Automation Robotics & Vision (ICARCV), 2012 New Jersey, pp. 1165-1169, 2012.
- [4] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation.", In CVPR, 2008.
- [5] K.Duan, D.Parikh, D.Crandall, and K.Grauman, "Discovering Localized Attributes for Fine-grained Recognition", In CVPR, 2012.
- [6] Y.Sun, "Attribute Based Object Identification", Liefeng Bo and Dieter Fox, Robotics and Automation (ICRA), 2013 IEEE International Conference on.
- [7] G.Tsagkatakis, and A.Savakis. "Sparse representations and distance learning for attribute based category recognition." Trends and Topics in Computer Vision. Springer Berlin Heidelberg, 2012. 29-42.
- [8] J.Joo, S.Wang, and S. Zhu, "Human Attribute Recognition by Rich Appearance Dictionary", International Conference on Computer Vision (ICCV), 2013.
- [9] N.Kumar, AC Berg, PN Belhumeur, and SK Nayar, "Attribute and Simile Classifiers for Face Verification", ICCV 2009.
- [10] A. Gallagher and T. Chen, "Estimating age, gender and identity using first name priors.", In CVPR, 2008.
- [11] L.bourdev, S.maji, J.malik, "Describing People: Poselet-Based Approach to Attribute Classification", Iccv the 2,011th.

## Performance Improvement in MapReduce via Overlapping of Mapper and Reducer

Saurabh Gupta, Manish Pandey

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

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

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

### I. INTRODUCTION

In day to day life, the amount of data being generated from numerous essential zones, including e-business, finance, hospitality data, CCTV cameras, education and environment, is colossal. To obtain important and relevant information, it is important to process this bulk of data in order to make business decisions or improve end user services. In recent years, a large number of computing frameworks [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] have been developed for big data processing. Among these frameworks, MapReduce [1] (Hadoop) is the most extensively used framework because of its simplicity, effortlessness and scalability. Hadoop framework is suitable for variety of algorithms, including large scale image processing [11], relational query evaluation [12] and web scale document analysis [13]. Apart from these domains, there are many applications i.e. PageRank[14], internet traffic analysis[15], social network analysis[16], neural network analysis[17], clustering[18], recursive relational queries[19] and Hypertext Induced Topic Search (HITS)[20] which require iterative calculations. Hence it is necessary to process these applications in a very efficient way. Hadoop uses the distributed architecture and process this huge data in cluster environment [21]. Hadoop framework divided into two parts to handle this huge information:

- Hadoop Distributed File System
- MapReduce.

#### 1) Hadoop Distributed File System(HDFS)

The Hadoop distributed file system is a fault tolerant and scalable file system for the MapReduce framework that is proposed to run on commodity hardware. In this file system; input data is broken into small pieces known as blocks. The default size for the block is either 64MB or 128MB. For data availability, it also supports data replication. By

default, every block is replicated three times across the cluster. According to data replication scheme, as shown in figure 1, first replica is created on a random machine where the space is available. The second replica of data block is produced on the same rack but on the different machine. The third replica of block is created on the different rack from the first replica is available.

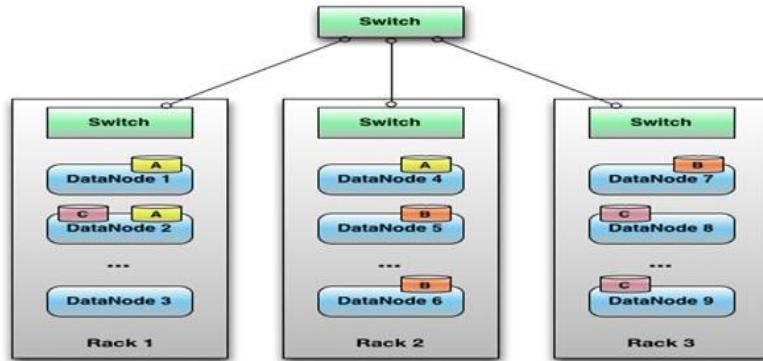


Fig. 1: Data replica concept

This replication technique is used to maintain the availability of our data at all times, because it is very rare to fail for machines of two different racks at the same time. HDFS supports master slave architecture in distributed environment and offers two daemon services:

- master daemon service
- slave daemon service

Both the NameNode and the SecondaryNameNode are included in master daemon service. As shown in figure 2, the responsibility of NameNode is to execute the file system operations like opening, closing and renaming the directories and the files. NameNode holds the metadata for file system like block and namespace information etc. When NameNode is in running state, all the information is in main memory, but it is also kept in secondary storage.

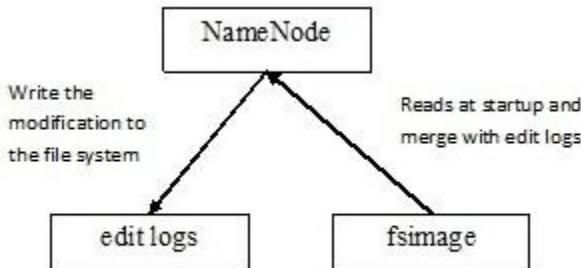


Fig. 2: Working of NameNode

There are two files where the updatons must be required by NameNode:

- fsimage – It is snapshot of HDFS when NameNode is started.
- edit logs – After starting of NameNode, the modifications made to file system is kept in the edit logs file.

In real world, it is very rare to restart the NameNode; hence the size of edit logs may be very big when NameNode runs for long period. In the case of system crush, more information can be lost because there is no updation in the

fsimage file (i.e. update fsimage file when NameNode restarted). So to avoid these situations SecondaryNameNode is used. Hence, as shown in figure 3, it is responsibility of SecondaryNameNode to merge edit logs with fsimage.

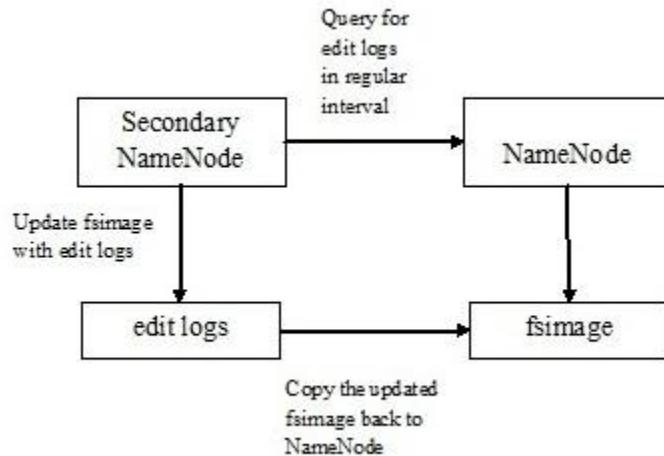


Fig. 3: Working of SecondaryNameNode

The slave service includes TaskTracker and DataNodes and these two can communicate with each other. The responsibility of DataNode and TaskTracker is to store the data in its local machine and to process the data that is stored in the local machine respectively. NameNode of the master service can only communicate with the DataNode of the slave service and vice versa. Similarly, JobTracker of the master service can only communicate with the TaskTracker of the slave service and vice versa.

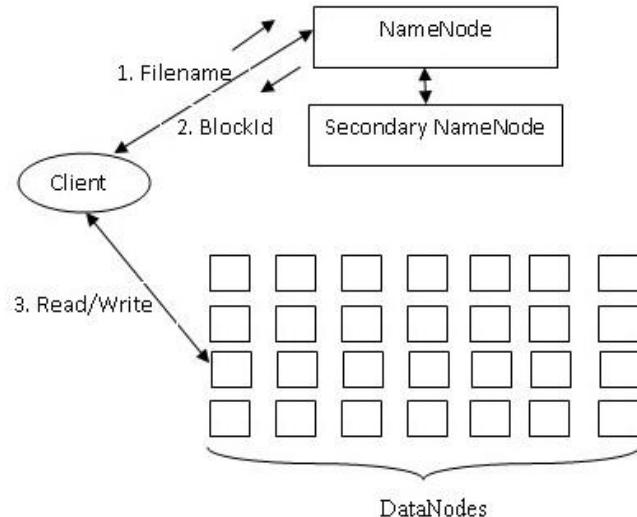


Fig. 4: Architecture of HDFS

Figure 4 illustrates the working of HDFS. When client sends a request to NameNode for storing the data, NameNode gives the location of free spaces where the client can store his/her data. After getting the location of these free spaces, client stores the data on these locations and NameNode maintains the information where the data is stored for this particular client. When the client submits its request to process the data, then JobTracker communicates with NameNode for the location where data is stored for that particular client. After getting the location, JobTracker assigns the job to TaskTracker. TaskTracker processes the requested data and the results are given back to the client.

## 2) MapReduce

After storing the data in HDFS, it needs to be processed. MapReduce processing is used to process this huge data. MapReduce consists of two functions, map function and reduce function. Map function takes the `< key; value >` pairs as input and produces `< key; value >` pairs as output. In the figure 5, input data is split into small pieces of information known as blocks. A record reader module reads the data from the blocks as input and produces the corresponding `< key; value >` pairs as output. Only one record can be processed by record reader at a time. After the record reader processed the input data, `< key; value >` pairs are given to map function as input for the further processing. The mapper function processes input `< key; value >` pairs and produces the corresponding output `< key; value >` pairs. Output `< key; value >` pairs produced by mapper function are known as intermediate data which is then sorted and shuffled. In shuffling stage, the transmission of data is done from mapper to reducer. After the shuffling stage, reducer function process these `< key; value >` pairs and produces the desired output which is stored by record reader in the HDFS.

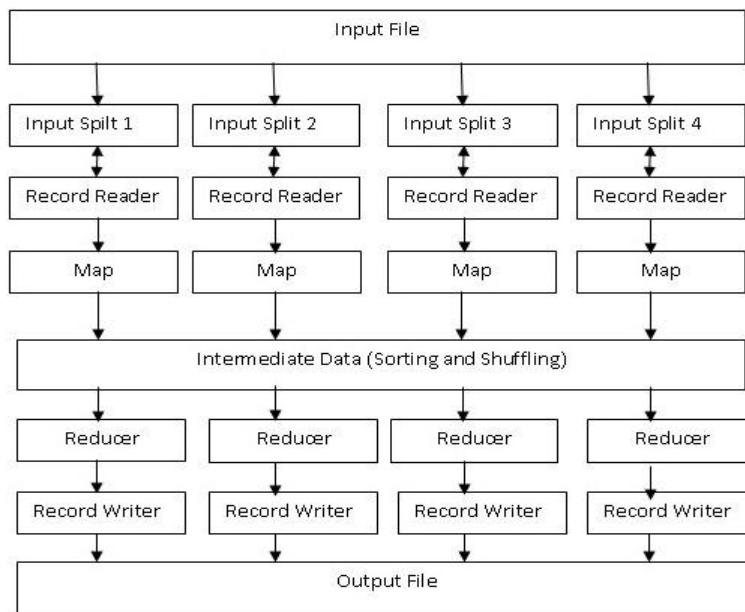


Fig. 5: Data flow in MapReduce

Traditional MapReduce implementation i.e. Hadoop forces the reducer task to start its execution after all the mapper task finishes its execution. In the heterogeneous environment machines have different processing capabilities. This barrier causes the underutilization of resources because faster machine finishes its map task and waits for the other machines to complete their's. The other work i.e. reduce task cannot be assigned to faster machine until the completion of all map tasks by every machine. In turn, this barrier between mapper and reducer task degrades the performance and also uses the system resources inefficiently. To overcome these limitation of traditional Hadoop, two approaches are proposed:-

- Overlapping of Mapper and Reducer task
- Hierarchical Reduction

This paper presents the improvement in MapReduce framework by using above two approaches. Three algorithms i.e. WordCount, Kmeans and PageRank are implemented on traditional and improved MapReduce model and the experimental results shows the speedup of improved MapReduce over traditional MapReduce.

## II. RELATED WORK

There has been a significant measure of innovative work that have been made to improve the MapReduce framework. Jeffrey Dean et al. [1] described that the computation is done via several phases such as map phase, shuffle phase and reduce phase. Figure 6 illustrated, these phases are executed in sequential manner.



Fig. 6: Data processing in MapReduce

MapReduce forces the shuffle phase to start its execution after all the map tasks are completed. Similarly, the reducer phase will perform its data processing only when shuffle phase finishes transmission of intermediate data. Benjamin Heintz et al. [22] proposed two techniques to overcome the performance limitation of traditional MapReduce. In the traditional MapReduce, the push-then-map approach is followed in which input data is stored in the DataNodes before the starting of computation. This approach faces two problems: first, performance degradation due to the dependency across the different phases i.e. all the tasks have to wait until slowest communication link completes its task. Second problem, map task cannot demand the work based on their capacity because of the separation of push and map phase. To avoid these limitations, Benjamin Heintz et.al.[22] proposed two solutions: first, overlapping of the push and map phases and second, collect locality information at runtime and make scheduling decisions based on this information. Two step techniques are as follows:

### Map-aware Push (Overlapping Push and Map Phases)

Two data sources (DS) having the data 15GB and two mappers (Map1 and Map2) are assumed. The bandwidth of link and computation speed of mapper is illustrated in the following figure 7. If optimize push distribution is used then DS1 pushes 10GB to Map1 and 5GB to Map2 and DS2 pushes 3GB to Map1 and 12GB to Map2. The DS1 will take 500 sec to store the data on the DataNodes whereas DS2 will take 300sec.

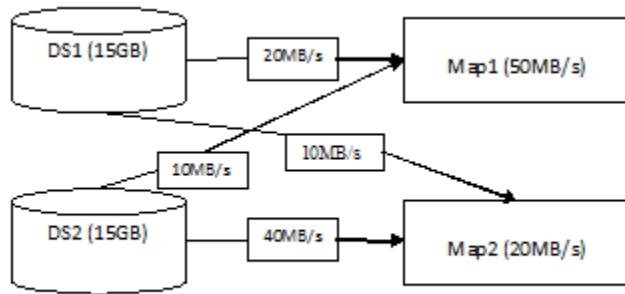


Fig. 7: Network with two data sources and two mappers

In the traditional MapReduce, the computation will start after the push phase completes i.e. after 500 sec. But this article, proposed a solution by overlapping of push and Map phases. In this approach, mapper will start its execution after it gets the data i.e. after 300sec DS2 starts its execution. The overlapping of phases avoids this unnecessary waiting for completion of communication link tasks.

### Shuffle-aware Map

In conventional MapReduce, the shuffling of data from mapper to reducer is an all-all operation as describe in figure 8. However, in cluster environment where the machine are having different configuration, a mapper with slow

outgoing communication link become a barrier in the shuffle phase. Due to this slow outgoing communication link the performance will be degraded. This article proposes map task scheduling based on the estimated shuffle cost from each mapper to enable faster shuffle and reduce execution.

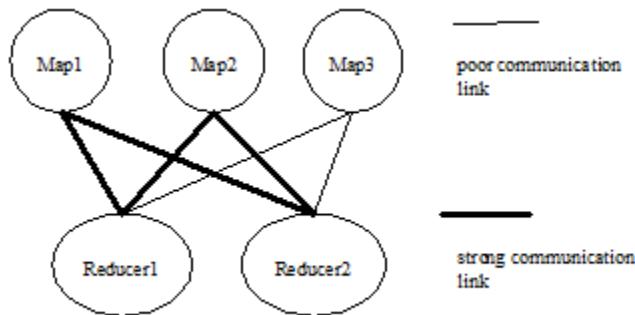


Fig. 8: Network where link from map3 to reducer1 and reducer 2 are barrier

Similar, but independent to our work, Minghong Lin et al.[23] proposed the overlapping tandem queue model for phase overlapping in which a job can be processed simultaneously in map phase and shuffle phase that minimize the average response time. Usually, the MapReduce jobs spend the most of time in the map phase and shuffle phase. However, the overlapping of map phase and shuffle phase is exploited where a fraction of the job may arrived to the shuffle phase before the fraction of entire job finished at the map phase. After the creation of the pieces of intermediate data from the map phase, they have been available for processing in the shuffle phase. When the entire job has completed at the shuffle phase, the job transferred to the reduce phase.

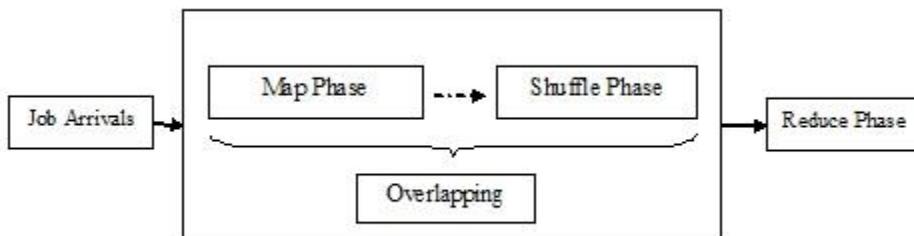


Fig. 9: Tandem queue model for phase overlapping in MapReduce

This paper focuses on the scheduling problem within the MapReduce architecture and presented the coordinated scheduling algorithm to minimize the average response time while the phase overlapping of map phase and shuffle phase is used.

### III. PROPOSED METHODOLOGY

The shortcoming of traditional MapReduce is that when all the mapper finishes their execution then only reducer starts its execution. If machines have different processing capabilities, the utilization of system resource is reduced which in turn impacts the performance. To overcome the above mentioned drawback of traditional MapReduce, the following solutions are proposed:

- Overlapping of Mapper and Reducer
- Hierarchical reduction

### A. Overlapping of Mapper And Reducer

The overlapping of mapper and reducer task is done by dynamically issuing reduce task to aggregate partially reduced results. As shown in figure 10, as soon as the numbers of predefined mapper tasks are completed, a new reducer task is created and starts its execution on an available data node. The reduce task is responsible for processing the data which are produced by mapper tasks that are just finished. The main challenge in improved MapReduce is to select the right threshold value that trigger the reducer task before the completion of mapper task. If threshold value is low then it wastes the CPU cycles and network bandwidth, while if its value is high then the prominent performance improvements will not be achieved. In our solution mapper and reducer are created in the same way as it is created in traditional MapReduce, hence fault-tolerant scheme of traditional MapReduce works with this approach too.

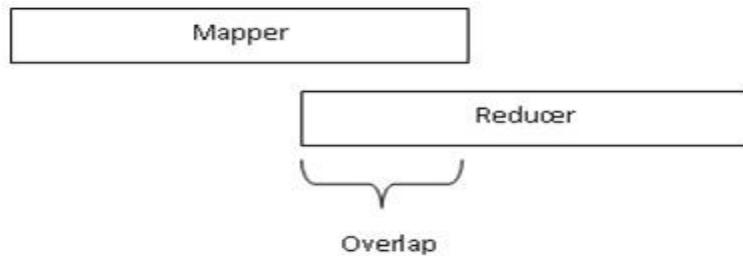


Fig. 10: Overlapping of mapper and reducer

### B. Hierarchical Reduction

In the traditional MapReduce, the reduction of key is always processed by one DataNode whereas in improved MapReduce, `<key; value>` pairs can be hierarchically reduced across the cluster as shown in figure 11. The data generated by mapper task is processed by corresponding reduce task. Reducer task starts its execution after the predefined number of mapper tasks is completed. When the processing of data is completed, another stage of reduce task is started. In this stage all the data processed by reduce task is processed again by another phase of reduce task and this will be continued until all the data is merged by reduce task. Therefore, the proposed approach is best suitable for those applications which need prominent reduce processing per key. However, this solution takes extra overhead because many time data is transferred to reduced task across the cluster. In particular, the JobTracker should keep track of all created reduce tasks. Whenever a reduce task fails, another same task should be created, and the appropriate tasks should be assigned again for reduction. In addition to the intercommunication cost, the number of writes to distributed and local file system are same (assuming the same number of mapper and reducer tasks) for both the traditional MapReduce and improved MapReduce. Therefore, improved MapReduce gives the better performance than traditional MapReduce when there is decent overlap between mapper and reducer tasks.

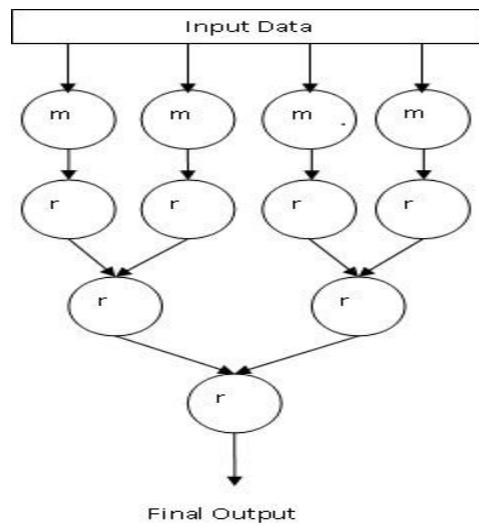


Fig. 11: Hierarchical Reduction

### III. RESULTS

In this section the performance analysis of our proposed solution is described. The configurations of system used in the experiment are given below along with the experimental results.

#### A. Experimental Platform

Three algorithms i.e. WordCount, PageRank and Kmeans are implemented on both the traditional and improved MapReduce. These algorithms are executed on the cluster of four machines having one master node and three data nodes. The machine used for the processing having the specifications described as follows

Table I: Configuration of cluster machines

Properties	Descriptions
Processor	HP i7 processor having 3.4GHz
RAM	4GB/machine
Graphics Card(GPU)	No GPU
Operating System	Linux/Ubuntu14.04
Hard Disk	500GB
Topology	Connected by gigabit Ethernet cable

#### B. Dataset

The description of dataset used for three algorithms i.e. WordCount, PageRank and Kmeans are described in the following Table II

Table II: Dataset Description

Algorithms	Descriptions	Size
WordCount	Wiki media dump data	977.5MB
PageRank	Nodes-4,847,571 1.1GB Edges-68,993,773	1.1GB
Kmeans	6000000 points each with 2 dimensions	1.2GB

### C. Performance Metrics

Counters are used to figure out whether and how frequently a specific event happened during a job execution. The performance metrics that are considered for the evaluation of MapReduce are as follows:

#### 1) File System Counter

It defines the attributes related to file system which are as follows:

- FILE BYTES READ : -Total number of bytes read from local file system
- FILE BYTES WRITTEN : - Total number of bytes writes to local file system
- HDFS BYTES READ : - Total number of bytes read from HDFS
- HDFS BYTES WRITE : -Total number of bytes writes to HDFS

#### 2) Map Reduce Framework Counter

It characterizes the attributes related to memory and processing time which are as follows:

- GC time elapsed: - Total time spent during garbage collection
- CPU time spent: - Total time spent by all task on CPU
- Physical Memory (in bytes): - Total physical memory bytes consumed by all task
- Virtual memory (in bytes): - Total number of virtual memory bytes consumed by all task (RAM and swap)
- Heap memory (in bytes): - Total number of heap memory bytes consumed by all task

#### 3) Job Counter

It specify following parameters:

- Job Counter (megabyte-second) for map: - total time taken by all the map tasks.
- Job Counter (megabyte-second) for reduce: - total time taken by all the reduce tasks
- Total Execution time (sec): - total time taken by the job i.e. time taken by map phase, shuffle phase, sort phase and reduce phase.

### D. WordCount Algorithm Analysis

WordCount is a non-iterative algorithm that is used to tally the recurrence of word happening in the document. Each mapper function process one line at a line and breaks every line into corresponding `<key; value>` pairs where key represents the word and the value represents the occurrence of that word in a line i.e. 1. The reducer function process all the `<key; value>` pairs based on the same key and produces the desired output. The instance of MapReduce has been outlined for large text file where we may wish to find the occurrence of each word in the file:

Pseudo Code of Mapper and Reducer

```
Map(String key1 , String value1)
 //key1 : file name
 //value1 : file content
 1) For each word w1 in value :
 2) EmitIntermediatePairs(w1,1);
```

```
Reduce(String key1, Iterator values1)
 //key1 : a word
 //value1 : a list of counts
 1) int r =0;
 2) for each v1 in values:
 3) r1+=parseInt(v1);
 4) Emit(AsString(r1));
```

### 1) Comparison of Counters for WordCount Application

This section presents the comparative analysis of WordCount application for different counters as shown in Table III and Table IV. As shown from experimental results, total time taken by the WordCount application in traditional and improved MapReduce is 313.033sec and 293.904sec respectively. Hence we get 6.5% speedup over traditional MapReduce for WordCount application. The performance analysis of WordCount application is illustrated in figure 12 and figure 13.

Table III: WordCount counter analysis for traditional MapReduce

Counter Name	Description	Map	Reduce	Total
File System Counters	FILE_BYTES_READ	1430150068	1430149894	2860299962
	FILE_BYTES_WRITTEN	2861145546	1441428669	4302574215
	HDFS_BYTES_READ	977535907	0	977535907
	HDFS_BYTES_WRITE	0	46326054	46326054
MapReduce Framework Counter	GC time elapsed	17037	538	17575
	CPU time spent	234740	77680	312420
	Physical Memory(in bytes)	2152783872	178675712	2331459584
	Virtual memory(in bytes)	15494946816	1942507520	17437454336
	Heap memory(in bytes)	1670381568	65536000	1735917568
Job Counter (megabyte-second)		1049775104	151477248	1201252352
<b>Total Execution Time (sec)</b>				313.033

Table IV: WordCount counter analysis for improved MapReduce

Counter Name	Description	Map	Reduce	Total
File System Counters	FILE_BYTES_READ	1430150068	1430149894	2860299962
	FILE_BYTES_WRITTEN	2861145546	1441428669	4302574215
	HDFS_BYTES_READ	977535907	0	977535907
	HDFS_BYTES_WRITE	0	46326054	46326054
MapReduce Framework Counter	GC time elapsed	8394	10387	18781
	CPU time spent	169140	54790	193930
	Physical Memory(in bytes)	2288754688	201609216	2490363904
	Virtual memory(in bytes)	15514386432	1930567680	17444954112
	Heap memory(in bytes)	1664614400	180879360	1845493760
Job Counter (megabyte-second)		484061184	69598208	533659392
<b>Total Execution Time (sec)</b>				293.904

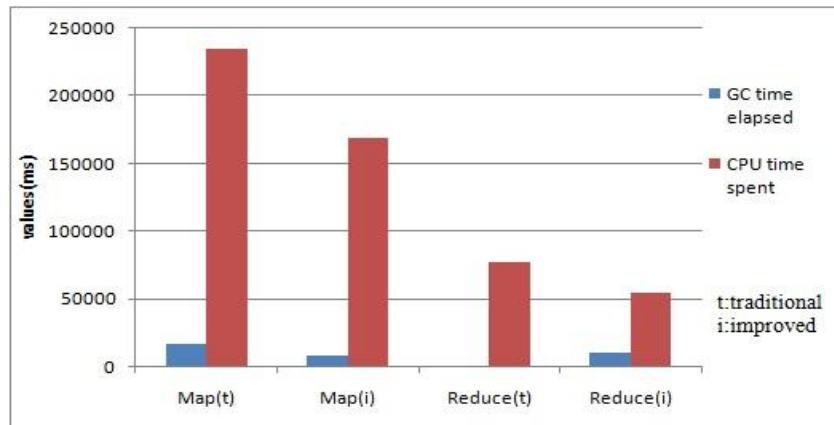


Fig. 12: Analysis of WordCount application for CPU vs. GC

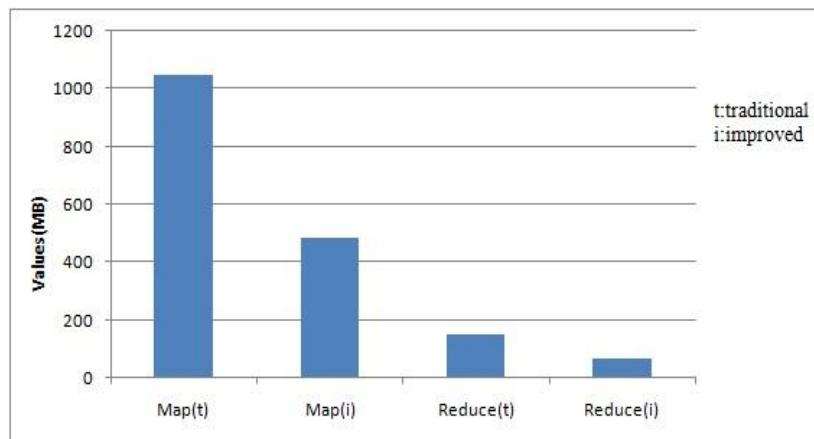


Fig. 13: Analysis of WordCount application for Job Counter

## E. PAGERANK ALGORITHM ANALYSIS

PageRank algorithm is used to analyze the significance of web pages on web (World Wide Web). It analyzes the link on the web by assigning numerical weight to each web page and measures web pages importance[24][25]. It is extensively used in many application areas such as web search, link prediction. Similar algorithm like rooted PageRank is used in many areas such as personalized news and video recommendation systems. The PageRank vector is defined over a directed graph in which each vertex has a numerical weight (rank or score). Initially each vertex has weight  $1/|V|$  and updates its score (rank or weight) iteratively as follow

$$\text{PageRank}^{i+1}(n) = (1 - d) + \sum_{u \in V^-(n)} \frac{d \cdot \text{PageRank}^i(u)}{V^+(u)}$$

Where  $V^-(n)$  is the set of vertices pointing to node n,  $V^+(n)$  is the set of nodes that n points to, and d is a constant representing the damping factor. This iterative process continues for a fixed number of iteration or till the convergence criterion is satisfied.

$$\sum_n |PageRank^i(n) - PageRank^{i-1}(n)| < Thresh_{val}$$

In mapper phase vertex i sends value  $R_{i,j} = \frac{R_i}{N_i}$  to all its out neighbors j where  $|N_i|$  represents number of is out neighbors. In reducer phase vertex j updates  $R_j$  by summing the  $R_{i,j}$  received from all its in neighbors i. PageRank algorithm's pseudo code for mapper and reducer phase are given below:

**Map Phase Pseudo Code:**

*Input:*  $\langle i, \frac{N_i}{R_i} \rangle$

*Output:*  $\langle i, N_i \rangle$

1. for all j in  $N_i$  do
2.  $R_{i,j} = \frac{R_i}{|N_i|}$
3. output  $\langle j, R_{i,j} \rangle$
4. end for

**Reduce Phase Pseudo Code:**

*Input:*  $\langle j, R_{i,j}, N_j \rangle$

*Output:*  $\langle j, \frac{N_j}{R_j} \rangle$

$$1. R_j = d + \sum_i R_{i,j} (1 - d)$$

### 1) Comparison of Counters for PageRank Application

This section presents the comparative analysis of PageRank application for different counters as shown in Table V and Table VI. As shown from experimental results, total time taken by the PageRank application in traditional and improved MapReduce is 677.701sec and 613.96sec respectively. Hence we get 10.38% speedup over traditional MapReduce for WordCount application. The performance analysis of PageRank application is illustrated in figure 14 and figure 15.

Table V: PageRank Counters analysis for traditional MapReduce

Data	Description	Map	Reduce	Total
File System Counters	FILE_BYTES_READ	170931424	170931406	341862830
	FILE_BYTES_WRITTEN	341968335	172372268	514340603
	HDFS_BYTES_READ	105901497	0	105901497
	HDFS_BYTES_WRITE	0	106116827	106116827
MapReduce Framework Counter	GC time elapsed	527	324	851
	CPU time spent	29510	17430	46940
	Physical Memory(in bytes)	307773440	175845376	483618816
	Virtual memory(in bytes)	1944756224	1940680704	3885436928
	Heap memory(in bytes)	209190912	89128960	298319872
Job Counter (megabyte-second)		24538112	18118656	42656768
<b>Total Execution Time (sec)</b>				677.701

Table VI: PageRank Counters analysis for improved MapReduce

Data	Description	Map	Reduce	Total
File System Counters	FILE_BYTES_READ	170931424	170931406	341862830
	FILE_BYTES_WRITTEN	341968335	172372268	514340603
	HDFS_BYTES_READ	105901497	0	105901497
	HDFS_BYTES_WRITE	0	106116827	106116827
MapReduce Framework Counter	GC time elapsed	445	266	711
	CPU time spent	27670	16130	43800
	Physical Memory(in bytes)	286212096	183296000	469508096

	Virtual memory(in bytes)	1938055168	1943711744	3881766912
	Heap memory(in bytes)	209190912	88080384	297271296
Job Counter (megabyte-second)		23681024	16898048	40579072
<b>Total Execution Time (sec)</b>			613.96	

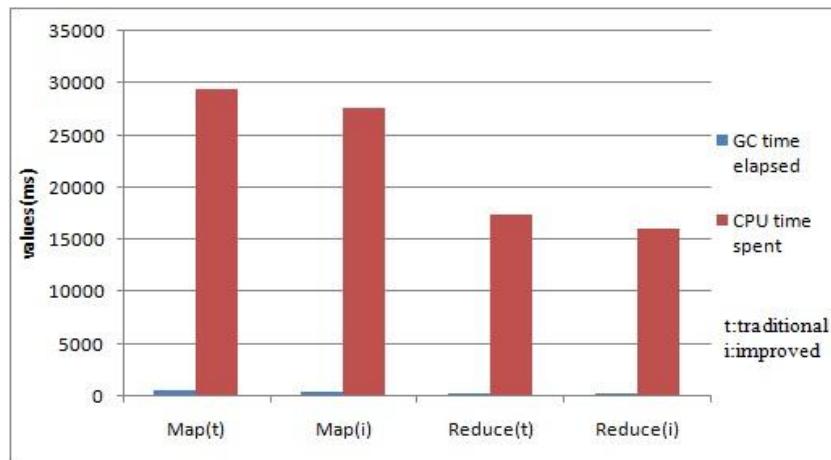


Fig. 14: Analysis of PageRank application for CPU vs. GC

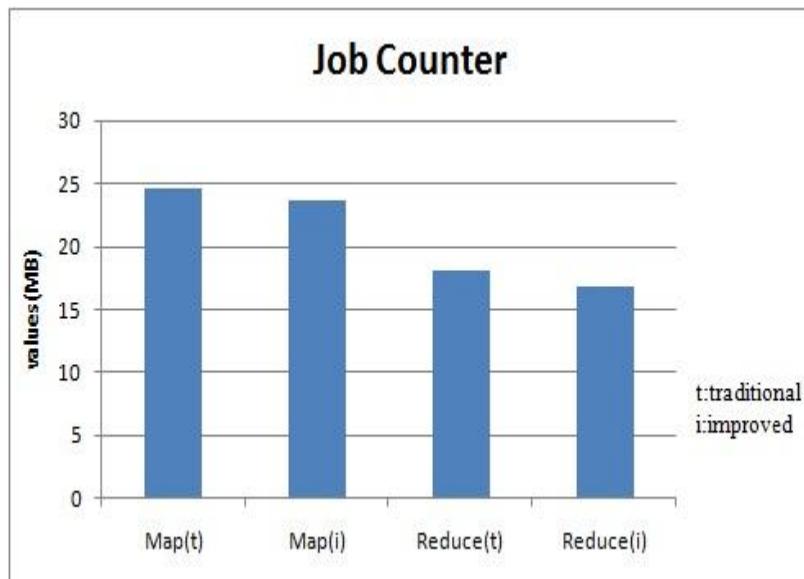


Fig. 15: Analysis of PageRank application for Job Counter

## F. Kmeans Performance

Kmeans algorithm is used to partition the points into k cluster. Each point belongs to one cluster with the lowest mean. It is widely used in many application areas such as cluster analysis, feature learning and vector quantization etc. Kmeans algorithm partition the point into k cluster. Each point has two attribute,  $p_{id}$  which is unique point id for each point and  $p_{val}$  which is point feature value. Initially k points are selected at random and defined as centroid of k clusters. Centroid points are defined by  $c_{id}$  that is unique point id for each centroid point and  $c_{val}$  which is feature

value of centroid point. In mapper phase, each point is assigned to the cluster based on the nearest centroid. In reduce phase, centroid cid updates the centroid of cluster by averaging all the value ( $p_{val}$ ) of assigned points in the cluster.

#### Map Phase Pseudo Code:

*Input: <  $p_{id}, p_{val}$  | <  $c_{id}, c_{val}$  >>*

*Output: <  $c_{id}, c_{val}$  >*

1.  $c_{id} = \text{find the nearest centroid of } p_{val} \text{ in } c_{id}, c_{val}$

#### Reduce Phase Pseudo Code:

*Input: <  $c_{id}, p_{val}$  >*

*Output: <  $c_{id}, c_{val}$  >*

1.  $c_{val} = \text{compute the average of } p_{val}$

#### 1) Comparison of counter For Kmeans Application

This section presents the comparative analysis of Kmeans application for different counters as shown in Table VII and Table VIII.

As shown from experimental results, total time taken by the PageRank application in traditional and improved MapReduce is 545.841sec and 509.992sec respectively. Hence we get 7.02% speedup over traditional MapReduce for WordCount application. The performance analysis of WordCount application is illustrated in figure 16 and figure 17.

Table VII: Kmeans Counters analysis for traditional MapReduce

Data	Description	Map	Reduce	Total
File System Counters	FILE_BYTES_READ	112132456	20095305	132227761
	FILE_BYTES_WRITTEN	119291569	91390343	210681912
	HDFS_BYTES_READ	1195795742	0	1195795742
	HDFS_BYTES_WRITE	0	2386674987	2386674987
MapReduce Framework Counter	GC time elapsed	24270	10234	34504
	CPU time spent	321235	152335	473570
	Physical Memory(in bytes)	1701532464	109190352	1810722816
	Virtual memory(in bytes)	15299192912	2133982640	17433175552
	Heap memory(in bytes)	722345576	152166808	874512384
Job Counter (megabyte-second)		658324480	198257625	856582105
<b>Total Execution Time (sec)</b>				545.841

Table VIII: Kmeans Counters analysis for improved MapReduce

Data	Description	Map	Reduce	Total
File System Counters	FILE_BYTES_READ	112132456	20095305	132227761
	FILE_BYTES_WRITTEN	119291569	91390343	210681912
	HDFS_BYTES_READ	1195795742	0	1195795742
	HDFS_BYTES_WRITE	0	2386674987	2386674987
MapReduce Framework Counter	GC time elapsed	21423	10890	32313
	CPU time spent	33123	131245	164368
	Physical Memory(in bytes)	162123421	1659523427	1821646848
	Virtual memory(in bytes)	16241422312	1223938072	17465360384
	Heap memory(in bytes)	712412461	116486867	828899328
Job Counter (megabyte-second)		487270400	17825792	505096192
<b>Total Execution Time (sec)</b>				509.992

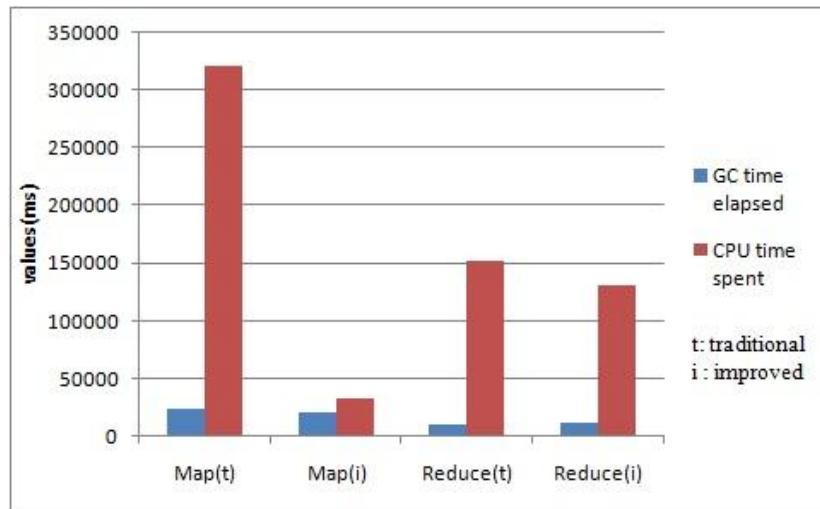


Fig. 16: Analysis of Kmeans application for CPU vs. GC

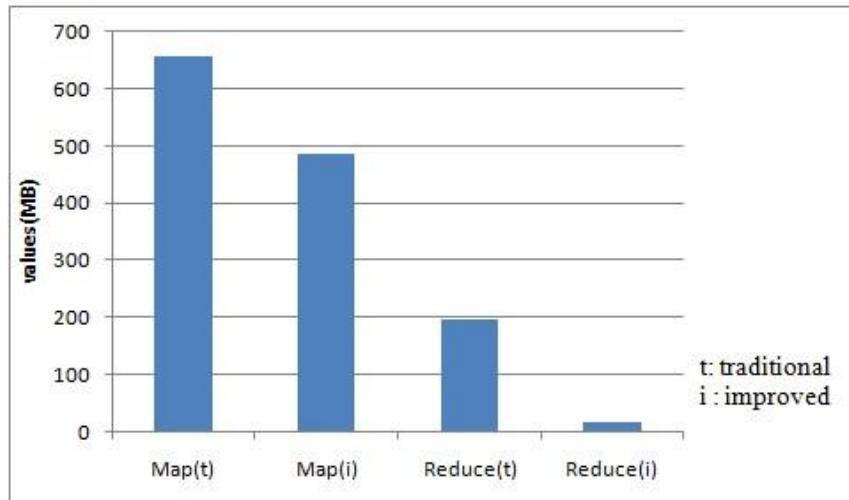


Fig. 17: Analysis of Kmeans application for Job Counter

#### IV. CONCLUSION

The old mechanism of mapper and reducer forces the system to waste its cycles. This happens because reducer waits for mapper to finish first i.e. faster machines completes its map task and waits for the other machine to complete their map task. Other work (reduce task) cannot be assigned to faster machine until the completion of all map task by every machine. In turn, this barrier between mapper and reducer task degrades the performance and also uses the system resources inefficiently. To overcome these limitations of traditional Hadoop, two improvements viz overlapping of mapper and reducer and hierarchical reduction when applied can achieve significant speedup for same read and write count. It can be seen from the experimental results the speedup of 6.5%, 7.02% and 10.38% for WordCount, Kmeans and PageRank applications respectively is obtained.

## V. FUTURE WORK

In future, this approach can be applied to broader class of algorithms for deep analysis. This approach can also be applied to those applications which are not iteratively reducible in the nature. Finally, the approach can be extended on real cloud environment.

## REFERENCES

- [1] J. Dean and S. Ghemawat, Mapreduce: Simplified data processing on large clusters, in Proc. 6th Conf. Symp. Oper. Syst. Des. Implementation, 2004, p. 10
- [2] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. henker, and I. Stoica, Resilient distributed datasets: A fault-tolerant abstraction for, in-memory cluster computing, in Proc. 9<sup>th</sup> USENIX Conf. Netw. Syst. Des. Implementation, 2012, p. 2.
- [3] R. Power and J. Li, Piccolo: Building fast, distributed programs with partitioned tables, in Proc. 9th USENIX Conf. Oper. Syst. Des. Implementation, 2010, pp. 114.
- [4] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski, Pregel: A system for large-scale graph processing, in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2010, pp. 135146
- [5] S. R. Mihaylov, Z. G. Ives, and S. Guha, Rex: Recursive, deltabased datacentric computation, in Proc. VLDB Endowment, 2012, vol. 5, no. 11, pp. 12801291
- [6] Y. Low, D. Bickson, J. Gonzalez, C. Guestrin, A. Kyrola, and J. M. Hellerstein, Distributed graphlab: A framework for machine learning and data mining in the cloud, in Proc. VLDB Endowment, 2012, vol. 5, no. 8, pp. 716727
- [7] S. Ewen, K. Tzoumas, M. Kaufmann, and V. Markl, Spinning fast iterative data flows, in Proc. VLDB Endowment, 2012, vol. 5, no. 11, pp. 12681279
- [8] Y. Bu, B. Howe, M. Balazinska, and M. D. Ernst, Haloop: Efficient iterative data processing on large clusters, in Proc. VLDB Endowment, 2010, vol. 3, no. 12, pp. 285296
- [9] J. Ekanayake, H. Li, B. Zhang, T. Gunarathne, S.-H. Bae, J. Qiu, and G. Fox, Twister: A runtime for iterative mapreduce, in Proc. 19th ACM Symp. High Performance Distributed Comput., 2010, pp. 810818
- [10] Y. Zhang, Q. Gao, L. Gao, and C. Wang, imapreduce: A distributed computing framework for iterative computation, J. Grid Comput., vol. 10, no. 1, pp. 4768, 2012
- [11] Wiley, K., Connolly, A., Krughoff, S., Gardner, J., Balazinska, M., Howe, B., Kwon, Y., Bu, Y.: Astronomical image processing with hadoop. In: Gabriel, C. (ed.) Astronomical Data Analysis Software and Systems (2010)
- [12] Hive. <http://hadoop.apache.org/hive/> (2010). Accessed 7 July 2010
- [13] Dean, J., Ghemawat. S.: MapReduce: simplified data processing on large clusters. In: OSDI, pp. 137150 (2004)
- [14] Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: bringing order to the web. Technical Report 1999-66, Stanford InfoLab (1999)
- [15] Moore, A.W., Zuev, D.: Internet traffic classification using bayesian analysis techniques. In: SIGMETRICS, pp. 5060 (2005)
- [16] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)
- [17] Hagan, M.T., Demuth, H.B., Beale, M.H.: Neural Network Design. PWS Publishing, Boston (1996)
- [18] Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. 31(3), 264323 (1999)
- [19] Bancilhon, F., Ramakrishnan, R.: An amateurs introduction to recursive query processing strategies. In: SIGMOD Conference, pp. 1652 (1986)
- [20] Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM 46(5), 604632 (1999)
- [21] Tom White. Hadoop: the definitive guide. O'Reilly, 2012.

- [22] Benjamin Heintz, Chenyu Wang, Abhishek Chandra, and Jon Weissman, "Cross-Phase Optimization in MapReduce" in Proc. ACM Int. Conf. Manage. Data, 2014, pp 277-302
- [23] Minghong Lin, Li Zhang, Adam Wierman, Jian Tan, "Joint Optimization of Overlapping Phases in MapReduce" in Proc. ACM SIGMETRICS Performance Evaluation Review, Vol. 41, No. 3, 2013
- [24] Song, Han Hee, Tae Won Cho, Vacha Dave, Yin Zhang, and Lili Qiu. "Scalable proximity estimation and link prediction in online social networks." In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, pp. 322-335, 2009
- [25] Brin, Sergey, and Lawrence Page. "The anatomy of a large-scale hypertextual Web search engine." Computer networks and ISDN systems 30, no. 1: pp. 107-117, 1998

# An Optimized Hybrid Multi-Digit BCD Adder Using QCA

D. Ajitha, Dept. of E.C.E, Research Scholar, JNTUA, Ananthapuramu, India.  
K. V. Ramanaiah, Dept. of E.C.E, Y.S.R Engineering College of Yogi Vemana University, Proddatur, India.  
V. Sumalatha, Dept. of E.C.E, JNTUA College of Engineering, Ananthapuramu, India.

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

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

## I. INTRODUCTION

Due to present serious challenges existing in conventional transistor technology, researchers are pointed to find an alternative to this technology. Amongst the upcoming technologies, quantum-dot cellular automata (QCA) is a suitable alternative technology that offers unique features such as smaller size and can operate at THz frequencies and room temperature [1, 2]. The essential elements in QCA are cells; each cell is composed of two mobile electrons that are located in opposite corners according to Columbic Energy, ensuring in two possible Polarizations (+1, -1) as shown in Fig. 1(a) [3]. Up to this time, many methods for fabrication of QCA basic cells are suggested such as metal island [4], magnetic [5], semiconductor, and molecular QCA [6]. As is discussed in [4-6], metal dot implementations have proven to be the most successful material systems which are based on ‘single-electron transistors’ fabrication techniques. Cowburn’s group firstly proposes the magnetic implementation and extended by the Porod group and the Bokar group. In the physical semiconductor implementation, the Cavendish group of Smith et al. proved QCA operation in GaAs/AlGa As hetero structures with confining top-gate electrodes and the group of Kern et al. demonstrated a silicon QCA cell by employing an etching technique to form the dots.

Furthermore, based on [7], the Fehlner and Lapinte groups have performed successful molecular synthesis in creating molecules that show the essential bistability. According to the columbic interaction between electrons in neighboring cells, the basic logic gates in QCA circuits (inverter and majority gates) are constructed.

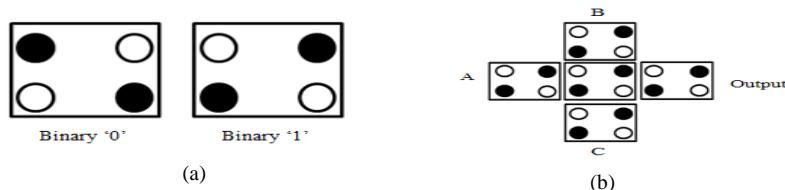


Figure 1: Basic logic cells and gates in QCA: (a) Two possible polarizations, (b) Three-input majority gate.

Clocking used in QCA for the purpose of controlling the data flow synchronization and supply of power to the cells. QCA have four clocks namely switch, hold, release, and relax [8]. There are two possible crossovers are available in QCA circuit design named as coplanar and multilayer crossovers. In the co-planar crossover, only one layer is used with normal and rotated cells, while in the multi-layer crossover, it uses two additional layers similar to conventional Integrated Circuit [9]. The structure of the vital element in QCA known as 3-input Majority gate is shown in Fig. 1(b). The basic AND, OR gates in QCA are obtained by fixing the polarization in the input of Majority gate to logic ‘0’ and logic ‘1’ respectively [10].

A parallel high-speed multi-digit decimal adder is significant in a multiprogramming future Nanocomputing environment. Particularly the multi-digit decimal adder plays a vital role in the applications that cannot endure the loss of precision that consequences industrial, financial, and Internet-based applications. Moreover, the decimal addition is the primary but the most essential unit among the decimal arithmetic processes in parallel multiplication to accumulate the partial products as well as in division also.

During the last decade, much interest has been paid to construct efficient BCD adder using QCA. The earlier work on Single-digit BCD adder designs using QCA were presented [11-15]. In recent times, two cost-efficient single-digit decimal adder designs named as Carry flow adder (CFA) and Carry look ahead (CLA)-based BCD adder is designed [15]. Even though there are several Single-digit BCD adder designs are implemented using QCA these designs are not further extended to perform the multi-digit BCD addition. Hence, in this paper, an attempt is made to construct the Multi-digit BCD adder to perform the multi-digit decimal addition in an optimized manner. The rest of the paper is organized as follows. Section II presents the proposed design. Section III presents performance analysis and simulation results obtained from QCADesigner tool to prove the functional correctness of the proposed design, and finally, Section IV concludes the paper.

## II. PROPOSED DESIGN

The fundamental purpose of this work is to present the multi-digit BCD adder design to perform the BCD/Decimal addition with less complexity and higher speed. A circuit which adds two BCD digits in parallel to produce BCD sum is known as BCD adder. The conventional ripple carry BCD adder design consists of three sections, the first section is a 4-bit binary adder in the upper portion as an initial stage, the correction logic in the second stage to validate the binary sum is equivalent to the Decimal sum or not. The correction logic generates output as logic ‘1’ if the binary sum greater than 9 and less than or equal to 19, or sum less than or equal to 9 with a carry out of Most significant bit (MSB) will be 1. If the correction logic generates output as logic ‘1’, then a ripple carry binary adder adds  $(0110)_2$  to the binary sum in the last stage to produce a valid Decimal sum else the binary sum appears as a Decimal sum in the final stage. The delay overhead in BCD adder is caused by either upper 4-bit binary adder or in the last part to produce the valid BCD sum. The subsections which impact primarily on the Decimal addition is the former one is upper 4-bit binary adder to generate the binary sum and later one is lower binary adder to produce the valid Decimal sum. This design is mainly focused on performing the Decimal addition by eliminating the delay caused by carry propagation in the final part of the design by introducing the 2:1 multiplexers produce the valid Decimal sum instead of a binary adder.

The block diagram of proposed HMDBA design is as shown in Fig. 2. The HMDBA is implemented by cascading Single-digit BCD adder (SDBA). The SDBA is the hybrid structure of CFA and CLA-based BCD adders. The upper part of the SDBA structure is similar to the CFA-BCD adder and in addition to the correction logic; some additional logic is added to generate the inputs to the multiplexers which resemble that of a CLA-based BCD adder. The additional logic is designed by observing the truth table of BCD adder [16].

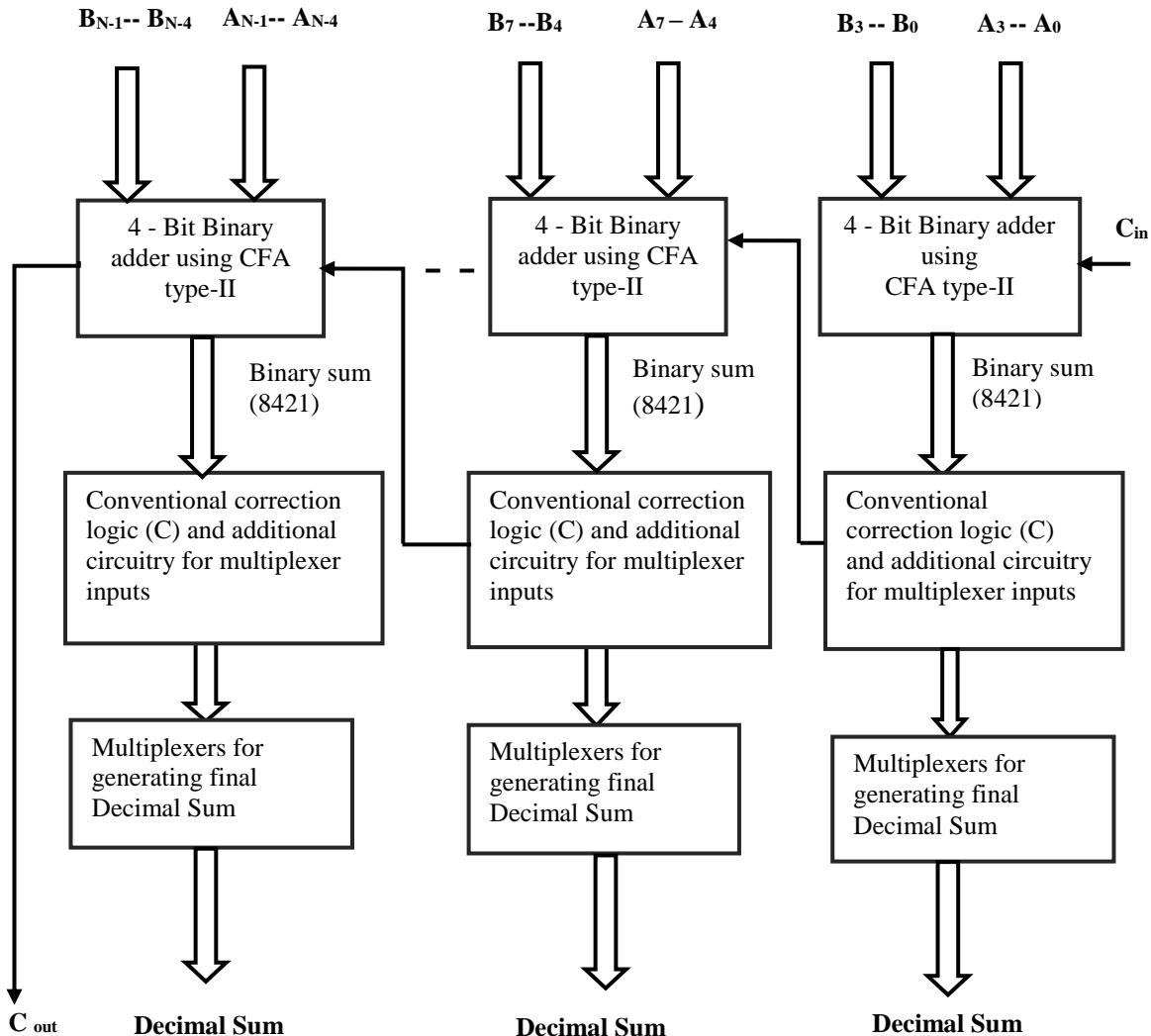


Figure 2. Block diagram of the proposed HMDBA design

In the top of the HMDBA as well as in each SDBA design, CFA type-II four-bit binary adder is used to generate the binary sum. The Design of the carry flow adder uses conventional carry propagation techniques but is optimized for QCA circuit layouts. The carry flow adder has three different adder types. By the maximum clock zone length limitation, either of the parallel CFA can be chosen. Type-I is a more compact design but requires more wire channels for input synchronization. Type-II reduces the wire channels, so the overall design has a smaller Area [17]. The rest of the design part in the SDBA structure is explained as follows.

The conventional correction logic implemented regarding majority gates is as follows.

$$C = M(K, M(Z_8, M(Z_4, Z_2, 1), 0), 1) \quad (1)$$

Where,

$K$  = Carry out of MSB from the binary sum,

$Z_8 Z_4 Z_2 Z_1$  – Four-bit Binary sum,

$CS_8 S_4 S_2 S_1$  – Decimal Sum.

The additional logic needed to provide the inputs to the multiplexers is derived from the truth table [16] by observing the binary sum combinations from  $(01010 \text{ and } 10011)_2$ . To generate the valid Decimal sum in the final stage the final 2:1 multiplexer part representation regarding the majority gates is given by the following expressions.

$$S_1 = Z_1 \text{ if } C = 0 \text{ or } 1 \quad (2)$$

$$S_2 = M(M(\bar{C}, Z_2, 0), M(C, \bar{Z}_2, 0), 1) \quad (3)$$

$$S_4 = M(M(\bar{C}, Z_4, 0), M(C, M(M(\bar{K}, (M(Z_8, M(Z_4, Z_2, 0, 0, 0), 0), 0), M(K, M(\bar{Z}_8, (\overline{M(Z_4, Z_2, 1, 0, 0, 1)}), 0), 1)), 1)) \quad (4)$$

$$S_8 = M(M(\bar{C}, Z_8, 0), M(C, M(K, Z_2, 0, 0, 1)) \quad (5)$$

The layout of SDBA adder is constructed by using CFA type-II in the upper part, and the second part, the final part is implemented by using (1)-(5). The proposed HMDBA layout design is achieved by extending the SDBA layout for n-number of decimal digits by cascading the SDBA structures as shown in the Fig. 2. The corresponding layouts of HMDBA design for performing 4-digit, 8-digit Decimal addition are displayed in Fig. 3 and Fig. 4.

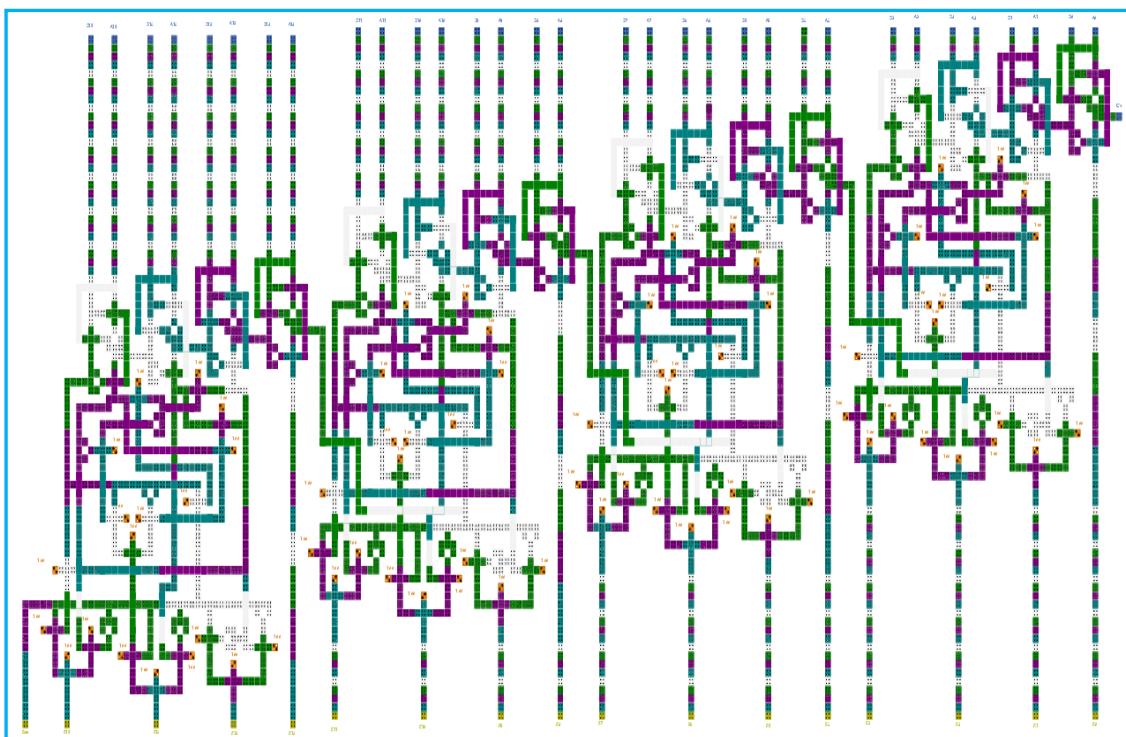


Figure 3. Layout of the proposed HMDBA 4-digit BCD adder

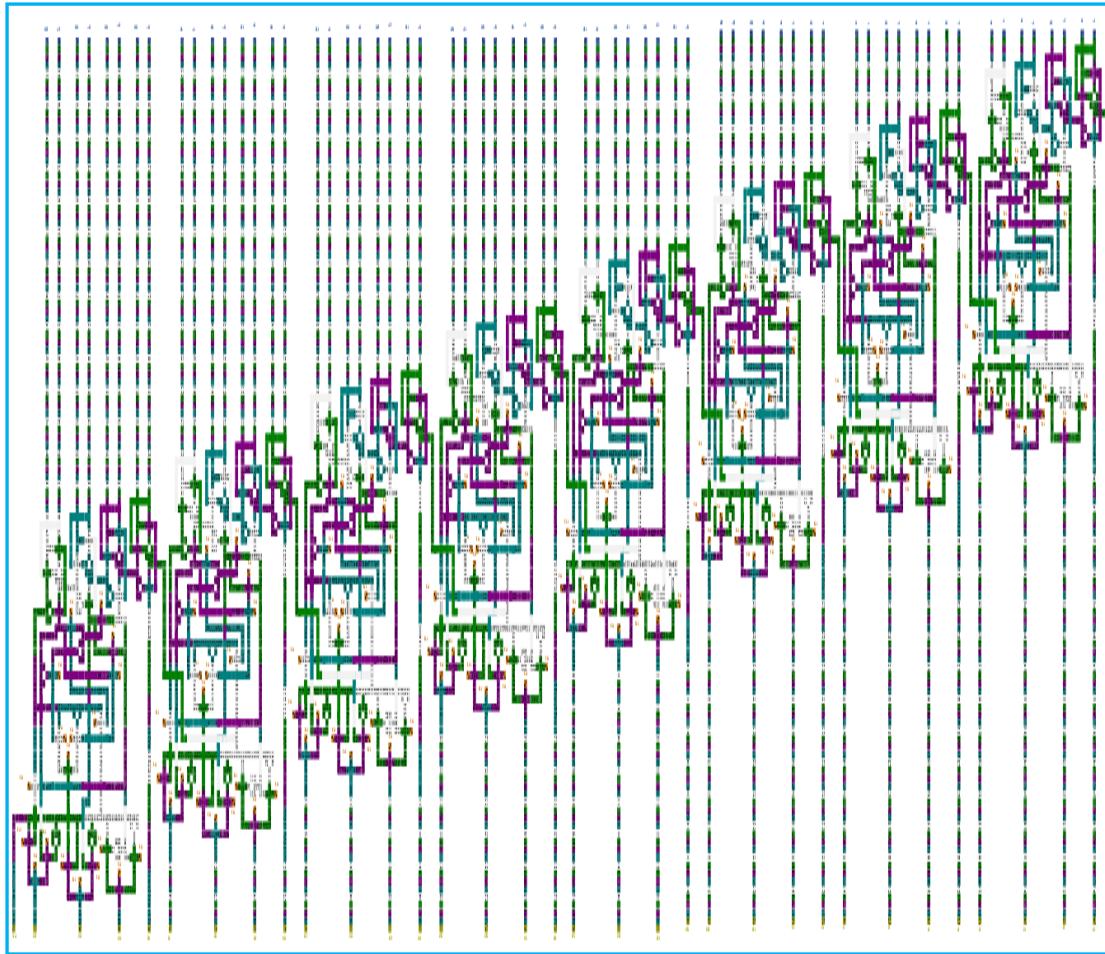


Figure 4. Layout of the proposed HMDBA 8-digit BCD adder

The proposed HMDBA design layouts are constructed with the help of QCADesigner tool [18-20]. The layout of 4-digit HMDBA is simulated with Bistable simulation Engine in QCADesigner tool to obtain the simulation results. All the design parameters are similar to that of earlier single digit BCD adder designs. The simulation results of 4-digit HMDBA design is as shown in Fig. 5. The first SDBA used to add the least significant decimal digit in the HMDBA requires 2.75 clock cycles. Out of these 2.75 clock cycles, two clock cycles are needed to produce the binary sum, correction logic as well as inputs to the multiplexers in the final part of the Decimal Adder and rest of the clock is utilized by the 2:1 multiplexers. To cascade the correction logic output to the next higher order digit input requires a delay of 2 clock cycles. Hence, the subsequent digit data inputs were delayed by 2 clock cycles for synchronizing all the inputs and outputs. Therefore, the 4-digit HMDBA design produces the significant output after 8.75 clock cycles. From the simulation results it is clearly displayed that the significant output appears after 8.75 clock cycles in clock 2 (clock zone 2) clearly marked in the Fig. 5.

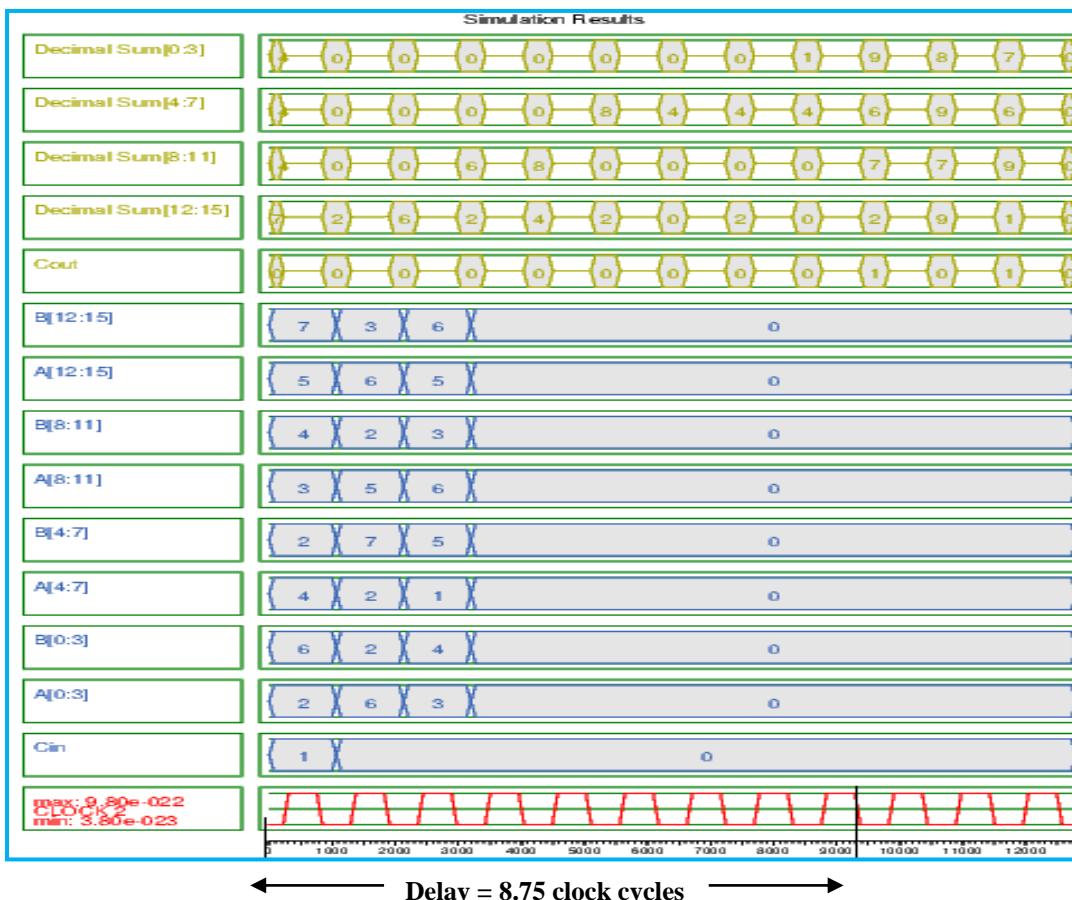


Figure 5. Simulation results of the HMDBA 4-digit BCD adder design

The following graphs give the clear idea about the number of cells (cell count), area, delay (latency) and the number of clock zones of HMDBA with existing QCA based Multi-digit BCD adders.

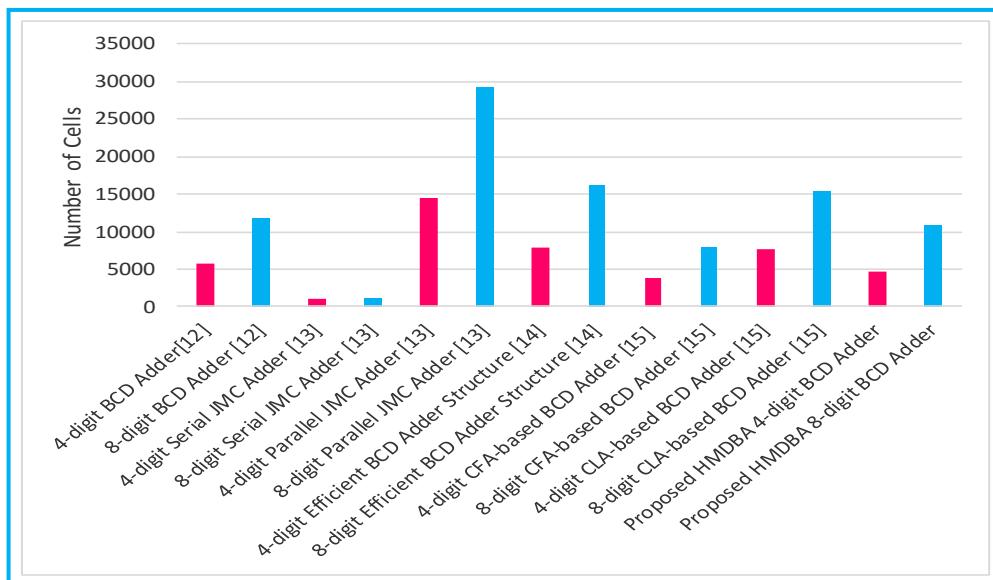


Figure 6. Comparison of the HMDBA performance about number of cells with existing methods

The performance comparison of the proposed HMDBA design regarding cell count is shown in Fig. 6. The existing single-digit BCD adders are constructed similar to that of conventional methodology with slight modifications. In the current designs, the final part of the BCD adder uses the 3-bit ripple carry binary adder structure instead of using 4-bit binary adder. Since according to correction logic no adder is required in the Least significant bit (LSB) position, therefore in the final part 3-bit adder is enough to produce the valid Decimal sum. Hence, while these designs are extended to perform the multi-digit BCD addition requires more area and delay [12], [14], [15]. Even though the Johnson–Mobius adder is a different technique adopted, but the serial JMC adder occupies less area with more delay, whereas the parallel JMC adder requires more hardware therefore again it leads to larger area requirements as well as more delay [13]. Even though the CLA-based design is distinct from the conventional structure, the circuit complexity is more due to generate and propagate functions, so it occupies more area compared to the SDBA used in the HMDBA. While, the proposed design is a combination of CFA based and CLA based design. Therefore it occupies less area, as well as speed, is competing with CLA based design.

The SDBA in the proposed HMDBA design needs 987 cells, with  $1.24 \mu\text{m}^2$  area and 2.75 clock cycles to perform single-digit BCD addition. The SDBA encompasses the best-optimized performance parameters amongst all the existing single-digit BCD designs. Hence by using this SDBA, the HMDBA is constructed to perform the multi-digit BCD addition in an efficient manner. The performance comparison of the proposed HMDBA design regarding area is shown in Fig. 7. It clearly shows that the proposed design occupies less area compared to all the existing designs. Whereas, it has a slight increase in the number of cells and area compared to the CFA-based BCD adder due to the additional circuitry to the multiplexer inputs.

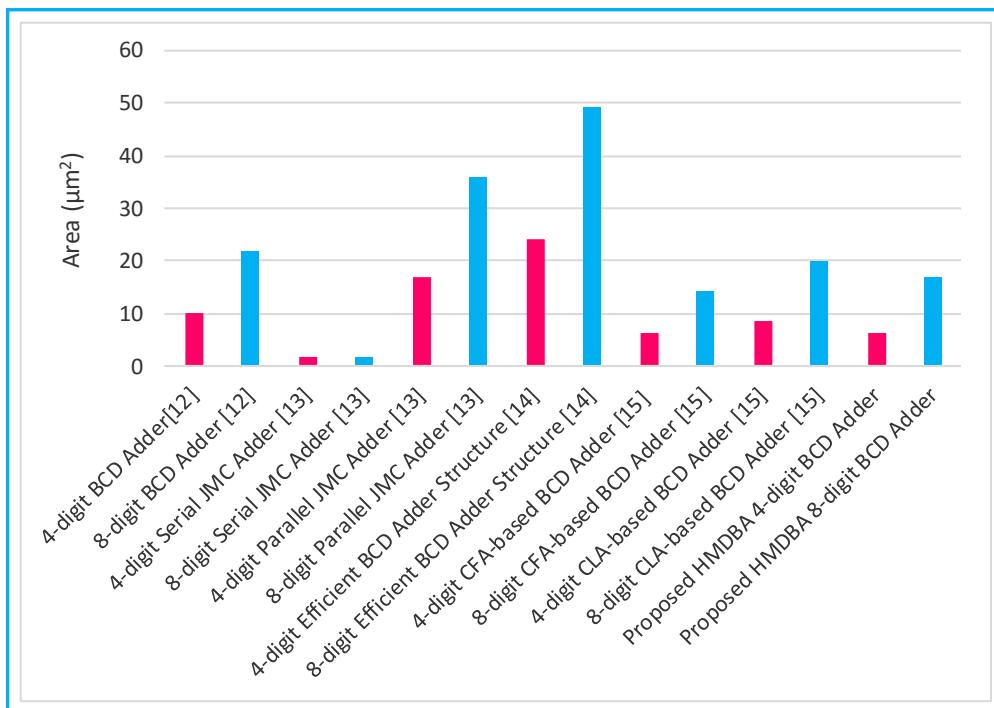


Figure 7. Comparison of the HMDBA performance concerning the area with existing methods

The performance of the proposed HMDBA regarding the Delay and clock zones are shown in Fig. 8 and Fig. 9 respectively. It clearly displays that the proposed design is having less related to the existing designs.

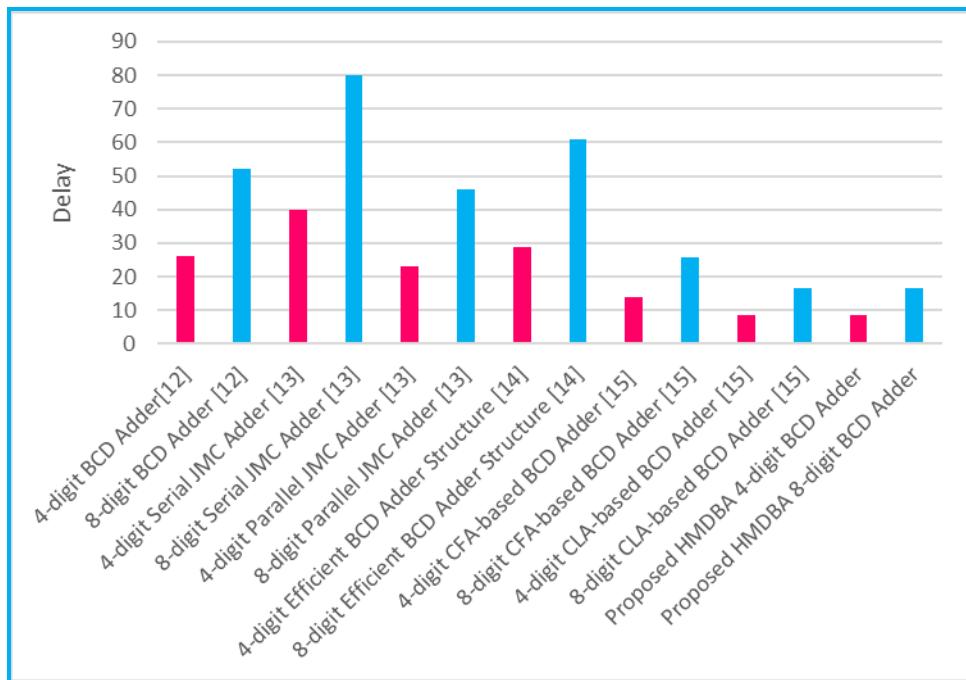


Figure 8. Comparison of the HMDBA performance regarding delay with existing methods

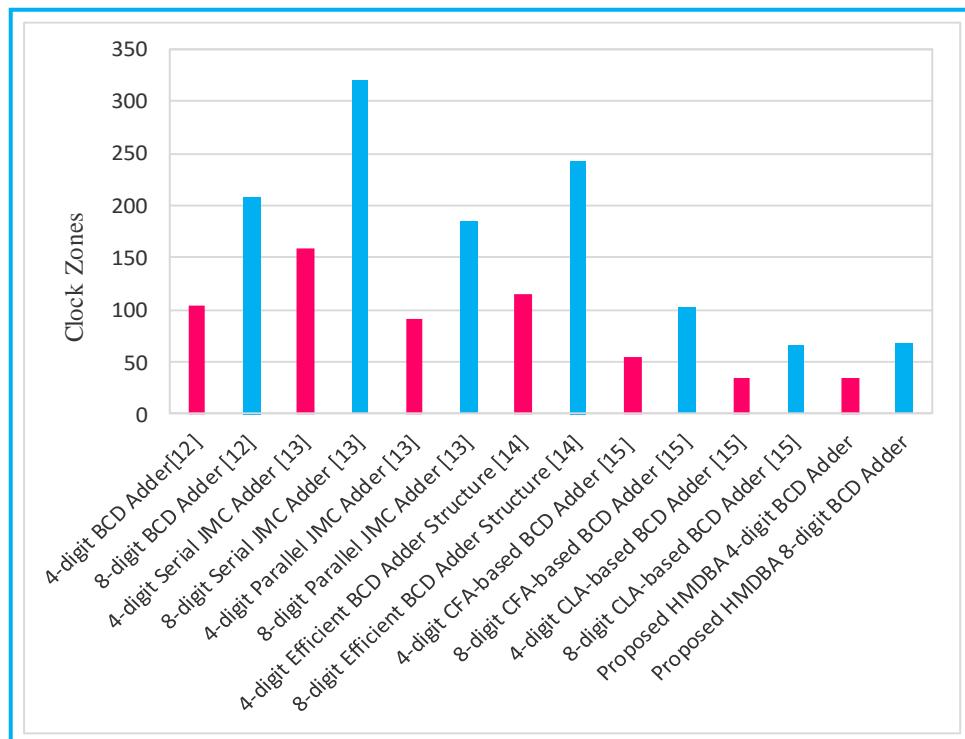


Figure 9. Comparison of the HMDBA performance regarding clock zones with existing methods

The CFA-based adder requires 9 clock zones (2.25 clock cycles) to produce the binary sum, whereas, the proposed design desires six clock zones (1.5 clock cycles) only. Further, the final part of the CFA-based BCD adder requires seven clock zones (1.75 clock cycles) to produce the valid decimal sum, however in the proposed design the 2:1 multiplexers requires three clock zones (0.75 clock cycle). While compared with the CLA-based BCD adder, the CLA-based adder requires seven clock zones (1.75 clock cycles) to produce the binary sum, correction logic, and to provide inputs to the multiplexers. However, the proposed design requires eight clock zones (2 clock cycles) with less area as well as less complexity to construct the circuit. The final part of the CLA-based BCD adder and proposed adder has the same delay. Therefore the HMDBA adder consumes less delay compared with CFA-based BCD adder and competing with CLA-based BCD adder. The overall cost function of the proposed HMDBA compared with existing designs is shown in Fig. 10.

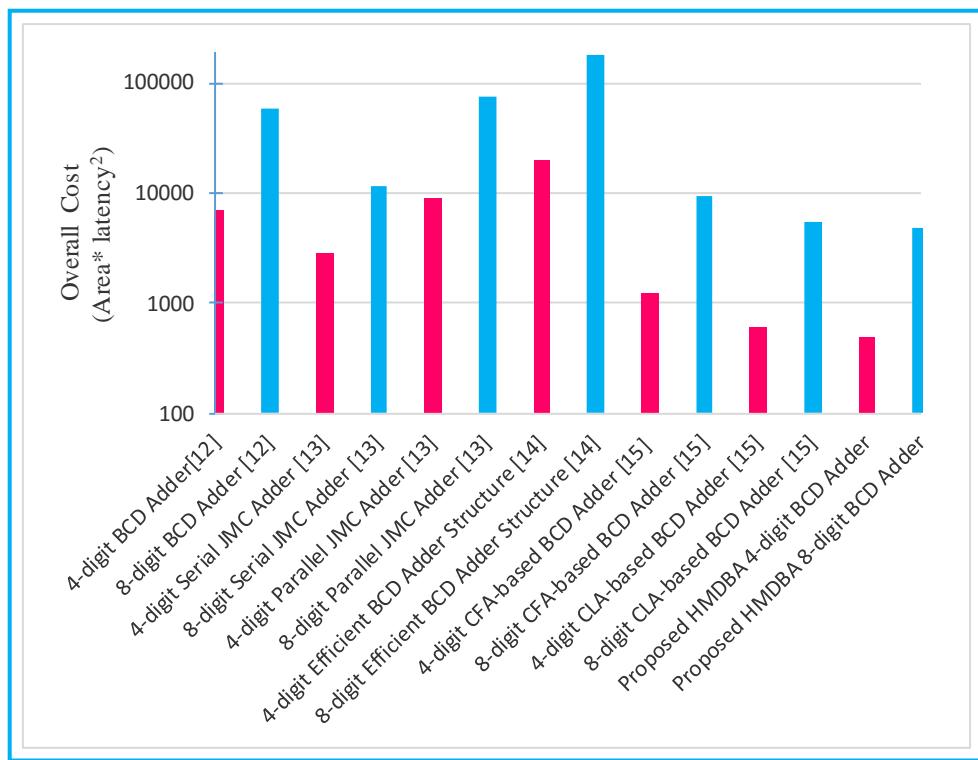


Figure 10. Overall cost of the HMDBA performance regarding delay with existing methods

The overall cost function of the circuit is given as Area \* latency<sup>2</sup> [21]. The overall cost of the proposed HMDBA design is less among all the existing designs as the HMDBA design occupies less area with optimized latency related to the current BCD adder designs.

#### IV. CONCLUSION

In this paper, a new model of BCD adder in combination with CFA-based binary adder and multiplexer is presented to perform the multi-digit BCD addition and implemented using QCA. The comparison of the HMDBA with the various existing methods clearly shows that the HMDBA offers better trade-off regarding speed and area between CFA and CLA-based BCD adders. The proposed HMDBA design significantly faster compared to CFA-

based BCD adder and utilizes less number of cells compared to the CLA-based BCD adder. From the obtained results it is resolved that N-digit proposed BCD adder design has a delay of  $(N-1)*2 + 2.75$  clock cycles (2.75 clock cycles is the delay of the first BCD adder) compared to the conventional design methods anticipated delay of  $N*One$  digit BCD adder. Moreover, the overall cost of the HMDBA is reduced by 50% compared to the CFA-based BCD adder. Hence, the proposed HMDBA design is an optimized design for Multi-digit BCD operations.

#### REFERENCES

- [1] J.M. Seminario, P. A. Derosa, L. E. Cordova, and B. H. Bozard, "A molecular device operating at terahertz frequencies: theoretical simulations," *IEEE Transactions on Nanotechnology*, vol. 3, no. 1, pp. 215–218, 2004.
- [2] R. P. Cowburn and M. E. Well and, "Room temperature magnetic Quantum Cellular Automata," *Science*, vol. 287, no. 5457, pp. 1466–1468, 2000.
- [3] C. S. Lent, P. D. Tougaw,W. Porod, and G. H. Bernstein, "Quantum Cellular Automata," *Nanotechnology*, vol. 4, no. 1, pp. 49–57, 1993.
- [4] G. H. Bernstein, A. Imre, V. Metlushko et al., "Magnetic QCA systems," *Microelectronics Journal*, vol. 36, no. 7, pp. 619–624, 2005.
- [5] G. L. Snider, A.O. Orlov, I.Amlani et al., "Experimental demonstration of Quantum-Dot Cellular Automata," *Semiconductor Science and Technology*, vol. 13, no. 8, pp.A130–A134, *International Scholarly Research Notices* 9, 1998.
- [6] P.D. Tougaw, Quantum Cellular Automata: computing with quantum dot molecules [Ph.D. thesis], University of Notre Dame, 1996.
- [7] N.G.Anderson and S. Bhanja, Field-Coupled Nano computing Paradigms, Progress, and Perspectives,vol 8280, Springer, Heidelberg, Germany, 2014.
- [8] Lent, Craig S., and P. Douglas Tougaw. "A device architecture for computing with quantum dots." *Proceedings of the IEEE* 85.4: 541-557, 1997.
- [9] Walus, Konrad, and Graham A. Jullien. "Design tools for an emerging SoC technology: Quantum-Dot Cellular Automata." *Proceedings of the IEEE* 94, no. 6, 1225-1244, 2006.
- [10] Porod, Wolfgang. "Quantum-dot devices and quantum-dot cellular automata." *International Journal of Bifurcation and Chaos* 7.10: 2199-2218, 1997.
- [11] M. Taghizadeh, M. Askari, K. Fardad, "BCD Computing Structures in Quantum Cellular Automata." International conference on Computer and Communication Engineering. 1042-1045, 2008.
- [12] F. Kharbash and G. Chaudhry, "The design of quantum-dot cellular automata decimal adder". 12th IEEE Int. Multitopic Conference. 71–75, 2008.
- [13] M. Gladshtean, "Quantum-dot cellular automata serial decimal adder". *IEEE Transactions in Nanotechnology*. 10, 1377-1382, 2011.
- [14] N.A. Shah, F. A. Khanday and Z. A. Bangi, "Quantum Cellular Automata Based Efficient BCD Adder Structure". *Journal of Communications in Information Science and Management Engineering*. 2, 11-14, 2012.
- [15] Weiqiang Liu, Liang Lu, Maire O'Neil and Earl E. Swartzlander, Cost efficient decimal adder design in Quantum-dot cellular automata, *IEEE International Symposium circuits and systems*. 1347-1350, 2012.
- [16] Ronald J. Tocci, Neal S. Widmer, Digital Systems:Principles and Applications, 3rd Ed., pp 245-255, (Prentice Hall ,2001).
- [17] Heumpil Cho, Adder and Multiplier Design and Analysis in Quantum-dot Cellular Automata. Doctoral Thesis, The University of Texas, Austin, 2006.
- [18] K. Wajus, T. Dysart, G. A. Jullien, and R. Budiman, "QCADesigner: a rapid design and simulation tool for Quantum-Dot Cellular Automata," *Trans. Nanotechnology*, vol. 3, no. 1, pp. 26-29, March 2004.
- [19] QCADesigner Documentation, <http://www.qcadesigner.ca/>.
- [20] K. Kim, K.Wu, and R. Karri, "Quantum-dot Cellular Automata design guideline," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E89-A, no. 6, pp. 1607–1614, 2006.
- [21] C. D. Thompson, "Area-time complexity of VLSI," in Proc. 7th Annual ACM Symp. Theory of Computing, pp. 81–88, 1979.

# Optimization and Traffic Management in IEEE 802.16 Multi-hop Relay Stations using Genetic and Priority Algorithms

Jawwad Ibrahim <sup>1</sup>, A. Rehman <sup>1\*</sup>, M. Saad Bin Ilyas <sup>1</sup>, Mohsin Shehzad <sup>2</sup>, Maryum Ashraf <sup>1</sup>

<sup>1</sup>*Department of Computer Science & Information Technology, The University of Lahore, Gujrat, Pakistan*

<sup>2</sup> *Huawei Technologies Co., Ltd. Islamabad, Pakistan*

\* Correspondence: azizur.rehman@cs.uol.edu.pk

## Abstract

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

## Index Terms

WiMAX, Multi-hop Relay Stations, Relaying Techniques, Genetic Algorithm, Priority Algorithm

## I. INTRODUCTION

Increasing demand for the wireless internet has made bandwidth utilization an indispensable concern for the telecommunication industry, making sure the bandwidth doesn't get wasted the problem of coverage holes must be addressed in the case of WiMAX. The areas with the low coverage or low signal strength are facilitated with the relay stations which helps to improve the signal strength that results in capacity enhancement. Coverage holes are one of the main reasons for the underutilization of bandwidth and low data rates.[1]

The best solution to such a problem is to install relay station in coverage holes and hence make use of cooperative relaying. Fig. 1, represents a cooperative network. Cooperative relaying in the multi-hop wireless network uses multiple antenna techniques and combines the relayed and direct signal and decodes it to maximize the capacity of the total network for given set of bandwidths. This is the difference between normal single hop network and cooperative relaying as the normal single hop system works on direct transmission only, the receiver in these systems receives and uses the information of direct signal only and treats relayed signal as interference.

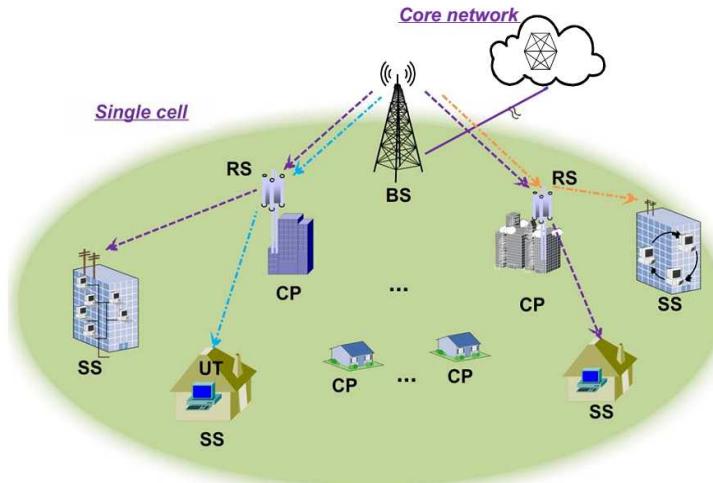


Fig. 1: Architecture of subscriber relaying [4]

Authors in [2], used relay channel with the main channel for transmission of bits. Channels upper bound capacity was found. The concept has recently been used when it is realized that some of the main issues of WiMAX can be solved on the basis of relaying. Some of these issues are given below

- Signal attenuation causes coverage limitation at the edge of cells
- Coverage holes are the reasons of poor signal reception
- An increase in the number of base stations adds in the network cost
- Mobile stations require more power when to communicate at long distance or high speed

Relay in IEEE 802.16j is defined as a set of equipment that depends on a multi-hop relay base station (MR-BS) that provides connectivity to other Relay stations or subscriber stations. The Relay station is a base station that requires line of sight (LOS) connectivity with another relay station or a base station. Relay station is not connected to the wired back-haul as shown in fig. 2. [3]

Broadband wireless networks which are mostly in place these days rely on LOS communication. As WiMAX is being targeted for providing higher data rates at large distances, so the coverage and the throughput of WiMAX should be better.

The difference between transmit and receive powers is so large that because of this relay cannot receive and transmit at the same time, as a result, the end to end transmission from base station to the mobile station and from the mobile station to base station systems do not require LOS. Hence for WiMAX, it is required to increase the number of base stations to provide LOS communication to all mobile stations. Increasing the base stations increases the cost of WiMAX deployment. A relay station is considered as an alternative to the problem described earlier, as relays work at lower powers and does not require a wired back-haul.

In the context of relaying techniques, we have several types available for this purpose. Relaying techniques can be categorized into three types:

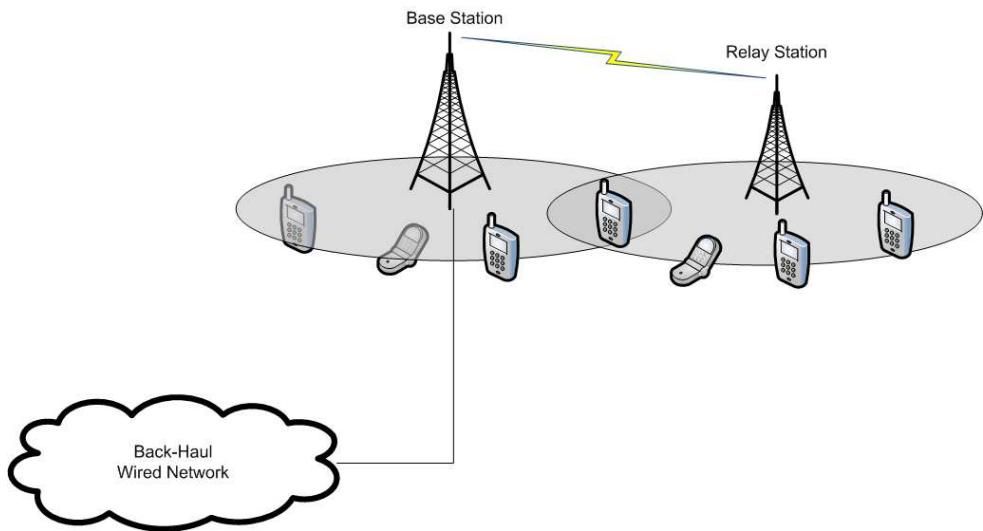


Fig. 2: Architecture of subscriber relaying

- **Amplify and Forward:** In this type of relaying technique, the relay station only re-transmits the received signal after amplifying it.
- **Decode and Forward:** These relays decode the data received in the first phase and perform error correction before forwarding. It only transmits data when it becomes error free.
- **Compress and Forward:** In this type of relaying, data which is received, is compressed in a lossy manner by making use of Wyner Ziv source coding and is forwarded thereafter. In this relaying it is assumed that the mobile station has an access to the direct transmission which is coming from base station [5], [6].

The rest of the paper is organized as follows. Section II covers the literature review with problem formulation. In Section III, solution method is proposed, whereas, Section IV, includes the Simulation Setup procedures. Section V shows the results and analysis details and last section VI, concludes the paper.

## II. LITERATURE REVIEW

By deploying Relay Station in cellular networks we can easily improve the system capacity as well as coverage of the network. In such a system which we can call relay assisted cellular network, a subscriber station (SS) can have made use of two possible links i.e. direct link to the base station (BS) and a link via a relay station (RS). This way data rate demands of SS can be easily met which otherwise is not possible.

By introducing RS to the network we can also enhance coverage radius of the cell by providing higher SNR to the SS which is located at the edges of the cell. Therefore it helps us to reduce the infrastructure cost of deploying more BS. The increase in coverage extension of the cell depends mainly on RS's placement in the cell. The reason behind this is that the locations of RS's affect the quality of communication of BS-RS and RS-SS links as well as inter cell interference from the neighboring cells. If an RS is located too close to the edge of the cell then there is a chance of low SNR on BS-RS link. Apart from this, it also causes higher interference to the neighboring cells.

On the other hand, if the RS is placed too close to the BS then the RS-SS link quality will be compromised and SS's at the edge of the cells won't benefit. Thus for a given set of system parameters, there is a need for optimal relay station placement to achieve maximum data rates and coverage.

Only a few researches have been made which have addressed the issue of optimal relay station placement in cellular networks. The authors in [7] and [8] analyze RS placement for wireless sensor networks, where the objective is to achieve maximum connectivity between pairs of adhoc relay nodes. In [9] and [10], the RS placement problem is analyzed from the perspective of Increasing system capacity rather than coverage radius extension. In [11], the same group of authors, developed an optimized framework that specifically considered solving the issue of Capacity Maximization RS Placement. According to this framework, it is aimed to achieve maximum capacity and satisfy the minimum traffic required for SS. The authors in [12] Consider dual relay architecture with cooperative RS pairs and proposes an algorithm to select the two best RS locations from a predefined set of candidate positions. In [13], an iterative RS placement algorithm is proposed which divides all points in the cell into good and bad coverage points and places RS's at the good points whose neighbors have bad coverage.

#### A. Problem Formulation

*1) Achievable Data Rates for Destination Node with D-F:* In decode and forward relaying the relay station receives the signal from the source decodes it and forwards it to the destination. It is assumed that the channels of the source to the relay node, source to destination and from the relay to the destination node are given as  $c_{21}e^{j\varphi 21}$ ,  $c_{31}e^{j\varphi 31}$ ,  $c_{32}e^{j\varphi 32}$ .

Let  $X_1$  and  $X_2$  are the channels at source and relay nodes transmitting information and  $Y_2$  and  $Y_3$  are the channels at relay and destination node receiving information.

The capacity of the broadcast channel from  $X_1$  to  $Y_2$  and  $Y_3$  with given  $X_2$  is

$$\max_{f(X_1, X_2)} I(X_1; Y_2, Y_3 | X_2) = \frac{1}{2} \log(1 + (1 - \beta)(c_{21}^2 + c_{31}^2)P_1) \quad (1)$$

$$R_1 = \max_{f(X_1, X_2)} \min\{I(X_1; Y_2 | X_2), I(X_1, X_2; Y_3)\} \quad (2)$$

Where the broadcast channel is reduced to the point-to-point channel because of decoding at the relay node, i.e.,  $I(X_1; Y_2, Y_3 | X_2)$  is reduced to  $I(X_1; Y_2 | X_2)$ . The capacity of the reduced broadcast channel is

$$\max_{f(X_1, X_2)} I(X_1; Y_2 | X_2) = \frac{1}{2} \log(1 + (1 - \beta)c_{21}^2 P_1) \quad (3)$$

Thus, the achievable rate is rewritten as

$$R_1 = \max_{0 \leq \beta \leq 1} \min\left\{\frac{1}{2} \log(1 + (1 - \beta)c_{21}^2 P_1), \right. \\ \left. \frac{1}{2} \log(1 + c_{31}^2 P_1 + c_{32}^2 P_2 + 2\sqrt{\beta c_{31}^2 c_{32}^2 P_1 P_2})\right\} \quad (4)$$

As discussed earlier that we are using Decode-Forward strategy for relaying towards the destination. The data rates for each destination have to be calculated through all possible candidate position. A relay subscriber association is shown below in fig. 3.

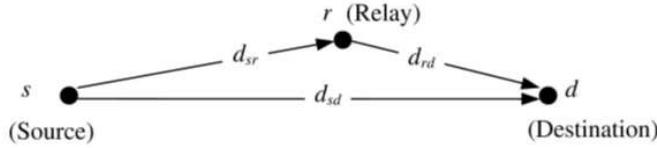


Fig. 3: Single subscriber relaying architecture

Calculation of data rate is divided into two parts in the first  $r^{(1)}$  part only rates from source to relay is considered and in second part  $r^{(2)}$  it has been added to rates of the relay to the destination and direct transmission. Where,

$$r^{(1)}(\theta) = C \left( \theta \frac{P_s}{N_o d_{sr}^\alpha} \right) \quad (5)$$

and,

$$r^{(2)}(\theta) = C \left( \frac{P_s}{N_o d_{sd}^\alpha} + \frac{P_r}{N_o d_{rd}^\alpha} + \frac{2}{N_o} \sqrt{(1-\theta) \frac{P_s P_r}{(d_{sd} d_{rd})^\alpha}} \right) \quad (6)$$

Where C is defined as Shannon capacity,

$$C(x) = \frac{1}{2} \log(1+x) \text{ for } x \geq 0$$

Where,  $\alpha$  is the path loss exponent ,  $P_s$  and  $P_r$  are the transmit power of source and relay.  $d_{sr}$ ,  $d_{sd}$  and  $d_{rd}$  are the distances from source to relay, source to destination and relay to destination respectively , which are calculated with Euclidian distance formula.

a) *Power Allocation:* One of the important factors which affect the achievable data rates directly is  $\theta$  parameter. It is the power allocation ratio at the source node between the “source relay” path and “source destination” path. Since  $r^1(\theta)$  is an increasing function of  $\theta$  and  $r^2(\theta)$  is a decreasing function of  $\theta$ . The value of  $\theta$  will be optimal when  $r^1(\theta) = r^2(\theta)$

$$\theta = \begin{cases} \frac{d_{sr}^\alpha}{P_s} \left( 2 \frac{(-d_{sd}^\alpha P_r + \sqrt{u})}{d_{sd}^\alpha d_{rd}^\alpha} + \frac{P_s}{d_{sd}^\alpha} + \frac{P_r}{d_{rd}^\alpha} \right) & \text{for } u \geq 0 \\ 1 & \text{for } u < 0 \end{cases} \quad (7)$$

Where,

$$u = P_r(d_{sd}^\alpha - d_{sr}^\alpha)(P_s d_{rd}^\alpha - P_r d_{sr}^\alpha) \quad (8)$$

b) *Final Data Rates:* Final data rates for the destination are the minimum of both the paths which in actual gives us the lower bound of capacity. The first part of this formula is an increasing function of  $\theta$  and the second part is a decreasing function of  $\theta$  but the final data rates that we get at the end are actually minimum of both.

$$r_{mn} = \min \left[ C \left( \theta \frac{P_s}{N_o d_{sr}^\alpha} \right), C \left( \frac{P_s}{N_o d_{sd}^\alpha} + \frac{P_r}{N_o d_{rd}^\alpha} + \frac{2}{N_o} \sqrt{(1-\theta) \frac{P_s P_r}{(d_{sd} d_{rd})^\alpha}} \right) \right] \quad (9)$$

The final data rates are simulated in MATLAB in the form of a matrix which is a M x N matrix. The first column in matrix ‘r’ in the fig. 4, shows that SS 1 is being relayed via all CP’s which are three. Similarly, the second and third column depicts that SS 2 and 3 are being relayed via all CP’s.

```

MATLAB 7.11.0 (R2010b)
File Edit Debug Parallel Desktop Window Help
Current Folder C:\Program Files\MATLAB\R2010b
Shortcuts How to Add What's New
Command Window
① New to MATLAB? Watch this Video, see Demos, or read Getting Started.

r1 =
2.3073    1.4415    0.5258
1.2080    1.2517    0.9354
2.8299    0.7757    0.7757

r2 =
2.4941 + 0.7085i  1.6587 + 0.5704i  1.0525
2.6834            1.2016 + 0.1215i  0.9886
3.0278 + 0.6769i  1.3197            0.9609

r =
2.3073          1.4415          0.5258
1.2080          1.2016 + 0.1215i  0.9354
2.8299          0.7757          0.7757

```

Fig. 4: Data rate simulation in MATLAB

2) *Capacity Maximization RS placement (CMRP) problem:* The system capacity C has to be increased by deploying fixed number of relay stations K and by allocating the bandwidth to subscriber stations by taking in account the location and traffic demands for subscribers.

Two sets of decision variables are defined, A and W.  $A = (a_{mn})$  MxN is an SS-CP incidence matrix or the location allocation matrix.  $a_{mn}$  indicates the index of the chosen CP serving a specific SS, such that  $a_{mn} = 1$  if  $SS_n$  is relayed via an RS located at  $CP_m$ ; otherwise,  $a_{mn} = 0$ .  $W = \{\omega_1 \dots \omega_n\}$  is the bandwidth allocation vector, where  $\omega_n$  is the amount of bandwidth assigned to  $SS_n$ .

$$\text{Maximize } C = \sum_{n=1}^M \sum_{m=1}^N a_{mn} \omega_n r_{mn} \geq \rho_n \quad (i)$$

$$\sum_{m=1}^M a_{mn} \omega_n r_{mn} \geq \rho_n \quad (ii)$$

$$\sum_{m=1}^M \max(0, 1 - 2(\sum_{n=1}^N a_{mn})) = M - K \quad (iii)$$

$$\sum_{m=1}^M a_{mn} = 1 \quad (iv)$$

$$\sum_{n=1}^N \omega_n \leq B \quad (v)$$

$$a_{mn} \in \{0, 1\} \quad (vi)$$

The objective function has following constraints

- i. The capacity increases in multi relay model
- ii. The minimum traffic demand is met
- iii. Satisfies the allocation of K relay stations on M candidate positions
- iv. Each subscriber is entertained by unique relay station
- v. The total bandwidth allocated to all stations should be less than the overall bandwidth of the cell
- vi. Station allocation matrix is binary

### III. PROPOSED SOLUTION

Due to the independencies of each point in solution set we have a mixed integer nonlinear problem which makes it a complex problem to solve for optimization. Solving the nonlinearity of the objective function becomes tough, the location allocation variables have discrete nature whereas bandwidth allocation is subjected to continuous nature. A solution is proposed which uses a Genetic algorithm to solve the problem.

#### A. Genetic Algorithm

Genetic Algorithms (GA's) are an automated searching algorithm which adapts to the solution itself. It is based on genetics and the evolutionary idea of natural selection. Genetic Algorithms are a part of artificial intelligence. "Survival of the fittest", Darwin's theory about evolution inspired this algorithm which makes it a part of evolutionary computing. GA's are used to solve optimization problems by performing a random search. GA's, perform a random search, but for the better performance they search the "search space" which has its history. GA's are close to nature, as in nature also the fittest and the strongest survives, so does in GA's, the fittest result is chosen. Solving problems mean looking for solutions, which is best among others. A search space has all the possible or partial solutions. This search space is usually searched to find the solutions in AI. The problem is first solved mathematically in terms of functions. Then the unknown parameters are then optimized to find the solutions.

GA is better than conventional AI results because they do not break in the presence of noise or even if the inputs are slightly changed. While searching the search space, GA's produces better results than heuristic approach etc [14],[15].

*1) Input Parameters:* First we define our input parameters which are SS's , BS , CP's , the number of generations, the size of initial population, the probability of cross over, probability of mutation. As explained earlier, optimization problem needs to be converted and molded in such a form that they can be solved through a genetic algorithm. After that, all the possible solutions are loaded into the code which is known as search space. This search space is a large number of possibilities which contains a combination of a number of relay stations to be deployed i.e.  $K < M$  where  $K$  is the number of relay station to be deployed and  $M$  is the total number of CP's. It also contains the permutations of these combinations up to the number of subscribers which are  $N$ .

2) *Initial Population:* After defining the population size a loop is placed which selects the number of solution sets which are equal to the population size. Random numbers are generated every time when the loop runs. This random number is basically a row number of the matrix (search space). That particular row gets selected and is placed into an initial population matrix which is represented by ‘j’. This loop goes from  $n = 1$  to assigned population size.

3) *Fitness Calculation:* For calculating fitness, a loop is placed which selects individual solution sets from the matrix ‘j’ every time it runs and it calculates the fitness by making use of the formula.

$$(\text{Achievable Data Rate} * \text{BW}) / \text{Traffic Demand}$$

Where BW = Bandwidth available

Before going into fitness calculation achievable data rates are loaded into the code with the name of fitness “calculation.mat”. With this formula fitness value of all the solution sets is calculated one by one and is placed in a matrix called ‘f’.

4) *Parent Selection:* For the selection of parents, two of the best solution sets are selected from the matrix ‘j’ based on the fitness value which is in matrix ‘f’. The best of out of these is named as elite.

5) *Cross Over:* A random number is generated if it fulfills the condition  $\text{Randn} \leq \text{Pc}$  and then cross over is performed between two parents and with the help of these two off springs are produced. Cross over point is randomized for every generation. Fitness value of these off springs is calculated one by one. If these values are more than the fitness value of the elite then it gets replaced by one of the off springs.

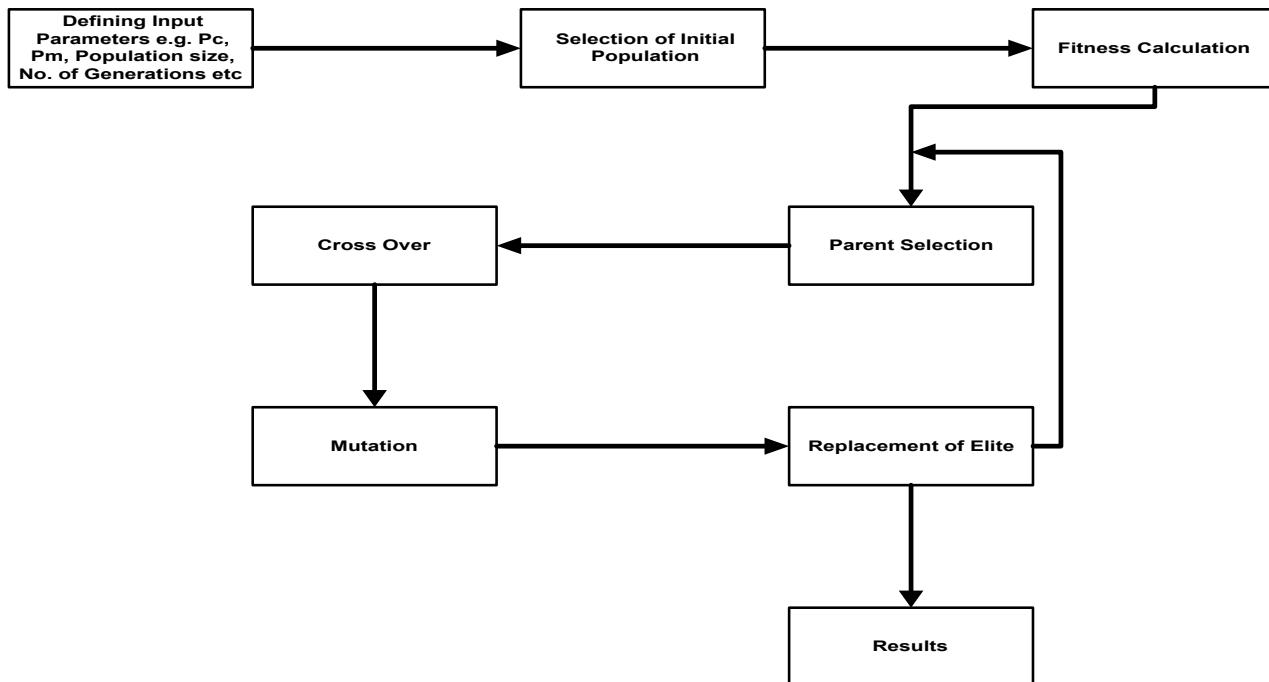


Fig. 5: Flow Chart MRMS problem

```
populationsize = 10

j =
4 1 1 1 1 4
3 4 4 4 4 4
4 4 2 2 2 2
4 4 3 3 3 3
3 3 3 2 3 3
1 1 2 2 1 2
3 3 3 4 4 3
4 3 3 4 4 3
2 2 2 4 2 2
4 2 2 2 4 2
```

Fig. 6: Population Size

```
crossoverpoint = 3

parent1 =
2 2 2 2 2 2

parent2 =
4 4 2 4 2 2

mutation

MUTATED =
2 2 2 2 2 2
```

Fig. 7: Crossover and parent selection

6) *Mutation:* Elite set is selected as the random number is generated if  $Randn \leq Pm$  then mutation within elite set will take place. Mutation points are randomized for every generation. Again fitness values are calculated of mutated elite if its more than the fitness value of elite then elite set will be replaced by mutated elite set. For next generation elite , from the last generation and one best solution set is selected from matrix ‘j’ and off spring is produced. This is how an optimal solution set is achieved. More number of generations will produce more optimized results. This method helps us to achieve optimality in a less time which otherwise takes a lot of time.

```

Elite =
2      2      2      4      2      2

parent1 =
2      2      2      4      2      2

parent2 =
4      4      2      2      2      2

```

Fig. 8: Elite and parent selection

7) *Flow Chart:* Here, a detailed flow chart of the algorithm is given as shown in fig. 9.

#### B. Priority Algorithm

Priority algorithm is another algorithm which we developed keeping in mind the cons of the previous algorithm. This algorithm gives better results as compared to GA as well as takes less computational time. Genetic algorithm fails in a scenario where subscribers having different or non uniform traffic demands. Here we are going to explain algorithm which deals in heuristic traffic demand, this algorithm converges more rapidly to optimal results in lesser time.

1) *Upper Bound of Capacity:* The upper bound of capacity is basically maximum cell capacity denoted by  $C_{ub}$ . This analytical bound is a constraint which helps in rapid estimating the capacity performance of cell in a particular network configuration. The network configuration includes the total bandwidth of cell locations of SS's and CP's for the deployment of RS's. The upper bound capacity is not affected by varying number of RS's.

$\omega_n = \omega_n^0 + \Delta_n$ , here  $\omega_n$  is the minimum bandwidth of SS's to achieve  $\rho_n$  by using the associated RS location information. We need to maximize the rate of SS's while minimizing  $\omega_n$  while achieving  $\rho_n$ .

So upper bound of the cell capacity can be derived as follows.

$$C = \sum_{m=1}^M \sum_{n=1}^N a_{mn} \omega_n r_{mn} = \sum_{m=1}^M \sum_{n=1}^N a_{mn} (\omega_n^0 + \Delta_n) r_{mn} \quad (10)$$

$$= \sum_{m=1}^M \sum_{n=1}^N a_{mn} \omega_n^0 r_{mn} + \sum_{m=1}^M \sum_{n=1}^N a_{mn} \Delta_n r_{mn} \quad (11)$$

$$= \sum_{n=1}^N \rho_n + \sum_{n=1}^N \Delta_n \sum_{m=1}^M a_{mn} r_{mn} \quad (12)$$

$$\leq \sum_{n=1}^N \rho_{n+\max} \left( \sum_{n=1}^N \Delta_n \left( \sum_{m=1}^M a_{mn} r_{mn} \right) \right) \quad (13)$$

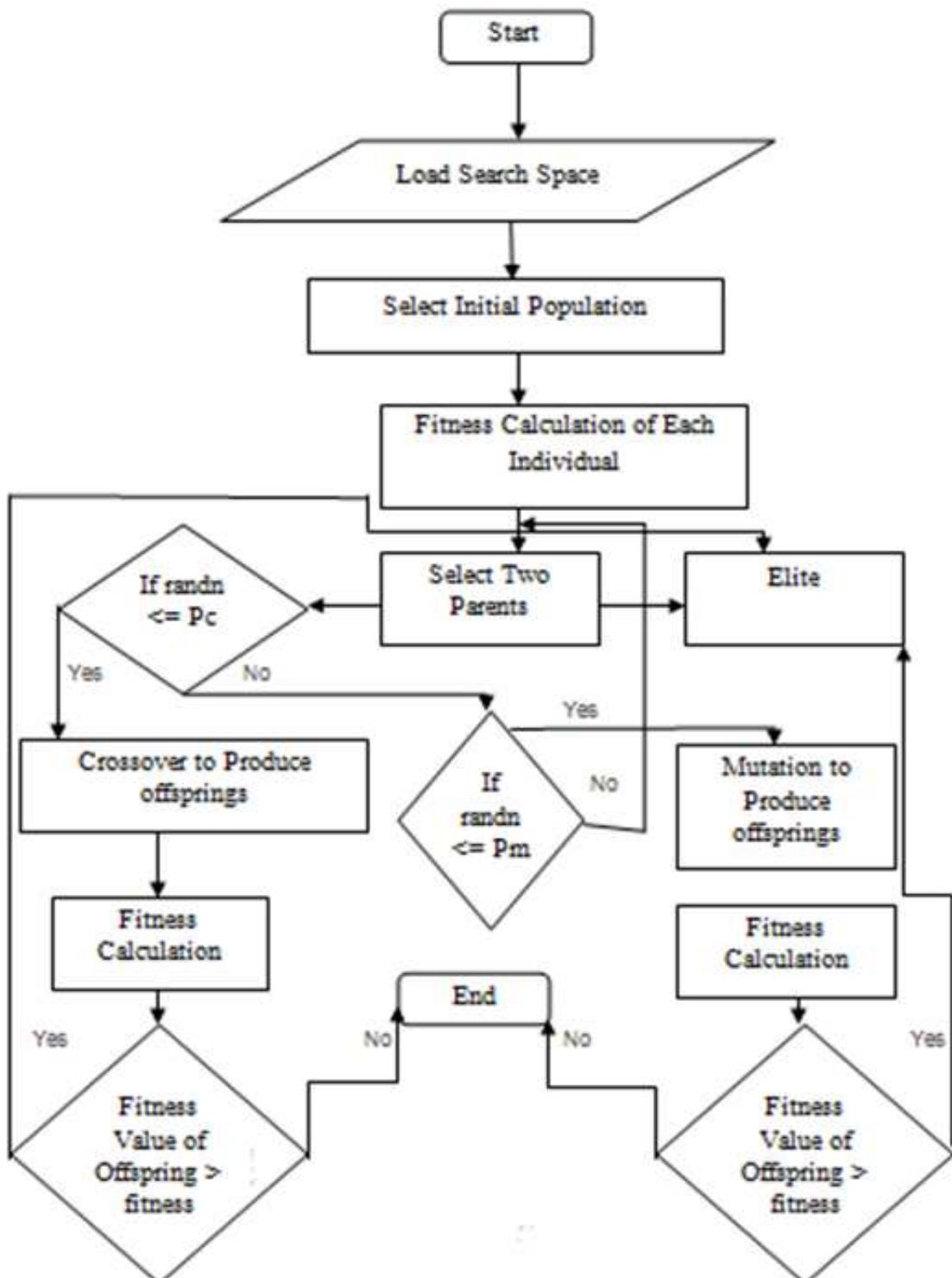


Fig. 9: Flow Chart of Proposed Algorithm

$$\leq \sum_{n=1}^N \rho_n + \left( BW - \sum_{n=1}^N \omega_n^0 \right) \max \quad (14)$$

The upper of cell capacity is expressed as

$$C^{ub} = \sum_{n=1}^N \rho_n + \left( BW - \sum_{n=1}^N \frac{\rho_n}{\max \forall_{mn} r_{mn}} \right) \max(r_{mn}) \quad (15)$$

The capacity requirement of the network will be fulfilled if the sum of all traffic demands of subscribers is less than the total capacity of the cell. To meet the capacity requirements either the network demands are to be upgraded or the cell capacity will be increased.

2) *Elaboration of Algorithm:* Achievable data rates  $r_{mn}$  are being calculated for SS's previously in GA. The upper bound of capacity that we have calculated above is being loaded in C into the code as a constraint which restricts the overall capacity of a cell. If the sum of the traffic demands of networks exceeds the upper bound capacity then no solution exists. If C is greater than the sum of minimum traffic demands of all subscribers then the subscribers are arranged in ascending order of their traffic demands and indexed in Nss. Then a loop is applied from 1 to total number of subscribers. In the first part of the loop, the number of SS's which are selected are less than or equal to the value of relay station to be deployed i.e.  $L < K$ . Then all the CP's are analyzed in search of maximum achievable data rate for those selected SS's. Once those CP's are selected their index number is stored in  $N_{rs}$ . In the second part of the loop, the achievable data rate for remaining SS's are being analyzed through the CP's indices which are being stored in  $N_{rs}$ . The SS's gets relayed through the index which provides maximum data rate. This is how all the SS's are relayed keeping in mind all the constraints.

#### IV. SIMULATION SETUP

The key step is the optimal placement of relay stations and the allocation of resources for the efficient association between the relay station and the subscriber to maximize the capacity of the cooperative relaying network and to achieve the minimum traffic demands of the subscribers.

A hexagonal cell as shown in fig. 10, centering a base station with green and red color shapes describes our problem model. The green circles illustrate the fixed subscriber locations, varying the size of the circle describes the varying traffic demands. Candidate positions are shown by the circles in red color where a relay station should be deployed. Drive testing is used to find the candidate positions after the gain calculations. The candidate position with the maximum gain has to be chosen as it is not possible to utilize all the candidate positions.

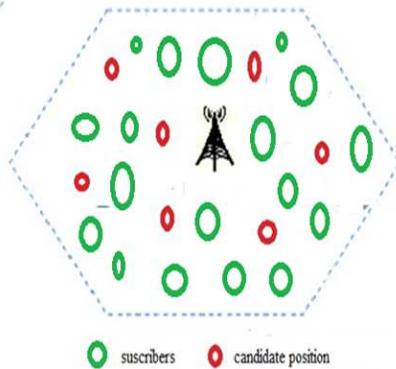


Fig. 10: Problem model for relay station placement

#### A. Implementing GUI Model

MATLAB tool “guide” is used to create GUI’s. After creating the GUI the tool allows the programmer to align, replace and edit the component properties such as color, size, and font etc. The basic steps for building a GUI are as follows:

- Select the elements required and define the function of elements.
- To give the name and set characteristics to each element MATLAB tool “property inspector” is used.
- There will be two files available after saving the created file. One file will contain the code and the other will contain the actual GUI.
- Call back function is used to write the code associated with each element.

#### B. Designed GUI Model

Designed GUI model consists of following main parts which make it easier for the end user to understand the proposed solution and find the results:

- Initialization
- Communication model
- Simulation and analysis
- Comparison
- Graph
- Processing time

The fig. 11, is a layout of our GUI which will be explained in detail.

1) *Initialize*: This section perform the actions of selection of a specified map, identify the candidate positions and the base stations. The upper bound capacity and it is determined that how many relay stations should be deployed. Following are the events that happen in this section:

a) *Select Map*: The purpose of this push button is to load the specified map of the location where the relay stations are to be deployed.

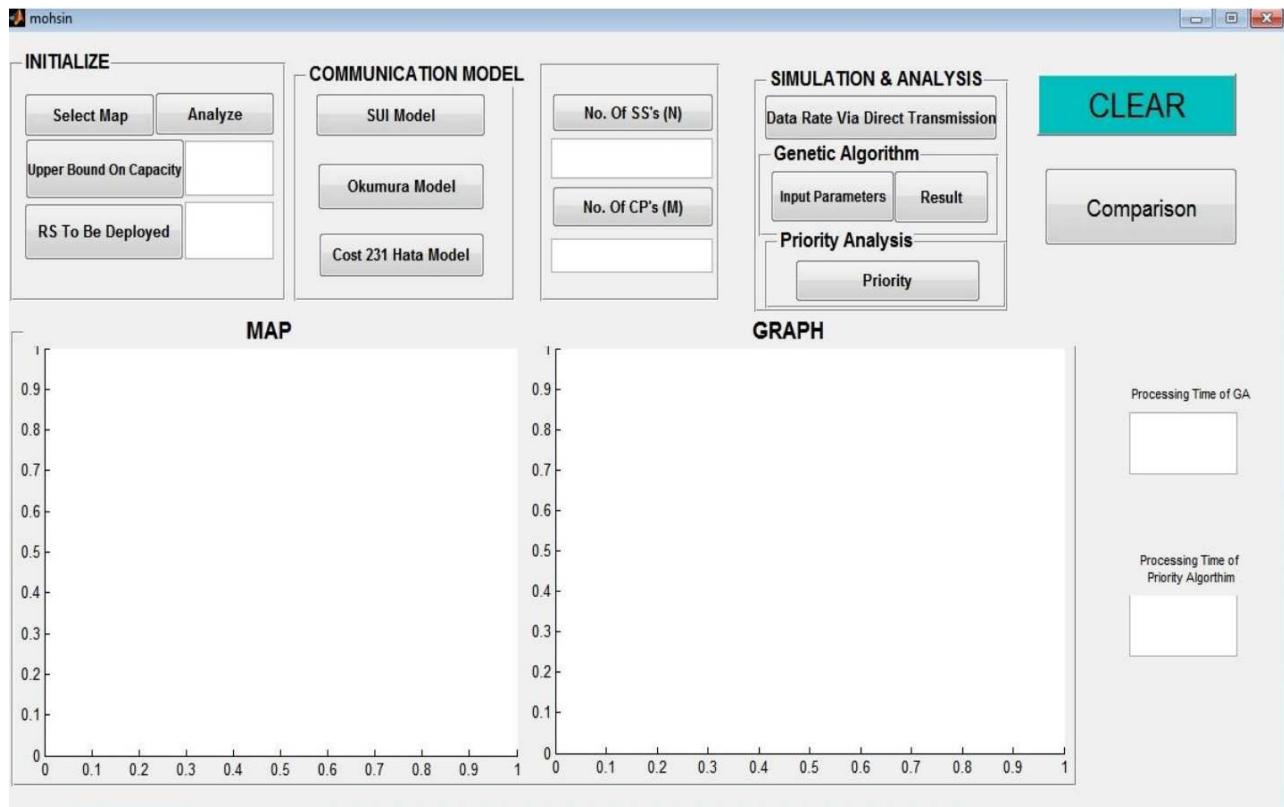


Fig. 11: Overview of GUI

b) *Analyze:* This section will calculate and find all the attributes by performing some image processing and working out different formulas discussed in the previous sections required for the further processing.

c) *Upper Bound capacity:* By clicking on this push button the upper bound capacity will be calculated with the help of formulas discussed earlier.

d) *RS to be deployed:* Number of Relay stations to be deployed will be given by the user in this part. If the number of relay stations exceed the number of candidate positions then the error will be shown.

2) *Communication Model:* The propagating constant alpha is calculated in this part of the program. As the propagation model differs from area to area i.e. propagation constant will be different for the rural areas and for city or green area. In the program, there are two models being introduced but the current program only works for the default value. The incorporated communication models are:

- Okumura Model
- SUI Model

3) *Simulation and Analysis:* This is the major part of the GUI where the main calculations are done. It performs following three vital functions:

- Calculates the direct transmission data rate i.e. from source to destination without relay station.

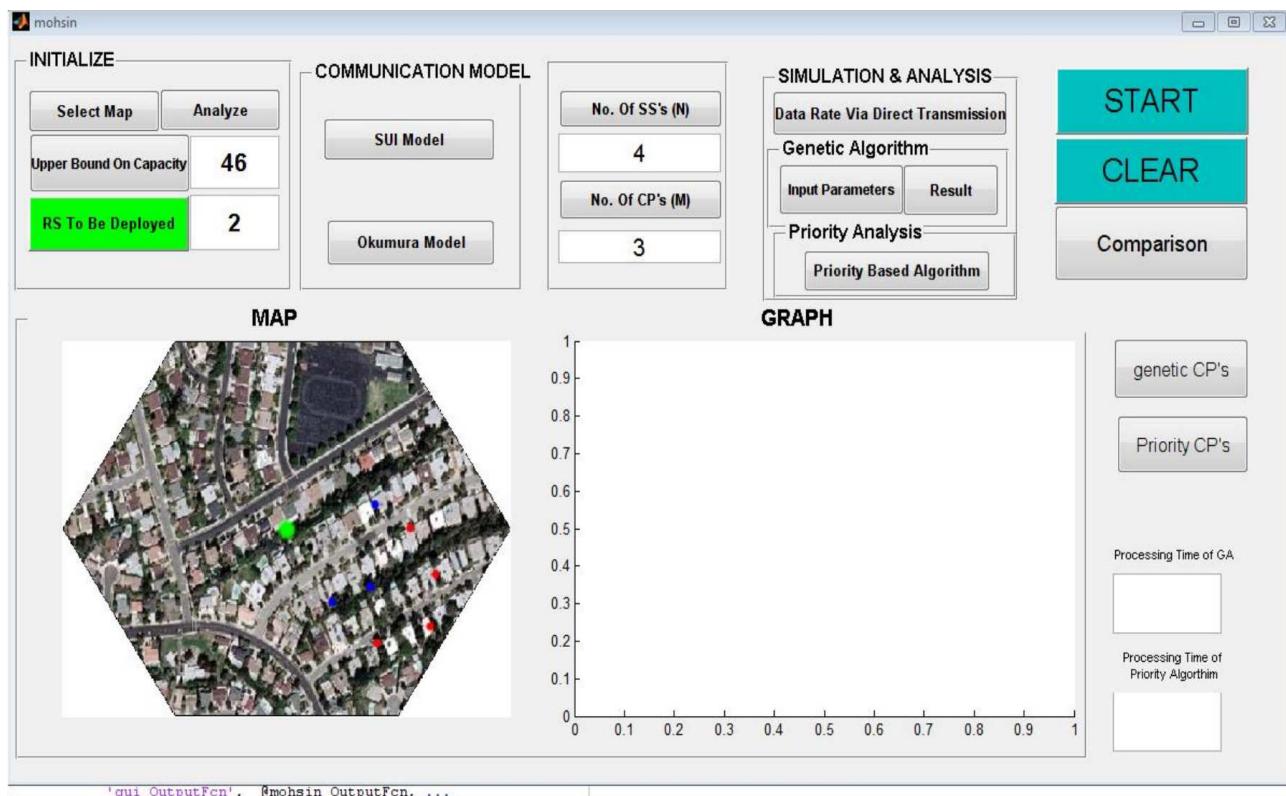


Fig. 12: Selection of Map and Image Processing

- Get the input parameters for the genetic algorithm and calculate the transmission data rate from source to relay and then to a destination on the basis of information given in the initialization part.
  - Incorporates the priority algorithm and show the results of an increase in the data rate on the graph.
- 4) *Processing Time:* Processing time highlights the processing time for the calculation and transmission of data from source to destination, when GA is performed, it performs the time calculation and shows the processing time. The Processing time of Priority algorithm shows the processing time taken while priority algorithm is used.

## V. RESULTS AND ANALYSIS

The numerical results are obtained by evaluating the convergence of Genetic Algorithm and improved Genetic Algorithm and by verifying the performance of relay stations. Cell with IEEE 802.16j using OFDM interference is considered. It is assumed that the external interference is constant. The transmission power for the base station is “1w” and for relay station is “0.5w”.  $\alpha$  (the path loss exponent) is 3.

In the initialization part, we have to first select map which is provided by the operator and contains the areas of interest where we have the subscribers and the candidate positions. The candidate positions are depicted by blue whereas, subscribers are depicted by red spots. We had use image processing commands and knowledge to extract the coordinates of the CP's and SS's.

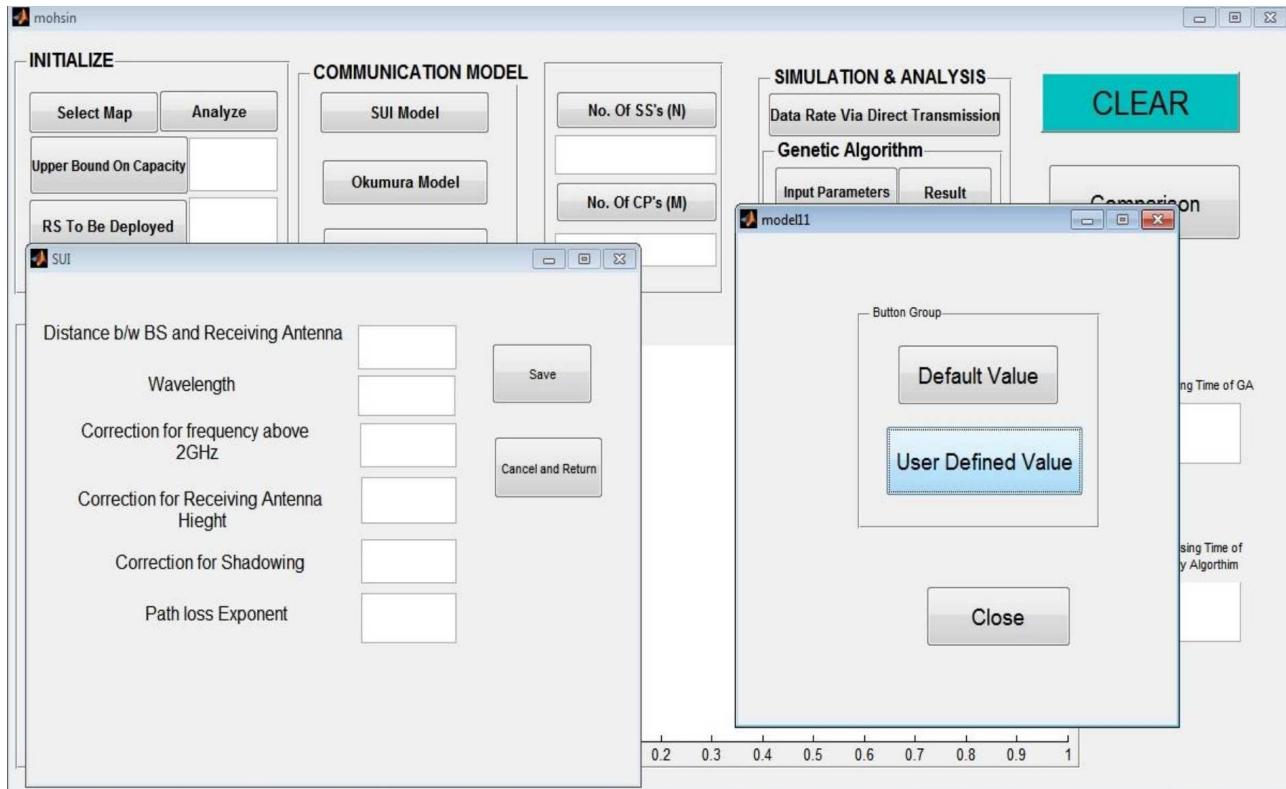


Fig. 13: Communication Model Analysis

After selecting the map and analyzing it we come to know about the coordinates of CP's as well as SS which are used in our MATLAB code, also we come to know about number of SS's and CP's. That is done when we click on the analyze button. After this, we calculate the upper bound of the capacity. We also input number of relay stations to be deployed and we keep a check on it as it cannot be more than the number of CP's. The whole procedure has shown in fig. 12.

In the next phase of Communication Model selection, we have two models which commonly used in wireless communications to calculate the values of the propagation constant depending on different factors like terrain, weather conditions, antenna height, LOS etc. We have set two options as shown in fig. 13.

- Default value
- Different parameters with different models

Till this point, we have gathered all necessary requirements for simulations to run. Fig. 14, shows the accumulative simulation results for all simulations. Where blue bar shows the results of Direct Transmission, green represents the results of Genetic Algorithm and the red one shows Priority Algorithm. Results clearly show that the algorithm that we have proposed is giving better results than the direct transmission, so it means that by placing RS's at optimal locations can enhance the capacity as well as coverage.



Fig. 14: Comparison of Simulations

We also see two blocks which are showing the run time of both the algorithms and after numerous simulations we found out that the priority time based algorithm takes less computational time than genetic algorithm making it more efficient.

## VI. CONCLUSION

The aim of this work was to enhance the overall capacity of the wireless network by using D-F cooperative relaying strategy as well as meeting the traffic demands of all subscribers. It is concluded through a number of experiments that the introduction of relay station resulted in significant increase in data rate as compared to the data rate achieved through direct transmission for every subscriber. The processing time also diminishes by using the priority. The proposed solution should play a vital role in the development, designing, and deployment of future wireless networks by providing guidelines for relay station placement and capacity planning.

## REFERENCES

- [1] Rehman, A., and Mustafa Shakir. Comparative Analysis of Scheduling Algorithms in IEEE 802.16 WiMAX. University of Bahrain: Scientific Publishing Center, 2014.
- [2] Cover, Thomas, and A. EL Gamal. "Capacity theorems for the relay channel." IEEE Transactions on Information Theory 25.5 (1979): 572-584.

- [3] Can, Basak, Hiroyuki Yomo, and Elisabeth De Carvalho. "Link adaptation and selection method for OFDM based wireless relay networks." *Journal of Communications and Networks* 9.2 (2007): 118-127.
- [4] Lin, Bin, et al. "Capacity enhancement with relay station placement in wireless cooperative networks." *2009 IEEE Wireless Communications and Networking Conference*. IEEE, 2009.
- [5] AlSanad, Abeer, Lulwah AlSuwaidan, and Mohammed Alnuem. "Strategies and Techniques for Relay Node Placement in Multi-hop Wireless Networks." *International Conference on Mobile Web and Information Systems*. Springer International Publishing, 2015.
- [6] Yang, Yang, et al. "Relay technologies for WiMAX and LTE-advanced mobile systems." *IEEE Communications Magazine* 47.10 (2009): 100-105.
- [7] Tang, Jian, Bin Hao, and Arunabha Sen. "Relay node placement in large scale wireless sensor networks." *Computer communications* 29.4 (2006): 490-501.
- [8] Liu, Hai, Pengjun Wan, and Xiaohua Jia. "On optimal placement of relay nodes for reliable connectivity in wireless sensor networks." *Journal of Combinatorial Optimization* 11.2 (2006): 249-260.
- [9] Lin, Bin, et al. "Optimal relay station placement in IEEE 802.16 j networks." *Proceedings of the 2007 international conference on Wireless communications and mobile computing*. ACM, 2007.
- [10] Meko, Sultan F., and Prasanna Chaporkar. "Channel partitioning and relay placement in multi-hop cellular networks." *2009 6th International Symposium on Wireless Communication Systems*. IEEE, 2009.
- [11] Lin, Bin, et al. "Capacity enhancement with relay station placement in wireless cooperative networks." *2009 IEEE Wireless Communications and Networking Conference*. IEEE, 2009.
- [12] Lin, Bin, et al. "Relay station placement in IEEE 802.16 j dual-relay MMR networks." *2008 IEEE International Conference on Communications*. IEEE, 2008.
- [13] Wei, Hung-yu, Samrat Ganguly, and Rauf Izmailov. "Ad hoc relay network planning for improving cellular data coverage." *Personal, Indoor and Mobile Radio Communications*, 2004. PIMRC 2004. 15th IEEE International Symposium on. Vol. 2. IEEE, 2004.
- [14] Golberg, David E. "Genetic algorithms in search, optimization, and machine learning." *Addion wesley* 1989 (1989): 102.
- [15] Mitchell, Melanie. "Genetic algorithms: An overview." *Complexity* 1.1 (1995): 31-39.

# A novel approach to Rank Authors in an Academic Network

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

Hikmat Ullah Khan<sup>\*1</sup>, Department of Computer Science, COMSATS Institute of Information  
Technology, Wah, Pakistan.

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

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

**Abstract-** Ranking the authors in an academic network is a significant research domain to find the top authors in various domains. We find various links based ranking algorithms and index based approaches to measure the productivity and impact of an author in a social network of authors. The research problem to rank the experts has vast applications such as advisor finding, domain expert identification. In this paper, we propose a novel approach to rank the scholars in the academic network of DBLP, a well-known computer science bibliography website. A huge data set is prepared covering the publications of more than 70 years. We propose AuthorRank and Weighted AuthorRank algorithms based on the state of the art ranking algorithms of PageRank and weighted PageRank algorithms respectively. For weighted algorithms, existing methods lack to provide diverse weights. We introduce the novel weights of h-index, g-index and R-index and elaborate their impact to identify the top authors in the scholarly network. The results confirm that the proposed algorithms find the top authors in an effective manner.

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

## I. INTRODUCTION

Social networks are created in the virtual communities of the social web. Social networks come into existence when people using internet interact with each other and share their common interest and work with each other and show liking and disliking about the items on the internet shared by these individuals [1]. These networks are of diverse nature, such as an academic social network like DBLP<sup>2</sup>, social tagging systems like Bibsonomy<sup>3</sup> and Delicious<sup>4</sup> and video sharing networks such as YouTube. A scholarly network is constituted by taking publication as a node and citation as a directed edge. The academic networks are based on publications, co-citation, and co-authorship [2]. There are a number of academic networks such as DBLP, Citeseer<sup>5</sup>, and Arnetminer<sup>6</sup> etc. In a scholarly network, the publications, authors, journals and conferences are considered as a node and the citation among these nodes act as links between them.

A number of research works explore and analyze the nature and behavior of the scholarly network in terms of impact and productivity of the scholars. In order to study the impact and productivity of the researcher, scientists proposed various indexing schemes [3]. These indexing and ranking methods are used to judge and compare the

---

<sup>\*1</sup> Corresponding author

<sup>2</sup> <http://dblp.org>

<sup>3</sup> <http://www.bibsonomy.org/>

<sup>4</sup> <https://delicious.com/>

<sup>5</sup> <http://www.citeseer.ist.psu.edu>

<sup>6</sup> <http://www.arnetminer.org>

impact and productivity of the authors. Usually these indices are used to measure the performance of the scholars in the academic institutions in order to offer scholarships and promotions to the faculty. All these indices use the traditional Bibliometrics like number of citations received by a scientist and the total number of publications of that scientist. It is difficult to measure the accurate impact of the research work, however, there are certain methods measure the quality of a scientist [4]. The measurement of impact and productivity have been carried out since past to compare the authors. The measures like Impact Factor, H-index and other indices have been proposed to measure the quality of the scholarly articles and indirectly that of the scholar. The related work Section reviews the relevant literature. .

In this research work, we have extracted the DBLP data for the period of 1936-2013. The ranking algorithm of PageRank is adapted and new weight of h-index, g-index and r-index and applied to rank the authors. Our main contributions are summarized as follows:

- Preparation of large data set of DBLP by extracting information, preprocessing it and then creating the academic network by taking authors, journal and conferences as nodes and citations as edge.
- Proposal of AuthorRank method based on PageRank algorithms.
- Proposal of Weighted AuthorRank method using novel weights of h-index, g-index and R-index.

The rest of the paper is organized as follows: In Section 2, existing works are reviewed, in Section 3, the proposed method is provided and in Section 4, results are discussed before concluding the paper.

## I. RELATED WORK

A number of research methods measure the quality and impact of a scholar. The basic metric to measure the quantity is the number of publications of a scholar and the basic metric to measure the quality is the number of citations. In 2001, Garfield proposed Impact Fact (IF) to measure the quality of an author [5]. But this metric has an issue that it measures average number of citations per paper, so an author may become reluctant to write more paper and may wait to increase average citation of existing papers. The next metric is the h-index, which is proposed by Hirsch [6] (6 J.E. Hirsch-2005). An author has the h-index of  $h$  value, if he/she has published  $h$  papers, each of which has been cited in other papers at least  $h$  times. H-index has become so significant metric to find the quality of a researcher that it has been applied in a number of other domains [7]. But, h-index has a limitation that once a paper is included in h-index, and then no further importance is given to that paper even if its citations become double. To counter this issue of h-core publication consideration for h-index computation, Egghe proposed g-index [8], which considers the h-core values and also every paper in the h-core has its importance.

Hypertext Induced Topic Selection (HITS) is a well-known link based ranking algorithm mainly introduced for ranking the web pages [9]. It gives the same importance to in-coming link (short as inlinks) as well as out-going link (short as outlinks) and thus, an outlink is not considered a better option for academic network. Another ranking algorithm that gives adequate importance to the inlinks is PageRank [10] (10 Page-1997,11 Brin-1998). The incoming links having high rank contribute accordingly. In a web environment, web pages have forward links that are called outgoing links and back-links called incoming links. The forward links of a web page can be easily

determined, but the back-links are difficult to find. It is not necessary that the links on a webpage are fixed and will not change. If there is collection of pages  $Q(k)$  and in this collection of pages, each page  $y$  is having a hyper link to page  $k$ , and then the formula for PageRank is given in equation 1:

$$PR(k) = (1 - d) + d \sum_{y \in Q(k)} \frac{PR(y)}{N(y)} \quad (1)$$

In the above equation,  $d$  denotes the damping factor. At start, all the incoming links and the outgoing links are treated equally and thus distributing the rank among them in an equal way, while in reality all links are not of the same importance, different links have different importance. Another drawback is that it is aimed for a random surfer but every user is not a random surfer, for example, it is not always going to provide good results to a researcher. In PageRank the older pages get a higher rank than the new one which is also a drawback of PageRank.

A weighted version of PageRank algorithm is presented by W. Xing et al [12] (12 Wenpu-2004) which gives weight to the pages that are using forward link as well as backward links instead of treating all the web pages equal initially. The number of forward links as well as backward links gives a good idea to assign weights. A larger weight is assigned to the web pages that are more popular as compared to the web pages that are less popular. The forward link as well as backward links both contributes to determine the popularity of each web page and the rank is allocated to a page according to its popularity. The limitation of weighted PageRank is that like original PageRank, it only consider forward link as well as backward links and does not consider the other factors like semantics of a web page. In (12 Wenpu-2004) the author has shown that weighted PageRank algorithm performs better than HITS and simple PageRank.

The weighted PageRank defined in (12 Wenpu-2004) is as follows:

$$PR(u) = (1 - d) + d \sum_{v \in L(u)} PR(v) W_{(v,u)}^{in} W_{(v,u)}^{out} \quad (2)$$

In above equation there are two important measures which are  $W_{(v,u)}^{in}$  and  $W_{(v,u)}^{out}$  are used to denote the back-links and forward-links to and from  $u$ , and  $u$  is a page for which the equation calculates the ranking results. Both  $W_{(v,u)}^{in}$  and  $W_{(v,u)}^{out}$  are calculated using equations 3 and 4 respectively:

$$W_{(v,u)}^{in} = \frac{I_u}{\sum_{p \neq u} I_p} \quad (3)$$

$$W_{(v,u)}^{out} = \frac{O_u}{\sum_{p \neq u} O_p} \quad (4)$$

In above equation  $I_u$  and  $O_u$  mean back-links and forward-links counts of  $u$ , and  $I_p$  and  $O_p$  mean back-links and forward-link counts of  $p$ . The topic-sensitive PageRank is proposed to rank the pages based on the topic discussed in the webpages [13]. (13 Taher-2003). Ying Ding et al., proposed weighted author rank algorithm (14 Ying-2009) to determine the reputation of researchers and used the equation 4 for weighted PageRank:

$$PR_{w(p)} = (1 - d) * \frac{w(p)}{\sum_{k=1}^n w(pk)} + d \sum_{p \in M(p)} \frac{PR_w(p)}{L(p)} \quad (5)$$

The proposed work is a continuation of our proposed frameworks [15] for finding more relevant information and feature centric, models [16,17] to find the top bloggers in a community of bloggers. This work is different that it uses the link based features while earlier approaches deals in content based factors.

### III. DBLP DATASET

In this research, DBLP<sup>7</sup> data are used. DBLP is a widely used bibliography portal of computer science listing more than 3.1 million articles<sup>8</sup> in October 2015. The downloaded data set is an XML file which has size 1.23GB and has 3818185 publications, 6598 conferences, and 1403 journals. The dataset contains data latest by December 2013. The XML file of the data set is imported in the Oracle database by developing an application to convert XML data into database format. A sample of data set records is given as follows:

```
<article mdate="2002-01-03" key="persons/Codd69">
    <author>E. F. Codd</author>
    <title>Derivability, Redundancy and Consistency of Relations Stored in Large Data
        Banks.</title>
    <journal>IBM Research Report, San Jose, California</journal>
    <volume>RJ599</volume>
    <month>August</month>
    <year>1969</year>
    <cdrom>ibmTR/rj599.pdf</cdrom>
    <ee>db/labs/ibm/RJ599.html</ee>
</article>
```

The dataset characteristics are provided in Table 1.

TABLE I  
THE DBLP DATASET CHARACTERISTICS

Characteristics	Value
Publications Period	1936-2013
Number of Publications	38,18,185
Number of Conferences	6,598
Number of Journals	1,403
Number of Authors	13,51,586

<sup>7</sup> <http://dblp.uni-trier.de/xml/>

<sup>8</sup> <http://en.wikipedia.org/wiki/DBLP>. Retrieved March 01, 2015.

#### IV. PROPOSED METHODOLOGY

In this research paper, PageRank and Weighted PageRank are adapted for ranking the authors. Adapted versions of PageRank and Weighted PageRank for authors are discussed below:

### A. AuthorRank

In our case, an author acts as node and an edge is created when one author in his/her publication cites the publication of another author. The AuthorRank is adapted from PageRank [16] as follows:

$$AR(k) = (1 - d) + d \sum_{c \in Q(k)} \frac{AR(c)}{N(c)} \quad \dots \quad (6)$$

In above equation  $k$  is for author,  $AR(k)$  represents AuthorRank of author  $k$ ,  $G(k)$  is the set of authors who are citing publications of author  $k$ ,  $c$  is for the author, for whom rank is being computed,  $M(c)$  is the set of outlinks of  $c$ ,  $d$  is the damping factor having a standard value of 0.85. The same symbols are being used for all the proposed approaches.

### B. WeightedAuthorRank using novel weights

The concept of assigning weights to the ranking entities is an important and significant contribution to attain better ranking results. We here propose novel weights of the use of author indexes. We posit here that as the indexes such as h-index, g-index, and R-index already play important role in ranking the authors in academic networks so the use to indexes as weight is significant weight that help to rank the authors. The adapted versions of weighted PageRank [10,1118] for author, journal and country are given below:

The weighted AuthorRank is:

$$ARW_{(a,b)} = (1 - d) \cdot \frac{w(a_b)}{\sum_{k=1}^n w(a_k)} + d \cdot \sum_{\text{examples } x \in a_b} \frac{ARW(a_k)}{L(a_k)} \quad (7)$$

Where  $AR_{w(a_i)}$  represents the weighted Author Rank,  $w(a_i)$  represents the weight of  $i^{\text{th}}$  author,  $a_x$  represents the set of authors having inlinks to publications of  $i^{\text{th}}$  author,  $L(a_x)$  represents the number of outlinks of  $a_i$ ,  $a_k$  represents the set of all the authors in the network of  $N$  authors,  $\alpha$  is the damping factor having value 0.85. In the equation 7, h-index, g-index and R-index have been used as weight.

## V. RESULTS DISCUSSION

The results of the proposed methods for ranking of authors, journals and conferences are discussed. AuthorRank is an adapted form of PageRank, in which, both in-links and out-links are used to calculate the rank of an author. The weights used for weighted AuthorRank are h-index, g-index and R-index. The top ten results of the AuthorRank and weighted AuthorRank are provided in table 2. F.F.Codd is ranked as top author by AuthorRank and Weighted AuthorRank algorithms. The variation in the results, measured by the use of variance measure, shows the impact of weight for calculating the rank of the author.

TABLE 2  
AUTORRANK AND WEIGHTEDAUTORRANK RESULTS FOR TOP 10 AUTHORS

Author	AuthorRank	WeightedAuthorRank (h-index weight)	WeightedAuthorRank (g-index weight)	WeightedAuthorRank (r-index weight)	Variance
E. F. Codd	1	5	1	1	4.00
Michael Stonebreaker	2	1	2	2	0.25
Jim gray	3	4	4	4	0.25
Jeffrey d. Ullman	4	2	3	3	0.67
Donald d. Chamberlin	5	9	7	7	4.25
Philip a. Bernstein	6	3	5	5	1.00
Raymond a. Lorie	7	6	6	6	0.25
Kapali p. Eswaran	8	22	22	10	68.00
Morton m. Astrahan	9	13	13	9	11.58
Peter p. Chen	10	16	11	13	14.25

Fig 1., shows that the variations of the top authors ranked by the AuthorRank. The ranking positions of a number of authors are close to each other. Also the variations in authors' ranking, using the proposed algorithms, are almost uniform, which reveals that ranking position of this author is stable for all ranking algorithms. Uniform ranking position in all the proposed algorithms verifies the correct ordering as well. When the ranking position of the authors increases in magnitude, then, the variation in the rank number of others algorithm also increases. For top five authors, ranking orders are more uniform which validates the correctness of the proposed algorithms. The variations in remaining five authors are understandable as it denotes the variations in the weights which are indexes which are different in calculation nature from one another.

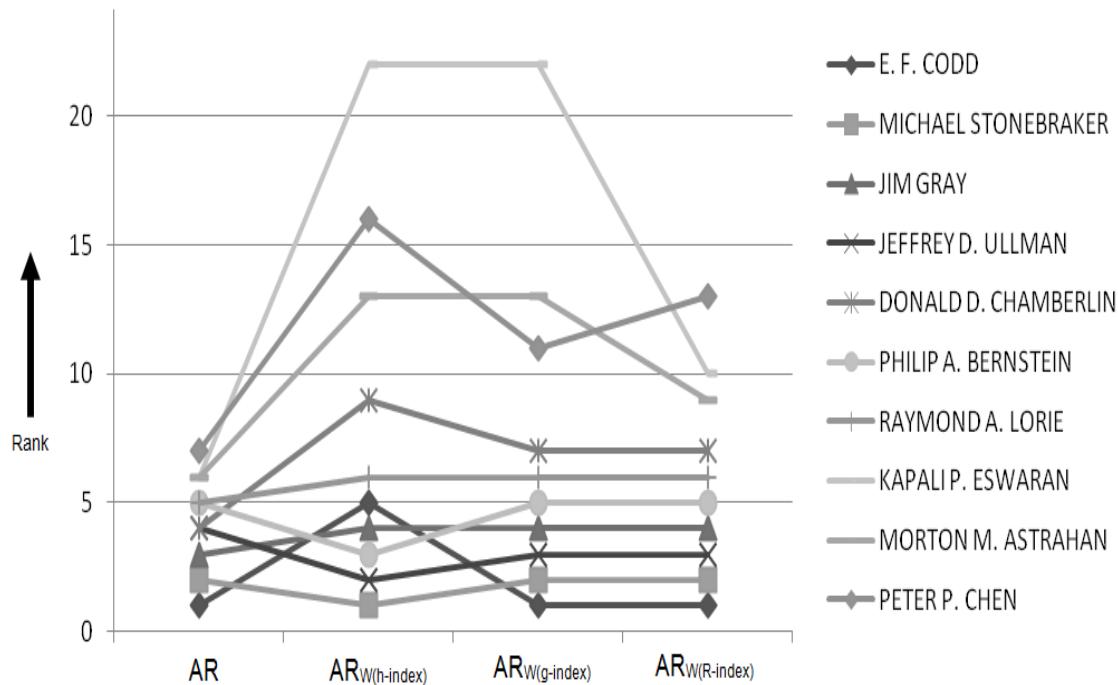


Fig 1. Top Ten Authors ranked using AuthorRank and Weighted AuthorRank

## VI. CONCLUSION

In this research work, we have prepared a large dataset of DBLP which is world famous computer science bibliography. We have prepared the dataset for all the records in the bibliography from year 1936 to 2013 which shows the significance of the work. PageRank, one of the most widely used ranking algorithms, has been adopted to rank the authors. In addition, the weighted version of PageRank using the novel weights of h-index, g-index and R-index has also been proposed. The results of top k have been discussed. The variations within the results have also been discussed. The top authors are similar but there are a lot of variations among the lower ranking authors, but there is very less variations in the results of journals and conferences which show the significance of our work. In future, we would like to further extend our work and would like to rank the journal, conferences and compare the proposed methods with the various index schemes.

## ACKNOWLEDGMENT

We are thankful to DBLP authorities for allowing us to extract the data for research purpose.

## REFERENCES

- [1] C. Haythornthwaite, ‘Social network analysis: An approach and technique for the study of information exchange”, *Library & Information Science Research*, vol. 18, no. 4, pp. 323–342, 1996.
- [2] Y. Sun, R. Barber, M. Gupta, C.C. Aggarwal and J. Han, “Co-author relationship prediction in heterogeneous bibliographical networks”, in *Proc. International Conference on Advances in Social Network Analysis and Mining*, 2011.
- [3] J. Panaretos and C.C. Malesios, “Assessing scientific research performance and impact with single indices”, *Scientometrics*, Vol 81, pp 635–670, 2009.
- [4] D. Bouyssou and T. Marchant, “Consistent bibliometric rankings of authors and of journals”, *Journal of Informetrics*, vol. 4, pp. 365-378, 2010.
- [5] E. Garfield, Impact factors and why they won’t go away, *Nature*, vol.411, no. 6837, pp. 522-522, 2011.
- [6] J.E. Hirsch, An index to quantify an individual research output, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 16569-16572, 2005.
- [7] S. Alonso, F.J. Cabreizo, E-Herrera-Viedma and F. Herrera, “h-index: a review focused on its variations, computations and standardization for different scientific fields.”, vol. 3, no. 4, pp. 273-289, 2009.
- [8] L. Egghe, and R. Rousseau, “An informetric model for the Hirschindex”, *Scientometrics*, vol. 69, no. 1, pp. 121-129, 2006.
- [9] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment.” *Journal of the ACM*, vol. 46, no.5, pp. 604-632, 1999.
- [11] S. Brin, and L. Page, “The anatomy of a large-scale hypertextual web search engine.”, *Computer Networks and ISDN Systems*, vol. 30, pp. 107-117, 1998.
- [11] L. Page, S. Brin, R. Motwani and T. Winograd, “The PageRank citation ranking: Bringing order to the web.” Technical Report, Standard InfoLab, 1999.
- [12] W. Xing, A. Ghorbani, “Weighted PageRank algorithm”, in *Proc. Second Annual Conference on Communication Networks and Services Research*, pp-305-314, 2004.
- [13] T.H. Haveliwala, “Topic-Sensitive PageRank: A context-sensitive ranking algorithm for web search”, *IEEE Transaction on Knowledge and Data Engineering*, vol. 15, No 4, pp. 784-796, 2003.
- [14] Y. Ding, E. Yan, A. Frazho and J. Caverlee, “PageRank for ranking authors in co-citation networks”, *Journal of the American Society for Information Science and Technology*, vol. 60, pp. 2229–2243, 2009.
- [15] H.U. Khan, S.M. Saqlain, M. Shoaib, M. Sher, “Ontology based semantic search in Holy Quran”, *International Journal of Future Computer and Communication*, Vol. 2, no 2, page 1-6, 2013.
- [16] H.U. Khan and A. Daud, T.A. Malik, “MIIB: A metric to identify top influential bloggers in a community”, *Plos One*, vol. 10, no. 9, 2015.
- [17] U. Ishfaq, H.U. Khan, K. Iqbal, “Modeling to find the top bloggers using sentiment features”, in *Proc. International Conference on Computing, Electronic and Electrical Engineering 2016*, pp. 227-233.

# A Semantic Multi-Agent Architecture for Multilingual Machine Translation

Salam Fraihat, Qusai Shambour and Mou'ath Hourani  
*Software Engineering Department, Faculty of Information Technology  
Al-Ahliyya Amman University, PO Box 19328, Amman, Jordan*

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

## I. INTRODUCTION

Machine Translation (MT) System is software system that produces translations, with or without human assistance, from one natural language into another. It produces translation throughout three main steps including the analysis of text of the source language, translation of text in the source language and generation of the text in the target language. These steps are essential to grasp the syntax, morphology and semantic features of two languages [1]. Although, MT systems can generate translations more cheaper and quickly than human translators; the quality of MT systems is less than the quality of human translation [2].

A software agent is an active and task-oriented software entity that performs tasks and cooperates with other agents to achieve their intended objectives. An agent has the capability to flexibly and autonomously perform in the environment where it is placed. Multi-agent system (MAS) is a network of software agents that are communicating and coordinating within each other to achieve specific goals that cannot be accomplished by a single agent. MAS exploits the power of knowledge sharing and exchange as key strategy for complex problem solving [3]. In terms of the development of MAS, there are currently a number of available general purpose frameworks and toolkits that provide standardization of the development of MAS and save time and effort of developers. Most of these frameworks are particularly developed to create simulations of complex systems and general purpose applications. Natural language processing, in particular, MT may benefit from the use of multiple agents that can work together to improve the translation process [1, 4]. However, to the best of our knowledge, there are limited researches in the area of using MAS in natural language processing, in particular, MT. Accordingly, much more research is needed to further develop and refine the area of using multi-agent technology for MT.

As defined by Tim Berners-Lee, “The Semantic Web is what we will get if we perform the same globalization process to Knowledge Representation that the Web initially did to Hypertext” [5]. The Semantic Web considered as an expansion of the existing Web in which information can give an unambiguous meaning, so that people and

computers can work and cooperate with each other effectively [6]. The semantic web is realized by adding semantics to the web in which it gives information a well-defined semantic meaning, so it makes it possible to facilitate information representing, interpreting, searching, sharing and reusing [7]. In particular, ontologies have been commonly employed as a way of providing the semantics to retain the retrieval information based on the intentional meaning more willingly than basically matching the search terms [3]. However, it has been proven that incorporating the semantic features of languages via ontologies in MT systems can effectively increase the quality of translations as such features can model relationships that are not realized from syntactic structures [2].

To this end, this paper reports the design of a Semantic Multi-Agent Architecture which is designed to develop Multilingual MT systems. In the proposed architecture, Multi-agent technology and ontologies will be incorporated to produce collaborative working environment for Multilingual MT. The automatic reasoning capacity of agents and their collaboration will improve the quality of translation process. Whereas, the use of ontologies has the benefit of translating the inputted text from the source to the target languages based on the intentional meaning rather than simply match the inputted terms based on their syntactic structure. The reminder of this paper is organized as follows: Section 2 presents an overview of the research background and related works. The proposed Semantic Multi-Agent Architecture is described in Section 3. Finally, conclusion and future works are discussed in Section 4.

## II. BACKGROUND AND RELATED WORKS

### A. Multilingual Machine Translation

Machine Translation (MT) systems are software systems used to produce text or speech translations from one natural language to another with or without human support. It is a challenging and demanding task to develop a MT system for a variety of languages with limited electronic tools and resources. MT systems can be: 1) bilingual systems which are particularly designed for two particular languages, or 2) multilingual systems which are particularly designed for more than a single pair of languages. A bilingual system may be either from one Source Language (SL) into one Target Language (TL), called unidirectional MT system, or may be from one Source Language (SL) into one Target Language (TL), and vice versa, called bidirectional MT system. Most bilingual systems are unidirectional, but multilingual systems are generally designed to be bidirectional [8].

MT systems can be categorized according to the use of traditional or modern technology and by which means they perform translation. MT systems can be classified into Direct MT, Rule based Translation, Corpus based Translation, and Knowledge based Translation [9, 10].

#### 1) Direct MT

In direct MT systems, a source language is translated directly into another language in which a direct word by word translation of the input source is performed with or without maintaining the sense of the word. Direct translation requires a bilingual dictionary as well as a morphological word analyzer. In direct MT, the translator replaces the corresponding word in the target language dictionary without taking into account any grammatical rules. The direct MT is unidirectional and it takes only one single pair of languages into account simultaneously [9, 10].

#### 2) Rule based translation

In rule-based MT systems, the source language text is parsed and an intermediate representation is created, then the intermediate representation is translated into the target language. Rule based MT system can be additionally classified into Transfer based MT and Interlingua based MT. In the Transfer based MT, a set of linguistic rules is used to translate text from source to target language to preserve the meaning of a sentence. When sentences go with one of the transfer rules, it is translated directly using a dictionary. Transfer rules are realized by analyzing the grammatical structures of both the source and target languages. In the Interlingua based MT, the source language is converted into an intermediary universal language, called Interlingua, then Interlingua is translated to more than one target language [9, 10].

### 3) *Corpus based translation*

Corpus based MT is promising and becoming a very popular translation approach in today's world. Corpus based MT requires a huge dataset since it is based on a corpus to perform a statistical analysis of source and target languages. Corpus based MT system can be additionally classified into statistical based MT and example based MT. The statistical based MT is based on the statistical models which are extracted from the corpus of both the source and target languages. The statistical models, for translating from the source language to the target language, are created using supervised or unsupervised statistical machine learning algorithms as well as some statistical information such as the characteristics of the sentences of the source and target languages. The example based MT preserves a corpus consisting of a set of translation examples between source and target languages. Firstly, it retrieves a comparable sentence and its translation from the corpus, then it adapts the retrieved translation to obtain the final proper translation [9,10].

### 4) *Knowledge based translation*

There was a limited semantic analysis in early MT systems which mostly consider the use of syntax. Language analysis using semantic based approaches have been recently introduced. Knowledge based MT requires a large knowledge base that consist of both lexical and ontological knowledge [9, 10].

Among the successful implementation of MT systems, a number of recent studies on Multilingual MT have been proposed [11-13]. Wibawa et al. [11] proposed a multilingual MT system which integrates the example-based MT (EBMT) and statistical MT (SMT) approaches. The proposed system helps Javanese<sup>1</sup> youths in translating between one of the four speech levels of Javanese to Indonesian language or vice versa. The experimental evaluation demonstrates that the Javanese-Indonesian translation is more accurate (average accuracy is 0.83% ) than the Indonesian-Javanese translation (average accuracy is 0.68% ). The authors argue that the accuracy of the proposed multilingual MT system can be improved by increasing the size of training data. On the other hand, Meera and Sony [12] presented a solution that relies on rule based direct MT system to translate a source English language into a target Malayalam and Hindi languages using a bilingual dictionary. In their implemented algorithm, they combined the first order predicate logic (FOPL) based semantic checking with a word sense disambiguation. The algorithm starts with defining the rules, then using Weka software to crate training sentences. The authors crated huge amount of training sentences for the purpose of increasing efficiency. The authors used the precision and recall formulas to evaluate the proposed algorithm. The results showed an accuracy of 74%. Authors suggested that the improvement

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Javanese\\_language](https://en.wikipedia.org/wiki/Javanese_language)

of the results can be achieved by more morphological inflections to the system. More recently, Choi et al. [13] developed a multilingual MT system to offer overseas tourists, in the 2018 Winter Olympic Games that will be held in Korea, with a multilingual speech translation service. The authors apply crowdsourcing translation using the existing large Korean-English corpus to create Korean-English-French and Korean-English-Spanish triangle corpora with minimum cost and time. However, authors notice very many translation errors from the triangle corpora. Therefore, a semi-automatic filtering of translation errors in large triangle corpora is proposed to resolve the translation loss caused by crowdsourcing translation.

### B. Multi-Agent Systems

A Multi-agent system is a software system that consists of a number of interacting intelligent agents and their environment. Coordination, and communication between agents create quality solutions that cannot be produced by a single agent in its *own capacity*. Multi-agent systems use search algorithms, procedural and functional approaches to solve problems that are very difficult for ordinary systems to solve. Knowledge and messages sharing, parse the messages, and understand messages are the features of multi-agents that enable them to socially communicate and behave in an intelligent way. Multi-agent interaction and knowledge sharing constitute a dynamic topology that changes each time to interact with the new environment to solve the challenging problem [14].

In terms of the development of Multi-agent System, there are currently a number of available general purpose frameworks and toolkits that provide standardization of the development of multi-agent system and save time and effort of developers. Examples of the standard Multi-agent System development frameworks are Jason<sup>2</sup>, JADE<sup>3</sup>, SeSAM<sup>4</sup>, and AgentBuilder<sup>5</sup>. Most of these frameworks are particularly developed to create simulations of complex systems and general purpose applications. However, none of the mentioned frameworks has been dedicated for the natural language processing research area, in particular, MT [1]. Accordingly, much more research is needed to further develop and refine the area of using multi-agent technology for MT.

The development of natural language processing applications, specifically MT systems, using the Multi-agent technology has gained little attention among the research community in which a limited number of studies have been proposed[1, 4, 15]. Aref [4] proposed multi-agent system techniques for natural language understanding. The purpose of using multi agent system (MAS) is to exploit the resources and capabilities of the interconnected agents. The author introduced decentralized MAS system that overcomes the single point failure essential to centralized systems and other resources and performance limitations. The system combined lexical and cognitive structural approach to design an understating multi-agent system. The system consists of two types of agents: the first is a lexical agent that contains the English vocabulary with all their linguistic information. The other is a cognitive structural agent that consists of six modules: Speech-to-text, Text-to-speech, Morphological, Semantic, Discourse, and Query Analyzers. These agents communicate with user input to simulate the answer for user questions. The authors implemented and tested the modules separately, and waived the integration between modules and user interface for future work. Minakow et al. [15] exploit the Multi-agent technology to develop a text understanding

<sup>2</sup> The Jason framework: <http://jason.sourceforge.net/wp/>

<sup>3</sup> The JADE framework: <http://jade.tilab.com/>

<sup>4</sup> The SeSAM framework: <http://www.simsesam.de/>

system for car insurance. The proposed Multi-agent system includes four steps: 1) morphological analysis, 2) syntax analysis, 3) semantic analysis and 4) pragmatics. The process starts by dividing the whole text into sentences, and then the first three steps are applied to each sentence for meaning extraction. Finally, the parsed text enters the final step – pragmatics. Hettige et al. [1] design and implement a Java based Multi-agent System for Machine Translation, known as MaSMT, which supports English to Sinhala languages machine translations. MaSMT includes two types of agents, specifically ordinary agents and manager agents. That control ordinary agents. A number of ordinary agents in the swarm are assigned to a specific manager agent to be within its control. Manager agents can directly communicate within each other and within its controlled ordinary agents. Ordinary agents can only directly communicate within each other in its own swarm and with its manager agent. The system framework implements object-object communication, MySQL database connectivity and XML-based data passing for message passing. Agent communication in the system framework conforms with the specification of the Agent Communication Language (ACL)<sup>6</sup> defined by FIPA (Foundation for Intelligent Physical Agents). Experimental results demonstrate that the MaSMT can be employed to develop successful natural language processing applications.

### C. Semantic Web and Ontologies

The Semantic Web is a framework that uses standards developed by the WWW Consortium to support common exchange protocols and data formats on the Web. Semantic web standardizing allows data to be reused and shared across application, enterprise, and community boundaries. The semantic web core is ontology, which is supported by languages such as RDF and OWL. In computer science, ontology define the specification of a conceptualization by facilitating explicitly the knowledge reuse [16]. There important aspects are needed to explicitly realize the ontology languages [16]:

- 1) Conceptualization: through selecting a suitable reference model such as entity-relationship model, it provides ontology construct to define the entities and relations and the relation between entities.
- 2) Vocabulary: the language should also include the syntax and grammars.
- 3) Axiomatization: rules, constraints and factual knowledge are required to capture the semantics for inference.

The Semantic Web architecture is composed of several layers. One of the most important is the ontology layer. Ontology is knowledge representations, containing words and statements that specify the semantics of a given knowledge domain in a given operational environment. The ontology takes the structure of a directed graph (nodes and arcs). Nodes represent concepts and edges represent the semantic relationships between those concepts. To exploit the ontology by a machine, certain rules must be respected, as the definition of a formal syntax and unambiguous semantics, and deduction of new knowledge presented implicitly in the ontology [17, 18]. In order to facilitate sharing activities and collaboration between agents, web based knowledge agrees on common and explicit ontologies [16]. Explicit ontologies facilitate the following features [16]:

- 1) Interoperability as specified by W3C.
- 2) Extensibility. This concept relies on developing the ontologies incrementally using reusability concept.

---

<sup>5</sup> The AgentBuilder framework: <http://www.agentbuilder.com/Documentation/Lite/>

<sup>6</sup> FIPA-ACL Specifications, Available at: <http://www.fipa.org/repository/aclspecs.html>

- 3) Visibility. To be able to understand the knowledge written in an unfamiliar language, common ontological on syntax and semantics is required.
- 4) Inferenceability. The ontology enables logical inference on facts through axiomatization.

The knowledge resulted in the ontology are conveyed using a number of components or brick base, which are mainly: 1) Concepts, 2) Relationship, 3) Functions 4) Axioms, and 5) Instances [19, 20].

The Resource Description Framework (RDF) and the Web Ontology Language (OWL) are two main standard languages. These languages facilitate the process of data and knowledge sharing and integration for easing the web-accessible information and services. The RDF is a language that facilitates information integration by allowing the use of Internationalized Resource Identifiers (IRIs) - a generalization of Uniform Resource Locators (URLs) - to refer to resources. Because of the RDF constraints (i.e., RDF doesn't have the ability to describe cardinality constraints), the need for more expressive ontology language is recognized. The OWL extends the RDF standard by processing the information content instead of just presenting information to humans. The OWL has more facilities in expressing meaning and representing machine interpretable content on the Web compared with other standard languages [16, 21]. To create, update, delete, maintain, communicate with ontology we can use the protégé tool, it is a free, open-source ontology editor and framework for building intelligent systems. The protégé supports OWL and RDF ontology languages [22].

There are a number of studies, that integrate the use of semantic web technologies with multi-agent systems, have been proposed [23-27]. In the paper by Williams [23], the DOGGIE agent is proposed to address the ontology problem in a multi-agent system made up of agents with diverse ontologies. Using a machine-learning algorithm, the author describes how the agents learn representations of their own ontologies and how the agents teach each other what their concepts mean. The author introduces a methodology based on agents on the semantic web that deals with the knowledge sharing issues resulting from diverse ontologies. More complex ontologies are planned to be carried out. Aref and Zhou [24] used the newly OWL standardized technique to construct the Multi-Agent communication and work. The authors concentrated on implementing a MultiAgent system that enables them to capture the meaning of the text using natural language understanding. The authors used four criteria to evaluate the developed MultiAgent system: representational adequacy, inferential adequacy, inferential efficiency and acquisitional efficiency. The results show that the developed system is not able to optimize all criteria. In spite of this, using OWL as a standard has a potential future in the field of language translation. Another research paper proposed by García-Sánchez et al. [25] discusses the main purpose of using semantic web technologies with multi-agent systems, in the biomedical research area, which is the increasing demand of data integration. The authors presented SEMMAS system, an ontology-based and domain independent framework that works on integrating intelligent agent and semantic web. The authors pointed out that improved data integration methods are required, and control mechanisms for data redundancy and inconsistency are needed. Lee & Wang [26] proposed an ontology-based computational intelligent multi-agent system for Capability Maturity Model Integration (CMMI) assessment. A CMMI ontology that represents the CMMI domain knowledge is predefined by domain experts to be utilized by the computational intelligent multi-agent system. The proposed system includes a natural language processing agent, an ontological reasoning agent and a summary agent. These agents cooperate with one another to accomplish the goal of effectively

summarizing the evaluation reports for the CMMI assessment. Hakansson et al. [27] presented a multi-agent system that uses ontology as an interface to search for a user request. Once the ontology found, the agent bring back its location and the systems complete the missing information. The multi-agents are coordinated by meta-agents. Meta-agents perform the mapping between ontologies and language translation and reason for the validity of the returned contents. The authors applied the presented MAS system in e-tourism domain. The testing results showed that the system is still at a primitive level, therefore the translation and domain definition must be handled by a reliable tool instead of using Google translate and general domain.

To summarize, there are very relatively little research effort in the area of using MAS and ontologies in MT. For that reason, this research is important to advance the current development of this area of research.

### III. THE PROPOSED SEMANTIC MULTI-AGENT ARCHITECTURE FOR MULTILINGUAL MT

The proposed architecture of Multilingual MT system, as shown by Fig. 1, consists of two major components: User Interface and Translation. Each component represents some of precise tasks realized by a number of autonomous agents. The process of the proposed system architecture starts by receiving a text (written in the source language) by the interface agent, thereafter, the text is subjected to different phases of analysis and communication with ontologies to generate a translation of the text in a target language.

#### A. *The user interface component:*

It contains two agents, as follows:

##### 1) **Interface Agent:** this agent has specific tasks, which are:

- Receives the user request which includes: the text to translate, the source language, and the target language (Step 1 in Fig. 1). Then, it transfers these information to the lexical analyzer agent (Step 2 in Fig. 1).
- Displays to the user the resulting translated text in a target language ( Step 7 in Fig. 1).

##### 2) **Language Recognition Agent**, this agent has specific tasks, which are:

- Recognizes the source language of a text, in case it isn't indicated in the user request, by matching the terms in the user request with the existing terms in a domain corpus or thesaurus of the source language (Step 1.1 in Fig. 1).
- Prompts the user for the target language in case it is not indicated in the user request.

#### B. *The translation component:*

It is the main part of the proposed architecture. It receives the user request and uses multi agents and source and target language ontologies to analysis the user text and translates it into the required target language.

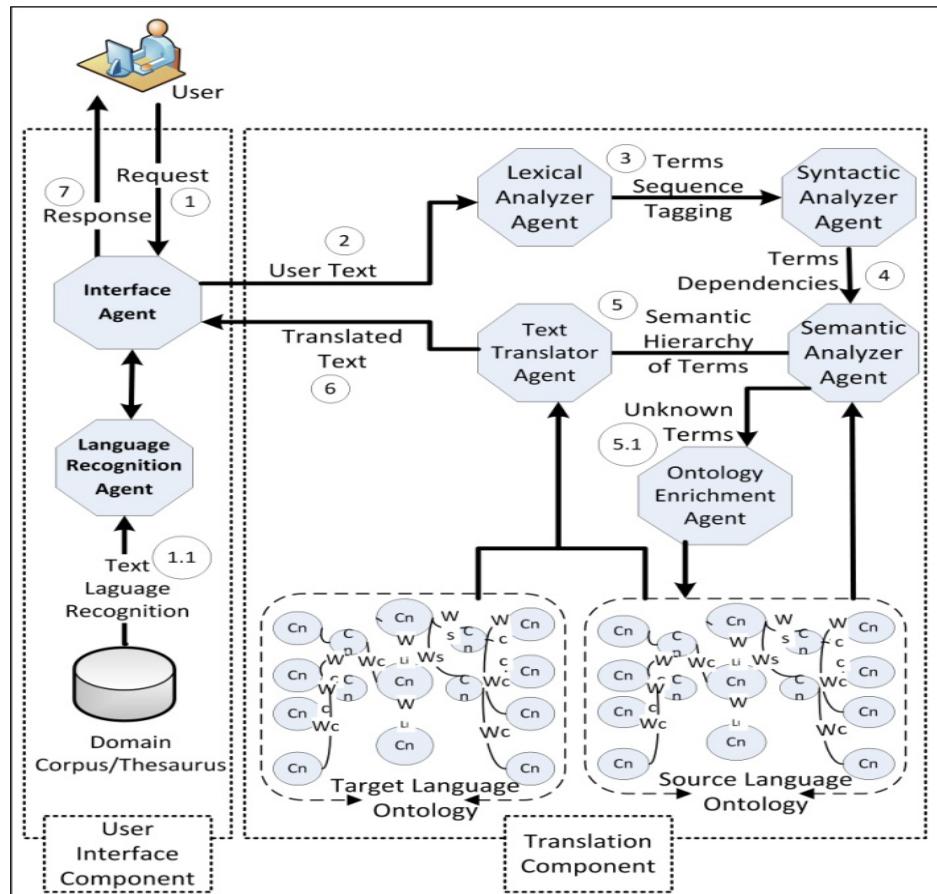


Fig 1. The Semantic Multi-gent Architecture for Multilingual MT

### An Illustrative Example:

We present a helpful example to better understand the purpose and the task realized by each agent in the proposed architecture. Suppose a user asks the system to translate a sentence:

***I am from Amman***

Written by a user in English (Source language is indicated by the user), and we need to translate it into French language (Target language is indicated by the user).

The translation component contains five agents, as follows:

**1) Lexical Analyzer Agent**, the main role of this agent is the extraction of a term sequence from the user text (on the source language). During the operation of the lexical analysis, the lexical analyzer agent eliminates all useless information: blanks, tabs, end of lines and line break...etc. Then it applies a segmentation process to separate words in order to extract the basic entities: name, verb, and adjective...etc. The TreeTagger software [28] can be used by this agent to accomplish such tasks. The main function of the TreeTagger is the lemmatization and automatic syntactical tag of the text. The output of the TreeTagger is a tagged terms sequence by basics entities: verbs, names..etc. (Step 3 in Fig. 1) (see example in Table1).

TABLE1: THE TAGGING PROCESS OF THE GIVEN EXAMPLE USING THE TREETAGGER TOOL. PP: PERSONAL PRONOUN, VRB: VERB, DET:  
DESCRIPTIVE / QUALIFICATIVE ADJECTIVE. NOM: COMMON AND PROPER NOUNS.

<b>Term</b>	I	am	from	Amman
<b>Tag</b>	PP	VRB:Present	DET	NOM

2) **Syntactic Analyzer Agent**, the main role of this agent is the verification of the conformity of the terms sequences of the basics entities, generated by the Lexical Analyzer agent, with the grammar rules that describe the source language. This agent generates as output a terms dependencies tree of the user text (Step 2 in Fig. 1) (see example in Fig. 2). To identify the syntactical relationships between terms, the syntactic analysis tool uses morpho-syntactic rules that describing the constraints in which a text must follow to be identified.

When a morpho-syntactic rule is identified in the text between two terms, it points out the presence of a relationship between these two terms.

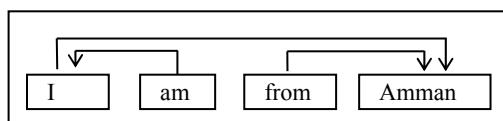


Fig. 2: Syntactic terms dependencies tree of the given example.

Syntactic Analyzer agent can use an automatic syntactic analysis tool, for example, the Stanford parser [29] that is a dedicated software to provide a dependency hierarchy in English language. It allows the automatic creation of syntactic terms dependencies tree based on a set of grammatical and syntactic analysis rules (see Fig. 2). The Stanford parser returns a complete syntactic terms dependencies tree in which the terms are connected by syntactic relations.

3) **Semantic Analyzer Agent**, the main role of this agent is to understand the sequence of terms based on the syntactic terms dependencies tree provided by the Syntactic Analyzer agent in source language (Step 5 in Fig. 1). This agent identifies the semantic concept, in the source language ontology, of each term in the dependencies tree. To better detect the semantic of a term and its corresponding semantic concept, the term context must be investigated. In general, ontologies are used to assure this type of process. A semantic analyzer tool [30, 31] can be used by the Semantic Analyzer agent. It receives the syntactic terms dependencies tree and determines the possible contexts of each term using the source language ontology

The semantic concept hierarchy is constructed via the mapping between the terms and their corresponding concepts, and the syntactic relationships of terms and their corresponding relationships, in the source language ontology. In case there are terms that don't correspond to any concept in the source language ontology, these terms are considered as unknown terms and they will be sent to the Ontology Enrichment agent (Step 5.1 in Fig. 1). As output, the Semantic Analyzer agent returns a concept semantic hierarchy of the inputted terms (see Fig. 3).

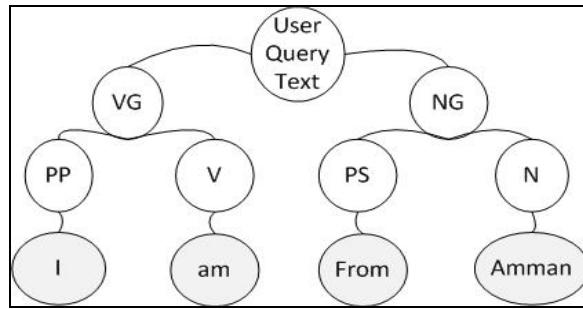


Fig. 3: Semantic concept hierarchy of the given example. VG: Verbal Group, PP: Personal Pronoun, NG: Nominal Group, PS: Preposition Spatial, V: Verb, N: Common and proper nouns.

**4) Text Translator Agent**, the main role of this agent is to receive a semantic concept hierarchy expressed on the source language, and produces as output a translated text in target language (Step 6 in Fig. 1). The agent can use a multilingual ontology such as the WordNet ontology. The WordNet is based on the theories of knowledge representation: memorizing terms and concepts in a hierarchical manner, using the inclusion relationships. It divides the data into four databases organized differently from one another, associated with the names of categories, verbs, adjectives and adverbs. The names and verbs are organized into hierarchies. Relationships such as hyponymy ("is a") and hyponymy link names and verbs with their "specializations" [32-35].

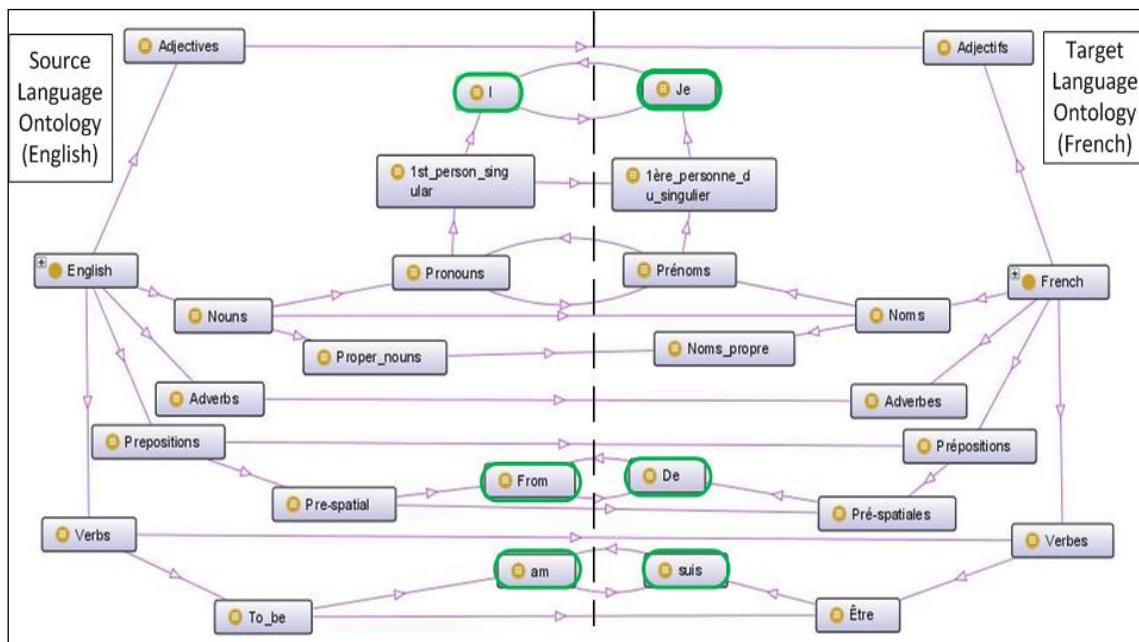


Fig. 4: The translation process of each concept, of the illustrative example, from the source to the target language using a multilingual ontology.

The main advantage of using the WordNet ontology is the fact that it is a multilingual representation of knowledge. This advantage allows the translation of the concepts and relationships from the source language to the target language. To translate the user text, the text translator agent must:

1. Translates each concept of the semantic concept hierarchy by its corresponding concept on the target language using a multilingual ontology such as the WordNet (see example in Fig. 4).
2. Transform the concept semantic hierarchy to a new translated text expressed using the target language. For example, the new translated text of the illustrative example in the French language will be:

**“Je suis d’Amman”**

3. Verify the lexical and syntactic syntax of the new translated text in accordance with the grammar rules of the target language.
4. Send the new translated text to the Interface agent.

5) **Ontology Enrichment Agent**, the main role of this agent is the enrichment of the source language ontology with the unknown terms extracted by the semantic analyzer agent from the text to translate (Step 5.1 in Fig. 1). There are several methods for automatic ontology enrichment in the literature [36, 37]. Such methods are based on a minimum generic base of rules between terms. Initially, non-redundant terms are extracted from a source domain (like thesaurus, corpus). Then, the approximation of terms is performed through the mapping between concepts of the ontology and the new terms. For more details on this, refer to [36, 37].

#### IV. CONCLUSION

This paper has reported the design of the Semantic Multi-Agent Architecture for a Multilingual Machine Translation system. The proposed architecture integrates the Multi-agent technology and ontologies to produce collaborative working environment in order to improve the quality of the Multilingual MT. It consists of two major components: the User Interface and the Translation. Each component represents some of precise tasks realized by a number of autonomous agents. The User Interface component contains two agents: Interface agent, and Language Recognition agent. The Translation component consists of five agents: Lexical Analyzer agent, Syntactic Analyzer agent, Semantic Analyzer agent, Text Translator agent, and Ontology Enrichment agent. These agents are communicating with each other and with source and target language ontologies to capture the syntactic, morphology and semantic features of the source and target languages. The process of the proposed system architecture starts by receiving a text written in the source language by the interface agent as an input, and completes when the text translator agent successfully produces a translated text in target language as an output. Future study will focus on the implementation of a semantic multi-agent system for multilingual machine translation, and on the evaluation the proposed architecture.

#### REFERENCES

- [1] B. Hettige, A. Karunananda, and G. Rzevski, “MaSMT: A Multi-agent System Development Framework for English-Sinhala Machine Translation,” International Journal of Computational Linguistics and Natural Language Processing (IJCLNLP), vol. 2, no. 7, pp. 411-416, 2013.
- [2] S. Alansary, M. Nagi, and N. Adly, “A Semantic Based Approach for Multilingual Translation of Massive Documents.” pp. 317-323.
- [3] U. Pakdeetrakulwong, and P. Wongthongtham, “State of the art of a multi-agent based recommender system for active software engineering ontology,” International Journal of Digital Information and Wireless Communications (IJDWC), vol. 3, no. 4, pp. 29-42, 2013.
- [4] M. M. Aref, “A multi-agent system for natural language understanding.” pp. 36-40.
- [5] T. Berners-Lee, J. Hendler, and O. Lassila, “The semantic web,” Scientific american, vol. 284, no. 5, pp. 28-37, 2001.
- [6] N. D. Rodríguez, M. P. Cuéllar, J. Lilius et al., “A survey on ontologies for human behavior recognition,” ACM Computing Surveys (CSUR), vol. 46, no. 4, pp. 43, 2014.
- [7] S. Fraihat, and Q. Shambour, “A Framework of Semantic Recommender System for e-Learning,” Journal of Software, vol. 10, no. 3, pp. 317-330, 2015.
- [8] P. Antony, “Machine translation approaches and survey for indian languages,” International journal of Computational Linguistics and Chinese Language Processing, vol. 18, no. 1, pp. 47-78, 2013.

- [9] S. Saini, and V. Sahula, "A Survey of Machine Translation Techniques and Systems for Indian Languages." pp. 676-681.
- [10] Sindhu D.V., Sagar B.M., and Rajashekhar M.S., "Survey on Machine Translation and its Approaches," International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 6, pp. 7317-7320, 2014.
- [11] A. Wibawa, A. Nafalski, A. E. Kadarisman et al., "Indonesian-to-Javanese Machine Translation," International Journal of Innovation, Management and Technology, vol. 4, no. 4, pp. 451-454, 2013.
- [12] M. Meera, and P. Sony, "Multilingual Machine Translation with Semantic and Disambiguation." pp. 223-226.
- [13] S.-K. Choi, J.-H. Shin, and Y.-G. Kim, "Semi-automatic Filtering of Translation Errors in Triangle Corpus." pp. 72-79.
- [14] L. Panait, and S. Luke, "Cooperative multi-agent learning: The state of the art," Autonomous Agents and Multi-Agent Systems, vol. 11, no. 3, pp. 387-434, 2005.
- [15] I. Minakov, G. Rzevski, P. Skobelev et al., "Creating contract templates for car insurance using multi-agent based text understanding and clustering," Holonic and Multi-Agent Systems for Manufacturing, V. Mařík, V. Vyatkin and A. W. Colombo, eds., pp. 361-370: Springer, 2007.
- [16] L. Ding, P. Kolari, Z. Ding et al., "Using ontologies in the semantic web: A survey," Ontologies, R. Sharman, R. Kishore and R. Ramesh, eds., pp. 79-113: Springer, 2007.
- [17] A. Harth, M. Janik, and S. Staab, "Semantic web architecture," Handbook of Semantic Web Technologies, J. Domingue, D. Fensel and J. A. Hendler, eds., pp. 43-75: Springer, 2011.
- [18] D. Fensel, Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce: Springer-Verlag New York, Inc., 2003.
- [19] T. R. Gruber, "A translation approach to portable ontology specifications," Knowledge acquisition, vol. 5, no. 2, pp. 199-220, 1993.
- [20] A. Gomez-Perez, M. Fernández-López, and O. Corcho-Garcia, "Ontological engineering," Computing Reviews, vol. 45, no. 8, pp. 478-479, 2004.
- [21] I. Horrocks, "Ontologies and the semantic web," Communications of the ACM, vol. 51, no. 12, pp. 58-67, 2008.
- [22] V. Jain, and M. Singh, "Ontology development and query retrieval using protégé tool," International Journal of Intelligent Systems and Applications (IJISA), vol. 5, no. 9, pp. 67, 2013.
- [23] A. B. Williams, "Learning to share meaning in a multi-agent system," Autonomous Agents and Multi-Agent Systems, vol. 8, no. 2, pp. 165-193, 2004.
- [24] M. M. Aref, and Z. Zhou, "The Ontology Web Language (OWL) for a multi-agent understanding system." pp. 586-590.
- [25] F. García-Sánchez, J. T. Fernández-Breis, R. Valencia-García et al., "Combining Semantic Web technologies with Multi-Agent Systems for integrated access to biological resources," Journal of biomedical informatics, vol. 41, no. 5, pp. 848-859, 2008.
- [26] C.-S. Lee, and M.-H. Wang, "Ontology-based computational intelligent multi-agent and its application to CMMI assessment," Applied Intelligence, vol. 30, no. 3, pp. 203-219, 2009.
- [27] A. Hakansson, R. Hartung, and J. J. Jung, "Using Multi-Agent System for Business Applications in Multilingual Ontologies," Innovation through Knowledge Transfer, R. J. Howlett, ed., pp. 157-166, Berlin, Heidelberg: Springer, 2010.
- [28] I. Korkontzelos, T. Zesch, F. M. Zanzotto et al., "SemEval-2013 Task 5: Evaluating Phrasal Semantics." pp. 39-47.
- [29] M. De Marneffe, and C. Manning, "Stanford Typed Dependency Manual revised for Stanford Parser v. 2.0. 4," Retrieved November, vol. 13, pp. 2012, 2012.
- [30] Y. Drissi, T. W. Grandison, C. G. Harrison et al., "Unified numerical and semantic analytics system for decision support," Google Patents, 2013.
- [31] A. J. Agrawal, and O. Kakde, "Semantic Analysis of Natural Language Queries Using Domain Ontology for Information Access from Database," International Journal of Intelligent Systems and Applications (IJISA), vol. 5, no. 12, pp. 81, 2013.
- [32] F. Bond, and R. Foster, "Linking and Extending an Open Multilingual Wordnet." pp. 1352-1362.
- [33] M. M. Boudabous, N. C. Kammoun, N. Khedher et al., "Arabic WordNet semantic relations enrichment through morpho-lexical patterns." pp. 1-6.
- [34] V. Hanoka, and B. Sagot, "Wordnet creation and extension made simple: A multilingual lexicon-based approach using wiki resources." pp. 1-6.
- [35] B. Sagot, and D. Fišer, "Automatic extension of WOLF." pp. 1-7.
- [36] N. Astrakhantsev, and D. Y. Turdakov, "Automatic construction and enrichment of informal ontologies: A survey," Programming and Computer Software, vol. 39, no. 1, pp. 34-42, 2013.
- [37] E. Blomqvist, Ontocase-automatic ontology enrichment based on ontology design patterns: Springer, 2009.

# Language and Security for None English Speakers

**Ibrahim Mohammed Alseadoon**

**Rabie A. Ramadan**

**Ahmed Y. Khedr**

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

## *Abstract*

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

## **1. Introduction**

People are very curious when dealing with money. They need to understand every word written or said to them. However, due to the lack of understanding the language, they may be victim to forgery. In fact, nowadays, businesses and money tracking as well as banking systems became easier in terms of accessing and more dangerous since everything can be done through online systems or SMS messages. Therefore, the security becomes one of the major issues for companies and banks. At the same time, we believe dealing with online systems in accessing educational systems might need to be secure as well. Career development and online courses also suffer from the security issues.

One of the recently major threats around the world is the phishing. Around 2 million customers of a major store in United States has result in exposing their confidential information (Dave, 2013). In fact, the number of phishing E-mails in 2015 reaches around 156 million E-mails per day (Safe, 2015). Although many software are produced against phishing, there are a large number of phishing E-mails go through the systems (Dhamija et al., 2006) (Sheng et al., 2010), (Alseadoon et al., 2015). Ashwini et al. (Ashwini, 2013) studied even the effect of grammar on the password that a user uses. The authors show that awareness of English grammar is might diminish the security of a password. This study is related to our study in terms of most of our people tend to learn the grammar even before they become excellent in English language.

In none English speakers' countries, the phishing problem is much deeper due to the lack of understanding the language. Most banking systems, educational systems, and merchandise systems use the English language in their interface and emails and SMS. Even, the websites URLs are written in English. None English speakers might not be able to distinguish between trusted URLs and none trusted ones. Therefore, it is easy on the attacker to set up traps for such users.

This paper is an effort towards examining the real impact of English language on the none-English speakers through a set of scientific experiments. The experiments involves people who deals frequently with online systems and their mother tongue is not English. The road map of this paper is as follows: in section 2, the used model is elaborated, section 3 shows the results and discussion, and the paper is concluded in section 4.

## 2. Problem Modeling

In order to come to a conclusion about the language on the security, a survey has been designed following using 5 points Likert scale Likert(1932). The survey is designed to be conducted by University students and employments in which they already familiar with their University accounts as well as Bank accounts. In addition, students/employments are familiar with websites URLs. Moreover, the survey is conducted anonymously in which a student/employments does not have to provide his/her name; so, he/she will be comfortable answering the survey questions. In addition, students/employments are asked to provide their level of English in order to be fair in judging their answer. The survey is divided into nine sections; each section tries to judge specific measuring criteria. The first section gathers information about the personality of the people conducting the survey. The second section is two questions identifying the level of English per person. The third section is four questions to identify the person's security awareness. The possibility of a person to be a victim is gained through three questions in section 4. Section 5 studies the awareness of the information and computer protection for the participants conducting the survey through answering 6 questions. Section 7 is about the practical experience of the participants in security by answering three questions. The eighth section is all about the trust which are three questions. The last section is 21 questions designed to measure the level of susceptibility of the person conducted the survey. The survey is written in Arabic language and was displayed at surveymonkey.com. Sample questions of the survey are shown in Figures 1 and 2 below.

تأثير اللغة الإنجليزية على أمن الحاسوب والمعلومات

الجنس

ذكر

إناث

الوظيفة التي تنتسب إليها هي

طالب

موظف

Figure 1: section 1 of the survey

The analysis of the survey was conducted using SPSS software version 23. The reliability factors were measured by gaining 0.7 to match the cut-off Hair et al.(2009). Those factors that did not match the cut-off are eliminated. Then, factor analysis was conducted to measure factors validity.

9. ما مدى موثوقية الصفحات التالية :

موثوق جدا	5	4	3	2	1	غير موثوق
-----------	---	---	---	---	---	-----------

Figure 2: Part of section 4 of the survey

### 3. Results and Discussion

Out of 250 persons were invited, 134 persons male and female are conducted the survey after removing the illegible ones due to incomplete survey. The results are given below.

For question 1 where the gender is expressed, as given in Figure 3, 36% of the participants were female and the rest were males. This is a reasonable percentage where males, in Arab countries, are more interactive in dealing with credit cards, bank accounts, and secure websites.

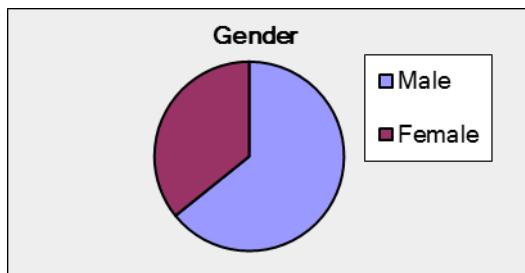


Figure 3: Gender percentage

Figure 4 is another statistics for question 2 in which the job is expressed in the survey. 76 % of the participants are working and the others are students. Again, it is an expressive percentage where employees are more into the critical business and they could be more careful when dealing with Internet, especially in handling things that cost money.

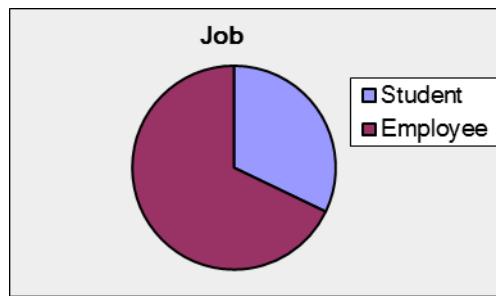


Figure 4: Job percentage

The number of years a participants studied English is presented in Figure 5. The chart shows the levels of English and its percentage. A diversity of levels is shown in the Figure 5 where 30% of the participants were studied English for more than 12 years. At the same time, 14% of the participants were studied English for less than 3 years. The beauty of this chart that almost equal percentage of English levels are represented in the survey. Therefore, the results will be more trusted and not biased. However, although a person could be studied English for long time but it might not be professional. A youth person might study English for may be two years and becomes more professional than an elder person who studied English for even more than 12 years. This fact is considered in the survey but adding another question for each participants to freely evaluate his excellence in the English language. Figure 6 is the results of the answers to the English Language fluency. As can be seen the distribution seems perfect for our survey where only 4% of the participants are replied as they are in the basic level of English. So, this low percentage will not affect our trust and confidence in the final results of the survey. However, only 20% of the participants who think they are excellent in English language.

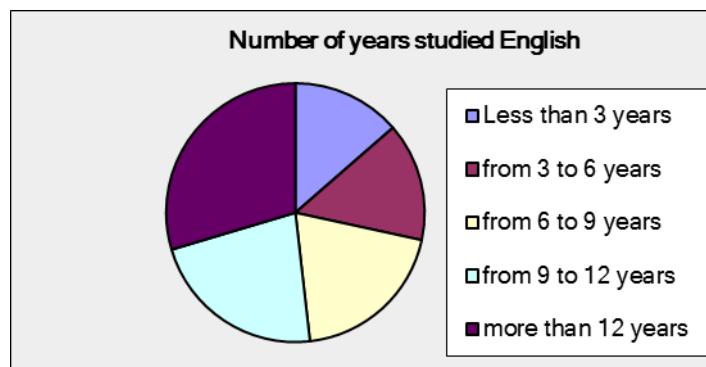


Figure 5: English levels

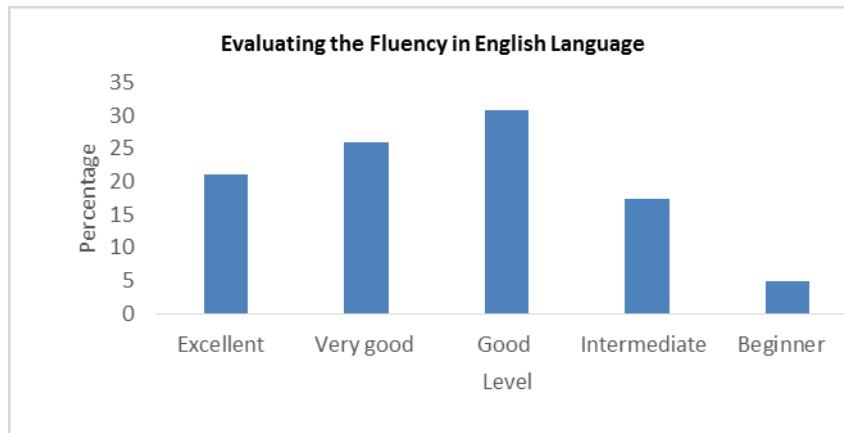


Figure 6: Evaluating the Fluency in English Language

The next questions that we would like to show is the ones related to security awareness. These questions examine the awareness of the participants to the important of securing their computer and information. In addition, it measures the security awareness of his/her responsibility to protect his/her own computer. Figure 7 depicts the answers to these questions. As can be seen, the security awareness among the participants is varied. However, few of the participants that have a doubt about the security importance and their responsibility towards securing their own computers and information. So, we are dealing on average with participants that aware that, at least, security is important thing to consider.

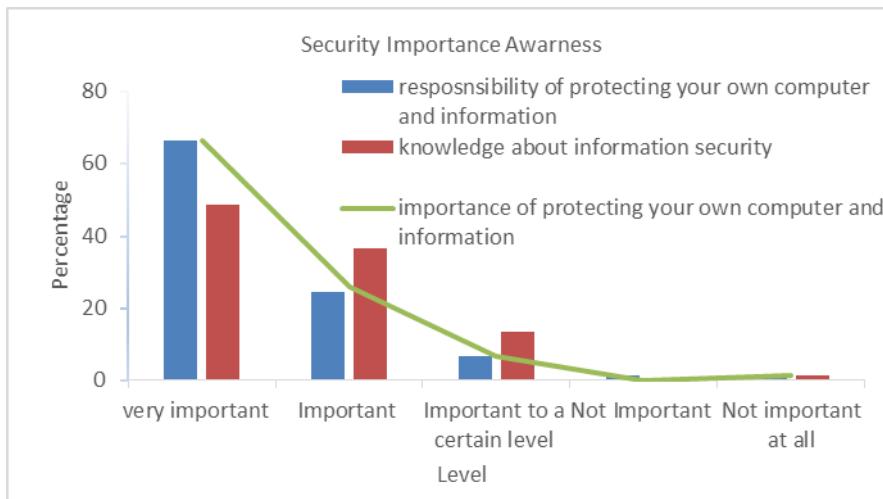


Figure 7: Security Importance Awareness

Another measure to the effect of the Language on the security is the usage of the Internet and online Banking System. Figure 8 shows the participants amount of time they spend on the Internet and the number of times they use the Banking System. Out of the Figure 8, we can say that at least 50% of the participants spend from two to four hours exploring the Internet either for work or fun. At the same time, almost 80% of the participants use the Internet banking from one to two times a month. The rest are accessing the online banking system more than four times a

month. However, a large percentage is using the Internet more than eight hours a day. This finding indicates that our participants are involved in activities related to online systems.

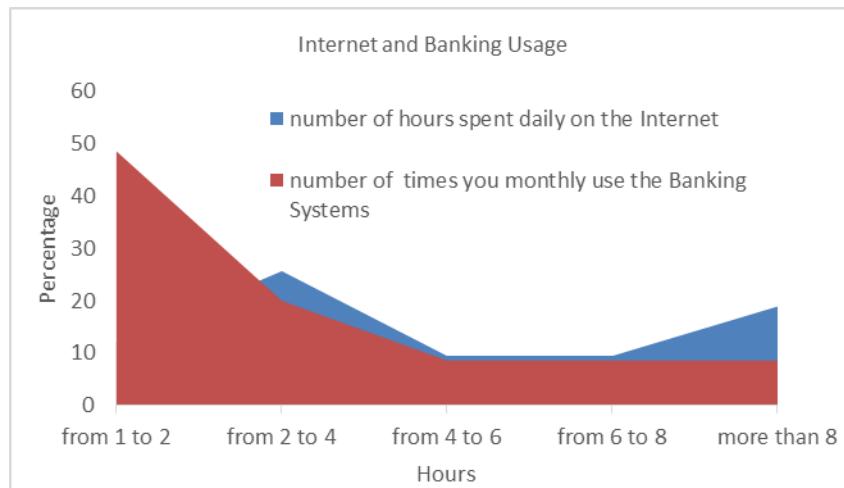


Figure 8: Internet and Banking usage

For E-mail richness, Figure 9 shows the results of the survey in this regard. A set of questions were designed to capture the richness of our participants in using the E-mail. The questions are listed as follows:

1. The E-mail allows me to communicate with other people and reply promptly
2. The E-mail allows me to communicate with other people in a way satisfying our personalities
3. The E-mail allows me and my partner to express our moods and feelings
5. The E-mail allows me and my partner to use rich language and variety of expressions

As can be seen in Figure 9, the average answers to these questions are 3.x. This is an indicator that our participants have E-mail awareness above the average and by looking at Table 1, the participants answer are leaning towards “Agree”.

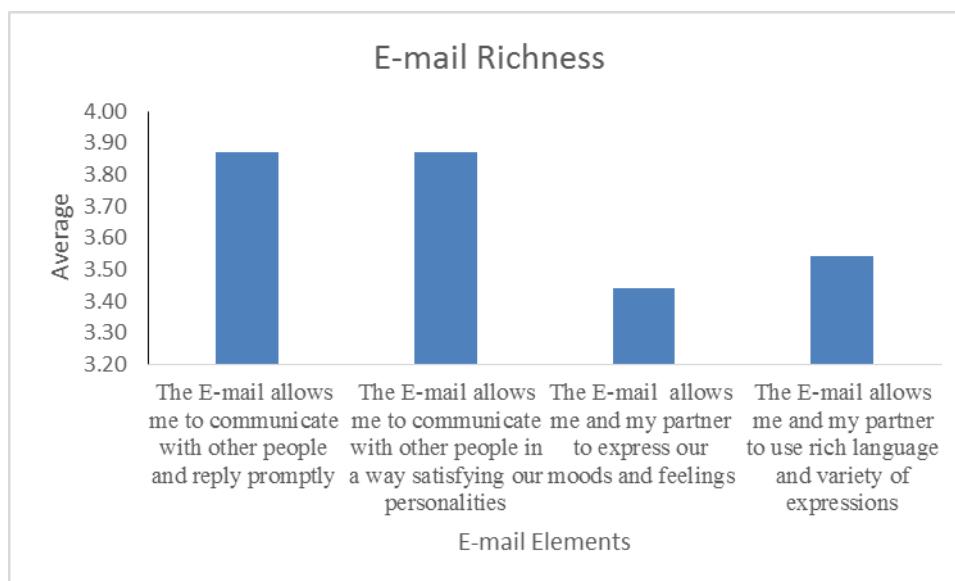


Figure 9: E-mail Richness

Table 1: E-mail Richness

Item	Average	Completely Agree	Agree	neutral	Do not agree	Completely do not agree
The E-mail allows me to communicate with other people and reply promptly	3.87	14	41	10	5	1
The E-mail allows me to communicate with other people in a way satisfying our personalities	3.87	17	35	13	5	1
The E-mail allows me and my partner to express our moods and feelings	3.44	9	30	18	11	3
The E-mail allows me and my partner to use rich language and variety of expressions	3.54	10	34	16	6	5

Another set of questions were design to examine the participants basic information about security such as:

- 1) Their password complexity and if they use letters and special characters
- 2) If they change the password frequently or not
- 3) If they make a backup from their information
- 4) If they have security programs on their computers or not

The results of this set of questions were orientated towards completely agree with average shown in Table 2. Therefore, most of our participants are aware of the basic security information which is good to our survey. However, there are some other questions included in the survey about the trust of others on the Internet. The average results were very high. Usually people who have high trust fall in phishing and stealing their information is easier than others who are very careful dealing with people on the Internet.

Table 2: Basic security information results

Item	Average /5
Their password complexity and if they use letters and special characters	4.58
If they change the password frequently or not	3.80
If they make a backup from their information	4.34
If they have security programs on their computers or not	4.54

There are some other questions about the participants interest in:

- Looking at the URL of the website when it require to enter information.
- Looking at the security certificates of the websites.
- Looking at the security lock appears before the URL of a website.

As can be seen in Figure 10, the average results are around 3.x out of 5; however, most of the answers fall under “mostly I do” column. This indicates that we have a good class of people who mostly carful about the security indictors.

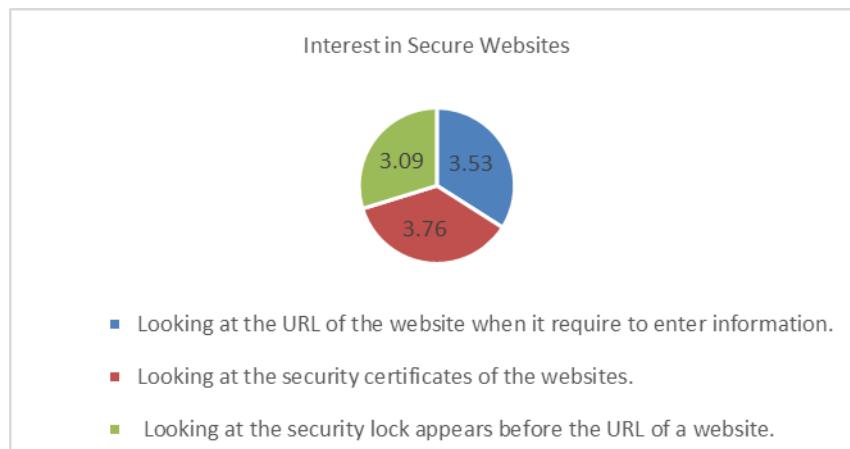


Figure 10: Interest in Secure Websites

Now, some questions were asked to participants regarding:

- 1) Fake online banking system
- 2) Real online banking system
- 3) Fake University website
- 4) Real University website
- 5) Fake link to a general website
- 6) Real link to a general website

The results of this section was astonishing to us in which almost 10% of the participants got it wrong. Therefore, our simple conclusion is that although people might be aware of basic security elements, they might fall in deceiving due to the language problems. Even the 20% of participants who said that they are excellent in English language, at least 10% of them were deceived. Others who think that they have moderate knowledge of English were also deceived. One more thing that we noticed is most of our participants are either student in a University level or an employee and also a University employee. This is a dangerous indictor in terms of they are considered highly ranked people in education and living. They are the ones who deal with the Internet in buying and selling. Therefore, our recommendations are as follows:

- The department of Computer Science should offer a program about security
- Banking system should change their websites languages to the country language
- The government and private sectors should offer awareness courses about security terms and risks
- The government should send frequently security awareness messages to people
- Youth should be encouraged to lean about the security risks

#### 4. Conclusion

This paper introduced an experiment testing the effect of English on the security for non-English speakers. The survey is scientifically designed to capture the behavior and the awareness of the participants towards security. The study concluded that non-English speakers can be easily deceived because of their illiteracy in English. Only 10% of the participants were passed the study although they stated that they are having a good English level. Even though, majority of the participants have indicated that they hold above average knowledge about security and behave securely. Therefore, we believe that the language has a great effect on dealing with security, especially Internet, E-mail, and Banking systems.

#### References

- [1]. Dave, P. (2013). *Email 'phishing' attacks by hackers growing in number, intensity* <<http://articles.latimes.com/2013/jul/25/business/la-fi-phishing-attacks-20130726>>. [cited 2016].
- [2]. Safe, G.C. (2015). *Phishing: How many take the bait?* <[http://www.getcybersafe.gc.ca/cnt/rsrcs/nfgrphcs/nfgrphcs\\_2012-10-11-eng.aspx](http://www.getcybersafe.gc.ca/cnt/rsrcs/nfgrphcs/nfgrphcs_2012-10-11-eng.aspx)>. [cited 2016].
- [3]. Alseadoon, I., Othman, M., & Chan, T. (2015). *What Is the Influence of Users' Characteristics on Their Ability to Detect Phishing Emails?*, in *Advanced Computer and Communication Engineering Technology*. Springer. p. 949-962.
- [4]. Likert, R. (1932). *A technique for the measurement of attitudes*. Archives of psychology.
- [5]. Dhamija, R., Tygar, J.D., & Hearst, M. (2006). *Why phishing works*. in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM.
- [6]. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, & L., Downs, J. (2010). *Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions*. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM.
- [7]. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis*.
- [8]. Ashwini Rao, Birendra Jha, and Gananand Kini. 2013. Effect of grammar on security of long passwords. In *Proceedings of the third ACM conference on Data and application security and privacy (CODASPY '13)*. ACM, New York, NY, USA, 317-324.

# A Proposed Technique for Preventing Criminal Attack

Gamal H. Eladl, PhD  
Mansoura University, Faculty of Computers and Information Sciences,  
Egypt, Mansoura, 35516

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

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

---

## 1. Introduction

Criminal attack means an illegal act for which someone does something activity against the law to set upon in a forceful, violent, hostile, or aggressive way, with or without a weapon [1]. The execution time of any crime is approximately less than one minute and it is a free-time and independent to happen. The malicious attackers are still thinking of new ways for executing their new crimes. On the other hand, researchers' duties should be oriented to serve the society against these crimes happens by proposing and applying new innovations those are based on the new trends of Information and Communication Technologies (ICTs). **Federal Bureau of Investigation** (FBI) includes an important event inside its yearly statistical report for crime Bank which focused on our point. It showed that more than 45 employees died on 2015 because the criminal attack has been done for different banks [2].

The content of this paper is organized as follows: Section 2 review the related work and existing template of protection schema. Section 3 introduces the importance of ANN in biometrics system. Section 4 presents the usage of smart cameras in pattern detection with the three common phases. The proposed technique workflow, block diagram and a software prototype of the proposed technique is depicted and discussed in section 5. A comparative study between existing model and proposed model is presented at section 6. Finally, the conclusions are drawn in section 7.

## 2. Related Work and Existing Template of Protection Schemas

There is no such thing as 100 per cent secure bank – unless it is a closed bank. For this reasons, we must look carefully at existing template controls and traditional ways of physical security characteristics for bank systems including [3]:

- A strong gate or electronic door
- Heavy walls to withstand objects thrown or projected against them
- A surrounding moat , to control access
- A drawbridge, to limit access to authorized people
- Gatekeepers, to verify that only authorized customers could enter the bank
- Security Cameras, to monitor the overall workflow of banks
- Men 's security guard

## 3. Importance of ANN in Biometrics system

As well, a soft-computing technique will be used in the novel technique related to ICT disciplines; especially ANN. ANN is a black box that solves non-linear/linear computational problem. It converts the known inputs into the determined outputs this done by calculating/training the desired weight. There are different types of ANN such as multilayered perceptron (MLP). It is one of the most ANN architectures which used in biometrics system. Its structure consists of hidden layers along with two layers (input, output). The hidden layers might be one or more according to the problem's types. Finding the accurate weights are the core of ANN training; those are represented the measurement of learning rate [4].

Additionally, the proposed technique will base on the biometric data of the bank teller in order to determine a bank teller's characteristic/action/motion/secret key. Moreover, it will be used as an automated method of recognizing an individual (bank teller) that is based on measurable & behavioral characteristics.

The face is one of the most acceptable biometrics because it is one of the most common of identification which humans use in their visual interaction. The bank teller can use his face with easy manner in order to create a special action/ unknown secret key. In fact, there are different types of biometric data such as (face-emotion, nose touch, eye movement, ear touch). Biometrics is typically collected by using a device called a sensor. This sensor called (smart camera) that is used to capture acquire the needed biometric visual interaction data. Smart cameras are not simply cameras for taking photos or recording videos and include different pattern recognition algorithms. Smart Camera can be used in our novel technique as a sensor in order to detect bank teller motion, measure, match his biometrics' object. It used for detecting the known action/secret code recognition and convert it into an SMS alert [5-6]. Figure 1 shows the basic component of any biometric system; it contains four major steps:

- 1) Acquiring biometric data from a person,
- 2) Extracting a feature set from the acquired data, and
- 3) Recording the feature set into a database or/and
- 4) Matching the feature set against feature set in the database.

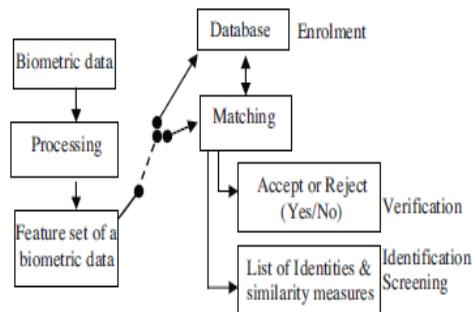


Figure 1: A generic biometric system

ANN can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. The advantages of ANN include:

- Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
- Self-Organization: ANN creates its own organization or representation of the information it receives during learning time.
- Fault Tolerance via Redundant Information Coding: However, some network capabilities may be retained even with major network damage [7].

#### 4. Smart Cameras Usage in Pattern Recognition

A smart camera is defined as a vision system in which the primary function is to produce a high-level understanding of the imaged sense and generate applications-specific data to be used in an intelligent system. A smart means it performs application-specific information processing (ASIP) which is usually not to provide better quality images for human viewing but to understand and describe what is happening in the images.

Smart cameras contain one or more highly performance microprocessors to run intelligent ASIP algorithms. There are different types and classification of vision systems and smart cameras such stand-alone, embedded and networked smart cameras. Our proposed technique will be based on a networked smart camera because it allows a digital image of any bank teller to be transferred directly from cameras to a PC [8].

Pattern recognition is the process of classifying input (e.g. image/data/pattern) based on the key feature in order to determine each object belongs to any one of several classes. Pattern recognition consists of three main phases as illustrated in figure 2, each including four levels. They are 1-signal level processing 2-feature level processing, 3-object level processing and 4-semantic-level processing [8].

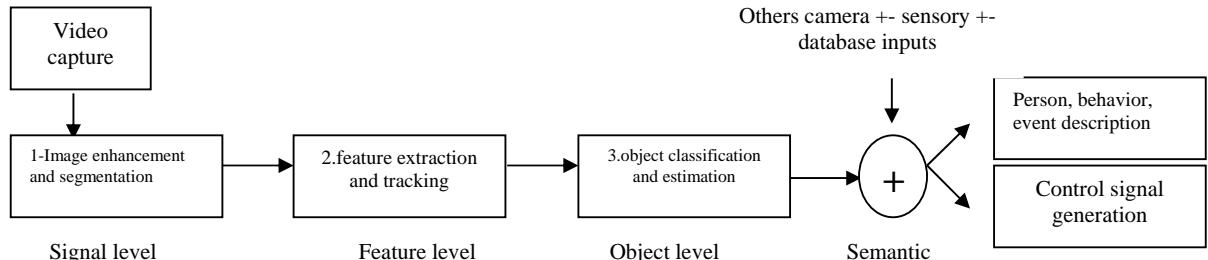


Figure 2: Processing flow of algorithms for pattern recognition inside smart cameras  
646  
<https://sites.google.com/site/ijcsis/>  
ISSN 1947-5500

## 5. The proposed Technique

The workflow of the proposed technique is initiated by firstly identify the type of biometric data such as (face-emotion, nose touch, eye movement, ear touch) that will be stored and learned to the proposed system. Secondly, it will be stored the agreement /determined action/hidden key/secret code for the intended bank teller that we need to protect him. Just the bank teller do the determined/unknown/secret action hence, the proposed technique start working to convert its input into programmable functions which include sending SMS as an emergency help/warning and calling the security departments or the existing bodyguard moreover taking a snapshot of the infected person or calling police office without his consideration or doing another backup/escape issues. Figure 3 depicts the block diagram of the proposed technique. It contains five essential parts. The steps of the proposed technique are summarized in figure 4.

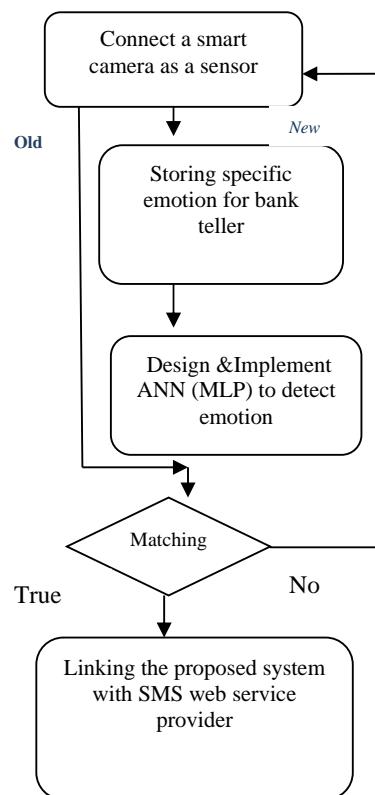


Figure 3 Block diagram of the proposed technique

- Step1:** Create a new account to the Bank teller
- Step2:** Select the biometrics object (Ear, Nose, mouse, etc)
- Step 3:** Connect a high quality smart camera
- Step 4:** Determine the protected text of SMS and SMS terminated
- Step 5:** Record and capture video from 1:2 seconds
- Step 6:** Design and Implement a MLP neural network (I/O)
- Step 7:** Train and Learn the MLP
- Step 8:** If the bank teller does the learned video then the proposed system directly connected to SMS web service provider and send a silent protected message to the bank security staff to safe him.

Figure 4 Steps of the proposed technique

Smart cameras have great advantages for applying the proposed technique. It embeds pattern recognition techniques that have a reflection for the security and preventing attacks. The role of ANN and its type MLP has an importance part of learning the desired neural network. A designed prototype is presented to support the proposed technique with a friendly and easily graphical user interface as shown in figure 5. Figure 5 contains the user interface of the proposed technique. It presented as a GUI prototype of the proposed model. The proposed system

should connect the IP of the smart camera for the intended bank teller. The bank teller should create a new account that store his video as a secret code/ hidden action based on the face emotion/biometrics (ear, nose, mouth ... etc).The input of MLP is the video of the bank teller. The MLP neural network should train itself by determining the desired weight.

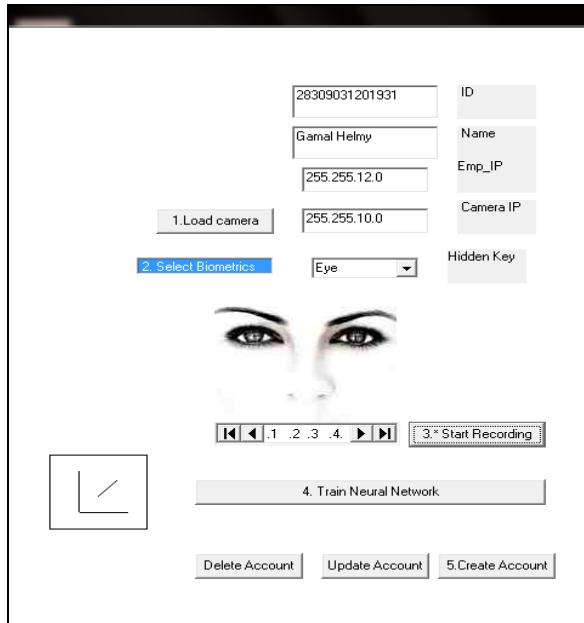


Figure 5 GUI for the proposed technique

## 6. Comparison of Existing Model and Proposed Model

Table 1and Table 2 show the comparison of existing methods of preventing and proposed model.

Table 1. A comparative study between the existing model and the proposed model

Method/Model	Cost	Fast Execution	Technical features	Prevent crime
Existing Models	High	Low	Semi	Half Secured
Proposed Model	Low	High	Fully Automation	Fully Secured

Table 2 comparison between the security level of existing models and proposed model

Method/Model	Security Level	ANN	Smart Camera	Face Recognition	Secret Key
Existing Models	Half secured	NA	Limited	N/A	NA
Proposed Model	Fully secured	Applicable	Fully Applied	Applicable	Applied

## 7. Conclusion

The proposed model has been developed for securing physical banking system with face recognition of bank teller face emotions. A novel technique is presented to prevent crime execution and makes the bank system more secure than others existing models. The integration of modern smart camera capabilities with MLP neural network is better than existing models and faster to take a critical decision from bank security guards. The death ratio will be decreased if all banks applied this proposed technique and preventing crime execution would be 100% approximately secured model. The piece of advice now for all bank systems is to embed their security systems with our proposed technique in order to get secured and safety surrounding environment.

## References

- [1] Criminal attack, definition available on [www.Wikipedia.com](http://www.Wikipedia.com), last access 1/Oct./2015
- [2] Online available on [www.FBI.com](http://www.FBI.com), Reports and publications , Bank crime statistical BCS last access 1/10/2015

- [3] Online available on [www.cvibes.net](http://www.cvibes.net), Last access Jan/2016
- [4] Alexeyy Minin," The Neural Network Analysis Data Filters", JASS, 2006 available on [www.eiic.cz](http://www.eiic.cz)
- [5] Chien Le,"A Survey of Biometrics Security Systems", online available on <http://www.cse.wustl.edu/~jain/cse571-11/ftp/biomet/index.html>
- [6] Yu Shi and F' abio Dias Real, Smart cameras: Fundamentals and Classification A.N. Belbachir (ed.), Smart Cameras, DOI 10.1007/978-1-4419-0953-4 2, Springer Science + Business Media, LLC 2010
- [7] Importance of ANN, online available on [www.abhinavjournal.com](http://www.abhinavjournal.com), last access 1/11/2015
- [8] Pattern Recognition definition and phases, smart cameras online available on [www.nicta.com.au](http://www.nicta.com.au), last access 1/Jan/2016.

# Modeling and Optimization the Four-Level Integrated Supply Chain: Sequential Quadratic Programming

Abolfazl. Gharaei (*Corresponding Author*), Ph.D. Candidate

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

Phone: +98 (21) 88830891, Fax: +98 (21) 88329213, e-mail: ab.gharaie@gmail.com

Seyed Hamid Reza. Pasandideh, Ph.D.

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

Phone: +98 (21) 88830891, Fax: +98 (21) 88329213, e-mail: shr\_pasandideh@khu.ac.ir

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

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

## I. INTRODUCTION AND RESEARCH LITERATURE

Supply chain management<sup>1</sup> is one of the most essential aspects of every business [1]. A supply chain is the set of entities which are involved in the design of new products and services, procuring raw materials, transforming them into semi-finished and finished products, and delivering them to the end customers [2]. SCM is efficient management of the end-to-end process starting from the design of the product or service to the time. This complete process includes product design, procurement, planning and forecasting, production, distribution, fulfillment, and after-sales support [1]. The integration of activities in a supply chain can be done either through forward mechanism that is from a supplier to a purchasing firm or through backward system going from a customer to a purchasing firm [3]. Supply chain integration<sup>2</sup> is expected to combine partners' resources and perspectives into a firm's value propositions thereby allowing all companies in the network to excel in performance. Previous studies highlighted the

---

<sup>1</sup> SCM

<sup>2</sup> SCI

importance of SCI by linking it with various outcomes like competitive advantage and supply chain performance [4, 5, 6]. Stank et al. (2001) suggested the significance of internal and external collaboration for successful supply chain [7]. Flynn et al. (2010) established the relationship between SCI (customer, supplier and internal integration) and their interactions on performance [8]. Gimenez et al. (2012) showed in their research that SCI increases performance moderated by a context variable like supply chain complexity [9]. Among an extensive research conducted on the SCI can mention the following recent works in chronological order. Ovalle et al. (2003) reviewed the literature to present the benefits of the CONWIP system in different supply chains. They carried on this analysis for different levels of demand volatility and material flow constraints along the chain and they optimized parameters to ensure the best performance of each control supply chain's policy in each scenario [10]. Jiu-Biing (2003) investigated an integrated supply chain-based spatial interaction model was formulated to determine facility locations of the transnational PC manufacturing centers and regional product distribution centers in China. The numerical results showed that the Shanghai municipality ranked highest for both the manufacturing and distribution centers for a transnational PC manufacturing enterprise [11].

Yao et al. (2004) considered an integrated supply chain model in which one vendor supplies items for the demand of multiple buyers. The objective of this model was to minimize the vendor's total annual cost subject to the maximum cost that the buyer might be prepared to incur [12]. Potter et al. (2004) studied, primarily using process-mapping techniques, the evolution of a case study steel supply chain within the UK over the past decade, drawing both previous work and current research. In fact, the approach taken in their paper showed that the changes that had occurred within the steel supply chain can be applied to previously described theory on supply chain evolution [13]. Dongsoo (2005) had designed and developed an integrated supply chain system. At first, he had analyzed hospital's business processes and reviewed system requirements for the efficient SCM. Vendor-Managed Inventory<sup>3</sup>, had adopted and implemented to improve material handling efficiency. Online procurement system developed for the departments that consumed drugs and placed orders [14]. SooWook (2006) examined the causal linkages among SCM practice, competition capability, the level of SCI, and firm performance. It was concluded that, in early stage, the emphasis on systemic SCI might be more crucial. Once SCI had been implemented, it might be advisable to focus on SCM practice and competition capability [15].

Stringer et al. (2007) provided a novel generic model of the integrated food supply chain for part of a major UK Food Standards Agency. In order to consider the types of breakdown throughout the food chain a generic approach adopted. This was constructed on a hierarchical basis. A hierarchical approach was used which described three levels: 5 stages, 9 operational steps and a series of 27 unit operations, each of which had defined [16]. Pitty et al. (2008) in part 1 from their two-part paper demonstrated that a dynamic model of an integrated supply chain could serve as a valuable quantitative tool that aided in such decision-making. In part 1, they presented a dynamic model of an integrated refinery supply chain. The model explicitly considered the various supply chain activities such as crude oil supply and transportation, along with intra-refinery supply chain activities such as procurement planning, scheduling, and operations management [17]. Cheng Hsiao (2008) investigated the integrated stochastic inventory problem for a two-stage supply chain consisting of a single retailer and a single supplier. By using batch shipment

---

<sup>3</sup> VMI

policy, the expected total cost reduced as significantly. Equally sized batch shipment models, controlled by both the reorder and the shipping points with sharing information, variable safety factors were constructed. The problem solved optimally by the proposed algorithms that determined the economic lot size, the optimal batch sizes, number of batches, and safety factor [18].

Ayoub et al. (2009) proposed an optimization model and solution approach for designing and evaluating integrated system of bio-energy production supply chains, supply chain, at the local level. A first part of this research introduced in earlier research work explaining the general Bio-energy Decision System (gBEDS). Two levels decision system for efficient planning and implementation of bio-energy production [19]. Peng et al. (2009) analyzed coal production, washing and processing, transportation and sales as an integrated coal supply chain and established a systematic optimization model that considers maximizing both the over-all profit and the satisfaction of the customers. So, it had achieved optimal system profit under conditions guaranteeing a certain level of customer satisfaction [20]. Tiwari et al. (2010) had mathematically modeled an integrated supply chain. Results indicated that the proposed model could not only search for optimal/near optimal solutions in large search spaces but also had good repeatability and convergence characteristics, thereby proving its superiority [21]. Anderson et al. (2010) studied the influence of supply chain structure on the equilibrium solution and hence on industry performance. They were able to give conditions under which decentralized supply chains were more profitable than integrated supply chains, and they investigated the impact of changes in the level of competition [22]. Mohammadi Bidhandi et al. (2011) proposed an integrated model and a modified solution method for solving supply chain network design problems under uncertainty. Model design of stochastic supply chain network was provided as a two-stage stochastic program where the two stages in the decision-making process correspond to the strategic and tactical decisions [23].

Lai'nez et al. (2012) provided a review that summarized the most recent concepts and structural components constituting the supply chain. Mentioned paper contributions were scarce, and the necessary negotiation among the responsible of the different supply chains lacks of an adequate representation framework [24]. Das et al. (2013) developed an integrated production inventory model of supplier and retailer that a delay in payment was offered by supplier to retailer for a constant deteriorating item and shortage was not allowed. In this model, the objective was to decide the position of the credit period and number of replenishment of retailer in finite time horizon in such a way that the integrated system got the optimum cost [25]. Sarkar et al. (2013) provided the paper that dealt with an integrated vendor-buyer supply chain model. Two models were constructed based on the probability distribution of the lead-time demand. The lead-time demand followed a normal distribution in the first model. In the second model, they considered the distribution free approach for the lead-time demand. For the second model, only mean and standard deviation were known [26]. Nasiri et al. (2014) modeled a three echelons supply chain with multiple distribution centers, production sites and suppliers. The main purpose of this paper was to model production, distribution and inventory control decisions simultaneously. For this supply chain, several commodities with defined items were produced [27]. Dong-Ping et al. (2014) considered a manufacturing supply chain with multiple suppliers in the presence of multiple uncertainties such as uncertain material supplies, stochastic production times, and random customer demands. They formulated the integrated inventory management policy for raw material

procurement and production control using the stochastic dynamic programming approach [28]. Zhang et al. (2014) investigated the optimal supply chain design for commodity chemicals (BTX, etc.) production via woody biomass fast pyrolysis and hydro processing pathway. Proposed approach in this paper was illustrated through a case study of Minnesota, with Minneapolis-St. Paul serving as the chemicals distribution hub [29].

Mohd Saleh et al. (2015) proposed a framework in investigating the influence of Supply Chain Relational Capital<sup>4</sup> on the execution of SCI by adopting relational capital theory. Without doubt, it was worthy to study what enabler factors that might facilitate the execution of integration as appeared to be a subject of interest among the academia, practitioners and policy makers who understand the critical application of SCI within SCM context [30]. Khalifehzadeh et al. (2015) aimed at designing a four-echelon supply chain structure including multiple suppliers, multiple producers, multiple distributors and multiple customers. In fact, in this paper, they considered a multi-objective multi-supplier, multi-plant, multi-DC, multi-product and multi-period PDP while shortage was authorized. The paper classified the customers based on their unit shortage costs. The high priority was given to customers with the highest unit shortage cost. They also considered several types of TSs where each pair of TS and route had its own reliability [31]. Alfalla-Luque et al. (2015) examined the relationship between employee commitment and SCI dimensions to explain several performance measures, such as flexibility, delivery, quality, inventory, and customer satisfaction. Their paper focused on the interrelationships between the different dimensions of SCI [32]. Diabat et al. (2016) developed a capacitated multi-echelon joint location-inventory model, according to which a single product is distributed from a manufacturer to retailers through a set of warehouses, the locations of which were to be determined by the model. They developed a genetic algorithm (GA) based heuristic to solve the problem [33]. Zhao et al. (2016) considered an integrated production-inventory-distribution planning problem faced by a multinational corporation managing a multi-stage supply chain over an infinite time horizon [34]. Najmi et al. (2016) addressed a large scale multi-period complementarity model for an advanced hydrocarbon biofuel supply chain integrated with existing petroleum refineries. This model simultaneously optimized the supply chain design and found the equilibrium quantity of feed stocks, crude oil and final products in the integrated supply chain [35]. Lemmens et al. (2016) reviewed the literature on model-based supply chain network design in order to identify the applicability of these models to the key issues of the design of a vaccine supply chain [36].

Jaber and Goyal (2008) investigated the coordination of order quantities amongst the players in a three-level supply chain. The first level of the supply chain consisted of multiple buyers, the second level of a vendor, and the third level of multiple suppliers. They developed a mathematical model in their paper which guaranteed the costs for the levels either remained the same as before coordination, or decreased as a result of coordination. Their paper was therefore a contribution along this line of research, because their suggested model achieved coordination among the members in a supply chain assuming common cycle time for all non-identical buyers. This matter facilitated the consolidation of orders by the vendor and subsequently by the suppliers. Consolidation of orders in a supply chain resulted in reducing the order processing costs of the chain members [37].

In this paper, we developed Jaber & Goyal's three-level supply chain model [37] to a four-level one with multi stage products and production capacity constraint. This chain consists of a supplier, a producer, a wholesaler, and

---

<sup>4</sup> SCRC

several retailers. These levels interact with each other in order to make an integrated chain to minimize the total inventory cost of chain. Ordering and holding costs, setup and production capacity costs are mentioned supply chain costs. In addition, we considered classic constraints as consisting of limitations on total procurement or production cost, maximum number of orders for products or items in each level, maximum space cost of products or items in each level, total number of stockpile for each level, maximum production time capacity for producing the products in producer level and maximum setup cost. The importance of present study can be seen in the cases in which collaboration between different companies in the supply chain is considered vital today for the success of each level in chain. If companies intend to remain competitive in the market, they may have to integrate with other levels of chains. Integration of levels with multi stage products in supply chains with the aim of cost minimization is one of the most important optimization problems in supply chain networks. So, in present study, we modeled inventory costs in a four levels of supply chain with multi stage products consisting of a supplier, a producer, a wholesaler and multiple retailers. In the following, we supposed levels as integrated and modeled total inventory cost of this chain. In the light of this, we further actualize the problem by assuming constraints containing limitations on procurement or production cost, number of orders, space costs, set up cost and production time capacity. Another realistic assumption, commonly ignored, is considering the limitation on the number of stockpiles in any SC level. The main goals in the paper are both number of agreed upon optimum stockpile of products and the agreed upon optimum period length of products upon which levels agreed to minimize total inventory cost of supply chain. So, it can be more applicable and widely used for chains which intend to be integrated, since the final goal of aforementioned chains is minimization of total inventory costs of chain. Also, chains tend to know the number of optimum stockpiles or optimum products batch size for products. Moreover, we used an exact proposed algorithm denoted by SQP. That is one of the most recently developed algorithms and perhaps one of the best optimization methods. This method with super-linear convergence rate helps us to find the optimum solution in large-scale nonlinear problems. The remainder of this paper organized as follows:

Section II describes assumptions and problem definition. Mathematical formulation of this model is provided in Section III. Section IV deals with the sequential quadratic programming (SQP) as solution method. Numerical examples are presented in section V. Finally, sensitivity analysis and conclusion are given in sections VI and VII, respectively.

## II. PROBLEM DEFINITION AND ASSUMPTIONS

We modeled a four-level integrated supply chain, the first level of which consists of a supplier, the second level of a producer, the third level of a wholesaler, and the fourth level of multiple retailers. These levels agree with each other on having the same period length and same number of stockpile for each product in order to make an integrated supply chain to minimize total cost of supply chain. In this problem, there are several multi stage products and constraints including: production time capacity for producer, procurement and production costs, number of orders, available space, number of stockpile and setup cost for producer. The goal is minimization of the total cost of four-level integrated supply chain while corresponding constraints are fulfilled. So, objectives are to find both the number of agreed optimum stockpile of products and the agreed optimum period length of products upon which

levels agreed to minimize total inventory cost of supply chain. The following assumptions are used for formulation of the problem:

- (1) There is a supplier, a producer, a wholesaler and multiple retailers that agreed with each other on having a same period length and same number of stockpile for each product in order to make an integrated supply chain to minimize total cost of supply chain.
- (2) There are " $l$ " items for each product in supplier level and there are " $n$ " products in each level of producer, wholesaler and retailers.
- (3) The planning horizon in this four-level integrated supply chain is infinite.
- (4) Shortage is not allowed in the model.
- (5) Discount is not allowed in the model.
- (6) Demand rate for products in each level of this model is known, deterministic and dependent.
- (7) The total procurement or production cost of products or items in each level is limited.
- (8) Maximum number of orders for products or items in each level is limited.
- (9) Maximum space cost of products or items in each level is limited.
- (10) The total number of stockpile for each product is limited and it is the same for each product at all levels.
- (11) Maximum production time capacity for producing the products in producer level is limited.
- (12) Maximum setup cost for producing the products in producer level is limited.
- (13) The period length of each product is unknown and it is same for each product at all levels.
- (14) Products shipment and delivery to the retailers must be done by wholesaler and according to the distance, procurement cost of per unit of product  $i$  for each of retailers is different from other retailers.

### III. MATHEMATICAL FORMULATION OF FOUR-LEVEL INTEGRATED SUPPLY CHAIN

Final objectives of this research are to find both number of optimum stockpile for products and optimum period length of products to minimize total cost of each level while the constraints are satisfied. The following notations for the parameters and the variables are used for each level.

#### A. Notations

$A_{is}$ : Fixed ordering cost related to items of  $i$ th product for supplier

$h_{jis}$ : Holding cost of  $j$ th item from  $i$ th product for supplier

$D_{is}$ : Demand rate of  $i$ th product of supplier

$\lambda_{is}$ : Number of optimum stockpile of  $i$ th product for supplier

$\lambda_s$ : Maximum number of stockpile for supplier's products

$T_{is}$ : The optimum period length of  $i$ th product for supplier

$TC_s$ : Total inventory cost of supplier

$f_{is}$ : The cost per unit of space for  $i$ th supplier's product

$S_s$ : Maximum space cost for supplier

$c_{is}$ : Supply cost per unit of product  $i$  that supplied by supplier

$C_s$ : Maximum supply cost for supplier

$O_s$ : Maximum number of orders for supplier

$A_{ip}$ : Fixed ordering cost per order of  $i$ th product of producer

$h_{jip}$ : Holding cost of  $j$ th item from  $i$ th product of producer

$h_{ip}$ : Holding cost of  $i$ th product of producer

$D_{ip}$ : Demand rate of  $i$ th product of producer

$a_{jip}$ : Fixed ordering cost per order of  $j$ th item from  $i$ th product of producer

$\lambda_{ip}$ : Number of optimum stockpile of  $i$ th product for producer

$\lambda_p$ : Maximum number of stockpile for producer's products

$T_{ip}$ : The optimum period length of  $i$ th product for producer

$TC_p$ : Total inventory cost of producer

$f_{ip}$ : The cost per unit of space for  $i$ th producer's product

$S_p$ : Maximum space cost for producer

$c_{ip}$ : Production cost for per unit of product  $i$  for producer

$s_{ips}$ : Setup cost for per unit of  $i$ th product in stage "s"

$c_{ips}$ : Production time capacity for per unit of  $i$ th product in stage "s"

$e_{ips}$ : The expense of per unit of production time capacity for per unit of  $i$ th product in stage "s"

$C_p$ : Maximum production cost for producer

$V_p$ : Maximum production time capacity for producing the products in producer level

$S_c$ : Maximum setup cost for producing the products in producer level

$O_p$ : Maximum number of orders for producer

$A_{iw}$ : Fixed ordering cost per order of  $i$ th product of wholesaler

$h_{iw}$ : Holding cost of  $i$ th product of wholesaler

$D_{iw}$ : Demand rate of  $i$ th product of wholesaler

$\lambda_{iw}$ : Number of optimum stockpile of  $i$ th product for wholesaler

$\lambda_w$ : Maximum number of stockpile for wholesaler's products

$T_{iw}$ : The optimum period length of  $i$ th product for wholesaler

$TC_w$ : Total inventory cost of wholesaler

$f_{iw}$ : The cost per unit of space for  $i$ th wholesaler's product

$S_w$ : Maximum space cost for wholesaler

$c_{iw}$ : Procurement cost per unit of product  $i$  for wholesaler

$C_w$ : Maximum procurement cost for wholesaler

$O_w$ : Maximum number of orders for wholesaler

$m$ : Number of retailers

$A_{irk}$ : Fixed ordering cost per order of  $i$ th product of  $k$ th retailer

$h_{irk}$ : Holding cost of  $i$ th product of  $k$ th retailer

$D_{irk}$ : Demand rate of  $i$ th product of  $k$ th retailer

$T_{irk}$ : The optimum period length of  $i$ th product for  $k$ th retailer

$TC_r$ : Total inventory cost of retailers

$f_{irk}$ : The cost per unit of space for  $i$ th product of  $k$ th retailer

$S_{rk}$ : Maximum space cost for  $k$ th retailer

$c_{irk}$ : Procurement cost of per unit of product  $i$  for  $k$ th retailer

$C_{rk}$ : Maximum procurement cost for  $k$ th retailer

$O_{rk}$ : Maximum number of orders for  $k$ th retailer

$n$ : Number of products

$l$ : Number of items which make up  $i$ th product

$s$ : Number of stage for producing the products which can be varied from one product to other product

$u_{ji}$ : Consumption rate of  $j$ th item in  $i$ th product

$f_i$ : Space occupied by each unit of product  $i$

$TC_T$ : Total inventory cost of four-level integrated supply chain

$\lambda_i$ : Number of agreed optimum stockpile of  $i$ th product upon which supplier, producer and wholesaler agreed to minimize total inventory cost of supply chain.

$T_i$ : The agreed optimum period length of  $i$ th product upon which levels agreed to minimize total inventory cost of supply chain.

#### B. The cost of four-level integrated supply chain

Total inventory cost of the supplier, including the ordering cost in the first term and holding cost in the second term is as (1):

$$TC_s = \sum_{i=1}^n \left( \frac{A_{is}}{\lambda_{is} \lambda_{ip} \lambda_{iw} T_{is}} + \frac{1}{2} \sum_{i=1}^n \left( \sum_{j=1}^l (h_{jis} u_{ji}) \right) D_{is} T_{is} (\lambda_{is} - 1) \lambda_{ip} \lambda_{iw} \right) \quad (1)$$

Total inventory cost of the producer, including the ordering cost in the first term, holding cost of produced product in the second term, holding cost of items in the third term, setup cost of products in the fourth term and production time capacity costs for producing the products in the fifth term is as (2):

$$\begin{aligned}
 TC_p = & \sum_{i=1}^n \frac{A_{ip} + (\sum_{j=1}^l a_{jip})}{\lambda_{ip} \lambda_{iw} T_{ip}} + \frac{1}{2} \sum_{i=1}^n h_{ip} D_{ip} T_{ip} (\lambda_{ip} - 1) \lambda_{iw} \\
 & + \frac{1}{2} \sum_{i=1}^n \left( \sum_{j=1}^l (h_{jip} u_{ji}) \right) D_{ip} T_{ip} (\lambda_{ip} - 1) \lambda_{iw} + \sum_{i=1}^n \left( \sum_{s=1}^s (s_{ips}) \right) D_{ip} T_{ip} \\
 & + \sum_{i=1}^n \left( \sum_{s=1}^s (c_{ips} e_{ips}) \right) D_{ip} T_{ip}
 \end{aligned} \tag{2}$$

Total inventory cost of the wholesaler, including the ordering cost in the first term and holding cost of products in the second term is as (3):

$$TC_w = \sum_{i=1}^n \left( \frac{A_{iw}}{\lambda_{iw} T_{iw}} + \frac{1}{2} \sum_{i=1}^n h_{iw} D_{iw} T_{iw} (\lambda_{iw} - 1) \right) \tag{3}$$

Total inventory cost of the retailers, including the ordering cost in the first term and holding cost of products in the second term is as (4):

$$TC_r = \sum_{k=1}^m \sum_{i=1}^n \frac{A_{irk}}{T_{irk}} + \frac{h_{irk}}{2} D_{irk} T_{irk} \tag{4}$$

Note that in (1) the ordering cost depends on the number of orders and number of each product stockpile for supplier, producer and wholesaler. Holding cost depends on the items which made up products and consumption rate items noted in products. In Equation (2), ordering cost in this level depends on the number of orders for products and items and number of each product stockpile for producer and wholesaler. In third term, holding cost of products depends on the items which made up products, consumption rate items noted in products and number of stockpile for producer and wholesaler. In both fourth term and fifth term, setup cost of products and production time capacity cost depend on period length of products. In Equation (3), ordering cost depends on the number of orders and number of each product stockpile for wholesaler. Also, holding cost depends on the number of stockpile for wholesaler. In Equation (4), the first term is the ordering cost of each retailer and second term is the costs associated with the holding for each of retailers; note that there are m retailers. Ordering cost depends on the number of each retailer orders.

As stated earlier, four levels of aforesaid supply chain interacted and agreed with each other on having the same period length and same number of stockpile for each product in order to provide an integrated supply chain to minimize total cost of supply chain. Constraints of this four-level integrated supply chain are:

- (1) The total procurement or production cost of products or items in each level is limited.
- (2) Maximum number of orders for products or items in each level is limited.

- (3) Maximum space cost of products or items in each level is limited.
- (4) The total number of stockpile for each level is limited.
- (5) Maximum production time capacity for producing the products in producer level is limited.
- (6) Maximum setup cost for producing the products in producer level is limited.

The objective function using (5) becomes:

$$\begin{aligned}
 \text{Min } TC_T &= \sum TC_s + TC_p + TC_w + \sum TC_{rk} \\
 &= \sum_{i=1}^n \left( \frac{A_{is}}{\lambda_{is} \lambda_{ip} \lambda_{iw} T_{is}} + \frac{1}{2} \sum_{i=1}^n \left( \sum_{j=1}^l (h_{jis} u_{ji}) \right) D_{is} T_{is} (\lambda_{is} - 1) \lambda_{ip} \lambda_{iw} \right) \\
 &\quad + \sum_{i=1}^n \frac{A_{ip} + (\sum_{j=1}^l a_{jip})}{\lambda_{ip} \lambda_{iw} T_{ip}} + \frac{1}{2} \sum_{i=1}^n h_{ip} D_{ip} T_{ip} (\lambda_{ip} - 1) \lambda_{iw} \\
 &\quad + \frac{1}{2} \sum_{i=1}^n \left( \sum_{j=1}^l (h_{jip} u_{ji}) \right) D_{ip} T_{ip} (\lambda_{ip} - 1) \lambda_{iw} + \sum_{i=1}^n \left( \sum_{s=1}^s (s_{ips}) \right) D_{ip} T_{ip} \\
 &\quad + \sum_{i=1}^n \left( \sum_{s=1}^s (c_{ips} e_{ips}) \right) D_{ip} T_{ip} + \sum_{i=1}^n \left( \frac{A_{iw}}{\lambda_{iw} T_{iw}} + \frac{1}{2} \sum_{i=1}^n h_{iw} D_{iw} T_{iw} (\lambda_{iw} - 1) \right) \\
 &\quad + \sum_{k=1}^m \sum_{i=1}^n \frac{A_{irk}}{T_{irk}} + \frac{h_{irk}}{2} D_{irk} T_{irk} \tag{5}
 \end{aligned}$$

All four levels interacted and agreed with each other on having the same period length and same number of stockpile for each product in order to make an integrated supply chain to minimize total cost of supply chain. Based on this cooperation and agreement between levels, we will have:

$$\lambda_{is} = \lambda_{ip} = \lambda_{iw} = \lambda_i \quad \text{and, } T_{is} = T_{ip} = T_{iw} = T_{is} = T_{irk} = T_i$$

Accordingly, by replacing  $\lambda_i$  and  $T_i$  parameters in (5) and constraints, the mathematical formulation of the problem becomes:

$$\begin{aligned}
 \text{Min } TC_T &= \sum_{i=1}^n \left( \frac{A_{is}}{\lambda_i^3 T_i} + \frac{1}{2} \sum_{i=1}^n \left( \sum_{j=1}^l (h_{jis} u_{ji}) \right) D_{is} T_i (\lambda_i - 1) \lambda_i^2 \right) + \sum_{i=1}^n \frac{A_{ip} + (\sum_{j=1}^l a_{jip})}{\lambda_i^2 T_i} \\
 &\quad + \frac{1}{2} \sum_{i=1}^n h_{ip} D_{ip} T_i (\lambda_i - 1) \lambda_i + \frac{1}{2} \sum_{i=1}^n \left( \sum_{j=1}^l (h_{jip} u_{ji}) \right) D_{ip} T_i (\lambda_i - 1) \lambda_i \\
 &\quad + \sum_{i=1}^n \left( \sum_{s=1}^s (s_{ips}) \right) D_{ip} T_i + \sum_{i=1}^n \left( \sum_{s=1}^s (c_{ips} e_{ips}) \right) D_{ip} T_i \\
 &\quad + \sum_{i=1}^n \left( \frac{A_{iw}}{\lambda_i T_i} + \frac{1}{2} \sum_{i=1}^n h_{iw} D_{iw} T_i (\lambda_i - 1) \right) + \sum_{k=1}^m \sum_{i=1}^n \frac{A_{irk}}{T_i} + \frac{h_{irk}}{2} D_{irk} T_i \tag{6}
 \end{aligned}$$

Subject to:

$$\sum_{i=1}^n c_{is} T_i D_{is} \leq C_s \quad (7)$$

$$\sum_{i=1}^n \left( \sum_{s=1}^s (s_{ips}) \right) D_{ip} T_i \leq S_c \quad (23)$$

$$\sum_{i=1}^n c_{iw} T_i D_{iw} \leq C_w \quad (9)$$

$$\sum_{i=1}^n c_{irk} T_i D_{irk} \leq C_{rk} \quad \text{for } k = 1, 2, \dots, m \quad (10)$$

$$\sum_{i=1}^n \frac{1}{T_i} \leq O_s \quad (11)$$

$$\sum_{i=1}^n \frac{1}{T_i} \leq O_p \quad (12)$$

$$\sum_{i=1}^n \frac{1}{T_i} \leq O_w \quad (13)$$

$$\sum_{i=1}^n \frac{1}{T_i} \leq O_{rk} \quad \text{for } k = 1, 2, \dots, m \quad (14)$$

$$\sum_{i=1}^n f_{is} f_i T_i D_{is} \leq S_s \quad (15)$$

$$\sum_{i=1}^n f_{ip} f_i T_i D_{ip} \leq S_p \quad (16)$$

$$\sum_{i=1}^n f_{iw} f_i T_i D_{iw} \leq S_w \quad (17)$$

$$\sum_{i=1}^n f_{irk} f_i T_i D_{irk} \leq S_{rk} \quad \text{for } k = 1, 2, \dots, m \quad (18)$$

$$\sum_{i=1}^n \lambda_i \leq \lambda_s \quad (19)$$

$$\sum_{i=1}^n \lambda_i \leq \lambda_p \quad (20)$$

$$\sum_{i=1}^n \lambda_i \leq \lambda_w \quad (21)$$

$$\sum_{i=1}^n \left( \sum_{s=1}^s (c_{ips}) \right) D_{ip} T_i \leq V_p \quad (22)$$

$$\sum_{i=1}^n \left( \sum_{s=1}^s (s_{ips}) \right) D_{ip} T_i \leq S_c \quad (23)$$

Constraints (7) to (10) guarantee procurement and production cost limitation for supplier, producer, wholesaler and retailers, respectively. Constraints (11) to (14) guarantee number of orders limitation for supplier, producer, wholesaler and retailers, respectively. Constraints (15) to (18) guarantee space costs limitation for supplier, producer, wholesaler and retailers, respectively. Constraints (19) to (21) guarantee limitation of number of stockpile for supplier, producer and wholesaler, respectively. Constraints (22) and (23) guarantee limitation of production time capacity and setup cost for producing the products in producer level, respectively.

The formulation given in (6) to (23) is a nonlinear-programming type. However, sequential quadratic programming (SQP) is a powerful and effective solution method for a wide range of nonlinear optimization problems [38]. It is an exact algorithm and its solution is optimum and does not require validation. SQP method will be used in the next section for optimization of research model.

#### IV. SOLUTION METHOD

SQP is an iterative method for nonlinear optimization. SQP method is used in problems for which the objective function and the constraints are continuously twice differentiable. SQP methods solve a sequence of optimization sub-problems each of which optimizes a quadratic model of the objective subject to a linearization of the constraints. If the problem is unconstrained, then the method reduces to Newton's method for finding a point where the gradient of the objective vanishes. If the problem has only equality constraints, then the method is equivalent to applying Newton's method to the first-order optimality conditions, or "Karush Kuhn Tucker" conditions, of the problem. In other words, SQP is one of the best methods for optimization of the solution of nonlinear programming problems with super-linear convergence rate [39]. Pasandideh et al. (2015) show that SQP has satisfactory performance for solving nonlinear problems in terms of optimum solutions, number of iterations to achieve the optimum solution, infeasibility, optimality error, and complementarity [40]. Also, Gharaei et al. (2015) confirmed satisfactory performance of this exact method for solving nonlinear problems [41]. In its purest form, the SQP algorithm replaces the objective function with the quadratic approximation and replaces the constraint functions by linear approximations as (24).

$$q_k(d) = \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla_{xx}^2 \mathcal{L}(x_k, \lambda_k) d \quad (24)$$

Recall the general form of a nonlinearly constrained optimization problem as (25), (26) or (27), (28).

*Minimize f(x)*

Subject to:

$$g_j + \nabla g_j^T \Delta X \leq 0 \quad ; \quad j = 1, 2, \dots, m \quad (25)$$

$$h_k + \nabla h_k^T \Delta X = 0 \quad ; \quad k = 1, 2, \dots, p \quad (26)$$

or:

*Minimize f(x)*

Subject to:

$$c_i(x) \leq 0 \quad \forall i \in I \quad (27)$$

$$c_i(x) = 0 \quad \forall i \in \varepsilon \quad (28)$$

The step  $d_k$  is calculated by solving the quadratic subprogram as (29).

$$\text{Min}\{q_k(d): c_i(x_k) + \nabla c_i(x_k)^T d \leq 0, i \in I; c_i(x_k) + \nabla c_i(x_k)^T d = 0, i \in \varepsilon\} \quad (29)$$

Also, at first, the Lagrange function is built in (30):

$$L = f(X) + \sum_{j=1}^m (\lambda_j g_j(X) + \sum_{k=1}^p (\lambda_{m+k} h_k(X))) \quad (30)$$

As in the case of Newton's method of unconstrained minimization, the solution vector  $\Delta X$  is treated as the search direction  $S$ , and the quadratic programming sub-problem (in terms of the design vector  $S$ ) is defined as follows:

Find  $S$  (search direction) such that:

$$Q(S) = \nabla f(X)^T S + \frac{1}{2} S^T [H] \cdot S \quad (31)$$

is minimized subject to:

$$\beta_j g_j(X) + \nabla g_j(X)^T S \leq 0 \quad ; \quad j = 1, 2, \dots, m \quad (32)$$

$$\bar{\beta} h_k(X) + \nabla h_k(X)^T S = 0 \quad ; \quad k = 1, 2, \dots, p \quad (33)$$

Constraint (32) and (33) are related to a local quadratic model of the Lagrangian  $\tilde{L}$  as the objective function which leads to the QP subproblem. In Equation (31) and constraints (32) and (33),  $[H]$  is a positive definite matrix which is taken initially as the identity matrix and is updated in subsequent iterations so as to converge to the Hessian matrix of the Lagrange function. Once the search direction, " $S$ " is found by solving the problem modelled in (30), the design vector is updated as shown in (34).

$$X_{j+1} = X_j + \alpha^* S \quad (34)$$

Where  $\alpha^*$  is step length and  $(\alpha^* S)$  is  $d_k$ . If the starting point  $x_0$  is sufficiently close to  $x^*$ , and the Lagrange multiplier estimating  $\{\lambda_k\}$  remains sufficiently close to  $\lambda^*$ , then the sequence generated by setting  $x_{k+1}=x_k + d_k$  converges to  $x^*$  at a second order rate. These assurances cannot be made in other cases. Indeed, codes based on this approach must modify the sub-problem when the quadratic  $q_k$  is unbounded below on the feasible set or when the feasible region is empty. The Lagrange multiplier estimates needed to set up the second-order term in  $q_k$  can be obtained by solving an auxiliary problem or by simply using the optimal multipliers for the quadratic sub-problem at the previous iteration. The strategy based on solving the quadratic sub-problem makes the decision about which of the inequality constraints appear to be active at the solution internally during the solution of the quadratic program. A somewhat different algorithm is obtained by making this decision prior to formulating the quadratic program. This variant explicitly maintains a working set  $w_k$  of apparently active indices and solves the quadratic programming problem as (35).

$$\text{Min}\{q_k(d) : c_i(x_k) + \nabla c_i(x_k)^T d = 0, i \in w_k\} \quad (35)$$

Also, we can find the step  $d_k$  as follows:

The contents of  $w_k$  are updated at each iteration by examining the Lagrange multipliers for the subproblem and by examining the values of  $c_i(x_{k+1})$  at the new iteration  $x_{k+1}$  for  $i \notin w_k$ . This approach is called the EQP (equality-based QP) variant of SQP to distinguish it from the IQP (inequality-based QP) variant described above. The sequential QP approach outlined above requires the computation of  $\nabla_{xx}^2 \mathcal{L}(x_k, \lambda_k)$ . Most codes replace this matrix with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) approximation  $B_k$ , which is updated at each iteration. An obvious update strategy (consistent with the BFGS update for unconstrained optimization) would be to define as (36).

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla_x \mathcal{L}(x_{k+1}, \lambda_k) \quad (36)$$

And update the matrix  $B_k$  by using the BFGS formula as (37).

$$B_{k+1} = B_k - \frac{B_k S_k S_k^T B_k}{S_k^T B_k S_k} + \frac{y_k y_k^T}{y_k^T S_k} \quad (37)$$

However, one of the properties that makes Broyden-class methods appealing for unconstrained problems is that its maintenance of positive definiteness in  $B_k$  is no longer assured, since  $\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*)$  is usually positive definite only in a subspace. This difficulty may be overcome by modifying  $y_k$ . Whenever  $y_k^T S_k$  is not sufficiently positive,  $y_k$  is reset to:  $S_k B_k (\theta_k - 1) + y_k \theta_k \rightarrow y_k$ , where  $\theta_k \in [0, 1]$  is the number closest to 1 such that  $y_k^T S_k \geq \sigma S_k^T B_k S_k$  for some  $\sigma \in (0, 1)$ . The convergence properties of the basic sequential QP algorithm can be improved by using a line search. The choice of distance to move along the direction generated by the sub-problem is not as clear as in unconstrained case, where we simply choose a step length which approximately minimizes  $f$  along the search direction. For constrained problems, we would like the next iteration not only to decrease  $f$  but also to come closer to satisfy the constraints. Often these two aims conflict. So it is necessary to weigh their relative importance and define a merit or penalty function, which can be used as a criterion for determining whether or not one point is better than another. Generally, SQP is a technique for the solution of nonlinear programming (NLP) problems with super linear convergence rate that requires the related QP sub-problem must be solvable per iteration [39]. In the next section, numerical examples are provided to illustrate the application of the proposed solution method for solving and optimizing the model of the four-level integrated supply chain.

## V. NUMERICAL EXAMPLES

Three numerical examples for this four-level integrated supply chain are solved in this section. In mentioned numerical examples, there are 1 supplier, 1 producer, 1 wholesaler and 2 retailers. There are 2 products in this chain for which product 1 in producer level has three-stages production process and product 2 has two-stages production process. Also, each product consists of 2 items which consumption rate of  $1^{st}$  and  $1^{st}$  items in  $1^{st}$  product ( $u_{ji}$  parameter) is 1. Consumption rate of  $1^{st}$  item in  $2^{nd}$  product is 2 and consumption rate of  $2^{nd}$  item in  $2^{nd}$  product is 1. Also, there is limitation on production time capacity for producing mentioned products in producer level. The

general data and parameters of all chain levels in three numerical examples are summarized in Tables (1) to (5). As stated before, the objectives are to find the number of optimum stockpile and optimum period length for each product in the levels of this large and nonlinear problem while the model constraints are satisfied.

The starting point of SQP method in the first example is:

$$\lambda_1 = 1.7, \quad \lambda_2 = 1.7, \quad T_1 = 0.12, \quad T_2 = 0.40$$

The starting point of SQP method in the second example is:

$$\lambda_1 = 1.5, \quad \lambda_2 = 1.5, \quad T_1 = 0.30, \quad T_2 = 0.50$$

The starting point of SQP method in the third example is:

$$\lambda_1 = 2, \quad \lambda_2 = 2, \quad T_1 = 0.6, \quad T_2 = 0.6$$

TABLE 1. GENERAL DATA AND PARAMETERS OF SUPPLIER IN THREE NUMERICAL EXAMPLES

Level	Parameters	Example 1	Example 2	Example 3
Supplier	$A_{1s}$	80	120	145
	$A_{2s}$	95	145	178
	$D_{1s}$	175	243	335
	$D_{2s}$	220	309	410
	$c_{1s}$	5	9	6
	$c_{2s}$	7	5	9
	$f_1$	0.6	1	0.9
	$f_2$	0.5	1.1	0.85
	$f_{1s}$	0.4	0.4	1
	$f_{2s}$	0.5	0.5	1.1
	$h_{11s}$	2	4	5
	$h_{21s}$	3	2	4
	$h_{12s}$	3	2	7
	$h_{22s}$	4	3	7
$\lambda_s$	7	6	15	
	$S_s$	5500	4800	6700
	$C_s$	44,000	41,000	85,200
	$O_s$	22	15	31

TABLE 2. GENERAL DATA AND PARAMETERS OF WHOLESALER IN THREE NUMERICAL EXAMPLES

Level	Parameters	Example 1	Example 2	Example 3
Wholesaler	$A_{1w}$	67	125	250
	$A_{2w}$	81	148	290
	$D_{1w}$	140	210	300
	$D_{2w}$	185	280	370
	$c_{1w}$	13	13	10
	$c_{2w}$	12	10	13
	$f_1$	0.5	0.8	0.7
	$f_2$	0.5	0.7	0.8
	$f_{1w}$	11	13	10
	$f_{2w}$	10	11	9
	$h_{1w}$	14	10	17
	$h_{2w}$	9800	12500	17100
	$\lambda_w$	91000	87500	165000
	$S_w$	40	25	78

TABLE 3. GENERAL DATA AND PARAMETERS OF PRODUCER IN THREE NUMERICAL EXAMPLES

Level	Parameters	Example 1	Example 2	Example 3
Producer	$A_{1p}$	70	100	180
	$A_{2p}$	85	139	197
	$D_{1p}$	150	238	320
	$D_{2p}$	200	295	395
	$c_{1p}$	7	11	8
	$c_{2p}$	10	7	11
	$f_1$	0.6	0.5	1.2
	$f_2$	0.7	0.5	0.7
	$f_{1p}$	10	6	8
	$f_{2p}$	12	7.5	9
	$h_{1p}$	3	2.5	10
	$h_{2p}$	7	6.5	11
	$h_{11p}$	4	4.6	5
	$h_{21p}$	8	7.5	9
	$h_{12p}$	30	25	50
	$h_{22p}$	40	30	57
	$a_{11p}$	35	33	65
	$a_{21p}$	50	45	92
	$a_{12p}$	10	12	25
	$a_{22p}$	9	10	15
	$s_{1p1}$	11	12	19
	$s_{1p2}$	8	10	14
	$s_{1p3}$	10	11	18
	$s_{2p1}$	50	55	95
	$s_{2p2}$	45	43	82
	$c_{1p1}$	30	28	60
	$c_{1p2}$	40	39	58
	$c_{1p3}$	35	37	78
	$c_{2p1}$	3	4.5	6
	$c_{2p2}$	7	5.5	9
	$\epsilon_{1p1}$	5	7	5
	$\epsilon_{1p2}$	6	8	5
	$\epsilon_{1p3}$	4	7	6
	$\epsilon_{2p1}$	100000	78500	210000
	$\epsilon_{2p2}$	30000	28500	57500
	$V_p$	9	8	6
	$S_c$	7000	6800	15100
	$\lambda_p$	55000	47000	89500
	$S_p$	35	27	65

TABLE 4. GENERAL DATA AND PARAMETERS OF RETAILER 1 IN THREE NUMERICAL EXAMPLES

Level	Parameters	Example 1	Example 2	Example 3
Retailer 1	$A_{1r1}$	50	66	125
	$A_{2r1}$	55	70	140
	$D_{1r1}$	60	101	150
	$D_{2r1}$	75	134	170
	$c_{1r1}$	40	15	25
	$c_{2r1}$	70	16	29
	$f_1$	0.3	0.9	0.8
	$f_2$	0.2	1.1	0.7
	$f_{1r1}$	6	9	9
	$f_{2r1}$	5	11	7
	$h_{1r1}$	8500	7300	17000
	$h_{2r1}$	43000	41000	68800
	$S_{r1}$	18	11	35

TABLE 5. GENERAL DATA AND PARAMETERS OF RETAILER 2 IN THREE NUMERICAL EXAMPLES

Level	Parameters	Example 1	Example 2	Example 3
Retailer 2	$A_{1r2}$	50	72	115
	$A_{2r2}$	55	65	138
	$D_{1r2}$	57	59	142
	$D_{2r2}$	71	76	168
	$c_{1r2}$	40	53	19
	$c_{2r2}$	70	65	33
	$f_1$	0.6	1	0.9
	$f_2$	0.5	1.1	0.85
	$f_{1r2}$	0.2	1	0.75
	$f_{2r2}$	0.25	0.9	0.8
	$h_{1r2}$	4	8	10
	$h_{2r2}$	6	10	8
	$S_{r2}$	7900	7600	18100
	$G_{r2}$	41500	38700	65100
	$O_{r2}$	19	10	32

The examples are solved using the SAS 9.2 computer software with its embedded SQP method. Table (6) shows the optimum solution, number of iterations to achieve the optimum solution, infeasibility, optimality error, and complementarity obtained for the three numerical examples.

TABLE 6. OPTIMAL RESULTS FOR NUMERICAL EXAMPLES

Example No	Optimum solution		Objective function value	Number of iterations	Infeasibility	Optimality error	Complementarity
Example 1	$\lambda_1$	1.8477	24587.306313	45	6.910668E-10	2.737262E-10	6.910668E-10
	$\lambda_2$	1.5819					
	$T_1$	0.10498					
	$T_2$	0.118					
Example 2	$\lambda_1$	1.3989	73052.178182	50	6.848921E-11	4.6568918E-9	6.848921E-11
	$\lambda_2$	1.3246					
	$T_1$	0.19947					
	$T_2$	0.20053					
Example 3	$\lambda_1$	2.2955	71826.043381	76	1.0480719E-9	8.0735187E-9	1.0480719E-9
	$\lambda_2$	1.8264					
	$T_1$	0.056149					
	$T_2$	0.075814					

In Table (6),  $\lambda_i$  and  $T_i$  are number of agreed optimum stockpile of  $i$ th product and agreed optimum period length of  $i$ th product, respectively. The "objective function value" column is calculated based on (6). The number of steps taken by the SQP method to achieve the optimal solution is shown in the column denoted as "number of iterations." As the computation is carried out in a finite-precision environment, rounding errors prevent the algorithm to obtain a solution which satisfies the preceding condition exactly (showed in infeasibility column). Instead, we terminate the algorithms at some small threshold values which can be measured in an absolute or relative sense. The sixth column in Table 5 refers to the "optimality error" and the last column ("complementarity") is the relative infeasibility defined as the maximum amount of constraint violation relative to the one in the "infeasibility" column. As stated earlier,

calculation is carried out in SAS software. The SAS programming code developed to solve the first example is given in Appendix A, as a representative for all SAS programming codes written to solve all problems.

## VI. SENSITIVITY ANALYSIS

A sensitivity analysis on the change rate of the objective function based on change rate of the number of agreed optimum stockpile for products (Lambda Parameter) is performed in this section only for Example #1 with the initial data shown in Tables (1) to (5). It involves increasing or decreasing Lambda parameter at  $\pm 10$ ,  $\pm 20$ ,  $\pm 30$ ,  $\pm 40$  and  $\pm 50$  percent, Table (7) represents the results. It can be seen from Table 7, increase or decrease the rate of the parameter " $\lambda_i$ " leads to decrease in rate of the objective function. Increasing the rate of parameter " $\lambda_i$ " (greater than 1), leads to decrease rate of the objective function with less steep and decreasing the rate of the parameter " $\lambda_i$ " (less than 1), leads to decrease rate of the objective function with more steep.

TABLE 7. EFFECTS OF " $\lambda_i$ " CHANGES ON THE OPTIMAL RESULTS OF EXAMPLE #1

The rate of increase/ decrease parameter "Lambda" in chain levels	Objective function value of chain	Change rate of objective function	The rate of increase/ decrease parameter "Lambda" in chain levels	Objective function value of chain	Change rate of objective function
1.5 " $\lambda_i$ "	26,328	-0.0708	0.9 " $\lambda_i$ "	24,685	-0.00398
1.4 " $\lambda_i$ "	25,737	-0.04677	0.8 " $\lambda_i$ "	25,023	-0.01773
1.3 " $\lambda_i$ "	25,260	-0.02737	0.7 " $\lambda_i$ "	25,717	-0.04595
1.2 " $\lambda_i$ "	24,900	-0.01273	0.6 " $\lambda_i$ "	27,003	-0.09826
1.1 " $\lambda_i$ "	24,670	-0.00337	0.5 " $\lambda_i$ "	29,415	-0.19636

## VII. CONCLUSIONS

In this paper, we modeled and optimized four-level integrated supply chain with "n" multi stage products. There are "l" items for each product in supplier level and "n" products in each level of producer, wholesaler and retailers. Levels of this chain consist of a supplier, a producer, a wholesaler and multiple retailers which interacted with each other for having the same period length and same number of stockpile for each product in order to minimize total cost of chain. Products in producer level have multi-stage production process. Accordingly, the SQP was utilized as an excellent exact solution in order to find the number of optimum stockpile and optimum period length for each product in the levels. The results of SQP implementations on numerical examples showed optimum performance of SQP exact method for solving nonlinear and large problems such as provided model. Finally, results of the sensitivity analysis confirmed excellent performance of SQP optimum method. For future research, we are going to consider this model in shortage condition with two objective functions, first: minimization of total inventory cost of chain and second: maximization of total reliability of mentioned four-level integrated supply chain.

## REFERENCES

- [1] L.X. Lu and J.M. Swaminathan, "Supply chain management," Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. pp. 709-713, 2015.
- [2] J.M. Swaminathan, " Supply Chain Management," International Encyclopedia of the Social and Behavioral Sciences, Elsevier Sciences, Oxford, England, 2001.
- [3] P.D. Cousins and B. Menguc, " The implications of socialization and integration in supply chain management," Journal of Operations Management. Vol. 5, pp. 604-620, 2006.
- [4] H.L. Lee and C. Billington, " Managing supply chain inventory: pitfalls and opportunities," Sloan management review. Vol. 33, No. 3, 1992.
- [5] R. Narasimhan and S.W. Kim, " Effect of supply chain integration on the relationship between diversification and performance: evidence from Japanese and Korean firms," Journal of operations management. Vol. 20, pp. 303-323, 2002.

- [6] D. Prajogo and J. Olhager, "Supply chain integration and performance: The effects of long-term relationships, information technology and sharing, and logistics integration," *International Journal of Production Economics*. Vol. 135, pp. 514-522, 2012.
- [7] T.P. Stank, S. Keller, P. Daugherty, " Supply chain collaboration & logistical service performance," *Journal of Business Logistics*. Vol. 22, pp. 29-48, 2001.
- [8] B.B. Flynn, B. Huo, X. Zhao, " The impact of supply chain integration on performance: A contingency and configuration approach," *Journal of Operations Management*. Vol. 28, pp. 58-71, 2010.
- [9] C. Gimenez, T. Van der Vaart, D.P. Van Donk, " Supply chain integration and performance: the moderating effect of supply complexity," *International Journal of Operations & Production Management*. Vol. 32, pp. 583-610, 2012.
- [10] O.R. Ovalle and A.C. Marquez, " Exploring the utilization of a CONWIP system for supply chain management. A comparison with fully integrated supply chains," *International Journal of Production Economics*. Vol. 83, pp. 195-215, 2003.
- [11] S. Jih-Biing, " Locating manufacturing and distribution centers: An integrated supply chain-based spatial interaction approach," *Transportation Research Part E*. Vol. 39, pp. 381-397, 2003.
- [12] M.J. Yao, C.C. Chiou, "On a replenishment coordination model in an integrated supply chain with one vendor and multiple buyers," *European Journal of Operational Research*. Vol. 159, pp. 406-419, 2004.
- [13] A. Potter, R. Mason, M. Naim, C. Lalwani, "The evolution towards an integrated steel supply chain: A case study from the UK," *International Journal of Production Economics*. Vol. 89, pp. 207-216, 2004.
- [14] k. Dongsoo, "An integrated supply chain management system: A case study in healthcare sector," *Lecture Notes in Computer Science*, Vol. 3590, pp. 218-227, 2005.
- [15] K. Soo Wook, "Effects of supply chain management practices, integration and competition capability on performance," *Supply Chain Management: An International Journal*. Vol. 11, pp. 241-248, 2006.
- [16] M.F. Stringer and M.N. Hall, "A generic model of the integrated food supply chain to aid the investigation of food safety breakdowns," *Food Control*. Vol. 18, pp. 755-765, 2007.
- [17] S.S. Pitty, W. Li, A. Adhitya, R. Srinivasan, I.A. Karimi, "Decision support for integrated refinery supply chains Part 1," *Dynamic simulation, Computers and Chemical Engineering*. Vol. 32, pp. 2767-2786, 2008.
- [18] Y. Cheng Hsiao, " Integrated logistic and inventory model for a two-stage supply chain controlled by the reorder and shipping points with sharing information," *Int. J. Production Economics*. Vol. 115, pp. 229-235, 2008.
- [19] N. Ayoub, E. Elmoshi, H. Seki, Y. Naka, "Evolutionary algorithms approach for integrated bio-energy supply chains optimization," *Energy Conversion and Management*. Vol. 50, pp. 2944-2955, 2009.
- [20] H.J. Peng, M.H. Zhou, M.Z. Liu, Y. Zhang, Y.B. Huang, " A dynamic optimization model of an integrated coal supply chain system and its application," *Mining Science and Technology*. Vol. 19, pp. 0842-0846, 2009.
- [21] M.K. Tiwari, N. Raghavendra, S. Agrawal, S.K. Goyal, "A hybrid Taguchi-Immune approach to optimize an integrated supply chain design problem with multiple shipping," *European Journal of Operational Research*. Vol. 03, pp. 95-106, 2010.
- [22] E.J. Anderson, Y. Bao, "Price competition with integrated and decentralized supply chains," *European Journal of Operational Research*. Vol. 200, pp. 227-234, 2010.
- [23] H. Mohammadi Bidhandi and R. Mohd Yusuff, "Integrated supply chain planning under uncertainty using an improved stochastic approach," *Applied Mathematical Modeling*. Vol. 35, pp. 2618-2630, 2011.
- [24] J.M. Lai'nez and L. Puigjaner, "Prospective and perspective review in integrated supply chain modeling for the chemical process industry," *Current Opinion in Chemical Engineering*. Vol. 1, pp. 430-445, 2012.
- [25] D. Das and P.A. Dutta, "System dynamics framework for integrated reverse supply chain with three way recovery and product exchange policy," *Computers & Industrial Engineering*. Vol. 66, pp. 720-733, 2013.
- [26] B. Sarkar and A. Majumder, "Integrated vendor-buyer supply chain model with vendor's setup cost reduction," *Applied Mathematics and Computation*. Vol. 224, pp. 362-371, 2013.
- [27] G.R. Nasiri, R. Zolfaghari, H. Davoudpour, "An integrated supply chain production-distribution planning with stochastic demands," *Computers & Industrial Engineering*. Vol. 77, pp. 35-45, 2014.
- [28] S. Dong-Ping, D. Jing-Xin, X. Jingjing, "Integrated inventory management and supplier base reduction in a supply chain with multiple uncertainties," *European Journal of Operational Research*. Vol. 232, pp. 522-536, 2014.
- [29] Y. Zhang, G. Hu, R.C. Brown, "Integrated supply chain design for commodity chemicals production via woody biomass fast pyrolysis and upgrading," *Bio-resource Technology*. Vol. 157, pp. 28-36, 2014.
- [30] Z. Mohd Saleh and R. Mohd Roslin, "Supply chain integration strategy: a conceptual model of supply chain relational capital enabler in the Malaysian food processing industry," *Procedia - Social and Behavioral Sciences*. Vol. 172, pp. 585 - 590, 2015.
- [31] S. Khalifehzadeh, M. Seifbarghy, B. Naderi, "A four-echelon supply chain network design with shortage: a mathematical modeling and solution methods," *Journal of Manufacturing Systems*. Vol. 35, pp. 164-175, 2015.
- [32] R. Alfalla-Luque, J.A. Marin-Garcia, C. Medina-Lopez, "An analysis of the direct and mediated effects of employee commitment and supply chain integration on organizational performance," *Int. J. Production Economics*. Vol. 162, pp. 242-257, 2015.
- [33] A. Diabat and R. Deskoores, "A hybrid genetic algorithm based heuristic for an integrated supply chain problem," *Journal of Manufacturing Systems*. Vol. 38, pp. 172-180, 2016.
- [34] S.T. Zhao, K. Wu, X.M. Yuan, "Optimal integer-ratio inventory coordination policy for an integrated multi-stage supply chain," *Applied Mathematical Modelling*. Vol. 40, pp. 3876-3894, 2016.
- [35] A. Najmi, H. Shakouri, S. Nazari, "An integrated supply chain: A large scale complementarity model for the biofuel markets," *Biomass and Bioenergy*. Vol 86, pp. 88-104, 2016.
- [36] S. Lemmens, C. Decouttere, N. Vandaele, M. Bernuzzi, "A review of integrated supply chain network design models: Key issues for vaccine supply chains," *Chemical Engineering Research and Design*. Vol. 109, pp. 366-384, 2016.
- [37] M.Y. Jaber and S.K. Goyal, "Coordinating a three-level supply chain with multiple suppliers, a vendor and multiple buyers," *International Journal of Production Economics*. Vol. 116, pp. 95-103, 2008.
- [38] P.E. Gill and E. Wong, "Sequential quadratic programming methods," UCSD Department of Mathematics, Technical Report NA. Vol. 10, 2010.
- [39] Z. Zhu, "An efficient sequential quadratic programming algorithm for nonlinear programming," *Journal of Computational and Applied Mathematics*. Vol. 175, pp. 447-464, 2005.
- [40] S.H.R. Pasandideh, S.T.A. Niaki, A. Gharaei, "Optimization of a multiproduct economic production quantity problem with stochastic constraints using sequential quadratic programming," *Knowledge-Based Systems*. Vol. 84, pp. 98-107, 2015.

[41] A. Gharaei, B. Naderi, M. Mohammadi, "Optimization of rewards in single machine scheduling in the rewards-driven systems," Management Science Letters. Vol. 5, pp. 629–638, 2015.

APPENDIX A: THE SAS PROGRAMMING CODE TO SOLVE THE FIRST NUMERICAL EXAMPLE

```

proc optmodel;
    var L {1..2} >= 0;
    var T {1..2} >= 0;
    minimize obj =
80/(L[1]^3*T[1]) + 95/(L[2]^3*T[2]) + 437.5*T[1]*(L[1]-1)*L[1]^2 + 1100*T[2]*(L[2]-1)*L[2]^2 + 140/(L[1]^2*T[1]) + 170/(L[2]^2*T[2]) + 1500*T[1]*(L[1]-1)*L[1] + 2800*T[2]*(L[2]-1)*L[2] + 97044*T[1] + 80000.5*T[2] + 67/(L[1]*T[1]) + 81/(L[2]*T[2]) + 770*T[1]*(L[1]-1) + 925*T[2]*(L[2]-1) + 100/T[1] + 110/T[2];

    con cons1:
        875*T[1] + 1540*T[2] <= 44000;
    con cons2:
        1050*T[1] + 2000*T[2] <= 55000;
    con cons3:
        1820*T[1] + 2220*T[2] <= 91000;
    con cons4:
        2400*T[1] + 5250*T[2] <= 43000;
    con cons5:
        2280*T[1] + 4970*T[2] <= 41500;
    con cons6:
        (1/T[1]) + (1/T[2]) <= 22;
    con cons7:
        (1/T[1]) + (1/T[2]) <= 35;
    con cons8:
        (1/T[1]) + (1/T[2]) <= 40;
    con cons9:
        (1/T[1]) + (1/T[2]) <= 18;
    con cons10:
        (1/T[1]) + (1/T[2]) <= 19;
    con cons11:
        42*T[1] + 55*T[2] <= 5500;
    con cons12:
        54*T[1] + 70*T[2] <= 7000;
    con cons13:
        42*T[1] + 46.25*T[2] <= 9800;
    con cons14:
        10.8*T[1] + 7.5*T[2] <= 8500;
    con cons15:
        6.84*T[1] + 8.88*T[2] <= 7900;
    con cons16:
        L[1] + L[2] <= 7;
    con cons17:
        L[1] + L[2] <= 9;
    con cons18:
        L[1] + L[2] <= 14;
    con cons19:
        18750*T[1] + 15000*T[2] <= 100000;
    con cons20:
        4500*T[1] + 3600*T[2] <= 30000;

/* starting point */
L[1] = 1.7;
L[2] = 1.7;
T[1] = .12;
T[2] = .40;
    solve with SQP / printfreq = 4;
    print L[1] L[2] T[1] T[2];
quit;

```

# Different languages Classification Engine

Khalaf F. Khatatneh  
Faculty of Graduate Studies  
Al-Balqa'a Applied University  
Al-Salt, Jordan

Samer S. Khanfar  
Faculty of Graduate Studies  
Al-Balqa'a Applied University  
Al-Salt, Jordan

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

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

## I. INTRODUCTION.

Every language in the world has complexities and complications, and most languages converge in most similar cases that drive them to difficulties. Such as Arabic and English languages, the Arabic language has 28 letters from the viewpoint of the human, but from the viewpoint of the computer, it has until this moment 38 letters without counting diacritics. It is an open set of letters. Perhaps the number of letters will increase in the future. Also, the English language has 26 letters from the viewpoint of the human, but it has 80 letters from the viewpoint of the computer, taking in the consideration capital letters, small letters and diacritics (Greek & Latin diacritics) [1].

### A. Issues of languages.

Some complexities, complications and issues in both languages must treated, such as, diacritics, the orthography, the morphology, the feminine, the number of nouns, the number of verbs, pronouns and the similarity (Fuzzy). In addition, the phonemic representation and the homophone differentiation belonging to the English language [1].

### B. Encoding techniques.

The advent and growth of the revolution of software made each operating system to have its own coding system. Hence, the need for a unified coding system appeared. Finally, the Unicode Transformation Format 8 (UTF-8) was been certified as a unified format for all operating systems, each language has its own Unicode charts ranges, see Table I and Table II [2].

Table I  
ARABIC UNICODE CHARTS RANGES

No	Name	Range
1	Arabic.	0600 – 06FF
2	Arabic Supplement.	0750 – 077F
3	Arabic Extended-A.	08A0 – 08FF
4	Arabic Presentation Forms-A.	FB50 – FDFF
5	Arabic Presentation Forms-B.	FE70 – FEFF
6	Arabic Mathematical Alphabetic Symbols.	1EE00 – 1EFFF

Table II  
ENGLISH UNICODE CHARTS RANGES

No	Name	Range
1	Basic Latin (ASCII).	0000 - 07FF
2	Latin-1 Supplement.	0080 - 00FF
3	Latin Extended-A.	0100 - 017F
4	Latin Extended-B	0180 - 024F
5	Latin Extended-C	2C60 - 2C7F
6	Latin Extended-D	A720 - A7FF
7	Latin Extended-E	AB30 - AB6F
8	Latin Extended Additional.	1E00 - 1EFF
9	Latin Ligatures.	FB00 - FB06
10	Fullwidth Latin Letters.	FF00 - FF5E
11	IPA Extensions.	0250 - 02AF
12	Phonetic Extensions.	1D00 - 1D7F
13	Phonetic Extensions Supplement.	1D80 - 1DBF

### C. The interdependence of sciences.

Several sciences exploited to achieve the desired goal. Such as, the statistics which dropped on the computer science and produced a science of Data Mining (DM). Data Mining (DM) is a Knowledge Discovery from Data (KDD) [3]. The Text Mining (TM) is a branch of the DM. TM cares of unstructured data; usually these data are textual data. The preprocessing phase is a major phase in the DM life cycle, which used to process textual data. Some of preprocessing operations are preparation, tokenization and Natural Language Processing (NLP). The Artificial Intelligence (AI) had dropped on MLCE also, especially by using the Machine Learning (ML) field, specifically by using supervised learning approach. In addition, classification algorithms which listed under statistics.

### D. The MLCE construction.

The MLCE construction can be done by using sciences, which have mentioned above. Those sciences have translated as modules, and can inserted into any phase in the MLCE life cycle, and they work serially or on the demand. More details in the following sections.

## II. RELATED WORKS.

Actually, no previous works similar to this work have done. Previous works explained methods, which used in the MLCE in a separate fashion. Such methods like fetching data, preparation, normalization, tokenization, stemming, training data and classification. Moreover, there are several works talked about automatic text classification system. This work characterized by treating several methods and shaping them as a single unit. An implementation was done and desired results have gotten. In following lines, related works of methods will described separately.

Considering fetching and preparing data, [4] explained a Mercator application that has programmed by Java. It is a web crawler depends on a multithreads technique. It has run on several machines differs in their properties, it has completed 77.4 million of HTTP requests, 1.8 million requests of total requests have no response, just 2.3% of total requests. 65.8 million Documents have downloaded, 80% of downloaded documents were between 1K and 32K bytes in size. Motaz Saad in his thesis [5] uses sequential steps to build a ready categorized corpus. Steps are: (1) Collecting web documents from internet using the open source offline explorer, “httrack”. (2) Encoding the corpus that consists of html/xml files into UTF-8 encoded text files by using the Text Encoding Converter by

“WebKeySoft” (no more available). (3) Trimming html/xml tags by developing a java program. A resulted corpus named “OSAC” (Open Source Arabic Corpus).

Considering the stemming, there are three types of stemming, dictionary-based stemming, root stemming and light stemming. [6] produced many light stemmers for Arabic language, Last light stemmer version named “Light10”. It mentioned that no a need to the morphological analysis when using the light stemming. It defined a set of prefixes and a set of suffixes. Light10 removes those prefixes and suffixes according to defined priorities and conditions. Likewise, [7] introduced “Ruled-Based Light Stemmer” for Arabic language that depends on stripping a set of prefixes and suffixes from words, taking into consideration problems of a compatibility and ambiguity that belongs to prefixes and suffixes. It depends on Arabic patterns, suffixes of the dual, suffixes of the plural, pattern of the breaking plural and the length of the word that has stripped. At the level of the English language, [8] its work depends on the conditional trimming for suffixes, using a list of suffixes and some condition codes to identify both the suffix be deleted and how to produce a valid stem. It produced a genius algorithm that used widely in several search engines and indexing systems until these days.

Considering training data, there are four approaches of training, supervised, semi-supervised, unsupervised and active learning. [9] Tested the performance of the Naïve Bayes classifier in Arabic text categorization. The classification process applied by both supervised and semi-supervised techniques in an attempt to compare between them. A light stemming performed here. The accuracy of the classifier measured using recall, precision, fallout and error. The results show that semi-supervised learning can significantly improve the classification accuracy of Arabic text.

Considering the classification, “Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts” [3]. There are several classification algorithms in the science of statistic, these algorithms can be used in the computer science and it can gives amazing results. [10] Invented a system for browsing a collection of about 237K Arabic news articles, which should be applicable to other Arabic news collections as well. It developed a new stemmer named “P-Stemmer” that tested by three classifiers and produce better results than other stemmers did. It used SVM classifier as automatic classification method. Evaluation techniques applied here. In [9] publication, the Naïve Bayes classifier (NB) has used in categorizing Arabic text using both supervised and semi-supervised techniques in an attempt to compare between them and see if the classification accuracy will increase as a result of using the semi-supervised technique. It shows that semi-supervised approach can improve the classification accuracy of Arabic documents, but this improvement becomes more difficult as the error rate becomes smaller. Other important research in [11] that classifies Arabic documents by using Maximum Entropy (ME) method after the process of the preprocessing that done on those documents.

Considering classification systems, there are several tries to build special purpose classification systems, which mostly does not base on scientific methods mentioned before. Such as, [12] developed MUDABlue. MUDABlue is a special purpose classification system that categorizes software automatically; it can classify the software to multi-classes and generates classes automatically. A user interacts with MUDABlue through a Web interface to browse and to search categories of software systems. Likewise, [13] Developed ArabDox. It does organizing, classifying, storing, indexing and retrieving documents in three languages Arabic, English and French.

### III. COMPONENTS OF THE MLCE.

In this section a descriptions about main components of the MLCE, these components appear in several modules which in stages of the life cycle of the MLCE. These components are the fetching, the preparation, the tokenization, the stemming and the classification, they shown in Figure 1.

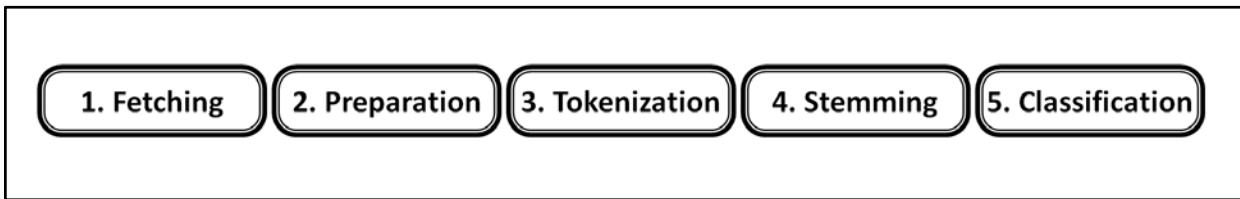


Figure 1. Components of the MLCE.

#### A. *The Fetching Tool.*

The fetching tool has abilities to integrate data that stored in any format into the database. Maybe data that have to be treated have been stored as xml, text, pdf, html, database, audio, images, videos and other types of files. Therefore, there is a need to a tool that has ability to integrate all those types of files into the database of the MLCE. This tool is useful in two cases; the first case is the uploading of data that already have classified into the database as a dataset that may use in training data or/and testing data. The second case is the listening to incoming requests, which ask to discover subjects of documents.

#### B. *The Preparation.*

The preparation procedure is the process the responsible of making the text document ready to the next step, it guarantees the produced text to be a single paragraph, free of numbers , words separated by single space, free of special characters, free of diacritics and just characters of the selected language only are allowed to pass. [1] developed A Preparation\_Normalization algorithm that used here in this work.

#### C. *The Tokenization.*

It is a way to pick up all words in the prepared document. Each picked word should sent to the next step. Words separated by single space. Talking about the tokenization here is at word level (not a character level, not a sentence level and not a paragraph level). This step mixed with the stemming process.

#### D. *The Stemming.*

It is a trimming operation, which removes prefixes and/or suffixes from a given word and transfers it to its stem or its root. Porter-Stemmer is advisable for the English language. Light-stemming approach is advisable for the Arabic language. [1] developed A Discretization-stemming algorithm that used here in this work.

#### E. *Behaviors of the preparation and the stemming.*

Both processes of the preparation and the stemming are external sub-programs, which can called from inside the MLCE and giving parameters, after sub-programs accomplish its tasks, they returns results as parameters to the MLCE. This way allows selecting any stemmer for a specific language, likewise for the sub-program of the preparation as shown in Figure 2.

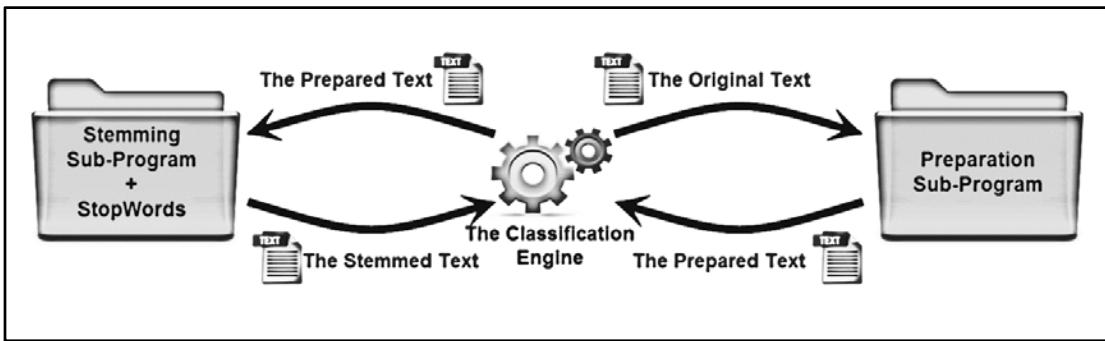


Figure 2. Behaviors of the preparation and the stemming.

#### F. The Classification.

The classification is a statistical operation. It does a lot of mathematical operations and computations to establish a model that helps in the prediction process and distinguishes between previously defined classes. There are many classification algorithms, such as, Naïve Bayesian (NB), Support Vector Machines (SVM), Maximum Entropy (ME) and more.

Multinomial Naïve Bayes Classifier (MNBC) applied here in this work. It is a simple probabilistic classifier and gives a height accuracy. It based on (1) to compute  $P(X|C_i)$  that in (2).

$$P(X|C_i) \approx N! \times \prod_{i=1}^k \frac{P_i^{n_i}}{n_i!} \quad (1)$$

The MNBC is an enhanced edition of NBC, which is suitable for TC (or documents categorization). MNBC is not like the simple NBC in the dealing with the appearance of words as a logical attribute. It takes into account frequencies of words in the training dataset and in documents. It determines the class of the inputted document not just by determining occurs of words but also by determining the frequencies of these occurs [14].

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (2)$$

#### IV. THE LIFE CYCLE OF THE MLCE.

The life cycle of the MLCE has four main components: the setting, the fetching, training data and classifying (see Figure 3).

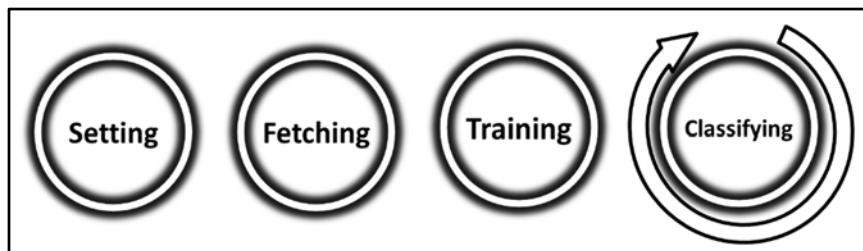


Figure 3. Components of the MLCE lifecycle.

A. *The Setting.*

During the setting phase is the establishment of the MLCE of parameters that describes its behavior, such as the number of languages, the preparation, the stemmer, the threshold of word frequencies and a classifier for each language. This phase is done one time only before beginning in any other phase, because this phase affects on results that given by the execution of next procedures.

B. *The Fetching.*

The phase of the fetching must visited one time at least. It transfers data into the MLCE database to make them ready for the next phase (training data).

C. *The Training.*

The training phase establishes subjects (classes) and their statistical calculations. In addition, it performs a sequence of procedures that applied on each textual document in the batch of documents, which has been chosen to be data for the training. Procedures that applied on each document shown in Figure 4.

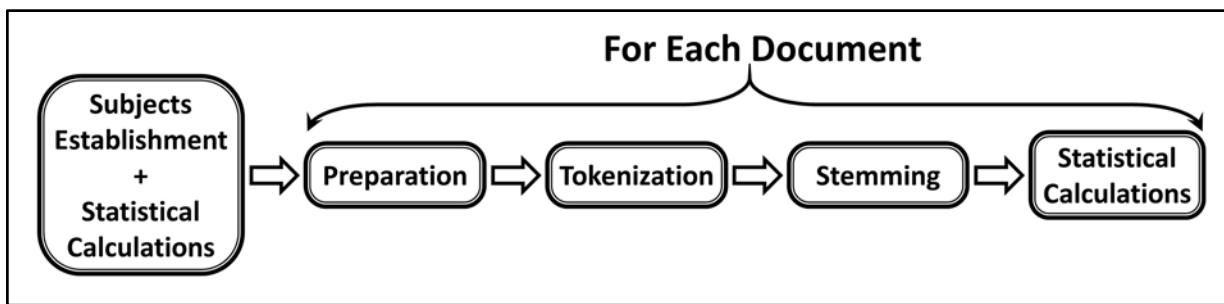


Figure 4. Procedure of the Training Phase

D. *The Classifying.*

The classifying phase is a continuous phase and listens to incoming jobs as shown in Figure 5; it is a memory resident program. Each job has the textual document as string and the language identifier, the job asks for the discovering of a subject. The classifying phase generates a proper thread for each incoming job depending on the language identifier; the generated thread has a sequence of procedures that do several operations and statistical calculations based on known classifiers, these procedures: the preparation, the tokenization, the stemming, statistical calculations and using a classifier, as shown in Figure 6. During this phase, all results after all procedures have to be stored in the database for future analysis and researches.

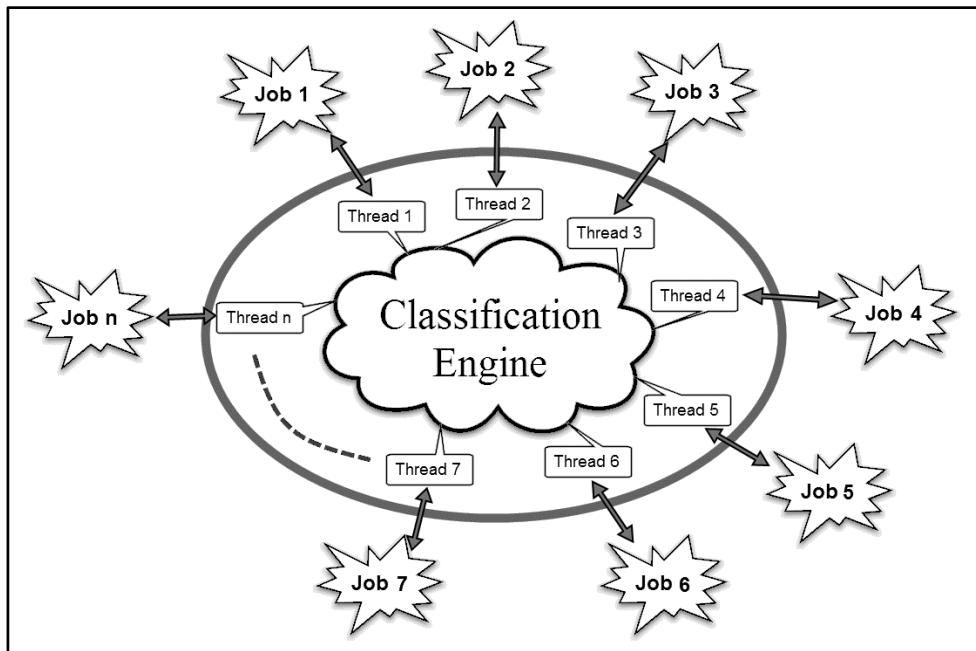


Figure 5. The Classifying Phase.

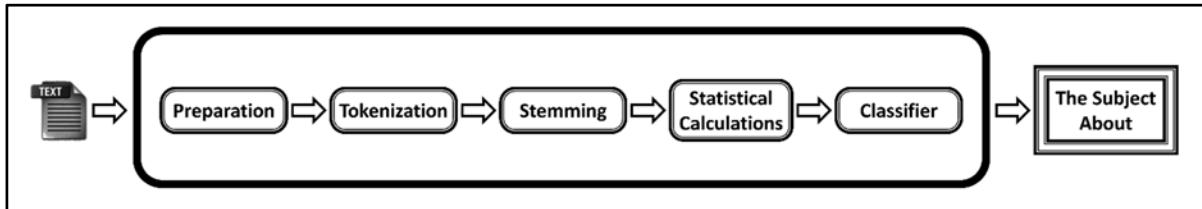


Figure 6. The Generated Thread.

## V. EXPERIMENTS AND RESULTS.

Some of the design of the MLCE that mentioned above has programmed by using VB.NET and MSSQL to test the performance and accuracy of the proposed design. All treated data had stored in the database, and results had extracted by using commands of the SQL.

### A. The Arabic and the English Corpora.

Open Source Arabic Corpus (OSAC) built by [15], modified by adding a new subject called “إرهاق” that contains 13 documents only, it became a collection of 22,442 documents already classified within 11 subjects (classes) Table III. For the English language, Wikipedia-categories (AFEWC) is a collection of xml and text files in three languages: English, French and Arabic. It has been downloaded from (sourceforge.net). It is free of charge; it is a collection of 8,288 documents, which already classified within 13 subjects Table III. All documents of both languages had stored in the database by using the fetching tool. After running the fetching tool, the database contains 30,730 documents.

Table III  
THE ARABIC AND THE ENGLISH CORPORA.

English			Arabic		
No.	Subject	#Documents	No.	Subject	#Documents
1	belief	26	1	ارهاب	13
2	culture	750	2	اقتصاد	3102
3	economy	176	3	تاريخ	3233
4	geography	387	4	تربية و اسرة و مرأة	3608
5	health	234	5	دين و فتاوى شرعية	3171
6	history	2328	6	رياضة	2419
7	mathematics	137	7	صحة	2296
8	medicine	252	8	فلك	557
9	nutrition	39	9	قانون	944
10	politics	355	10	قصص	726
11	religion	275	11	وصفات و اكلات	2373
12	science	1458			
13	sport	1871			
<i>Total</i>		8288	<i>Total</i>		22442

#### B. The Running of the Training.

After execution the training process, certain changes happened on the database. The training process has applied on a part of stored data for both languages. Table IV shows results of statistical calculations that have applied on English documents, which previously selected. Likewise,

Table V shows results of statistical calculations that have applied on Arabic documents, which previously selected. Table VI shows data for both two languages about aggregated summaries of subjects and words of vocabularies. Table VII shows aggregated data about quantity of words for all training documents.

Table IV  
RESULTS OF STATISTICAL CALCULATIONS FOR ENGLISH DOCUMENTS.

English			
Subject	#of words in the subject	#of trained documents	Prior probability
belief	40936	26	0.011938
culture	252593	200	0.091827
economy	269457	176	0.080808
geography	179059	200	0.091827
health	297091	200	0.091827
history	230759	200	0.091827
mathematics	152867	137	0.062902
medicine	193697	200	0.091827
nutrition	62527	39	0.017906
politics	270262	200	0.091827
religion	363400	200	0.091827
science	181553	200	0.091827
sport	259716	200	0.091827
<i>Total</i>	2753917	2178	

Table V  
RESULTS OF STATISTICAL CALCULATIONS FOR ARABIC DOCUMENTS.

Arabic			
Subject	#of words in the subject	#of trained documents	Prior probability
إرهاق	5870	13	0.005173
اقتصاد	64985	250	0.099483
تاريخ	266975	250	0.099483
تربية و اسرة و مرأة	368819	250	0.099483
دين وفلواني شرعية	155650	250	0.099483
رياضة	67531	250	0.099483
صحة	175379	250	0.099483
فلك	77240	250	0.099483
قانون	181898	250	0.099483
قصص	181983	250	0.099483
وصفات وأكلات	53620	250	0.099483
Total	1599950	2513	

Table VI  
AGGREGATED SUMMARIES OF SUBJECTS AND WORDS OF VOCABULARIES.

Language	# Vocabulary words (Unique)	Training Documents	# Subjects
English	95688	2178	13
Arabic	27114	2513	11

Table VII  
AGGREGATED DATA ABOUT QUANTITY OF WORDS FOR ALL TRAINING DOCUMENTS.

The Language	# of words in original documents	# of words in prepared documents	# of words in stemmed documents
English	5682215	5548909	2753917
Arabic	2303103	2171715	1599950

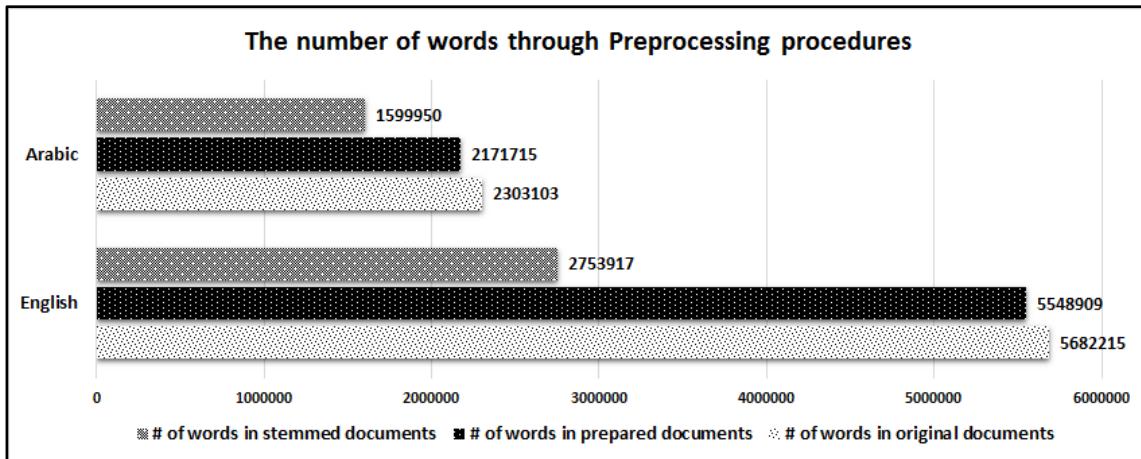


Figure 7. Number of words thought the preprocessing phase.

### C. The Experiment No. I.

The experiment no.1 has done by using the same data that had used in the training process, (see Table IV and

Table V). It is a reclassification operation. The MLCE reclassified all documents that had passed. Some documents had reclassified in other classes. Table VIII and Table IX show results for both languages Arabic and English.

Table VIII  
RESULTS OF EXPERIMENT NO. 1 ON ENGLISH LANGUAGE.

The Subject	# Training Documents	The Same Subject	% of similarity	Other Subject	% of different
<b>belief</b>	26	3	11.54%	23	88.46%
<b>culture</b>	200	168	84.00%	32	16.00%
<b>economy</b>	176	164	93.18%	12	6.82%
<b>geography</b>	200	182	91.00%	18	9.00%
<b>health</b>	200	162	81.00%	38	19.00%
<b>history</b>	200	149	74.50%	51	25.50%
<b>mathematics</b>	137	132	96.35%	5	3.65%
<b>medicine</b>	200	168	84.00%	32	16.00%
<b>nutrition</b>	39	35	89.74%	4	10.26%
<b>politics</b>	200	175	87.50%	25	12.50%
<b>religion</b>	200	179	89.50%	21	10.50%
<b>science</b>	200	145	72.50%	55	27.50%
<b>sport</b>	200	187	93.50%	13	6.50%

In the subject “belief” as shown in Table VIII, only 11.54% of documents reclassified in the subject itself. This refers to several reasons, one of them is that the subject “belief” is a weak subject and must not found or mixed with other. Other reason is that the operations used through the training process or the classification process needs to reconfigure correctly by using other algorithms. Addition some reason refers to terms that in the “belief” subject appear in other subjects more. On other hand, the “mathematics” subject has 96.35% similarity. Same reasons mentioned before maybe effects on this result. In addition, maybe the subject has training documents should not be in training data.

Table IX  
RESULTS OF EXPERIMENT NO. 1 ON ARABIC LANGUAGE.

The Subject	# Training Documents	The Same Subject	% of similarity	Other Subject	% of different
ارهاب	13	8	61.54%	5	38.46%
اقتصاد	250	250	100.00%	0	0.00%
تاريخ	250	242	96.80%	8	3.20%
تربية و اسرة و مرأة	250	249	99.60%	1	0.40%
دين وفتاوی شرعية	250	239	95.60%	11	4.40%
رياضة	250	248	99.20%	2	0.80%
صحة	250	248	99.20%	2	0.80%
فلك	250	249	99.60%	1	0.40%
قانون	250	248	99.20%	2	0.80%
قصص	250	230	92.00%	20	8.00%
وصفات وأكلات	250	250	100.00%	0	0.00%

As shown in Table IX, the subject “وصفات وأكلات” has a similarity percentage of 100%. This result indicates to an idea that there are terms found in the subject make a document (with those terms) classified under this subject. The subject “دين وفتاوی شرعية” has 4.40% as a percentage of different, 11 documents caused this percentage. 2 documents classified as “تربية و اسرة و مرأة” and 9 documents as “تاريخ“.

D. The experiment No. 2.

In this experiment, 100 documents as maximum (if available) from each subject classified by the engine. Those documents from the collection in the database and not in training data.

Table X  
RESULTS OF EXPERIMENT NO. 2 ON ENGLISH LANGUAGE.

The Subject (S)	# New Test Documents (TD)	# of TD Labelled in S	% Labelled in S	# of TD Labelled not in S	% Labelled not in S
belief	<NA>	<NA>	<NA>	<NA>	<NA>
culture	100	43	43.00%	57	57.00%
economy	<NA>	<NA>	<NA>	<NA>	<NA>
geography	100	74	74.00%	26	26.00%
health	34	23	67.65%	11	32.35%
history	100	35	35.00%	65	65.00%
mathematics	<NA>	<NA>	<NA>	<NA>	<NA>
medicine	52	32	61.54%	20	38.46%
nutrition	<NA>	<NA>	<NA>	<NA>	<NA>
politics	100	79	79.00%	21	21.00%
religion	75	61	81.33%	14	18.67%
science	100	38	38.00%	62	62.00%
sport	100	96	96.00%	4	4.00%

As shown in Table X, the “history” subject has 35% the least percentage, this means that most of documents not classified in a correct manner. This refers to several reasons, maybe putting this subject with available subjects make it very weak. Maybe training data not selected correctly. Maybe preprocessing operations needs more conditions. Maybe classification algorithm have to change.

The “sport” subject has 96% the highest percentage; it proves that everything working well. Actually, sports subjects are different in the conceptual side, like in other subjects. In other subjects may there is a similarity in the conceptual side; for example, "belief" and "religion" subject are close together in the conceptual side. In addition, "health" and "medicine" subjects. And so on.

Table XI  
RESULTS OF EXPERIMENT NO. 2 ON ARABIC LANGUAGE.

The Subject (S)	# New Test Documents (TD)	# of TD Labelled in S	% Labelled in S	# of TD Labelled not in S	% Labelled not in S
ارهاب	<NA>	<NA>	<NA>	<NA>	<NA>
اقتصاد	100	100	100.00%	0	0.00%
تاريخ	100	96	96.00%	4	4.00%
تربيۃ و اسرة و مرأة	100	93	93.00%	7	7.00%
دين وفتاوی شرعیة	100	100	100.00%	0	0.00%
رياضة	100	99	99.00%	1	1.00%
صحة	100	100	100.00%	0	0.00%
فلک	100	94	94.00%	6	6.00%
قانون	100	94	94.00%	6	6.00%
قصص	100	64	64.00%	36	36.00%
وصفات وأكلات	100	100	100.00%	0	0.00%

The experiment on the Arabic language gave different results, some subjects reach to 100%, and the least percentage was 64% it was for the “قصص” subject. Subjects in the Arabic collection are not similar in the viewpoint

of conceptual, except the subject “قصص”，this type of subjects may effect on other subjects like “تاريخ” and “اسرة و مرأة” subjects. So, the correct selection of subjects and choosing training data in a very careful manner, may effect on the behavior and results of the classification engine.

#### REFERENCES

- [1] Khanfar, S.S., Discovering Subjects Through the Design of a Multi-Lingual Engine That Classifies Documents, in Faculty of graduate studies. 2016, Al-Balqa' Applied University. p. 97.
- [2] Unicode. The Unicode Consortium 2015 June 18 [cited 2015 December 2]; Available from: [www.unicode.org](http://www.unicode.org).
- [3] Han, J., M. Kamber, and J. Pei, Data mining: concepts and techniques: concepts and techniques. 2011: Elsevier.
- [4] Heydon, A. and M. Najork, Mercator: A scalable, extensible web crawler. World Wide Web, 1999. 2(4): p. 219-229.
- [5] Saad, M.K., The Impact of Text Preprocessing and Term Weighting on Arabic Text Classification, in Computer Engineering. 2010, The Islamic University - Gaza: ResearchGate. p. 113.
- [6] Larkey, L.S., L. Ballesteros, and M.E. Connell, Light stemming for Arabic information retrieval, in Arabic computational morphology. 2007, Springer. p. 221-243.
- [7] Ababneh, M., et al., Building an Effective Rule-Based Light Stemmer for Arabic Language to Improve Search Effectiveness. International Arab Journal of Information Technology (IAJIT), 2012. 9(4).
- [8] Porter, M.F., An algorithm for suffix stripping. Program, 1980. 14(3): p. 130-137.
- [9] Khatatneh, K., Classified Arabic Documents Using Semi-Supervised Technique. Journal of Systems and Software, 2015.
- [10] Kanan, T. and E.A. Fox, Automated Arabic Text Classification with P-Stemmer, Machine Learning, and a Tailored News Article Taxonomy. 2015, Department of Computer Science, Virginia Polytechnic Institute & State University.
- [11] El-Halees, A., Arabic text classification using maximum entropy. The Islamic University Journal (Series of Natural Studies and Engineering), 2007. 15(1): p. 157-167.
- [12] Kawaguchi, S., et al., Mudable: An automatic categorization system for open source repositories. Journal of Systems and Software, 2006. 79(7): p. 939-953.
- [13] Sakhr, S.S.C. ArabDox Knowledge Management. 2015 [cited 2015 December 27]; Available from: <http://www.sakhr.com/>.
- [14] Witten, I.H. and E. Frank, Data Mining: Practical machine learning tools and techniques. 2005: Morgan Kaufmann.
- [15] Saad, M.K. and W. Ashour. OSAC: Open Source Arabic Corpora. in 6th International Symposium on Electrical and Electronics Engineering and Computer Science, Cyprus. 2010.

# Level Set Segmentation of Oil Spill Images with Non-Separable Wavelet Transform

Ganta Raghootham Reddy<sup>1</sup>, Ramudu Kama<sup>2</sup>, R.Srikanth<sup>3</sup> and Rameshwar Rao<sup>4</sup>

<sup>1</sup>Research Scholar, Department of ECE, OUCE, Osmania University, Hyderabad, <sup>2,3</sup>Dept.of ECE, KITS-Warangal-15,

<sup>4</sup>Rtd.Professor, Department of ECE, Osmania University, Hyderabad, rameshwarrao@hotmail.com

## Abstract

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

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

## 1. Introduction

**N**ASA developed a long term earth observatory for observing environment changes in land surfaces, atmosphere, biosphere, water bodies and ocean of the earth. The cyclone, oil spills, fire etc are the earth observation images. Each image shows the details of the area affected or covered under it. The Study and analysis of such images are tedious tasks including the Image enhancement, segmentation, description, and image representation etc. In this paper, our work is mainly focusing on oil spills segmentation of ocean SAR images.

The goal of image segmentation is nothing but partition a image in to its constituent parts, resulting in various homogeneous regions. Every region is a constituent part or object in the entire scene. The important area of interest within this category is to determine the edges of a digital image. It is typically used to locate objects and boundaries (lines, curves, etc.) in images. Image segmentation based on Level set algorithms can be classified into two categories. First category based on edge based segmentation and the second category is the segmentation based on regions in the images called region segmentation. In the first category mostly preferred edge detection method is Geodesic Active Contour (GAC) model for efficient detection of edges in an image [1]. The GAC model extracts the edges in an image, the contour locks around the boundaries of the object by utilizing the gradient operator to obtain the of edge stopping function (ESF). The balloon forces are used in most of the active contours, which is difficult to design for evolving the contour. The sign determines the direction of the balloon force for evolving of the contour. The positive sign in the balloon force is for shrinking and the negative sign for expansion of the contour. The contour evolve faster for large balloon force but simultaneously also may pass through weak edges in

the images and the contour is not allowed to evolve for small balloon forces. The detection of interior and exterior boundaries fails using edge detection models. The drawback of the active contour model is the problem of local minima because the targeted object differs from defined initial contour.

Most of the disadvantages discussed in the edge based models is overcomes by Region based ACMs. Region based segmentation gives better performance of images with weak or without edges and less sensitivity to noise because information utilized is within or outside the contour. This model has the advantage to detect the internal and external boundaries of an image if the location of the initial contour is far away from the desired objects. Chan & Vese [2] implemented one popular model for the above concept “active contours without edges”. This particular model is defined to minimize the function of Mumford-Shah [3].

The most popular transform is wavelet transform [4] is a best tool for researches and non-stationary signals [5]. The typical examples for the non-stationary signals are real world images; in this paper, we considered the 2D oil spill image which is a good example. The singularities in the basic separable wavelets can give the three directions only i.e vertical, horizontal and diagonal information of an image, the quincunx wavelet transform can provide more directional information, and the one we introduced in this paper is a non-separable wavelet [6–7]. The study and analysis are reported on the 2D oil spill images using the non-separable quincunx wavelet decomposition.

We introduced a novel method for oil spill segmentation of ocean SAR images. This proposed quincunx wavelet transform using level set model algorithm is efficient for segmenting accurately the oil spills. The non-separable wavelet transform is the best tool for enhancing, compression of image by adjusting the quincunx coefficient with proper and suitable levels[8-12].

This paper is organized as follows. The background of conventional models discussed in Section II. In Section III the pre-processing step non-separable quincunx wavelet decomposition for the proposed model is described. The proposed levels set model in Section IV. Experimental results presented and discussed briefly in section V.

## 2. Literature Review of Level Set Model

In 1970's the first level set method introduced by Dervieux and Thomasset [13][14], but this work is not of much attention. After that new level set model introduced by Osher and Sethian in 1987[12], this work received popularity and research on tracking by moving interfaces in image processing, computational geometry, computer vision. These models can solve the problems in implicit manner for propagation of surface or curve. The main idea is to introduce signed function for evolving the curve; the actual contour is represented by the zero level set function on to the image. Derive an implicit surface flow when it is applied to the zero level set according to the motion equation of the contour. The most basic and important edge and region- based level set models explained in proceeding topics

### 2.1 The GAC Model: Edge based segmentation

Image segmentation, using the Geodesic Active Contour (GAC) based on connection of classical “snakes” with energy minimization, and geometric active contours depends on curve evolution. Along edges of regions, Let C is the planer curve in  $\omega$ . Let  $\Omega$  be the bounded open subset of  $R^2$  and F can be defined as  $F:[0,a] \times [0,b] \rightarrow R$  is an image in hand for segmentation. The GAC model can be represented in a equation by minimizing the energy functional as given below

$$E(c) = \int_0^1 g(|\nabla F(C(q))|) |C'(q)| dq \quad (1)$$

Whereas g is the edge stopping function as given in equation. 2. The main function of g is used to stop the evolving curve when it arrives to the objects boundaries in the given image [13] is given as below

$$g(\nabla F) = \frac{1}{1 + |\nabla G_\sigma * F|^2} \quad (2)$$

The term  $G_\sigma * I$  in the denominator describes smoothed versions of image, which is result of convolution of image  $I$  with a Gaussian kernel with standard deviation given as  $\sigma$ . The modified solution for geodesic problem of level set will give Euler Lagrange Equation by [1]

$$C_t = g(|\nabla F|)k \vec{N} - (\nabla g \cdot \vec{N}) \vec{N} \quad (3)$$

where the curvature of the contour is given as  $k$  and the normal inward to the curve is specified as  $\vec{N}$ . To speed up the propagation speed a constant velocity term  $\alpha$  is added in equation. By adding the constant velocity parameter  $\alpha$  in the above equation can be represented new one as below

$$C_t = g(|\nabla F|)(k + \alpha) \vec{N} - (\nabla g \cdot \vec{N}) \vec{N} \quad (4)$$

The problem of object detection can be solved by zero level set maintaining the flow steady state ( $\phi = 0$ ), and take  $k=0$ , and this model still converges for slower motions. The major advantage of this type of model is that it is obtained with less parameter. The corresponding level set formulation is given as below.

$$\frac{\partial \phi}{\partial t} = g |\nabla \phi| \left( \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) + \alpha \right) + \nabla g \cdot \nabla \phi \quad (5)$$

The shrinking and expanding function of Level set is steering by balloon force indicated as  $\alpha$

Choosing of edge stopping element  $g$  is crucial for the GAC model, in the case of ideal edges this particular approach of object detection via geodesic computational method does not depend on the selection of parameter  $g$ , as long as  $g$  is maintained positive, strictly decreasing function and  $g(r) \rightarrow 0$  as  $r \rightarrow \infty$ . Actually real time images does not contain ideal edges. This is an initial level (zero level set function) "edge detector" operation, which resembles to the same given in the previous active contours models. The use of this better edge detectors, as for instance energy ones can immediately improve the results.

## 2.2 The Chan and Vese (CV) model: Region Based Segmentation

Chan and Vese developed a novel model for a special case of model of Mumford-Shah functional. Let  $c_1$  and  $c_2$  are the mean pixel values of inside and outside the contour of image  $F(x,y)$  in  $\Omega$  domain. Where  $C$  is the evolving curve element in  $\Omega$  changes along with inward and outward forces with reference to the regions in image as given in below figure 1. Let  $\phi$  is the level set function with negative values outside and positive values inside the curve and zero on the curve  $C$  as depicted in figure 1

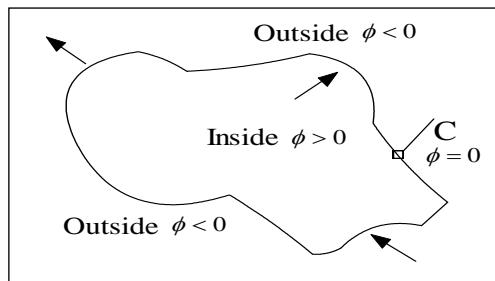


Fig 1. Propagating Curve'  $C'$  in different directions

$$\phi(x, y) = \begin{cases} 1 & \text{inside}(c) \\ -1 & \text{outside}(c) \\ 0 & \text{otherwise} \end{cases} \quad x, y \in \Omega$$

The C-V method converted into equations in terms of using minimizing energy parameter and by adding some important controlling terms as given below

$$E(c_1, c_2, C) = \lambda_1 \int_{inside(c)} |F(x, y) - c_1|^2 dx dy + \lambda_2 \int_{outside(c)} |F(x, y) - c_2|^2 dx dy, \quad x, y \in \Omega \quad (6)$$

The evolving curve and can given as  $\phi$ . Keeping term  $\phi$  fixed and minimizing the energy  $E(c_1, c_2, \phi)$  with respect to  $c_1$  and  $c_2$ , these can be expressed as below

$$c_1(\phi) = \frac{\int_{\Omega} F(x, y) \cdot H(\phi) dx dy}{\int_{\Omega} H(\phi) dx dy} \quad (7)$$

$$c_2(\phi) = \frac{\int_{\Omega} F(x, y) \cdot (1 - H(\phi)) dx dy}{\int_{\Omega} (1 - H(\phi)) dx dy} \quad (8)$$

Where  $H(\phi)$  the function is given as below

$$H(\phi) = \begin{cases} 1, & \text{if } \phi \geq 0 \\ 0, & \text{if } \phi < 0 \end{cases}$$

$H(\phi)$  is the Heaviside function greater than zero then the curve has nonempty interior in  $\Omega$  and  $1 - H(\phi)$  is greater than zero value for the curve nonempty exterior in  $\Omega$ . By including the elements length and area energy terms in the Equation. 6 and solving for purpose of minimizing them the corresponding variational level set formulation is given below

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[ \mu \nabla \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - v - \lambda_1 (1 - c_1)^2 + \lambda_2 (1 - c_2)^2 \right] \quad (9)$$

Where the controlling parameters  $\lambda_1, \lambda_2, \mu$  and  $v$  are the fixed ones such that  $\lambda_1, \lambda_2 > 0$  &  $\mu, v \geq 0$ . Where  $\mu$  and  $v$  are used for controlling the smoothness of zero level set (initial level set) function and increasing the propagation of speed respectively and  $\nabla$  is the gradient operator and  $\delta(\phi)$  is the Dirac function as given below

$$\delta(\phi) = \frac{d}{d\phi} H(\phi)$$

### 3. Non-Separable Quincunx Wavelet Decomposition

We propose a new method for detection of oil spill regions in the non-separable wavelet domain. Each denoised quincunx wavelet coefficient is measured by introducing the discrimination power (DP) for better detection of the oil spill regions in each scale. The Discrimination Power provides more information such as edge information, for region information we consider the oil spills and background. The probability measure of the wavelet coefficient belonging to the oil spills surface is calculated based upon the derived interscale relationship [16]

### 3.1 Image de-noising by LMMSE

To suppress the noise in the proposed non-separable wavelet domain, we performed the de-noising based on the Linear minimum mean squared error estimation (LMMSE) filtering applied on to the quincunx wavelet coefficients. Whenever a non- separable undecimated quincunx wavelet coefficients is applied on to the observed image  $f(x,y)$ , all singularities due to noise and image components are captured.[25]

$$T[F(x,y)=T[F'(x,y)]+T[N(x,y)] \dots \dots \dots (10)$$

Assumed the noise free quincunx wavelet coefficients are conditionally not depends on zero mean and their variances  $\sigma^2$ . The noise is assumed in this case is to be an AWGN noise with variance  $\sigma_n^2$ . Thus, compute the noise free quincunx wavelet coefficients using the formula.

$$\hat{WT}_i^j = \frac{\hat{\sigma}^2(j)}{\hat{\sigma}^2(j) + \sigma_n^2} WT_i^j \dots \dots \dots (11)$$

Where  $\hat{\sigma}^2$  represents the maximum likelihood Estimator (MSE) of local variance  $\sigma^2$  and the wavelet coefficients are computed as follows [18].

$$\hat{\sigma}^2(k) = \max \left\{ 0, \frac{1}{\#_{\Omega_k}} \sum_{\beta \in \Omega_k} (WT_i^\beta)^2 - \sigma_n^2 \right\} \dots \dots \dots (12)$$

Where  $WT_i^\beta$  represents the wavelet coefficients at scale ‘i’ location  $\beta$ ,  $\Omega_k$  are the set of coefficients in square neighborhood window centered at location k and # represents the number of elements in given set. The SD of noise  $\sigma_n$  can be calculated using the as below [16]

$$\sigma_n = \frac{\text{median}(|WT_i^j|)}{0.6745}, j=1 \dots \dots \dots (13)$$

where  $|.|$  represents the absolute value.

### 3.2 DP of the quincunx wavelet coefficients

In the wavelet domain, the singularities can be captured by the wavelet coefficients [17] of the images. The singularities in the images i.e interior regions and edges can be calculated by the wavelet coefficients  $\hat{WT}_j^i$ . The sparse and high wavelet coefficients gives the edges of the oil spills in image along with in background strong edges are present, similarly the low valued quincunx wavelet coefficients gives the singularities of the images due to the faint edges, interiors regions of the oil spills or the background. The power of the quincunx wavelet coefficients is the ‘Discrimination Power’ (DP). Quantitatively we measure the Discrimination Power for distribution of different wavelet coefficients. The Discrimination Power of quincunx wavelet coefficient can be defined as:

$$DP(w) = \int_0^w H(x)dx \dots \dots \dots (14)$$

Here  $H(x)$  is the histogram of the wavelet coefficients. The histogram is a distribution function of the modulus of the argument ‘x’. Mallat [17] found experimentally for estimation of the histogram curve formula by:

$$H(x) = ke^{-\left(\frac{|x|}{\beta}\right)^{\gamma}} \dots \dots \dots (15)$$

Where K is a constant,  $\beta$  is the measures of the variance and  $\gamma$  is peak decreasing rate. From the above definition of DP, still the discrimination of all the quincunx wavelet coefficients is a tedious task. The wavelet coefficients can be assumed to be the normal distribution of the edges and interior region of the oil spills, finally, the normal distribution of edges and interior regions can be achieved by the modified function, say  $H'$  (.), can be defined as

$$H'(x) = Ke^{-x^2} \dots \dots \dots (16)$$

i.e.  $\beta = 1$  and  $\gamma = 2$ . So, finally the modified definition of Discrimination Power (DP) of wavelet coefficients are follows

$$DP(WT_j) = \int_0^{\hat{WT}_j^i} h'(x)dx = k \int_0^{\hat{WT}_j^i} e^{-x^2} dx \dots \dots \dots (17)$$

The parameter K is constant and it is equal to  $\frac{2}{\sqrt{\pi}}$ . This value of K to make  $\int H'(x)dx = 1$ . The DP is measured in terms of error function (erf) is defined as follows:

$$DP(WT_j) = erf(\hat{WT}_j^i) \dots \dots \dots (18)$$

The DP has two main properties as follows:

First one is Fixed Range: for any given wavelet coefficient at any scale falls within the similar range of  $[-1,1]$  for measured DP after that it can be compared easily across scales.

Second property for Enhancement: The DP value is measured to not only enhance the power of the low wavelet coefficients but also suppress the power slightly for the larger wavelet coefficients.

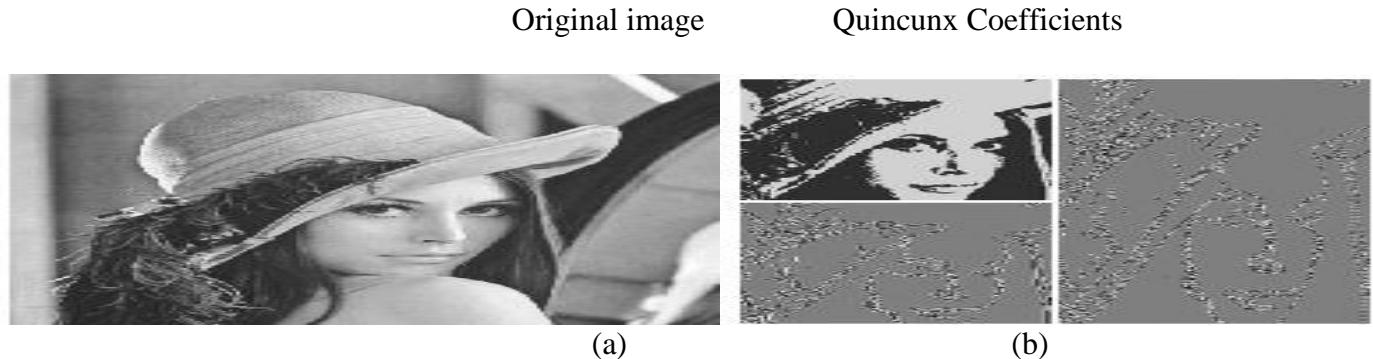


Fig2. Wavelet coefficients for the quincunx sub sampling with an example for  $J = 2$  iterations for Compact representation.

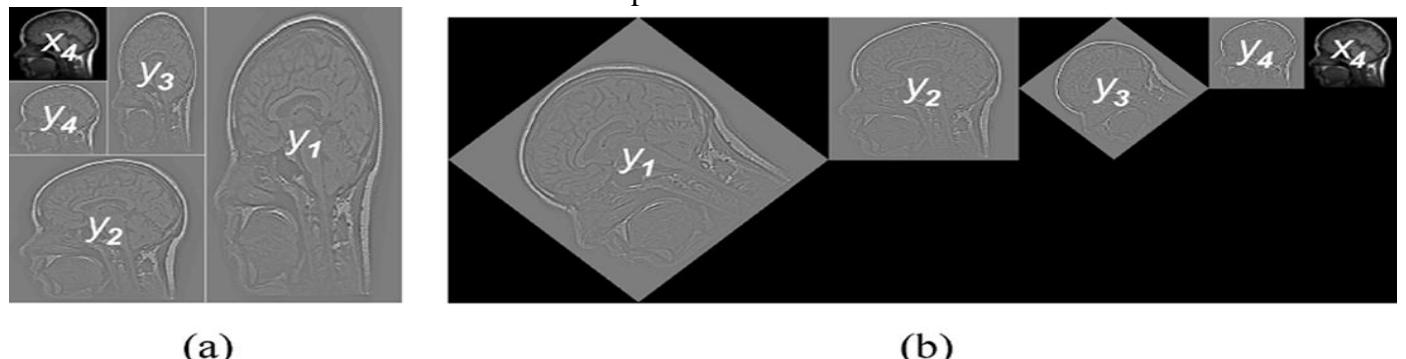


Fig 3. Wavelet coefficients can be arranged in two ways s. An example in this case for  $J = 4$  iterations. (a) Compact form.(b) Classic form

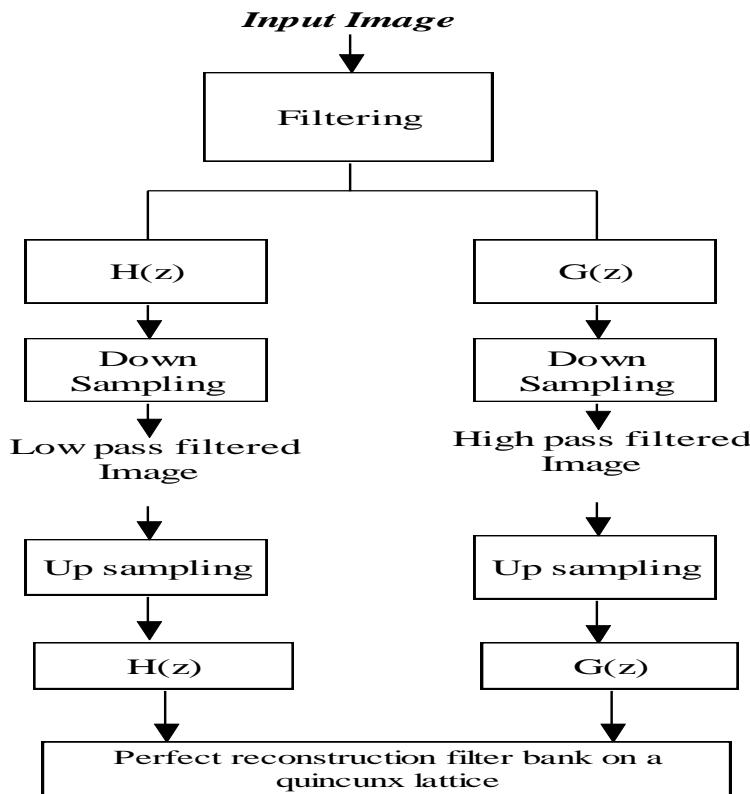


Fig.4 Quincunx wavelet decomposition and reconstruction process

## 4. Proposed Method

### 4.1 Implementing Quincunx Wavelets into Level set Model:

The non-separable quincunx wavelet decomposed images are implemented using level set model resulting in accurate segmentation and faster results. The decomposed results are of less size without loss of actual information and it gives more directional information. The same is applied with new term called SPF function. This segments the edges and boundaries of oil spill regions accurately.

The contour initialized in the region can take values in the range [-1, 1] and smaller regions within the Region Of Interest (ROI) [18]. The region function modulates according to the signs of the pressure forces inside and outside of the ROI. When the object is outside the contour shrinks and if it is inside the object, the contour expands. This external balloon forces are called signed pressure forces. This SPF function is very helpful for solving the boundary leakage problems generally appearing at weak edges.

Consider a non-separable wavelet, which makes the segmentation difficult and may get improper segmentation results. We derived two mean intensities of decomposed images are  $c_1$  and  $c_2$  in the region  $\Omega$  and the mean intensities  $c_1$  and  $c_2$  are given by the following equations.

$$C_1 = \frac{\int_{\Omega} F_{wd}(x, y) \cdot H(\phi) dx dy}{\int_{\Omega} H(\phi) dx dy} \quad (19)$$

$$C_2 = \frac{\int_{\Omega} F_{wd}(x, y) \cdot (1 - H(\phi)) dx dy}{\int_{\Omega} (1 - H(\phi)) dx dy} \quad (20)$$

Where  $F_{wd}(x, y)$  be the wavelet decomposed image of the  $F(x, y)$ . The new Signed Pressure Force (SPF) functions is designed based on the above analysis[18] as follows

$$SPF(F_{wd}(x, y)) = \frac{F_{wd}(x, y) - \frac{c_1 + c_2}{2}}{\max\left(\left|F_{wd}(x, y) - \frac{c_1 + c_2}{2}\right|\right)} , \quad x, y \in \Omega \quad (21)$$

The final new level set formula based on wavelet can be derived, substitute equation 21 in to equation 5 as follows

$$\frac{\partial \phi}{\partial t} = spf_{wd} |\nabla \phi| \left( \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) + \alpha \right) + \nabla spf_{wd} \cdot \nabla \phi \quad (22)$$

For segmenting the given oil spill images the parameter  $\alpha$  is adaptively changes with respect to  $spf_{wd}$  [19-25].

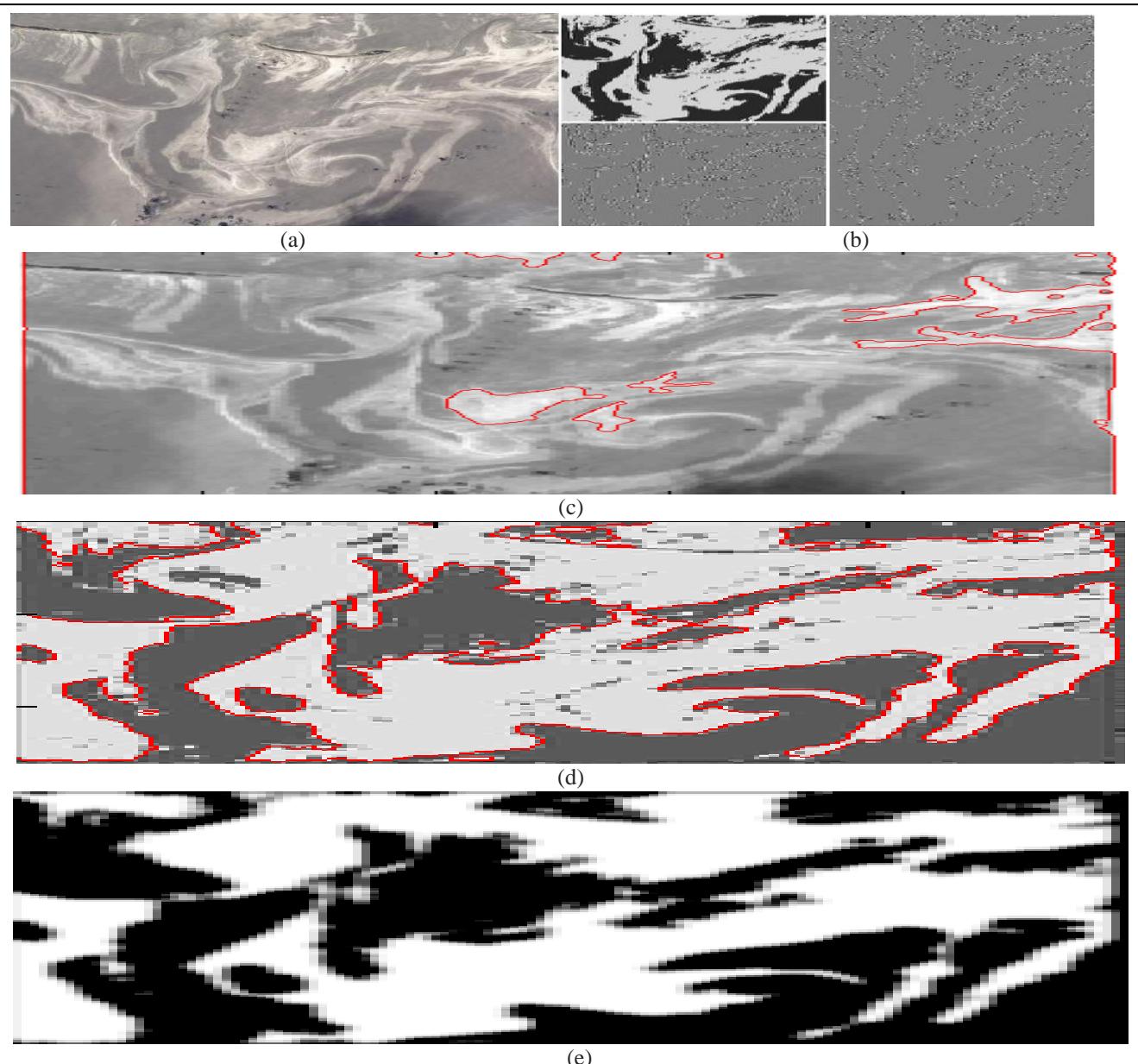


Fig.5. Results obtained with conventional and proposed methods: figure (a) is the original image, (b) is the non separable quincunx wavelet coefficients with '2' iterations (c) and (d) is the oil spill detection using conventional active contour model and proposed model respectively and last figure (e) represents the segmented image with proposed wavelet based level set method.

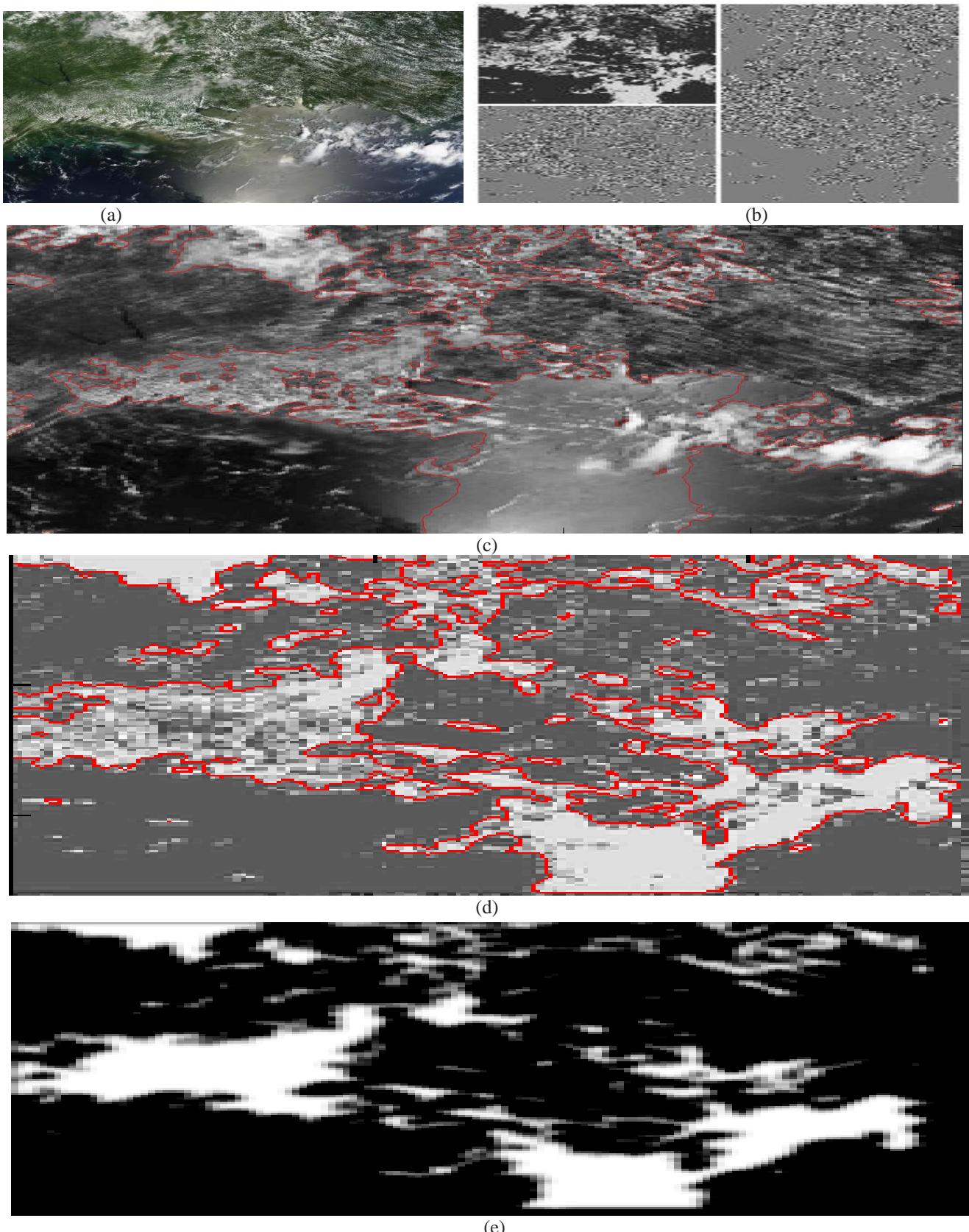


Figure 6. Results obtained with conventional and proposed algorithms: figure (a) the original image, (b) is the non separable quincunx wavelet coefficients with '2' iterations (c) and (d) is the oil spill detection with conventional active contour model and proposed model respectively and figure (e) depicts the segmented image with proposed wavelet based level set method

## 5.Simulation Results and Discussion

Implementation of suggested algorithm is performed on two unique images for segmenting oil spills regions of SAR images. These images are taken from data set of NASA earth observatory. The proposed results are analyzed and compared with the C-V active contour model as shown in figure 5 and 6. In figure 5 and 6 Original oil spill images are figure (a). Figure (b),(c), (d) and (e) are quincunx wavelet coefficients, previous method, proposed method and its final segmentation respectively.

All of SAR images shows that area of oil spills appear as white(brighter) areas in the Synthetic Aperture Radar(SAR) images[17][18], because the oil dampens the capillary waves on the sea surface. The original oil spill images were rescaled to 256x256 for contour evolution of the both proposed and conventional methods for preprocessed images.

The edge indicator function and contour fitting energy are taken into consideration from the conventional C-V method and results shown in figure (c) of figures 5 and 6. The conventional algorithms [19][20] captures wide area (which is not required) or fail to capture the needed region within the given input image, the proposed method can solve the above mentioned problems. Taking into consideration components of the quincunx wavelet coefficients the proposed algorithm of contour evolution is designed with non separable quincunx wavelet decomposition. Figure (d) in the above figures shows the results of the proposed work done. It shows clearly that the contour in the proposed method perfectly locks the specified areas with oil spills in the images .

The computed area which is covered by the contour is the exact region within the input image which indicates the spread of oil spills in the given image . By utilizing conventional C-V algorithm, there are few disadvantages namely, inefficient curve evolution which shows excess area; Cannot find edges of oil spills in the given images and apart from the above stated the conventional method requires assigning zero level contour and particular  $\mu$  value based on the image texture.

The suggested method works with a initial  $\mu$  value taken as 50. The initial contours for the two images are square shaped intial contour functions defined from 1 to (N-100)th pixels of an NxN sized input image. The desired output was obtained within the range of 70 – 100 iterations. The main advantage of the proposed method is the time required to compute the area coverage. With each iteration less time was required for same purpose as compared with previous methods. The proposed method the evolving contour can also penetrate in to the areas of oil spill where conventional methods fails to do so. For calculating the segmented area, after segmentation the contour maintains the values less than zero or -1. Accurate area calculated with the help of manually designed contour for the object boundaries.

This experimentation is done on Matlab 7.8 with RAM of 1GB and 3Ghz processor .To observe the accuracy and efficiency of the proposed model or algorithm, the performance measures computed in terms of area error measure and accurate area segmented is tabulated. Table 1 shows clearly the segmented area for specific mu values chosen to obtain desired results using conventional and proposed methods. Area covered by the predicted contour for both the algorithms is measured and evaluated the performance on the basis of area error computed. The accurate area of the oil spill calculated by manually defining the contour around the area of oil spill, there by areas of all the images is calculated, and shown in table 2.

The figure 7 depicts the graph of the performance of final contour segmented areas of different images, these values from table 2. The proposed algorithm covered more segmented area compared with previous algorithm. The graph is plotted based on the area error measure for conventional and proposed models; it is clearly visible that the proposed algorithm area error is less when compared to conventional model. It means that proposed model accurately segments the oil spills regions compared to conventional. The bar plot are as depicted in the Figure. 8 the first bar represent the area error measurement of the proposed with that of the accurate by taking initial  $\mu$  value 50.

**Table 1.** Comparative statement for time needed for evolution, number of iterations and error measure of area.

Images	conventional method			proposed method				
		iterations	evolution time	area(error)		iterations	evolution time	area error
Image1	$\mu=50$	200	25.32098sec	9861 mm <sup>2</sup>	initial $\mu=50$	100	10.1207sec	8798mm <sup>2</sup>
Image 2	$\mu=50$	150	23.334915sec	22350mm <sup>2</sup>	initial $\mu=50$	70	4.5894sec	5634mm <sup>2</sup>

**Table 2.** Comparative analysis of segmented area measures of different images

Image	Conventional	proposed
Image 1	14475	15538
Image 2	1986	18702

The other bar in the same figure describes the area of the conventional methods with  $\mu$  value 50 respectively. Here we can conclude that by taking a larger  $\mu$  value the estimated error is minimized. But if it is increased further, it fails to locks the required regions of interest and takes large time periods to evolve. For the proposed method after selecting the initial value the  $\mu$  value goes on changing as contour evolves.

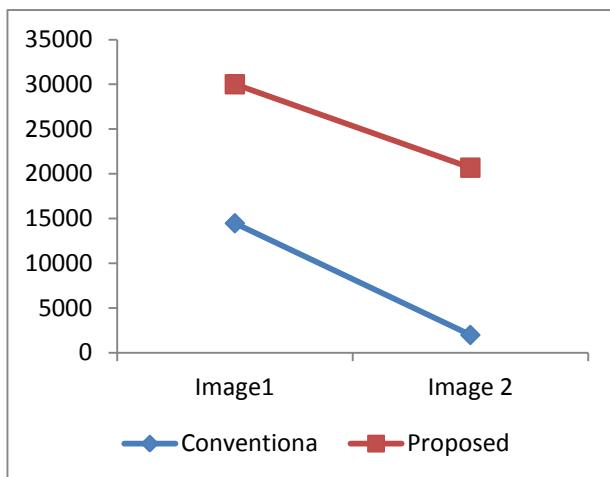


Figure 7. Performance of segmented area covered

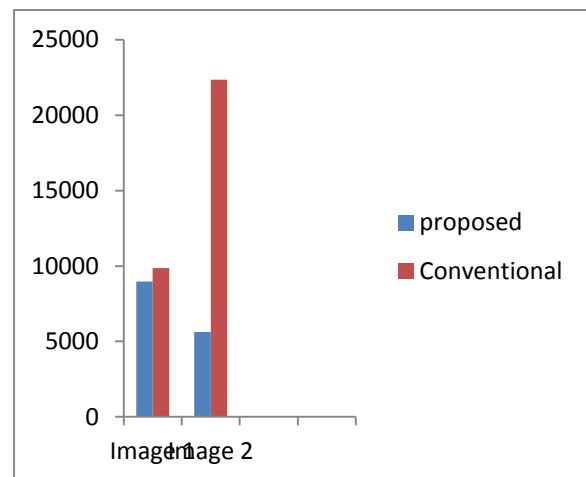


Figure 8. Area error measures for  $\mu_{ue} = 50$

## 6. Conclusion

We introduced a novel algorithm for the purpose of image segmentation of oil spills regions with level set function based on non-separable wavelet transform. Implemented for better computation of the interior regions and for finding the edges of images for oil spills applications. This method depends on a single threshold and is independent of the gray level profile of the image, overcomes inhomogeneities of the images, and is useful for finding the precise oil spill area at a faster rate with the proposed level set model.

## References

- [1] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours, in: Processing of IEEE International Conference on Computer Vision'95, Boston, MA, 1995, pp
- [2] T. Chan, L. Vese, Active contours without edges, *IEEE Transaction on Image Processing* 10 (2) (2001) 266–277.
- [3] D. Mumford, J. Shah, Optimal approximation by piecewise smooth function and associated variational problems, *Communication on Pure and Applied Mathematics* 42 (1989) 577–685.
- [4] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (7) (1989) 674–693.
- [5] O. Rioul, M. Vetterli, Wavelets and signal processing, *IEEE Signal Process. Mag.* 8 (4) (1991) 14–38.[24] M. Daszykowski, et
- [6] J. Kovacevic, M. Vetterli, Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for  $R^n$ , *IEEE Trans. Inf. Theory* 38 (2) (1992) 533–555
- [7] T. Chen, P.P. Vaidyanathan, Recent developments in multidimensional multirate systems, *IEEE Trans. Circuits Syst. Video Technol.* 3 (2) (1993) 116–137
- [8] J. Kovacevic, M. Vetterli, Nonseparable two- and three-dimensional wavelets, *IEEE Trans. Signal Process.* 43 (5) (1995) 1269–1273.
- [9] G. Aubert, P. Kornprobst, Mathematical Problems in Image Processing : Partial Differential Equations and the Calculus of Variables, Springer, New York, 2002.
- [10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2007.
- [11] C.Y. Xu, A. Yezzi Jr., J.L. Prince, On the relationship between parametric and geometric active contours, in: Processing of 34th Asilomar Conference on Signals Systems and Computers, 2000, pp. 483–489.2002.
- [12] S. Osher and J. Sethian, “Fronts propagating with curvature-dependent Speed: Algorithms based on Hamilton-Jacobi formulations,” *J. Computer. Phys.*, vol. 79, no. 1, pp. 12–49, Nov. 1988.
- [13] A. Dervieux and F. Thomasset, “A finite element method for the simulation of Rayleigh-Taylor instability,” *Lecture Notes Math.*, vol. 771, pp. 145–158, 1980.
- [14] A. Dervieux and F. Thomasset, “Multifluid incompressible flows by a finite element method,” *Lecture Notes Phys.*, vol. 141, pp. 158–163, 1980.
- [15] Caselles, V., Catte, F., Coll, T., and Dibos, F. 1993. A geometric model for active contours. *Numerische Mathematik*, 66:1–31.
- [16] Ratnesh Singh Sengara,, Ashutosh Kumar Upadhyaya, Manjit Singha, Vikram M. Gadreba “ Analysis of 2D-gel images for detection of protein spots using a novel non-separable wavelet based method” *Biomedical Signal Processing and Control* 25 (2016) 62–75. *Numerische Mathematik*, 66:1–31.
- [17] S. Mallat, W.L. Hwang, Singularity detection and processing with wavelets, *IEEE Trans. Inf. Theory* 38 (2) (1992) 617–643
- [18] Kaihua Zhang , Lei Zhang , Huihui Song , Wengang Zhou “Active contours with selective local or global segmentation: A new formulation and level set method “*Image and Vision Computing* 28 (2010) 668–676.
- [19] G. Raghatham Reddy, Syed Zaheeruddin, B. Narsimha and R. Rameshwar Rao “Illumination-Reflectance Based Novel approach for Level Set Evolution” proc of 2011 IEEE International Conference on Electro/Information Technology ( IEEE EIT- 2011), pp. 1-5.
- [20] Malladi, R., Sethian, J.A., and Vemuri, B.C. 1994. Evolutionary fronts for topology independent shape modeling and recovery. Proc. of the 3rd ECCV, Stockholm, Sweden, pp. 3–13.
- [21] H. S. Solberg, G. Storvik, R. Solberg, and E. Volden (1999): Automatic Detection of Oil Spills in ERS SAR Images, *IEEE Transactions on Geoscience and Remote Sensing*, 37, 4, 1916-1924
- [22] ESA-ESRIN Frascati, Italy, January 14–16 2004, available at <http://earth.esa.int/polinsar/pr.html>.
- [23] Gade, M., Scholz, J., & von Viebahn, C. (2000). On the detectability of marine oil pollution in European marginal waters by means of ERS SAR imagery. Proc. IGARSS 2000, vol. 6 (pp. 2510– 2512).
- [24] Curlander, J. C., & McDonough, R. N. (1991). Synthetic aperture radar, systems and signal processing. Wiley series in remote sensing. John Wiley & Sons.
- [25] G.R Reddy, K.Ramudu, A.Srinivas and Rameshwar Rao (2016). “Segmentation of medical Imagery using wavelet based active contour model” IJAER, Vol 11, No.5, pp no.3675-3680,ISSN:0973-4562.

## A Fast Priority-Flood Algorithm with Pruning for Depression Filling in Hydrologic Analysis

Ruchi Lawaniya, Manish Pandey

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

**Abstract—** The Digital elevation models are widely used spatial data source to incorporate the topographic information within geographical and hydrological applications. Depressions in DEM are lower areas surrounded by surface without any outlet. They interrupt or disconnect the flow path and create inaccurate drainage pattern. Subsequently, recognizing and removing the depression is a vital necessity for any hydrological study, which is commonly done prior to the use of DEM to conduct the hydrologic analysis. Usually, handling the depressions is a time consuming task for applications of huge terrain dataset with high resolution. This paper presents an improvement on priority-flood algorithm for recognizing and processing the depressions based on gridded digital elevation model in digital terrain analysis. The improvement on previous method is done by introducing a novel concept of pruning the dead cells from the priority queue. The priority queue cells that will never be used for further computation are considered as dead cells. Pruning of the dead cells can reduce the number of cells in the priority queue. Thus, the overall running time of *Insertion* and *Deletion* operations within the priority queue is asymptotically decreased. The proposed *PriorityFloodPruning* algorithm runs in  $O(K \log_2 K)$  time, where K is the number of cells present in the priority queue after pruning. The proposed *PriorityFloodPruning* algorithm shows 1.13x to 1.25x speedup.

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

### I. INTRODUCTION

The DEMs are digital representation of altitude of continuous ground surface in raster form. They are stored as a rectangular array where value of each cell represents a landscape height regarding any reference datum. The DEMs are widely used to derive relief features such as gradient and slope direction, compute flow simulation path, drainage pattern, watershed features and sub-catchment areas in hydrology. The automatic derivation of drainage network over a DEM is an essential requirement of many applications of GIS such as hydrologic analysis, pollution diffusion analysis and land erosion etc. To automate extraction of the fully connected drainage pattern in the raster environment, the flow path from every cell must be directed to an outlet on the border of DEM. But, the presence of spurious depressions in DEM interrupts and disconnects the flow path. The depressions are the lower elevation areas without outlet, they consist one or more connected cells of equal height which are surrounded by higher elevation cells of topographic surface. In the hydrological and geomorphic applications, identification and removal of the depressions is completed prior to the use of the DEM.

In the hydrology community, there has been a significant measure of innovative work that concentrates on in utilizing the raster environment for handling the depressions and extracting the flow path such as elevation-smoothing method [14], pit-filling algorithms [3], carving method [17], breaching method [11], [8] and hybrid algorithm combining procedures of carving and depression filling [18].

This paper proposed an improvement on the priority flood algorithm by introducing a new concept of pruning the dead cells from the priority queue. The DEM cells, whose neighboring cells are already traversed or processed, are considered as dead cells. As the computation of least-cost search progresses for flow path, the number of dead cells is increased. From these dead cells, new cells can't be grown. So there is no longer need to remain in priority queue. The removal of these cells decreases the number of cells in priority queue. The memory consumption and computation time required for insert, find minimum, delete minimum operations of priority queue is decreases. Thus, the computation time of *PriorityFloodPruning* algorithm  $O(K \log_2 K)$ , where K is the less than the total number of cells in DEM.

## II. LITERATURE SURVEY

Topography specifically impacts the spatial variation of land surface. The organization of topographic data within geographic applications has been accomplished using digital elevation method models (DEMs). DEMs are most generally used in hydrologic and geologic analysis, natural resources exploration, hazard observation, agricultural management and so forth. The DEMs are computerized representation of ground surface height with respect to typical base level usually for a planet (Earth, Moon or Asteroid) in raster structure. The DEMs are produced using height information from several points which may be spaced at regular or irregular interval. In early days, DEMs were used to be created from the contour data in the satellite images or topographic maps by using the interpolation operations. With the advancement of technology, today high resolution elevation datasets for a large surface area become available which are produced by remote sensing (ESRI, 2005) [2], [24] i.e. Interferometric Synthetic Aperture Radar (InSAR) and LIDAR techniques.

DEMs are the end result of numerous data collection, interpolation and processing steps (Kreveld et al.) [13]. The inherent data errors found during data capturing, interpolation defects during DEM generation, averaging of elevations within cells, smoothing effects caused by resampling or truncation of interpolated values to lower precision etc impact DEM quality and introduce the artificial or spurious depressions. The depressions comprise one or more than one connected cells of same height which are surrounded by higher elevation cells of land surface. The frequent presence of spurious pits in DEM interfere the flow path and create inaccurate and inward drainage network (Tarboton, 1997) [21].

Prior to use of the DEM in hydrologic studies, preprocessing of the DEM is completed to rectify the depressions and flat regions. In the hydrology community, a number of algorithms have been carried out in the raster environment over the past decade to preprocess the depressions and determine flow network. These algorithms can be categorized into two classes as follows:

### A. Conventional algorithms for depression filling

O'Callaghan and Mark [14] developed a smoothing method to predict the drainage pattern and watershed delineation in hydrological applications. The smoothing operation alters the elevation values of all cells of grid which causes the loss of information in non-problematic areas of DEM and interfere the originality of DEM.

Jenson and Domingue [3] have presented a most common method that is suitable for complex depressions and flat areas in the raster environment. Now, the algorithm (abbreviated to J&D algorithm) has been implemented in different types of geographic information system (GIS) software, such as RiverTools, ArcGIS and others. To handle the depressions, the J&D method includes two phases: In initial phase, fills the single cell depressions by increasing the altitude values to the pour point. Furthermore, in second phase, handle the complex depressions containing more than one cell. The running time of this method is  $O(N^2)$  that is not considerable for large area dataset. The main issue is that looping depressions located on a flat surface can't be rectified in this algorithm.

Martz and Garbrecht [9], [10], [11] presented a breaching method whereby the height value of one or two cells selected at the outlet of the spurious depression, is lowered to reduce the size of the depression.

Planchon and Darboux [15] have proposed an algorithm (abbreviated to P&D algorithm), that was initially designed to analyze the microrelief features of soil surface. Rather than filling the depression, this algorithm depends on two stages; first water-covering stage in which assign the infinite altitude to all DEM cells with the exception of the boundary cells and second is water-draining stage in which iteratively decrease the elevation of each DEM cell to guarantee that there is a flow path for every cell to outlet on the boundary of DEM. Its practical running time is asymptotically proportional to  $N^{1.2}$  and the hypothetical calculation time is asymptotically relative to  $N^{1.5}$  in all cases, except for some extraordinary ones with no practical interest.

An alternative method called carving was described by Soille et al. [17] in which every single pit is suppressed by making a descending path from it to the closest point having a minimum elevation value. Through a progressively flooding the DEM from all relevant stream outlets, carving paths are determined.

Soille et al. [18] have introduced a hybrid method where carving procedure is applied in conjunction with the depression filling procedures to positive and negative adjustments to be used to handle the depressions. By combining the pit filling and carving approaches, the cost of transforming an input DEM into pitless DEM is minimized.

Wang et al. [23] described a new method based on heuristic search strategy to extract the flow path. In this method, the outlet is searched using the heuristic information to remunerate insufficient searching information of different methods.

## B. Least cost search algorithms for depression filling

Ehlschlaeger [1] was the first to suggest the least-cost search algorithm to find the shortest path between a start location and a goal location. This method employs the  $A^t$  search algorithm that mimics the way a manual interpreter would mentally determine the flow path of water and location of drainage patterns.

Wang et al. [6] presented a priority-flood method to efficiently processing of the depression in large DEM by introducing the concept of spill elevation and utilizing the priority queue into the least-cost search algorithm. The running time of this method is  $O(N \log_2 N)$  where  $N$  is the total number of cells in DEM.

A variant of Wang et al. algorithm has been given by Liu et al. [7] whereas a sorted dictionary is used to hold equal elevation cells and unique elevation cells are processed by priority queue. This algorithm has  $O(N \log_2 k)$  time complexity where  $k$  is number of unique elevation cells of DEM.

Metz et al. [5], [12] described least cost path (LCP) method using a total ordered priority queue to extract the drainage network and watersheds analysis. This method yields accurate location of flow accumulation streams which are validated against GPS field control point and reference points digitized from satellite images for large DEMs.

Richard et al. [19] presented an improvement on the priority-flood algorithm by including the FIFO queue concept to process the depressions. To determine which cell is next to being flooded, this method work out by using a priority queue floods DEMs inwards from their boundary cells. The processing time of this method is defined as  $O(M \log_2 M)$  asymptotically, where  $M$  does not exceed the number of cells  $N$ .

## III. PROPOSED WORK

### A. Overview

Priority Flood algorithm proposed by Richard et al. [19] works in two phases. In the first phase of method, the edge cells of DEM are inserted into a priority queue with increasing order of elevation values and the cells are marked as processed. In second phase, the head of the priority queue  $C_x$  is removed first and all the adjacent neighbors which have not been processed are traversed. There are following three possible cases.

- i. If the adjacent neighbor is not traversed and its elevation is higher than the current cell  $C_x$  (under processing), it is inserted into the priority queue.
- ii. If the elevation value of unprocessed adjacent cell is equal to the elevation of current cell  $C_x$  than it is inserted into the FIFO queue.
- iii. If the adjacent neighbor is not traversed and its elevation is lower than current cell  $C_x$ , the adjacent cell's elevation is increased up to the current cell and then inserted into the FIFO queue.

The process is continued, until both priority queue and FIFO queue are empty (Fig.1).The running time of the algorithm mainly depends on the *Insert* and *Remove* operations of the priority-queue (Coremen, 2001) [20]. So, the time complexity of this method is  $O(M \log_2 M)$ , where  $M$  is the number of priority queue cells and it is much less than the conventional algorithm.

### B. Proposed *PriorityFloodPruning* algorithm for handling the depressions

In contrast but independent to the work proposed by Priority Food algorithm, the work proposed in this paper systematically eliminates the unwanted cells in the priority queue so that the effective count of cells in the priority queue remains less throughout the computation process. The proposed method tries to reduce the computational time of priority-flood algorithm by reducing the number of elements in the priority queue. One possible way to reduce the number of elements in the priority queue is to prune the dead-cells.

Fig.1: Pseudo code of Priority Flood algorithm

---

Input: A Raster Gridded DEM  
Output: Filled and Depressionless DEM

1. Let  $Pqueue$  is a priority queue
2. Let  $Fqueue$  is FIFO queue
3. Let  $Flag$  is a Flag matrix of same dimensions( $n \times n$ ) as that of DEM
4. Let  $Flag$  is initialized with false
5. for all edge cells ( $e$ ) of DEM {
6.     Insert(  $e$ ,  $Pqueue$  )
7.      $Flag(e) = true$
8. }
9. while ( ! (Empty( $Pqueue$ ) && Empty( $Fqueue$ ))) {
10. if ( ! Empty( $Fqueue$ ))
11.      $p = Remove(Fqueue)$
12. else
13.      $p = Remove(Pqueue)$
14. for all  $n = 8$ -adjacent cells of  $p$  {
15.     if ( $Flag(n) == true$ )
16.         continue
17.     else {
18.         if ( $ELV(n) \leq ELV(p)$  )
19.              $ELV(n) = ELV(p)$
20.             Insert( $n$ ,  $Fqueue$ )
21.              $Flag(n) = true$
22.         else
23.             Insert( $n$ ,  $Pqueue$ )
24.              $Flag(n) = true$
25. }}

---

As the computation progresses, the neighboring cells of center cell are inserted into priority queue or FIFO queue. These neighboring cells may be processed or unprocessed. The cells whose 8- neighboring cells have been already processed, no need to remain in the priority queue because they will never be used in further computations. New cells cannot be grown from these cells. These cells are known as dead-cells. The pruning of dead-cells from priority queue will decrease the elements in the priority queue and therefore, reduce the overall computation time and memory requirements. The flow chart in Fig.2 illustrates the entire procedure.

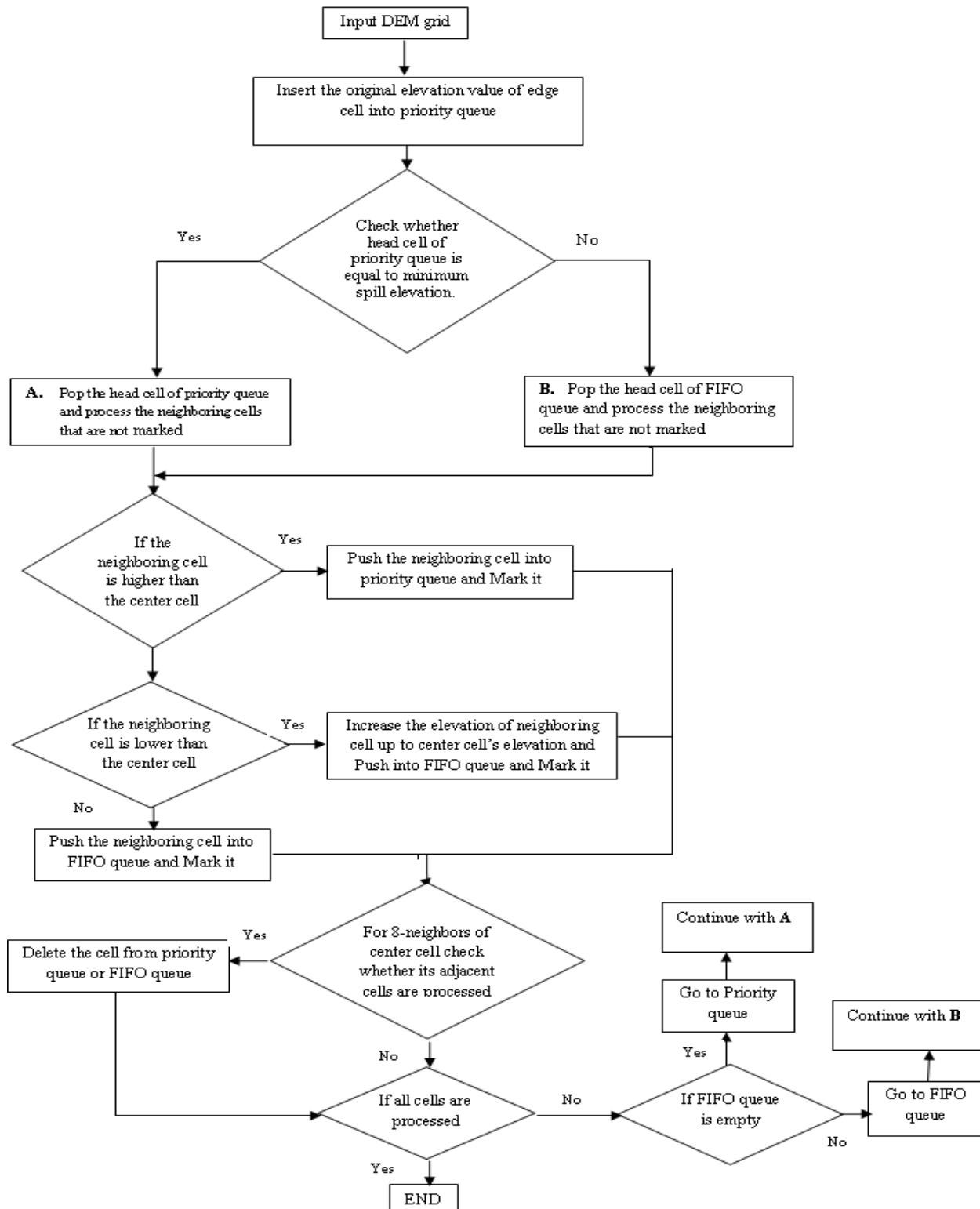


Fig.2: Data flow of *PriorityFloodPruning* algorithm

Let us assume for the sake of simplicity a square DEM of  $n \times n = N$ , where the number of rows and columns contain  $n$  cells each.

The number of edge cells,  $N_e = 4\sqrt{N} - 2$

The number of priority queue cells =  $N_x$ ,

The number of FIFO queue cells =  $N_y$

The number of pruned cells =  $N_p$

The computational procedure of *PriorityFloodPruning* is summarized into three steps:

*Step 1:* In initial step, start with handling of edge cells  $N_e$  by inserting them into a priority queue. To create a depression less DEM, every grid cell has a down slope path to an outlet of DEM. The edge cells of DEM are regarded as the potential outlets. We start processing with the outlet on boundary and progressively expand the flow path from these outlets to interior cells using upstream strategy.

*Step 2:* Next, processing of priority queue cells  $N_x$  and FIFO queue cells  $N_y$  is done to handle the depressions and flat surfaces. Here, the head cell of priority queue is removed as center cell  $C_x$  and the adjacent cells of  $C_x$  are processed by inserting them into priority queue or FIFO queue. The cells of depressions or flat surfaces need not to sort because they have same elevations, and thus such cells can be inserted into FIFO queue.

*Step 3:* The final step includes the pruning of dead cells  $N_p$  from priority queue or FIFO queue to reduce the number of elements in priority queue. The cells whose 8-neighboring cells have been already processed, no need to remain in the priority queue because they will never be used in further computations. New cells cannot be grown from these cells. These cells are known as dead-cells. After subsequent iterations, such cells become dead cells. So there is need to remove these cells from priority queue before the next iteration starts.

The total number of cells which needs to be processed =  $N - N_p$

$$= N_x + N_y - N_p$$

The priority queue has the distinct elevation value for each cell except the edge cells. Each priority queue cell is either process or pruned. And each FIFO queue is also either process or pruned. To process FIFO queue cells, O(1) time is required that is not considerable. So the time complexity depends on processing of priority queue cells.

$$= N_x - N_p$$

= K              i.e. K - elements are remained after pruning

and these cells will be used for further computation time.

### Computational procedure and Pseudo code

The detailed computational steps are described into pseudo code of Fig.3. The priority queue is declared as *Pqueue* and FIFO queue is declared as *Fqueue*. The cells in priority queue are stored with their elevation values and their location including row number (n) and column number (n) in DEM. A flag matrix is declared as the same dimensions ( $n \times n$ ) of DEM of  $N$  cells. The Flag value of all cells is initialized to false. When the cell has been processed or inserted into *Pqueue* or *Fqueue*, the flag value of that cell is marked as true. The *Insert* and *Remove* operations are used to add or delete the cells from *Pqueue* and *Fqueue*.

The *Pqueue* is initialized in *for* loop of line 5 to 8 of pseudo code (Fig. 3). The edge cells of DEM are inserted in the *Pqueue* and marked with *true* in the *Flag Matrix*. The elevation values are arranged according to the min heap property. In line 9 to 13 the head cell of *Pqueue* or *Fqueue* is considered as center cell and removed.

Fig.3: Pseudo code of *PriorityFloodPruning* Method

---

Input: A Raster Gridded DEM  
Output: Filled and Depressionless DEM

```

1  Let Pqueue is a priority queue
2  Let Fqueue is FIFO queue
3  Let Flag is a Flag matrix of same dimensions ( $n \times n$ ) as that of DEM
4  Let Flag is initialized with false

    //Initialization of priority queue
5  for all edge cells ( $N_e$ ) of DEM {
6      Insert( e, Pqueue)
7      Flag(e) = true
8  }

//processing of depressions and flat surface
9  while ( ! (Empty(Pqueue) && Empty(Fqueue))) {
10     if( ! Empty(Fqueue))
11         p = Remove(Fqueue)
12     else
13         p = Remove(Pqueue)

14     for all n = 8-adjacent cells of p {
15         if(Flag(n) == true)
16             continue
17         else {
18             if ( ELV(n)  $\leq$  ELV(p) )
19                 ELV(n) = ELV[p]
20                 Insert(n, Fqueue)
21                 Flag(n) = true
22             else
23                 Insert(n, Pqueue)
24                 Flag(n) = true
25         }
26     //Pruning the cells
26     for all n = 8-adjacent cells of p {
27         if all 8-adjacent cells of n == true{
28             If(n  $\notin$  Pqueue)
29                 Remove(Pqueue, n)
30             If(n  $\notin$  Fqueue)
31                 Remove (Fqueue,n)
32         } }
32 }
```

---

In line 14 to 25, the 8-adjacent cells of center cell are examined. If the elevation value of neighboring cell is higher than the center cell elevation, the cell is inserted into *Pqueue* and marked *true* in *Flag matrix*. If the elevation value of neighboring cell is equal to the elevation of current cell being processed, then it clearly indicates that it is a flat surface and the cell is inserted into the *Fqueue* and the corresponding entry in the *Flag Matrix* marked *true*. If the elevation value of neighboring cell is lower than the elevation of current cell being processed, then it is increased up to the current cell elevation and inserted into the *Fqueue*. Its corresponding entry in *Flag matrix* is set to *true*. In lines 26 to 32 of pseudo code, for each adjacent cells of current cell, check whether its neighboring cells are already processed or not, if yes, then no need to examine them. All such cells can be removed from *Pqueue* and/or *Fqueue*. The process is continued until all the adjacent cells are examined. The following Fig. 4 illustrates the process of depression filling using example. The difference of elevation value between the original DEM and depressionless DEM shows the distribution of land surface.

*PriorityFloodPruning* algorithm uses two types of data structures:

- i. Priority queue
- ii. FIFO queue

The priority queue is implemented using min-heap structure. The fundamental operations required in the method are find minimum, find dead cell, insertion in priority queue or FIFO queue, deletion of minimum elevation cell and dead cell from priority queue or FIFO queue.

*Step 1:* The priority queue is initialized with the edge cells using BuildHeap method. The number of edge cells  $N_e = 4\sqrt{N} - 2$  in worst case. The computation time for initialization of priority queue=  $(O(\sqrt{N}) = O(K)$  in worst case, K is much larger than the number of edge cells  $\sqrt{N}$ .

47	36	35	43	50	67
50	37	29	39	52	69
56	41	30	32	44	59
63	50	37	28	32	45
71	61	51	36	29	38
80	72	62	46	33	37

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

47	36	35	43	50	67
50					69
56					59
63					45
71					38
80	72	62	46	33	37

1	1	1	1	1	1
1					1
1					1
1					1
1					1

(a)

(b)

47	36	35	43	50	67
50					69
56					59
63					45
71			36	33	38
80	72	62	46	33	37

1	1	1	1	1	1
1					1
1					1
1					1
1			1	1	1

47	36	35	43	50	67
50					69
56					59
63			33	33	45
71			36	33	38
80	72	62	46	33	37

1	1	1	1	1	1
1					1
1					1
1			1	1	1
1			1	1	1

(c)

(d)

47	36	35	43	50	67
50					69
56		33	33	44	59
63		37	33	33	45
71		51	36	33	38
80	72	62	46	33	37

1	1	1	1	1	1
1		1	1	1	1
1		1	1	1	1
1		1	1	1	1
1		1	1	1	1

47	36	35	43	50	67
50	37	33	39	52	69
56	41	33	33	44	59
63	50	37	33	33	45
71	51	36	33	38	
80	72	62	46	33	37

1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1

(e)

(f)

47	36	35	43	50	67
50	37	33	39	52	69
56	41	33	33	44	59
63	50	37	33	33	45
71	51	36	33	38	
80	72	62	46	33	37

1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1

47	36	35	43	50	67
50	37	33	39	52	69
56	41	33	33	44	59
63	50	37	33	33	45
71	61	51	36	33	38
80	72	62	46	33	37

1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1

(g)

(h)

*Step 2:* The minimum elevation cell in priority queue can be found in  $O(\log_2 K)$  time because after finding the element adjustment of values (Heapify) will be done. An adjacent cell of  $C_x$  is inserted into K - elements priority

queue,  $O(\log_2 K)$  comparison are required. And minimum elevation cell is removed from priority queue using  $\log_2 K$  comparisons. So both the insert operation and the delete operation of the priority queue take  $O(\log_2 K)$  time. In the FIFO queue, both insert and remove operations on a DEM cell take the  $O(1)$  time.

The time for processing of all DEM cells is calculated on the basis of computation time to perform insert and remove operation on priority queue elements. The K number of cells currently in priority queue will always be lower than the total number of cells N in DEM. To process the depressions and flat surfaces, will require  $O(K \log_2 K)$  where  $K = N_x - N_p$  i.e. the number of cells in priority queue after pruning

*Step 3:* To check whether the cell is dead cell or not, 8- comparisons are required. And pruning of a dead cell from priority queue is required  $O(\log_2 K)$  time. And pruning of cell from FIFO queue is required  $O(1)$  time.

The overall computation time = Initialization of priority queue + Processing of DEM cells + Pruning of dead cells

$$= O(K) + O(K \log_2 K) + O(\log_2 K)$$

However, the time complexity of this method is  $O(K \log_2 K)$  in worst case, where  $K = N_x - N_p$  i.e. the number of cells in priority queue after pruning. The running time of Priority Flood algorithm is  $O(M \log_2 M)$ , where  $M = N_x$  i.e. the number of priority queue cells. In comparison with the processing time of Priority Flood algorithm, the proposed algorithm is slightly better in practice.

#### IV. RESULTS

In this section the performance analysis of proposed *PriorityFloodPruning* algorithm is described. The configurations of system used in the experiment are given below along with the experimental results.

##### A. Dataset

In the present study, the four different DEM areas are used to performance analysis of algorithms. The test areas are sampled from Bhopal (Madhya Pradesh), Shimla (Himachal Pradesh), Kedarnath (Uttarakhand) and Mount Abu (Rajasthan) DEM. The Cartosat DEMs are used as dataset with 1 and 3 arc-sec (30m and 90m spacing at equator) which are produced by ISRO. The ASTER GDEM is also used as best available option for the study area. The locations and topographic descriptions of test areas are listed in Table I. To provide the quantitative description of land surface, the Minimum Elevation, Maximum elevation and Local relief (range) are included.

Table I: Topographic description of test areas

Data Set	Regions	Location	Data sources	Grid spacing	Min Elevation	Max Elevation	Local relief
A	Bhopal	Central part of India	ASTER GDEM	30m	311	600	456
B	Kedarnath	Mandakini river valley in India	ASTER GDEM	90m	210	7817	4013
C	Shimla	Northern India	Cartosat DEM	30m	450	6800	3625
D	Mount Abu	Western India	Cartosat DEM	90m	36	692	364

### B. Experimental Setup

The *PriorityFloodPruning* algorithm, Priority Flood algorithm are implemented in the java environment using Cartsat DEMs and ASTER DEMs of different size. The hardware specifications required to implement this algorithm are described in Table II.

Table II: Hardware and Software details

Properties	Descriptions
Processor	HP i3 processor having 2.2GHz and 4-core
RAM	4GB
Graphics Card(GPU)	No GPU
Operating System	Windows 7
Hard Disk	500GB
Software	NetBeans IDE 8.0.2

### C. Comparative Analysis

The results of the speed comparisons between *PriorityFloodPruning* algorithm and Priority Flood algorithm are displayed in Table III. The processing time includes the time cost of the original DEM and the output DEM. The each value of processing time is taken as average of ten attempts on every dataset. As showed in Fig. 5, the *PriorityFloodPruning* algorithm performs up to 13% ~25% faster than the Priority Flood algorithm.

Table III: Comparison in processing time of Priority Flood algorithms

Name of Dataset	Size of DEM			<u>Processing Time In Sec</u>	
	No of Rows	No of Columns	Total No of Pixels	<i>PriorityFloodPruning</i> Algorithm	Priority Flood Algorithm
A	100	100	10000	8	10
	300	300	90000	227	264
B	500	500	250 000	1096	1301
	500	1000	500000	3364	4147
C	800	800	640000	4741	5462
	800	1000	800000	7867	9546
D	1000	1000	1000000	11703	13569
	1000	1500	1500000	16268	18543

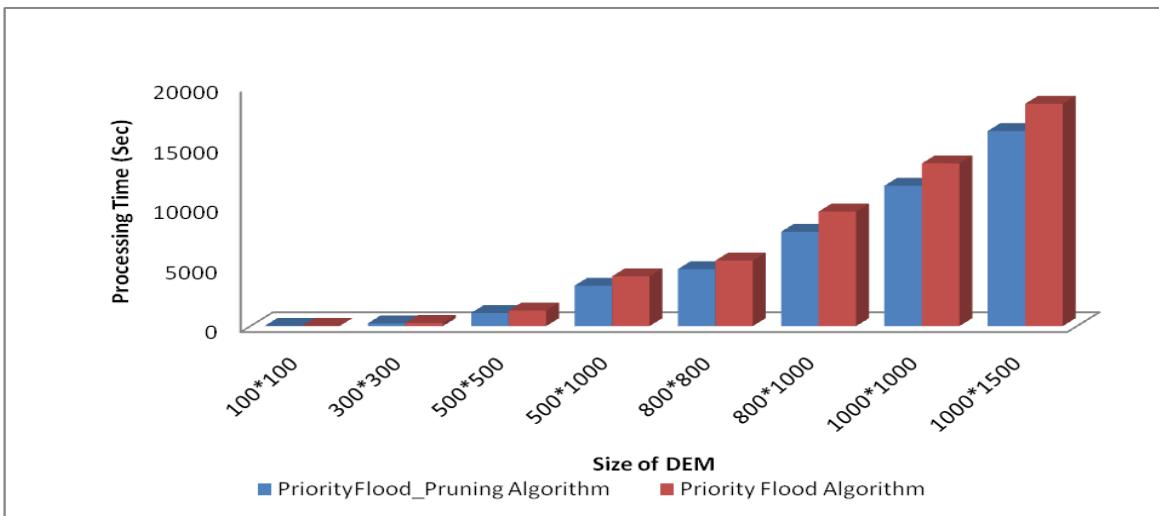


Fig. 5: Comparative analysis of *PriorityFloodPruning* algorithm and Priority flood algorithm

## V. CONCLUSION

The *PriorityFloodPruning* algorithm presented here uses a new concept of pruning the dead-cells from queues that reduces the time for priority queue operations and improves the efficiency of priority flood algorithm for filling the depressions. The comparative analysis of proposed *PriorityFloodPruning* algorithm are performed with the priority-flood algorithm using differently sized DEMs and on the basis of test statistics, our method presented in this paper works faster than priority-flood algorithm. The use of a priority queue and FIFO queue structures makes the algorithm easier to understand for users and programmers. Pruning of unwanted cells will save the processing time of depression filling algorithms. The advantages of this priority queue-based algorithm are its simplicity and clarity, which make it very easy to write correct program code quickly.

## REFERENCES

1. C. Ehlschlaeger, "Using the at search algorithm to develop hydrologic models from digital elevation data." In: *International Geographic Information Systems (IGIS) Symposium*, vol.89, pp.275–281, 1989.
2. ESRI (Ed.), "Arcgis 9. Geoprocessing Commands Quick Reference Guide." *ESRI Press*, Redlands, pp. 175, 2005.
3. S.K Jenson and J.O Domingue, "Extracting topographic structure from digital elevation data for geographic information system analysis." *Photogrammetric Engineering and Remote Sensing*, vol. 5, no-11, pp. 1593–1600, 1988.
4. W. Zhang and C. Liu, "A novel method for Filling the Depressions in Massive DEM Data", *IEEE Geoscience and Remote Sensing Symposium*, pp. 4080 – 4083, 2007.
5. M. Metz, H. Mitasova, R. Harmon, "Accurate stream extraction from large, radar-based elevation models." *Hydrology and Earth System Sciences Discussions*, vol. 7, pp. 3213-3235, 2010.
6. L. Wang and H. Liu, "An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modeling," *Int. J. Geogr. Inf. Sci.*, vol. 20, no. 2, pp. 193–213, Feb. 2006.
7. LIU Yong-He, ZHANG Wan-Chang, and XU Jing-Wen—"Another Fast and Simple DEM Depression-Filling Algorithm Based on Priority Queue Structure", *Atmospheric and Oceanic Science letters*, vol. 2, no. 4, pp. 214–219, 2009.
8. J. Lindsay, I. Creed, "Removal of artifact depressions from digital elevation models: towards a minimum impact approach." *Hydrological processes*, vol. 19, no-16, pp. 3113-3126, 2005.
9. L.W. Martz and J. Garbrecht, "An outlet breaching algorithm for the treatment of closed depressions in a raster DEM." *Computers & Geosciences*, vol. 25, pp. 835–844, 1999.
10. L.W. Martz and J. Garbrecht, "Automated extraction of drainage network and watershed data from digital elevation models." *Water Resources Bulletin*, vol. 29, no-6, pp. 901–908, 1993.

11. L.W. Martz and J. Garbrecht, "The treatment of flat areas and depressions in automated drainage analysis of raster digital elevation models." *Hydrological Processes* vol. 12, pp. 843–855, 1998.
12. M. Metz, H. Mitasova, R. Harmon, "Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search." *Hydrology and Earth System Sciences* vol. 15 no-2, pp-667, 2011.
13. M. Kreveld, J. Nievergelt, T. Roos, and P. Widmayer, "Lecture Notes in Computer Science (tutorials)" *Algorithmic Foundations of Geographic Information Systems*, no-1340, pp. 37–78. Springer-Verlag, Berlin, 1997.
14. O'callaghan and D.M. Mark, "The extraction of drainage networks from digital elevation data." *Computer Vision, Graphics and Image Processing*, vol. 28, pp. 323–344, 1984.
15. O. Planchon And F.Darboux,"A fast, simple and versatile algorithm to fill the depressions of digital elevation models." *Catena*, vol. 46, pp. 159–176, 2001.
16. J. Lindsay, and I. F. Creed, "Sensitivity of digital landscapes to artifact depressions in remotely-sensed DEM," *Photogrammetric Engineering and RemoteSensing*, vol. 71 no- 9, pp. 1029–1036, 2005.
17. P. Soille, J. Vogt, and Roberto Colombo, "Carving and adaptive drainage enforcement of grid digital elevation models", *Water Resources Research*, Vol. 39, No- 12, December 2003.
18. P. Soille "Optimal removal of spurious pits in grid digital elevation models", *Water Resources Research*, vol. 40, W12509, 2004..
19. R. Barnesa, C. Lehmanb, D. Mulla, "Priority-Flood: "An Optimal Depression-Filling and Watershed-Labeling Algorithm for Digital Elevation Models", *Computers & Geoscience* , vol. 62, pp. 117-127, July 2013.
20. H. Cormen, E. Leiserson, L. Ronald and C. Stein, "Introduction to Algorithms, Second Edition, Section 6.5: Priority queues." *MIT Press and mcgraw-Hill*, ISBN 0-262-03293-7, pp. 138–142, 2001.
21. D. Tarboton, "A new method for the determination of flow directions and upslope areas in grid digital elevation models." *Water Resources Research*, vol. 3, no-2, pp. 309–319, 1997.
22. Wang et al. " An Efficient Algorithm for Depression Filling and Flat-Surface Processing in Raster DEMs." *IEEE Geoscience and Remote Sensing letters*, vol. 11, no-12, December 2014 .
23. W. Yang, K. Hou, F. Yu, Z. Liu, T. Sun, "A novel algorithm with heuristic information for extracting drainage networks from raster DEMs." *Hydrol. Earth Syst. Sci. Discuss*, vol. 7, pp. 441–459, 2010.
24. USGS, "Digital raster elevation data and related products" /<http://www.eros.usgs.gov/products/elevation.html>S 2007.
25. Y. Liu, W. Zhang, J. Xu, " Another fast and simple DEM depression-filling algorithm based on priority queue structure." *Atmospheric and Oceanic Science*, vol. 2, pp.-214-219, 2009.

# Content-Based Video Browsing: semantic similarity and personalization

Jamel SLIMI, Anis Ben AMMAR, Adel M. ALIMI

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

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

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

## I. INTRODUCTION

Multimedia data provide large amounts of information for users. They cover most of the daily events; and they assure the awfully various requests of the people (sport, cultural, political...) [1]. This wealth of information is based on the perceptual human appearance. In fact, an image can be expressed as 1000 words.

Interest in the treatment of this type of data is caused by the large amount of information generated daily [2]. This increase in production, especially for video data, is due to the technological progress in video production and the availability of digital cameras [3].

Therefore, the size of the video data collections has become very large, either in personal database or in the database of dedicated Web sites [4]. It generates the birth of several evaluation companies as TRECVID which is interesting to several research areas taking video data as a data source such as indexing, classification, concepts extraction, etc.

These were well treated by several researchers. Consequently, the subject of exploration and navigation in multimedia databases is not yet considered in the right way. In fact, a simple exploration and fast access to documents in the large video corpus has become an urgent need.

Video data visualization systems need to overcome several problems. The first problem resides in the difficulty of the video semantic concepts extraction. Semantic description of the video data offers improved classification

[5]. Based only on low-level descriptors (color, shape, texture) limits of the video browsing's performance. Existing systems are based essentially on low-level descriptors [6, 7].

The second problem related to video data visualization system resides in the representation of documents in visualization space. The visualization space is too limited and cannot post all the documents of the corpus. Moreover, the visualization space depends on the degree of the user's perception. It cannot regroup too much information in the same time. This causes an organization problem of the keyframes representing video data in the visualization space.

The third problem, and the important, consists in the selection of video document which will be posted to the user. In fact, users are different and do not have the same preferences and the same expectations. Personalization of the access to information is very important for visualization systems. No visualization system treats the appearance personalization; all have the same representations for all users. The user needs a system to suit its own needs and their own preferences.

To treat the problems mentioned in the preceding section, we develop an intelligent video browsing system based on semantic classification and allowing an adequate interface to the user preferences. Semantic classification is build based on calculation of semantic similarity distance between descriptors vectors of the documents composing the large video collection. Descriptors vectors are composed by a set of concepts. These concepts are the result of multimodal indexing of video document. We combine video text concept, visual features and audio features to construct video descriptors vectors. The obtained classes will be projected in the visualization space in the form of a network. This network is composed by nodes (keyframes representing documents) and colors edges (representing the similarity distance between video data). The representation of documents in the visualization space is inspired from the functioning of the biological neuron. Essentially, to improve our video browsing performances, we have integrated a personalization module. This module is based on user profile which is constructed by collecting his/her preferences. User profile is composed by static indicators and dynamic indicators. User model is based on fuzzy logic. Fuzzy logic allows flexibility and suppleness of user profiles. The personalization module permits a faster access to the desired documents and an appropriate interface to the user preferences.

## II. PREVIOUS WORKS AND PROBLEM POSITION

Multimedia data browsing is a research area incorporating several other areas such as indexing, classification and man-machine interface. Large collections of video data need methods and tools for easy navigation and quick access to the document. Visualization tools are based on the human visual system and aim to facilitate the extracting knowledge task from large collection of documents.

Existing data visualization approaches are coupled or with annotation tools [8] or multimedia retrieval engines [9]. Video browsing process is based on classification step and data representation step. Its objective consists in an adequate representation allowing simple navigation and fast retrieval document process. Many video browsing systems exist in literature. They can be regrouped by techniques used for the organization, or by type of visualization interface. In what follows, we present some examples of them.

“Zoomslider” [10], video browsing system, permit a quick overview of video data corpus. Its limits consist in the disagreeable aspect caused by the poor visual feedback, and the lack of scalability for long video.

“FutureViewer” [8] permits annotation and video exploration. It is characterized by 2D Cartesian plane. Axes correspond to the feature selected by user. User can annotate and navigate in data collection.

MediaMill team propose video search engine [9, 11, and 12] composed by four browsing tools for video data visualization: Galaxy browser, Sphere browser, Cross browser and Rotor browser. Galaxy browser allows to the user to annotate video data using 2D similarity-based visualization of keyframes. This tool can advance the efficiency of the relevance feedback process by mixing relevance feedback [13, 14]. The Sphere Browser uses semantic conceptual similarity and a novel representation interface. The Sphere Browser display the time-line of the current video data on the horizontal axis, and for each video it displays the relevant threads on the vertical axis. The Cross Brower is based on a linear ordering to rank video data. The vertical axis represents the selected concept. The horizontal one is used to visualize video program in a time-line from which a keyframes is selected. The Rotor browser represents an extended Cross Brower based on more than two axes.

“Instant video browsing” proposed by [15] offers parallel view and tree view of video data collection. It divides video data into many parts of equal length which will be posted in the visualization screen. Its limits reside in pretreatments of video data collection such as indexing and segmentation.

In [16, 17, and 18], the authors present a simple tool for storing images based on their color similarities. Just using the color descriptor decrease the system performance. The organization of the visualization space is based on 3D cylinder form.

“ReGIMVIZ” [19] is a video browsing tool based on semantic classification and novel representation interface. The semantic classification is based on semantic correlation between concepts vectors representing video data composing video collection. This tool integrates a personalization module allowing an adequate interface to the user preferences.

In [20, 21 and 22], the authors propose a video browsing based on semantic concepts network and V form visualization interface. Documents are organized in two axes based on semantic similarity between them. The selected document is placed in the center of the axes.

All proposed systems are based on data clustering and intuitive representation. Those tools treat one type of video documents (news, sport...). Also, they based their pretreatments on low-level descriptor which reduces their performances. Those tools are not centered users and do not take into consideration their preferences. They have the same representation for all users.

Our video data visualization system bases the semantic classification on high-level semantic concepts extraction. The obtained classes are projected in the visualization space, documents organization is inspired from biological neuron function. The advantage of our video browsing consists in the integration of personalization module allowing a visualization interface adequate to the user preferences. Following section describes our framework details.

### III. OUR FRAMEWORK

Data visualization process is composed by three main tasks: semantic classification, data representation and personalization (figure 1). Classification is a key task in the data visualization process. Concerning our tool to implement semantic classification, we rely on the extraction of semantic concepts from video data. Each video is described by a set of semantics concepts which compose the descriptor vector. To obtain video semantic classes of video data, we apply a set of rules for calculating similarity between descriptors vectors of videos collection. After the classification phase, the obtained classes will be projected into the visualization space. The most important task of our work consists in the integration of personalization module in our video browsing system. The main object of personalization is to allow a visualization graph adequate to the user's expectations.

Personalization module is based on selection of the user's preferences extracted from his interaction with the system. Following sections detailed each parts of visualization process.

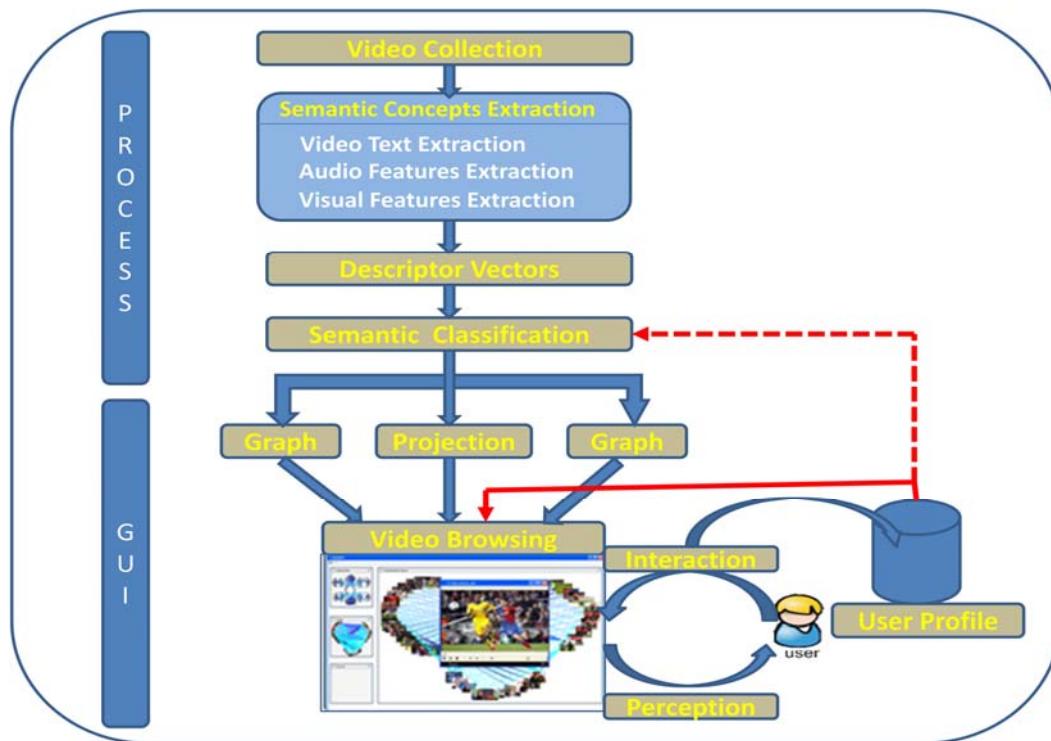


Figure1. Framework description

#### IV. VIDEO INDEXING

Video indexing is an important step to start video data visualization process. In our work, we use a multimodal video indexing based on video text extraction, audio features extraction and visual features extraction. Video data must be segmented on shot to facilitate indexing process. Shot is composed by one or more related frames that represent a continuous action in time and space [23, 24]. For ease of use, a shot is represented by a keyframe [25]. Generally, the keyframe is the central frame of a shot [26]. Each shot is described using combination between text concepts, audio concepts and visual concepts. Based on obtained concepts, we construct concepts vectors describing video data (figure 2).

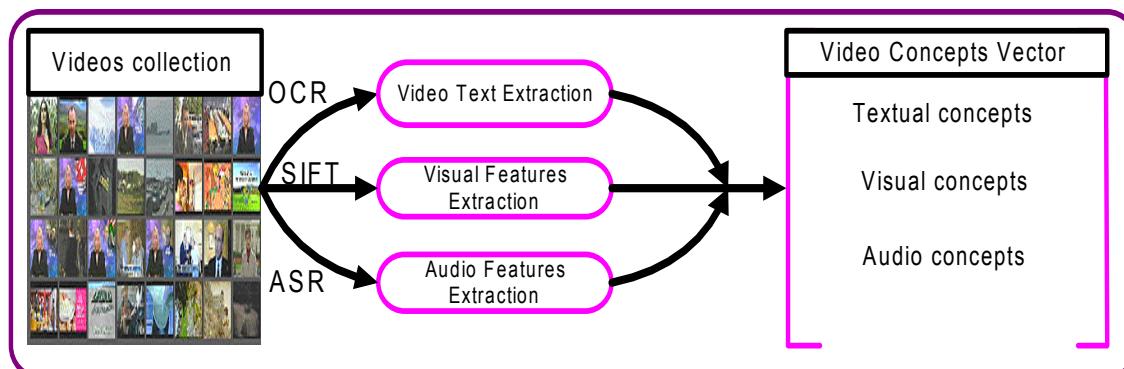


Figure2. Video Multimodal Indexing

#### 4.1. Video Text Extraction

Many types of videos contain important and useful information in their embedded text such as news. To extract text concept we must, firstly, divided video data on frames. We have used Java Media Framework (JMF) to segment video data on frames. Secondly, frames are sent to an Optical Character Recognition (OCR) for text recognition (figure 3). Video text extraction involves detection, localization, enhancement and recognition of the textual content in the video frame [27].

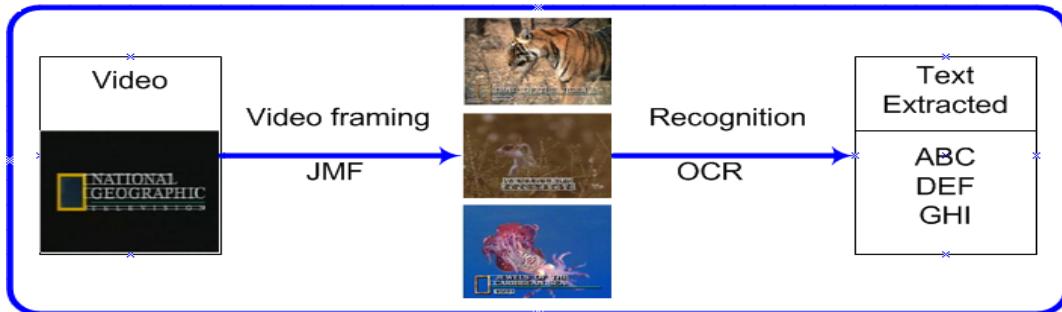


Figure3. Video Text Extraction

#### 4.2. Visual Feature Extraction

The objective of visual features [28, 29, 30, 31 and 32] consists in improving video analysis. Two parts compose visual descriptor. General information descriptors are based on low-level descriptors (colours, shapes, regions, textures); and specific domain information descriptors based on semantic concepts.

*Colour:* distribution of colour in the keyframe is represented by colour histogram [30]. Hue, Saturation and Volume define the space colour. Many tools describe colour such as Dominant Colour Descriptor (DCD), Scalable Colour Descriptor (SCD), Colour Structure Descriptor (CSD) and Colour Layout Descriptor (CLD).

*Texture:* The texture descriptor observes region homogeneity and histograms of these region borders. Homogeneous Texture Descriptor (HTD), Texture Browsing Descriptor (TBD) and Edge Histogram Descriptor (EHD) are used to describe texture.

*Shape:* Two ideas to segment keyframe: the first consists in grouping pixels based on color homogeneity, texture or both [33, 34, and 35]; the second is done by connecting edges lines [36]. Shape descriptors are: Region-based Shape Descriptor (RSD), Contour-based Shape Descriptor (CSD) and 3-D Shape Descriptor (3-D SD). Important semantic information can be extracted from shape descriptors. Many objects can be recognized through their forms.

*Region:* the Scale-Invariant Feature Transform (SIFT) descriptor measures the region around a key point and describes each region using an edge orientation histogram. SIFT descriptor is the best performing interest region descriptor. SIFT describes the local shape of the interest region using edge histograms [37]. The interest region is divided into a 4x4 grid and every sector has its own edge direction histogram. The grid is aligned with the dominant direction of the edges in the interest region to make the descriptor rotation invariant.

In our system, we use 130 concepts given by TRECVID 2010. We based on SIFT to detect these concepts in video document. The result of this step is represented by the following figure which illustrates some of used concepts (left) and an XML document presenting videos that contains the concept number 130 (right).

TV10_ID	LSCOM_Name	
001	Actor	</videoFeatureExtractionFeatureResult>
002	Adult	<videoFeatureExtractionFeatureResult fNum="130">
003	Airplane	<item seqNum="1" shotId="shot10028_1"/>
004	Airplane_Flying	<item seqNum="2" shotId="shot10028_2"/>
005	Anchorperson	<item seqNum="3" shotId="shot10028_3"/>
006	Animal	<item seqNum="4" shotId="shot10028_4"/>
007	Asian_People	<item seqNum="5" shotId="shot10028_5"/>
008	Athlete	<item seqNum="6" shotId="shot10028_6"/>
009	Basketball	<item seqNum="7" shotId="shot10028_7"/>
010	Beach	<item seqNum="8" shotId="shot10028_8"/>
011	Beards	<item seqNum="9" shotId="shot4781_32"/>
012	Bicycles	<item seqNum="10" shotId="shot4781_33"/>
013	Bicycling	<item seqNum="11" shotId="shot4781_34"/>
014	Birds	<item seqNum="12" shotId="shot4781_35"/>
015	Boat_Ship	<item seqNum="13" shotId="shot4781_39"/>
016	Boy	<item seqNum="14" shotId="shot4781_40"/>

Figure4. Visual features extraction

#### 4.3. Audio Feature Extraction

Earlier researches on classification and video indexing are essentially based on visual descriptors. Recently, researchers have noted that the audio descriptors are also very important. Audio descriptors facilitate video semantic content detection. So, it is easier to detect people laughing or people crying based on audio descriptors than using visual descriptors.

Most audio features extraction methods are based on automatic speech recognition (ASR). The importance of ASR method comes from its reliability, robustness and inductance in the topic of video data. Generally, description results from sampling audio data in different size. The increased number of samples leads to a better semantic description of an audio signal.

#### 4.4. Descriptors Fusion

A video data is composed by different modalities, such as visual, audio and text, an important step in multimedia indexing consists in fusion [38]. This last is defined as the study of efficient methods allowing transformation of collected information from different sources into a representation providing effective support for following tasks. A general information fusion definition is given in [39] as “Information fusion is to combine information from multiple heterogeneous sources to improve decision making”.

Video document is represented by three sources of information such as visual, audio and text [40]. So, to improve video data processing, we must base their indexing process on all of media composing video documents. Given that a video document consists of different modalities, multimodal techniques are clearly necessary to exploit all available information to benefit video browsing [41]. In fact, the idea is to communicate semantic information by combining information channels. The channels can be visual, auditory or textual. Basing video indexing on multimodalities causes gives an important improvement in video semantic description which improves all following research task such as classification and visualization. In our approach, video descriptor vectors are composed of different concepts extracted from different channel: visual, audio and text.

#### 4.5. Video Descriptor Vector

Each document is characterized by a set of semantic concepts. These concepts will constitute the video data concepts vector. The objective of the combination between textual concepts, visual concepts and audio concepts is to improve the semantic classification. Thus, we can facilitate video collection navigation and accelerate video data access. The main objective of our system is to provide automatic content analysis and an intelligent tool for retrieval and browsing in large video data collection. Firstly, we extract different concepts embedded in each video and then construct video descriptor Vector based on combination between different extracted concepts (figure 4). Figure 4 illustrate an example of video descriptor vector.

```
- <video Name="TRECVID2010_10001">
  <concept Name="Building" Weight="0.0607" shot_representatif="shot10001_10" />
  <concept Name="Ground_Vehicles" Weight="0.0034" shot_representatif="shot10001_24" />
  <concept Name="Highway" Weight="0.0017" shot_representatif="shot10001_33" />
  <concept Name="House_Of_Worship" Weight="0.0052" shot_representatif="shot10001_26" />
  <concept Name="Motorcycle" Weight="0.0034" shot_representatif="shot10001_24" />
  <concept Name="Outdoor" Weight="0.0607" shot_representatif="shot10001_10" />
  <concept Name="Road" Weight="0.0017" shot_representatif="shot10001_33" />
  <concept Name="Roadway_Junction" Weight="0.0017" shot_representatif="shot10001_33" />
  <concept Name="Sports" Weight="0.0034" shot_representatif="shot10001_24" />
  <concept Name="Streets" Weight="0.0069" shot_representatif="shot10001_23" />
  <concept Name="Suburban" Weight="0.0069" shot_representatif="shot10001_23" />
  <concept Name="Vehicle" Weight="0.0034" shot_representatif="shot10001_24" />
</video>
```

Figure 5. Video Descriptor Vector

## V. SEMANTIC CLASSIFICATION

Semantic classification is based on multimodal indexing step. Indexing step is based on semantic concepts extraction. Concepts are composed by text video concepts, visual features and audio features. By combining collected concepts, we compose video vectors descriptors. Based on those, we calculate semantic similarity distance between documents of the corpora. This phase allow building Concepts Semantic Network (CSN). CSN is a network regrouping concepts based on their semantic similarity distance. Following sections show how to calculate the similarity distance between documents of the corpus.

### 5.1. Semantic similarity and context similarity

To calculate semantic similarity distance, we adopt combination between Leacock and Chodorow semantic similarity (LCH) [42] and Flicker Context Similarity (FCS) [43]. The combination between these similarities improves the computation of the similarity distance.

LCH is a semantic similarity measure incorporating edge-based approach. Based on the length of shortest path connecting two concepts ( $C_i, C_j$ ) and the maximum depth of WordNet hierarchy (D); we calculate semantic similarity between two concepts as follow:

$$LCH (ci, cj) = - \log \frac{L(ci, cj)}{2D} \quad (1)$$

Where  $L (C_i, C_j)$  denotes the path length between two concepts, and D is the maximum depth of ontology. The following figure illustrates the ontology used connecting concepts.

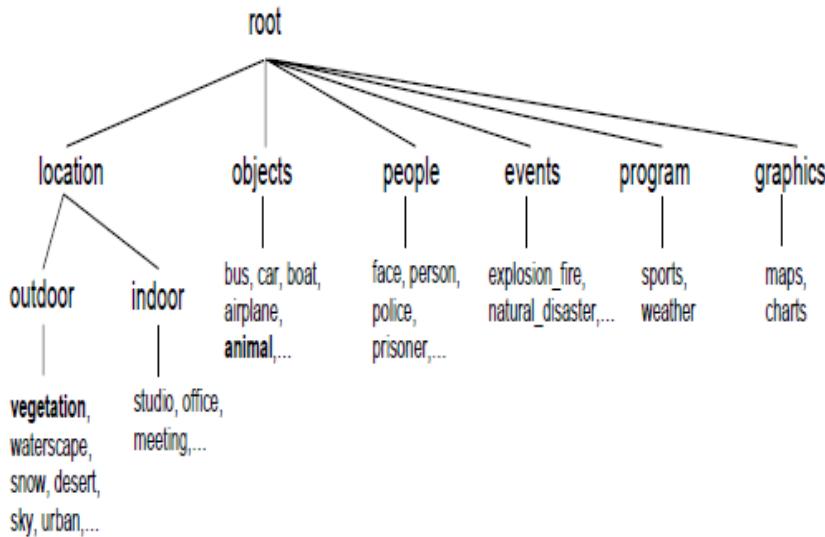


Figure 6. fragment of used ontology

FCS allows the contextual relationships measure between two concepts  $C_i$  and  $C_j$ . FCS measure is based on NGD measure which calculate based on the following formula:

$$NGD(C_i, C_j) = \frac{\max\{\log f(C_i), \log f(C_j)\} - \log f(C_i, C_j)}{\log N - \min\{\log f(C_i), \log f(C_j)\}} \quad (3)$$

Where  $f(C_i)$  (resp.  $f(C_j)$ ) denotes the number of shots containing concept  $C_i$  (resp.  $C_j$ ).

$F(C_i, C_j)$  denotes the number of shots containing both concepts  $C_i$  and  $C_j$ .

$N$  denotes the total number of shots of the video corpus.

In result, FCS is calculated as follow:

$$FCS(C_i, C_j) = \exp(-NGD(C_i, C_j)) \quad (2)$$

### 5.2. Concepts Semantic Network (CSN)

Based on the preceding formulas, we combine FCS and LCH to calculate the conceptual and contextual similarity between two concepts  $C_i$  and  $C_j$  as follows:

$$Sim(C_i, C_j) = FCS(C_i, C_j) + LCH(C_i, C_j) \quad (4)$$

Then, we add two variables  $\alpha$  and  $\beta$  to moderating LCH and FCS. With  $\alpha + \beta = 1$

$$Sim(C_i, C_j) = \alpha FCS(C_i, C_j) + \beta LCH(C_i, C_j) \quad (5)$$

In our work, Semantic similarity interval is  $[0, 1]$ . For this reason, we use the inverse of formula 5. When two concepts have a strong connection, semantic similarity distance tends towards 0; and if they don't have any relation semantic similarity distance tends towards 1.

$$Sim(C_i, C_j) = \frac{1}{1 + (\alpha FCS(C_i, C_j) + \beta LCH(C_i, C_j))} \quad (6)$$

Figure 5 presents the result of the application of formula 6. The result is a semantic concepts network which connects concepts based on semantic similarity distance.

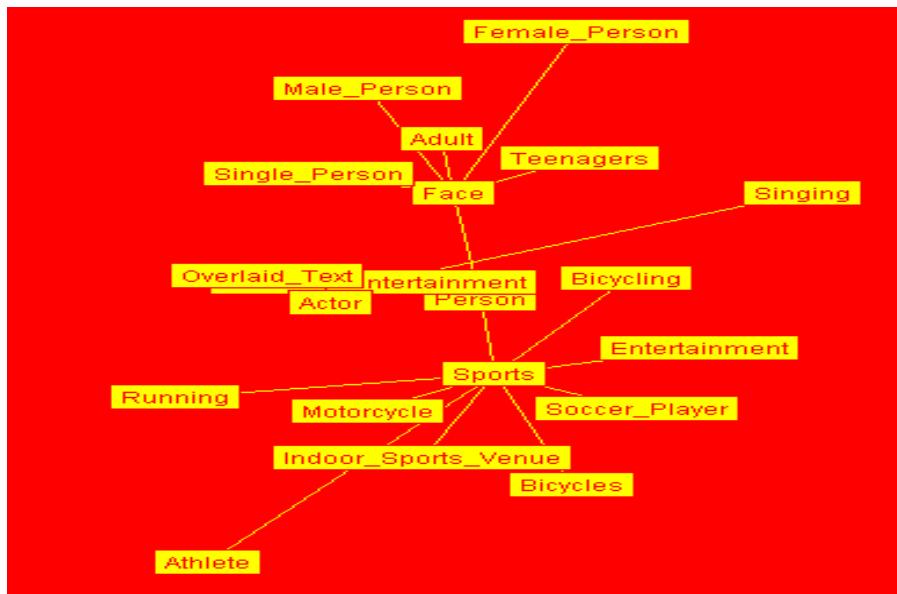


Figure 7. Concepts Semantic Network.

### 5.3. Suggestion Module

#### 5.3.1. Semantic Suggestion

Semantic suggestion consists in finding all concepts which have a strong relation with a selected concept (Sc). The objective is to rank concepts in the semantic network SCN based on their relevance and importance to Sc.

$$D = \arg \max Sim (Sc, cj) \quad cj \in SNC \quad (7)$$

Where D denotes concepts selected for semantic suggestion and Sim (Sc,Cj) is the similarity between Sc and a concept cj in CSN that has been computed via formula (7).

#### 5.3.2. Visual Suggestion

Near-duplicate keyframes (NDK) is very important methodologies to measure video data similarity. In our work, we exploit CSN to measure video data similarity. Figure 6 presents an example of application of visual suggestion. The semantic similarity between two video data is computed as follow:

$$S = \arg \max Sim (Ks, Kj) \quad Kj \in C \quad (8)$$

Where Ks is a selected keyframe and Kj is one of the documents of the collection. Sim (Ks, Kj) is computed as follow:

$$Sim (Ks, Kj) = \sum_{i=1}^M \sum_{s=1}^N Sim (C_i^{Ks}, C_s^{Kj}) \quad (9)$$

M and N denote the number of concepts that are associated to KS and kj respectively, and Sim(C<sub>i</sub><sup>KS</sup>, C<sub>s</sub><sup>Kj</sup>) is computed via formula (6).

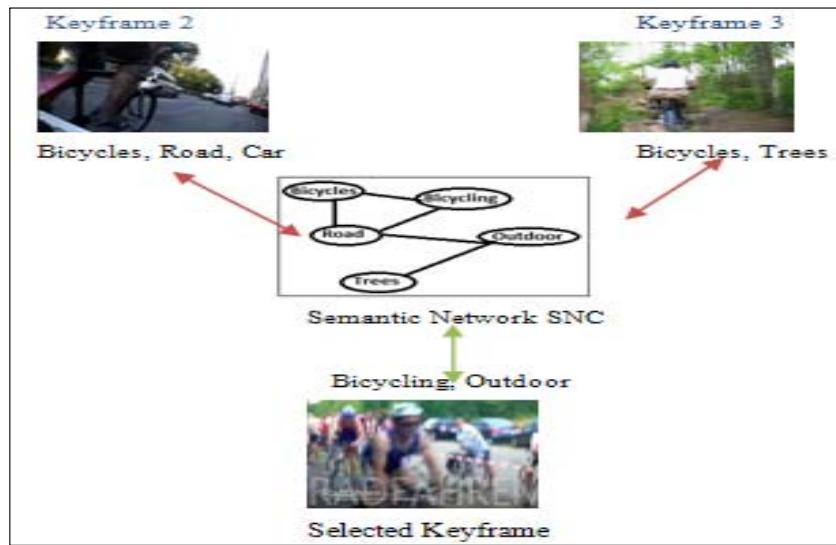


Figure 8. NDK Retrieval with semantic similarity.

Using proposed semantic similarity distance allows the improvement of semantic classification. Conceptual and contextual similarity improves precision of semantic distance. Application of semantic similarity distance on descriptors vectors of video data of the corpus allows the building of semantic classes. These classes will be projected in the visualization space like described in the following section.

#### VI. VISUALIZATION GRAPH

Visualization interface is the most important part of the Video data browsing system. The interface plays a crucial role for the judgment of this type of systems. Visualization space provides an overview of the video corpus to the user. Visualization is the result of the projection of the obtained classes in the visualization space. Using metaphor in data visualization system renders the system more meadows to the user thought and more comprehensible. Several systems use a metaphor to post data collection in their visualization space such as galleries metaphor, building metaphor...

Our visualization graph is in the form of a network. It gathers a whole of nodes and arcs. So we take as a starting point the operation of the biological neuron to use it as a metaphor. Following section is dedicated to explain biological neuron function.

##### 6.1. Neuronal metaphor

Biological neuron or nerve cell is an electrically excitable cell. Its objective is to transmit information based on chemical and electrical signaling. Neurons are connected to form neuronal network.

Biological neuron is composed by cell body (soma, core), axon and dendrites (figure 7). The body gives rise to multiple dendrites and one axon. This last is a special cellular extension that travels for a distance and arises from the cell body at a site called the axon hillock.

By analogy, the core represents the nodes, the axons represent the edges connecting the classes and dendrites are the edges connecting the documents. The click on a keyframe is considered as a stimulus activating others related keyframes.

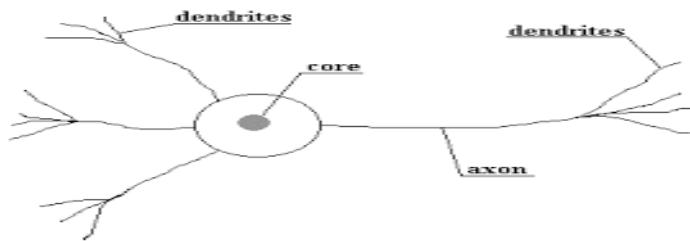


Figure9. Biological Neuron

### 6.2. Visualization interface

Visualization interface is limited by the visualization space and the degree of the perception of user. For this reason, we can't post all the documents of a large corpus in the same time. Network representation and neuron metaphor aim to be adequate to surmount this problem.

The visualization graph is represented by colored edges and nodes. Colored edges are the semantic relations between classes and between documents of the corpus. The red color means that the resemblance between two videos is more than 75%, the yellow color means that the resemblance is more than 50% and the white color means that the resemblance degree is less than 50%. Nodes or keyframes represent the documents in the visualization space. Our proposed Visualization graph, in the form of network, permits an effective and simple browsing of video data collection. User can extract knowledge easily from the large corpus of video data. This is done by explicit semantic relation between documents on the visualization graph (figure 10).

Limited visualization space causes a big problem which consists in what are the documents to be posted. In a limited space we can't post all the documents of a large corpus of video. So, posted documents should be well-chosen. For this reason, we integrate a personalization module in our video browsing system. Personalization allowed an adequate interface to user preferences. This module will be described in the following section.

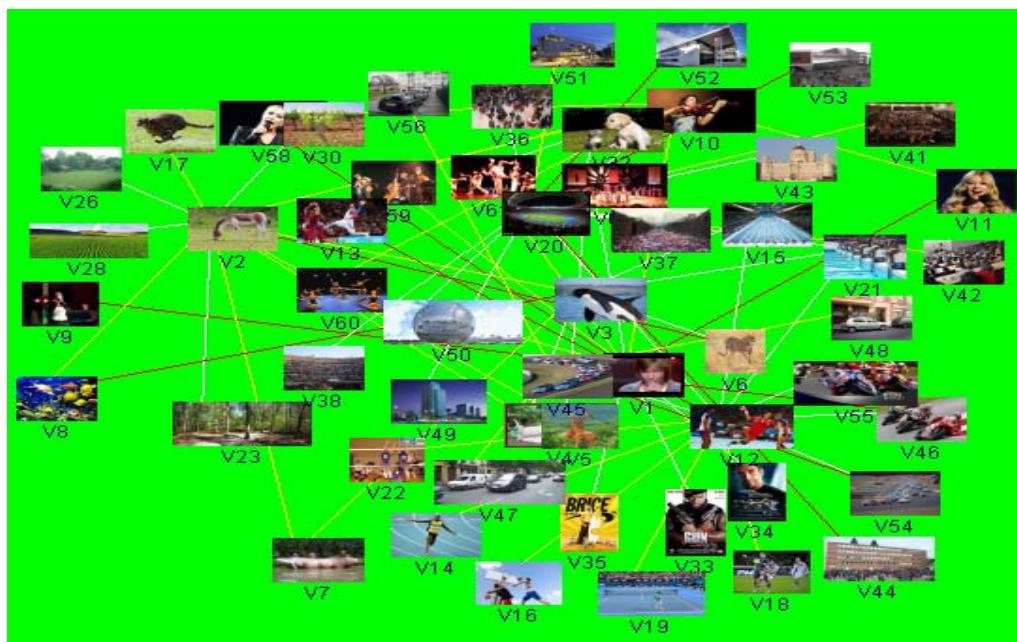


Figure10. Visualization Graph

## VII. PERSONALIZATION

The users do not have the same waiting and same profiles. Each user has his own preferences. The user interacts with the system to obtain needed information [44]. The objective of personalization module is to provide the user with an appropriate environment to his profile. Personalization makes data access process faster and visualization interface more adequate to user's preferences.

User's profile construction is based on its interaction with video data visualization system. Before the user's input, the first step consists in posting a global overview of data collection. This overview is conditioned by revealing the possible maximum range of existing concepts in video data corpus. After the user's input, the second step, the video visualization system must collect various user interest centers.

User's profile is based on static indicator and dynamic indicators. User's personal information presents static indicators (first name, last name, age...). Dynamic indicators are extracted from user interaction and data manipulation such as duration of consultation and frequency of consultation. The dynamic indicators are the duration of consultation, the frequency of consultation and user navigation history.

In our work, we based on fuzzy logic to model multidimensional user profile. Fuzzy logic permits a great flexibility and suppleness to the user profile. In fact, user profile is not static and can be changed throughout navigation. Also, user's profile can be varied in the same session or different sessions. In this case, video visualization system must post all concepts related to the user profile.

User profile model inputs are the duration of consultation and the frequency of consultation. Based on those inputs, we can predict user profile. The output of the user's model is the relevance degree of video data. Following figures illustrate various function of user model.

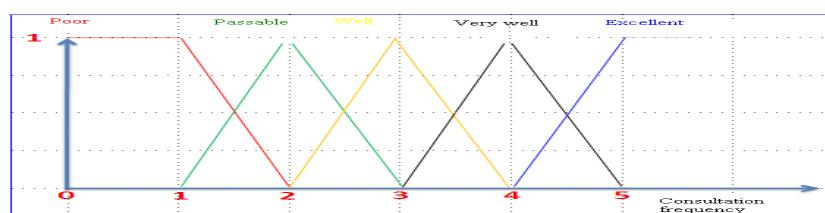


Figure11. Frequency consultation function

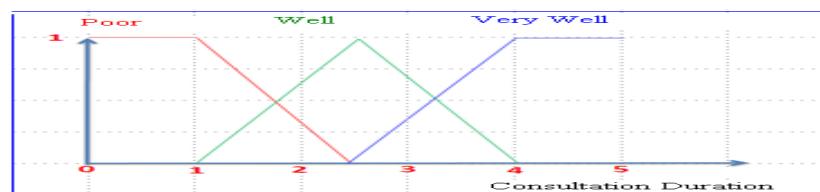


Figure12. Duration consultation function

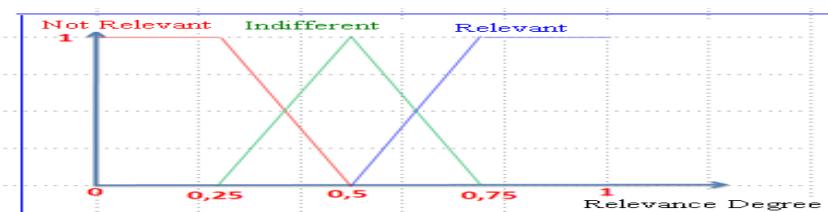


Figure13. Relevance degree function

Starting navigation in video data corpus triggers the collection process of user preferences. User profile will influences visualization interface by eliminating not related video data to the chosen one and posting all video documents strongly related to the selected one. The user can view semantic video distance between connected video data. None can predict the following actions of the user. For that the video data visualization system must keep a global overview of the video documents collection (figure 14).

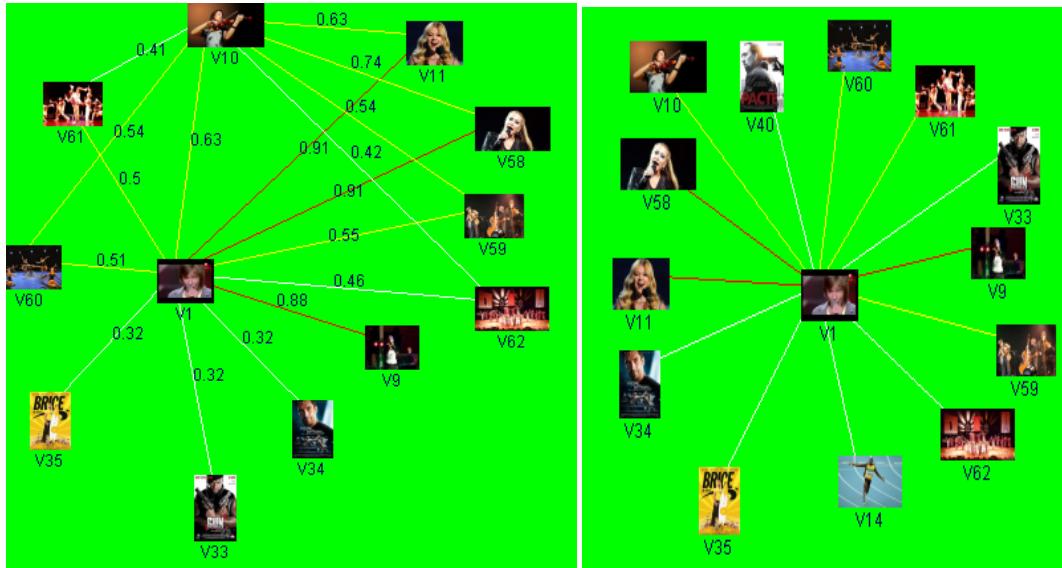


Figure14. System navigation

The performance of the current video data visualization system still needs to be improved. Improvements will relate to the semantic classification part by augmenting the number of used concepts and to the personalization module by integrating users' emotion. In fact, when the user is tired (sad), the system will post automatically relaxing (happy) video data.

### VIII. DISCUSSION AND EXPERIMENTATION

The contributions of our system compared to existing systems reside in three points. Firstly, existing systems based their classification on low-level descriptors but our approach is based on high-level concepts to compute semantic similarity distance between documents of the collection. Secondly, we treat different type of video data but existing system focused on one type of video data (sport, news...). Thirdly, the major important point is that our system integrates a personalization module allowing navigation on demand. Personalization is based on fuzzy logic modeling user profile. Another point consists in visualization interface metaphor. Our system inspired metaphor from the function of biological neuron and the form of neuron network. Existing systems based their metaphor only on the form such as sphere, city and cylinder browser. In the following table, we compared our systems to some existing video browsing systems. The comparison will be done between the services offered by each system also between their main features.

TABLE I  
Comparison of our system with existing systems

Systems	Navigation	Quer-yng	Meta-phor	Personalization	Main Features
Fudan university hierarchical Video Retrieval [45]	Yes	YES	NO	NO	- video decomposed into shots - textual description from audio transcript - hierarchical video browsing scheme - visual similarity search, transcription based search
MKLab System [46]	YES	YES	NO	NO	- videos decomposed into shots - MPEG-7 visual descriptors (color, texture) - interactive video retrieval using MPEG-7 descriptors - 2D grid web interface mimics functionality of shopping cart encountered in e-commerce sites
Zoomslider [10]	YES	YES	NO	NO	- videos decomposed into shots - quick overview of video data corpus. - indexing is based on low-level descriptors - 2D visual interface
Fork browser [9]	YES	YES	NO	NO	- videos decomposed into shots - supervised learning of concepts using BoW and SVM - visual similarity using BoW - multiple thread-based video navigation - number of threads adjustable to the type of user - fast keystroke based navigation
Future Viewer [8]	YES	YES	NO	NO	- videos decomposed into shots - permits annotation and video exploration. - indexing based on visual features - visual interface is characterized by 2D Cartesian plane. Axes correspond to the feature selected by user. - User can annotate and navigate in data collection.
Our System	YES	YES	YES	YES	- videos decomposed into shots - multimodal indexing based on video text features, audio features and visual features - network form of visualization space allowing explicit similarity distance between videos based on color edges. - neuronal metaphor facilitating navigation in large video data corpus.

Compared to existing systems, our video browsing tool offers a personalization module which makes its interfaces more appropriate to the expectations of different users. Personalization allows acceleration of desired document access process. Also, the visualization interface in the form of a network facilitates the navigation in large corpus of video data. This representation, based on colored edges, simplifies the knowledge extraction task. Our tool still needs some improvement as the integration of a human facial expression recognition module allowing posting the appropriate videos to the psychological state of the user. For example, the system must post relaxant videos if the user is angry (comedy videos if he/she is sad).

### 8.1. User based evaluation

User based evaluation consists in filling a form which contains five important questions. Questions concern system design, simplicity of use, knowledge extraction, navigation and video data similarity. The questions will be note between 1 and 5. The number of participants in this survey is 100.

- Q1: what about system design?
- Q2: what about simplicity of use?
- Q3: what about data access process?
- Q4: what about navigation?
- Q5: what about video data similarity?

TABLE II

User Study

Notes	1	2	3	4	5	Average
Q1	0	10	19	25	46	<b>4.07/5</b>
Q2	2	6	17	20	55	<b>4.2/5</b>
Q3	6	12	15	23	44	<b>3.87/5</b>
Q4	1	5	15	30	49	<b>4.21/5</b>
Q5	1	4	13	13	60	<b>4.00/5</b>

By analyzing the results, we note the satisfaction of the users of our video data visualization system.

### 8.2. Similarity Measurement Evaluation

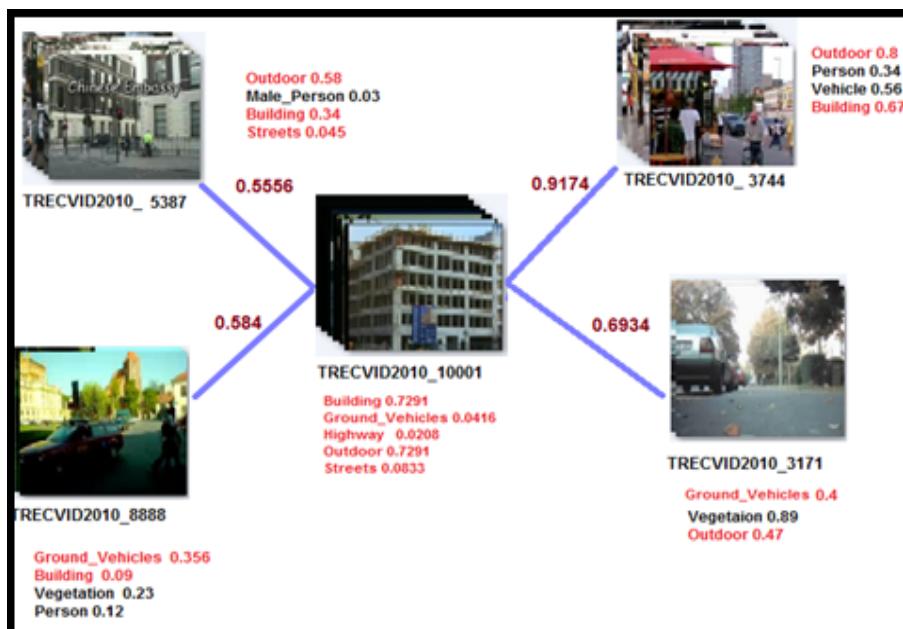


Figure15. Example of video similarity

To evaluate similarity measurement, we chose arbitrarily 10 video documents. Similarity measurement evaluation consists in comparison between manually similarity and system similarity. Effectiveness of our system is deduced by similarity precision which will be calculated based on formula 10.

$$\text{Precision} = \frac{\text{Correct Similar Shots}}{\text{Returned Shots}} \quad (10)$$

Following figure illustrate similarity measurement precision. We deduce that the simple documents which contain a limited number of concepts have a high precision (Example s1 and s3). But the precision decreases for documents containing various concepts (Example s5 and s7).

Considering relation between concepts in a fuzzy framework, as in [47], will probably enhance the results. In fact, similarity degree, even in a contextual space, is not frozen and depends on user since proposed visualizations are personalized.

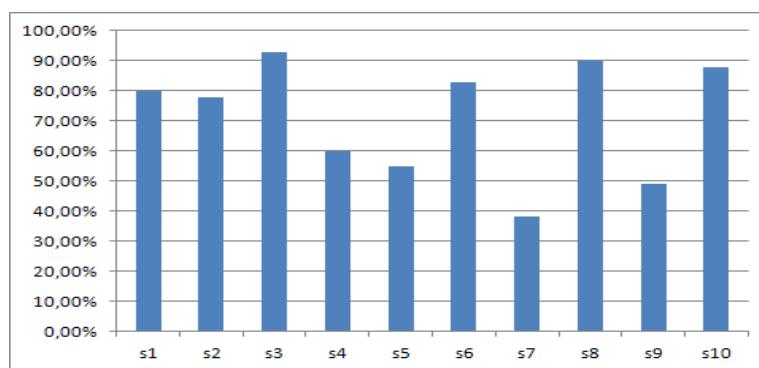


Figure16. Similarity measurement precision

## IX. CONCLUSION AND FUTURE WORK

In this paper we have described our video data visualization system based on semantic similarity and personalization. Our system is composed of three main parts. The first part consists in semantic classification based on semantic concepts extraction and concepts semantic network. CSN is constructed based on computation of semantic concepts similarity. The second step is devoted to the visualization interface which is in the form of a network and inspired from biological neuron function. The last step, and the most important part, consists in the integration of a personalization module in the video data visualization system. This module allows an adequate browsing to the user's preferences. User preferences, which are collected from interaction with the system, permit the construction of the user profile. In our work, experiments showed encouraging results that can be improved. Further works concern the fuzzy interpretation of contextual relationship between visual concepts in video. We also plan to integrate a 3D representation of the visualization graph.

## ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

## REFERENCES

- [1] Fan.J, Luo. H, & Elmagarmid. A. K, "Concept oriented indexing of video database toward more effective retrieval and browsing", IEEE Transactin on Image Processing, pp 974–992, 2004.
- [2] Y. Mu et al., "Multimedia semantic-aware query-adaptive hashing with bits reconfigurability". Int. Journal of Multimedia Information Retrieval (IJMIR), pp 59-70, 2012.
- [3] V. Vijayakumar, R. Nedunchezhian, A. Hanjalic , "A Study on Video data mining". Int. Journal of Multimedia Information Retrieval (IJMIR), pp 153-172, 2012.

- [4] C. Beecks, T. Skopal, K. Schoeffmann, & T. Seidl, “*Towards large-Scale Multimedia Exploration*”, 5th International Workshop on Ranking in Databases, 2011.
- [5] P. Wang, A. F. Smeaton, “*Semantic-based selection of everyday concepts in visual lifelogging*”. Int. Journal of Multimedia Information Retrieval (IJMIR), pp 87-101, 2012.
- [6] G. Amarnath, J. Ramesh, “*Visual information retrieval*”, Communications of the ACM, 1997 .
- [7] C. Town, “*Ontological inference for image and video analysis*”, Machine Vision and Applications, 17(2), 94-115,2006.
- [8] M. Campanella, R. Leonardi, P. Migliorati, “*Future viewer: an efficient framework for navigating and classifying audiovisual documents*”. In WIAMIS, Montreal, Canada, 2005.
- [9] C. Snoek, M. Worring, D. Koelma, A.W.M. Smeulders, “*Learned lexicon-driven interactive video retrieval*”. In Proceeding of the ACM International Conference on Image and Video Retrieval, pp 11,2006.
- [10] W. Hurst, P. Jarves, “*Interactive, dynamic video browsing with the zoomslider interface*”, in IEEE ICME, 2005.
- [11] M. Worring, C.G.M. Snoek, D.C. Koelma, G.P. Nguye, O. de Rooij, “*Lexicon-based browsers for searching in news video archives*”. In ICPR, pp 1256, 2006.
- [12] M. Worring, C.G.M. Snoek, D.C. Koelma, G.P. Nguye, O. de Rooij, W.M. Smeulders, “*The Mediamill semantic video search engine*”. In IEEE ICASSP, Honolulu, Hawaii, USA, 2007.
- [13] G.P. Nguyen, M. Worring, “*Interactive access to large image collections using similarity based visualization*”, Journal of Visual Languages and Computing, 2007.
- [14] G.P. Nguyen, M. Worring, “*Optimization of interactive visual similarity based search*”, ACM TOMCCAP, 2007.
- [15] M. Del Fabro, K. Schoeffmann, L. Boeszoeményi, “*Instant Video Browsing: A tool for fast Nonsequential Hierarchical Video Browsing*”. In Workshop of Interactive Multimedia Applications, 2010.
- [16] K. Schoeffmann, M. Taschwer, L. Boeszoeményi, “*The video explorer: a tool for navigation and searching within a single video based on fast content analysis*”. In ACM Multimedia Systems, 2010.
- [17] C. Müller, M. Smole, K. Schoeffmann, “*Demonstration of a hierarchical Multi-layout 3D video browser*”. In IEEE International Conference on Multimedia and Expo, 2012.
- [18] K. Schoeffmann, D. Ahlström, L. Bszormenyi, “*Video browsing with a 3D thumbnail ring arranged by color similarity*”. In Advances in Multimedia Modeling, pp 639-641, 2012.
- [19] J. Slimi, A. Ben ammar, A.M. Alimi, “*Video Data Visualization System: semantic classification and personalization*”, International Journal of Computer Graphics & Animation (IJCDA) Vol.2, No.2/3, 2012.
- [20] J. Slimi, S. Mansouri, A. Ben ammar, A.M. Alimi, “*Semantic browsing in large videos collection*”, In OAIR, Lisbonne, Portugal, pp 53-56, 2013.
- [21] J. Slimi, S. Mansouri, A. Ben ammar, A.M. Alimi, “*Video Exploration tool based on Semantic Network*”, In OAIR, Lisbonne, Portugal, pp 213-214, 2013.
- [22] J. Slimi, A. Ben ammar, A.M. Alimi, “*Interactive video data visualization based on semantic organization*”, In CBMI 2013, Hungary , 161-166, 2013.
- [23] Davenport. G, Smith. T.G.A, & Pincever. N, “*Cinematic principles for multimedia*”, IEEE Computer Graphics & Applications, vol. 11, pp. 67–74, 1991.
- [24] B. Solmaz, S.M. Assari, M. Shah , “*Classifying web videos using a global video descriptor*”, Machine Vision and Applications 24(7), 1473-1485, 2013.
- [25] V. Vijayakumar, R. Nedunchezhian, “*A study on video datamining*”. Int. Journal of Multimedia Information Retrieval (IJMIR), pp 153-172, 2012.
- [26] Brunelli. R. Mich. O, & Modena. C. M, “*A survey on the automatic indexing of video data*”, Journal of Visual Communication and Image Representation, vol. 10, pp. 78–112, 1999.
- [27] H. Karray et al., “*ReGIM at TRECVID (2008) “high-level features extraction and video search”*”, In TREC Video Retrieval Evaluation online proceedings, 2008.
- [28] K. Lee, N. Street, “*Model-based detection, segmentation, and classification for image analysis using on-line shape learning*”, Machine Vision and Applications 13(4): pp 222-233, 2003.
- [29] M. S. Lew, ed., “*Principles of Visual Information Retrieval*”. Advances in Computer Vision and Pattern Recognition, 2001.
- [30] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, & R. Jain, “*Content based image retrieval at the end of the early years*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 1349–1380, 2000.
- [31] T. Tuytelaars, K. Mikolajczyk, “*Local invariant feature detectors: A survey*”, Foundations and Trends in Computer Graphics and Vision, vol. 3, pp. 177–280, 2008.
- [32] K. Van deSande, T. Gevers, C. G. M. Snoek “*Evaluating Color Descriptors for Object and Scene Recognition*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, 2010.

- [33] C. Carson, S. Belongie, H. Greenspan, J. Malik , “*Blobworld: Image segmentation using expectation-maximization and its application to image querying*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 1026–1038, 2002.
- [34] Y. Deng, B. S. Manjunath , “*Unsupervised segmentation of color-texture regions in images and video*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 800–810, 2001.
- [35] M. A. Hoang, J. M. Geusebroek, A. W. M. Smeulders, “*Color texture measurement and segmentation*”, Signal Processing, vol. 85, pp. 265–275, 2005.
- [36] T. Gevers, “*Adaptive image segmentation by combining photometric invariant region and edge information*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 848–85, 2002.
- [37] J. C. Van Germert, J. M. Geusebroek, C. G. M. Snoek, M. Worring, A. W. M. Smeulders, “*the challenge problem for automated detection of 101 semantic concepts in multimedia*”, In proceeding of the ACM International Conference on Multimedia, pp. 421–430, 2006.
- [38] S. Romberg, R. Lienhart, E. Hörster, “*Multimodal Image Retrieval fusing modalities with Multilayer multimodal pLSA*”. Int. Journal of Multimedia Information Retrieval (IJMIR), pp 31-44, 2012.
- [39] I. Boloch, J. P. Lecadre, H. Matre, “*Approches probabilistes et statistiques*”. In I. Bloch, ed., Fusion d’informations en traitement du signal et des images, Trait IC2, chap. 6, 87–118, Hermès, 2003.
- [40] M. Halvey, J. M. Jose, “*Bridging the gap between expert and novice users for video search*”. Int. Journal of Multimedia Information Retrieval (IJMIR), pp 17-29, 2012.
- [41] S. Redunac, M. Larson, “*Leveraging visual concepts and query performance prediction for semantic-theme-based video retrieval*”. Int. Journal of Multimedia Information Retrieval (IJMIR), pp 263-280, 2012.
- [42] C. Leacock, M. Chodorow, “*Combining local context and wordnet similarity for word sense identification*”. In WordNet: An Electronic Lexical Database In C. Fellbaum ED, MIT Press, pp 265-283, 1998.
- [43] Y.J. Jiang, C.W. Ngo, S.F. Shang, “*Semantic Context Transfer across Heterogeneous Sources for Domain Adaptive Video Search*”, ACM Multi Media, Beiging, China, 2009.
- [44] B. Thomee, M.S. Lew, “*Interactive search in image retrieval: a survey*”. Int. Journal of Multimedia Information Retrieval (IJMIR), pp 71-86, 2012.
- [45] Z. Sun, Y. Song, Y. Zheng, H. Yu, C. Jin, H. Lu, and X. Xue, “*Fudan University:hierarchical video retrieval with adaptive multimodal fusion,*” Proc. ACM International Conference on Content-Based Image and Video Retrieval, pp. 549-550, 2008.
- [46] S. Vrochidis, P. King, L. Makris, A. Mountzidou, V. Mezaris, and I. Kompatsiaris, “*MKLab interactive video retrieval system,*” Proc. ACM International Conference onContent-Based Image and Video Retrieval, pp. 563-564, 2008.
- [47] N. Elleuch, M. Zarka, A. Ben ammar, and A.M. Alimi, “*A fuzzy ontology-based framework for reasoning in visual video content analysis and indexing*”. In eleventh international workshop on Multimedia Data mining MDMKDD, 2011.



**Jamel SLIMI (M' 09)** is a phd student in the National Engineering School of Sfax, Tunisia. Member of the IEEE student Brench in Sfax, Tunisia. Member of the IEEE computer society. Member of the ReGIM Laboratory (Research Groups on Intelligent Machines). He received the Master degree in computer science (2005) from the National Engineering School of Sfax. He is also a assistant in the computer science department of the Higher Institute of sciences and technologies of Mateur (Tunisia). His research interests include multimedia indexing, knowledge management, multimedia data retrieval, data visualization and personalization.



**Anis Ben AMMAR (M' 06, SM' 12)** is assistant professor in Higher Institute of Business Administration of Sfax, Tunisia. He is Member of the IEEE computer society. He obtained a master degree and then a PhD both in Computer 1999 and 2003 in Paul Sabatier-FRANCE University. Anis Ben Ammar is a member of REGIM Laboratoy (REsearch Group on Intelligent Machines). His research interests include applications of intelligent Multimedia processing including patter recognition, audiovisual analysis, indexing and retrieval. He focuses his research on information retrieval system especially for social WEB.



**Adel M. Alimi (S'91, M'96, SM'00).** He graduated in Electrical Engineering in 1990. He obtained a PhD and then an HDR both in Electrical & Computer Engineering in 1995 and 2000 respectively. He is full Professor in Electrical Engineering at the University of Sfax since 2006. Prof. Alimi is founder and director of the REGIM-Lab. on intelligentMachines. He published more than 300 papers in international indexed journals

and conferences, and 20 chapters in edited scientific books. His research interests include applications of intelligent methods (neural networks, fuzzy logic, evolutionary algorithms) to pattern recognition, robotic systems, vision systems, and industrial processes. He focuses his research on intelligent pattern recognition, learning, analysis and intelligent control of large scale complex systems. He was the advisor of 24 Ph.D. thesis. He is the holder of 15 Tunisian patents. He managed funds for 16 international scientific projects. Prof. Alimi served as associate editor and member of the editorial board of many international scientific journals (e.g. "IEEE Trans. Fuzzy Systems", "Pattern Recognition Letters", "NeuroComputing", "Neural Processing Letters", "International Journal of Image and Graphics", "Neural Computing and Applications", "International Journal of Robotics and Automation", "International Journal of Systems Science", etc.). He was guest editor of several special issues of international journals (e.g. Fuzzy Sets & Systems, Soft Computing, Journal of Decision Systems, Integrated Computer Aided Engineering, Systems Analysis Modelling and Simulations). He organized many International Conferences ISI'12, NGNS'11, ROBOCOMP'11&10, LOGISTIQUA'11, ACIDCA-ICMI'05, SCS'04ACIDCA'2000.

Prof. Alimi has been awarded with the IEEE Outstanding Branch Counselor Award for the IEEE ENIS Student Branch in 2011, with the Tunisian Presidency Award for Scientific Research and Technology in 2010, with the IEEE Certificate Appreciation for contributions as Chair of the Tunisia Computational Intelligence Society Chapter in 2010 and 2009, with the IEEE Certificate of Appreciation for contributions as Chair of the Tunisia Aerospace and Electronic Systems Society Chapter in 2009, with the IEEE Certificate of Appreciation for contributions as Chair of the Tunisia Systems, Man, and Cybernetics Society Chapter in 2009, with the IEEE Outstanding Award for the establishment project of the Tunisia Section in 2008, with the International Neural Network Society (INNS) Certificate of Recognition for contribution on Neural Networks in 2008, with the Tunisian National Order of Merit, at the title of the Education and Science Sector in 2006, with the IEEE Certificate of Appreciation and Recognition of contribution towards establishing IEEE Tunisia Section in 2001 and 2000.

He is the Founder and Chair of many IEEE Chapters in Tunisia section. He is IEEE CIS ECTC Education TF Chair (since 2011), IEEE Sfax Subsection Chair (since 2011), IEEE Systems, Man, and Cybernetics Society Tunisia Chapter Chair (since 2011), IEEE Computer Society Tunisia Chapter Chair (since 2010), IEEE ENIS Student Branch Counselor (since 2010), He served also as Expert evaluator for the European Agency for Research. since 2009.

# Creating and Configuring Cloud Computing Environment

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

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

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

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

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

## I. INTRODUCTION

According to National Institute of Standards and Technology (NIST) [1] Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models. The essential characteristics include on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The service models include IaaS, PaaS, and SaaS. The four deployment models include private cloud, community cloud, public cloud and hybrid cloud. Cloud computing is becoming an effective platform to deliver services to millions of users. The cloud services vary from the three basic services to various emerging services evolving with the passage of time. Most of the governments are switching their public services infrastructure onto the cloud to handle the demands of providing public services to increasing population. Cloud is becoming an effective solution to provide Anything as a Service (XaaS) to the users. In SaaS model, the user uses the services by using cloud application (App). The user downloads the App on their electronic gadget and interacts with the resources of the provider using the App. The

user uses a thin client like a web browser or mobile phone to interact with the cloud service provider (CSP). The Cloud provider uses the virtualization technology to handle the SaaS requests of millions of users simultaneously. The real cloud environment encompasses of various entities interacting with each other, millions of user requests with different requirements. Thus the real cloud environment becomes complex and dynamic environment. The testing of the real cloud environment is a tedious and difficult task. Moreover cloud platforms are generally proprietary which adds more difficulty to test the cloud. Simulation techniques can be employed to test the performance of cloud. Simulation tools can be used to create and configure a virtual cloud environment. By using simulation tools, Cloud testing can be done in a controlled and monitored way. In this article, we are going to explore the simulation strategy to create a virtual cloud. We would explore CloudAnalyst simulation tool to simulate large-scale cloud environment. Then we use the simulation tool to simulate a virtual cloud.

The rest of the paper is organized as follows: Related work is discussed in section II. Literature is reviewed within the context of the simulation techniques and cloud simulation. In section III, we explore the CloudAnalyst simulation tool. We discuss the various classes and packages of the simulation tool. In section IV we use the simulation tool to create and configure a virtual cloud environment. We collect the data pertaining to real cloud environment and analyze the results. Finally in section V, conclusion and future scope are listed.

## II. RELATED WORK

The real cloud environment is complex and difficult to monitor and test. Simulation tools and techniques can resolve this problem and simulate cloud environments which can be tested in a controlled and monitored way. A lot of research has been carried out on this topic.

In [2] authors have proposed CloudSim toolkit that enables modeling and simulation of cloud environments. The proposed toolkit can model the data centers, virtual machines and various resource provisioning policies. It supports modeling and simulation of single cloud and inter-networked clouds. It allows the users to implement their own policies and resource provisioning policies. Finally, the usefulness of the proposed toolkit is demonstrated by involving dynamic provisioning of applications in the hybrid federated cloud environment.

In [3] authors have devised a new system for grid and cloud services simulation. The system improves the efficiency of grid and cloud systems development. The system uses quality indicators of real environment and incorporates it into the design process. To achieve it, the simulation program is linked with the real monitoring system of cloud or grid by using a special database.

In [4] authors have discussed the importance of monitoring tools in the efficient management of cloud. They have identified the practical capabilities that an ideal monitoring tool should possess to serve the objectives. Based on the identified capabilities, a classification of monitoring tools is presented to determine the strength and weakness of the tools.

In [5] authors have proposed GreenCloud, a packet-level simulator for energy aware cloud computing data centers. The simulator is designed to capture the energy consumption by the various components of the data center i.e. servers, switches, and links as well as packet-level communications. The simulation results obtained shows the effectiveness of the simulator in managing power applied to computing and networking units.

In [6] authors have reviewed the basic concepts of cloud computing security and cloud computing simulators. They indicate that there are two types of cloud simulators. The first type of cloud simulators is based on software only. The second type of simulators is based on both software and hardware. They analyze and compare the features of various cloud simulators.

In [7] authors have reviewed and compare the features of cloud computing simulation tools. They explore and compare the features of SPECI, GreenCloud, OCT, Open Cirrus, GroudSim, NetworkCloudSim, EMUSIM, DCSim, iCanCloud, CDOSim, TeachCloud, MDCSim, and GDCSim. Finally, they conclude that every simulator has its pros and cons and choice of simulation tool depends on the requirements of the user.

In [8] authors have proposed a cloud simulation environment with a scalable network and message passing application model. It enables accurate evaluation of scheduling and resource provisioning policies. The proposed environment helps in optimizing the performance of real cloud environment.

In [9] authors have proposed a simulation tool CloudAnalyst, for evaluation and modeling of cloud environments and applications. The tool studies the behavior cloud applications under various deployment configurations. It helps the developers to analyze the behavior of distributed applications on cloud infrastructure and helps in the optimization of applications performance.

Most of the work has been focused on the study and comparison of various tools for cloud simulation, suitability of various simulation tools as per the requirements of the users and the deployment of different types of applications on the cloud infrastructure and their behavior once they are deployed. In this article, we extend this work by exploring the features CloudAnalyst simulation tool. We use the tool to create a virtual cloud to simulate the real cloud environment.

### III. CLOUDANALYST [10]

CloudAnalyst is a CloudSim based cloud simulation tool. It is used to simulate complex and dynamic cloud environment. To understand the behavior of large scale distributed applications, various simulation tools like GridSim and CloudSim etc. can be employed, but to analyze the behavior of large scale internet applications, CloudAnalyst can be the best option. It simulates large- scale internet application on the virtual cloud infrastructure and helps in optimizing the performance of applications before actual deployment of the application. As shown in figure 1. CloudAnalyst simulator is built on top of CloudSim toolkit by adding some new extensions and graphical user interface (GUI). The new extensions add the capability to model internet and internet application behaviors and GUI adds visualization capabilities to the tool and separates programming exercise from simulation experiment set

up exercise. It enables the researcher to execute simulations in repeated fashion by changing the parameters without complex programming exercise. Moreover, the results produced after the simulation are graphical and can be easily analyzed and enhanced.

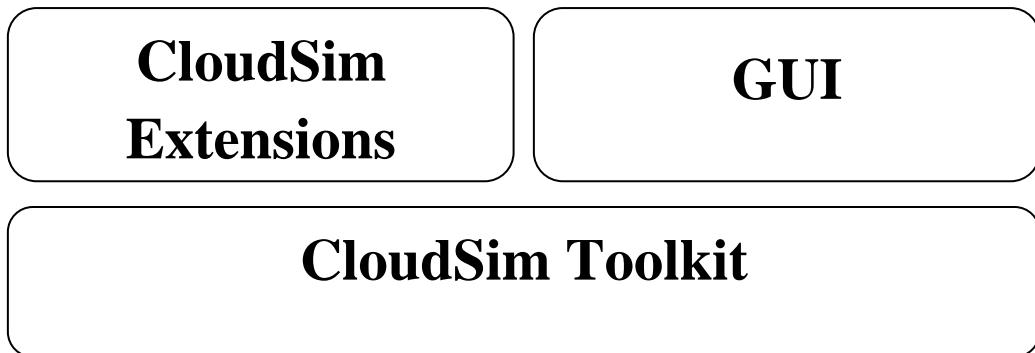


Figure 1. CloudAnalyst Simulator

The new extensions added to the CloudSim toolkit include various packages. These packages include classes and interfaces, which adds functionality to simulate internet applications and analyze its behavior in GUI form. The description of each package is as follows:-

Packages		
<b>Cloudsim.ext.datacenter</b>	<b>Cloudsim.ext.event</b>	<b>Cloudsim.ext.gui</b>
1. ActiveVmLoadBalancer.java 2. DatacenterController.java 3. RoundRobinVmLoadBalancer.java 4. ThrottledVmLoadBalancer.java 5. VirtualMachineState.java 6. VmLoadBalancer.java	1. BaseCloudSimObservable.java 2. CloudSimEvent.java 3. CloudSimEventListener.java 4. CloudSimEvents.java 5. CloudsimObservable.java	1. DataCenterUIElement.java 2. GuiMain.java 3. MachineUIElement.java 4. SimulationUIElement.java 5. UserBaseUIElement.java 6. VmAllocationUIElement.java
<b>Cloudsim.ext.gui.screens</b>	<b>Cloudsim.ext.gui.utils</b>	<b>Cloudsim.ext.servicebroker</b>
1. ConfigureSimulationPanel.java 2. InternetCharacteristicsScreen.java 3. ResultsScreen.java 4. SimulationPanel.java	1. AbstractListTableModel.java 2. CommunicationPath.java 3. MultilineTableHeaderRender.java 4. SimMeasure.java 5. SimGraph.java 6. SimpleTableModel.java	1. BestResponseTimeServiceBroker.java 2. CloudAppServiceBroker.java 3. DynamicServiceBroker.java 4. ServiceProximityServiceBroker.java
<b>Cloudsim.ext.stat</b>	<b>Cloudsim.ext.utils</b>	<b>Cloudsim.ext</b>
1. HourlyEventCounter.java 2. HourlyStat.java	1. CommPath.java 2. InternetEntityRegistry.java 3. IOUtil.java 4. MiscUtil.java 5. ObservableList.java 6. PdfExporter.java	1. Constants.java 2. GeoLocatable.java 3. Inernet.java 4. InternetCharacteristics.java 5. InternetCloudlet.java 6. Simulation.java 7. UserBase.java 8. WorldGeometry.java

The classes and interface defined in the various packages adds some new extensions to the CloudSim toolkit and enables the researcher to perform simulations in an easy way. A brief description of each class is given below. The list includes the name of the package, the classes, and interfaces in the package and the functionality of each class.

## Package : cloudsim.ext

Class / Interface	Brief Description
<b>Constants.java</b>	The Constants interface defines all the global constants to be used in various classes of the simulator. While creating a new VM load balancing policy or new service broker policy, the global constants have to be incorporated in this interface.
<b>GeoLocatable.java</b>	Any simulation entity that can be located physically in the simulation world e.g. datacenter should have to implement the GeoLocatable interface.
<b>Internet.java</b>	The Internet class emulates the real world internet. This class routes the request of UserBase to the destination DataCentreBroker and insert the transmission latency to the user request. It updates the progress of the simulation to the GUI.
<b>InternetCharacteristics.java</b>	The InternetCharacteristics class defines the behavior of internet i.e. it holds the data about network latency and bandwidth. It also includes methods to access the information.
<b>InternetCloudlet.java</b>	This class extends the Cloudlet class of CloudSim toolkit. It encapsulates the request and responses exchanged between the UserBase and DatcentreController.
<b>Simulation.java</b>	This class is the main class which acts as a controller of simulation.
<b>UserBase.java</b>	A User Base models a group of users that is considered as a single unit in the simulation and its main responsibility is to generate traffic for the simulation. This class extends CloudSim class and the body() method is used to periodically generate bursts of traffic representing the number of users online at the time. The number of users online at a particular time is calculated using the parameters peakHours, peakAvgUsers and offPeakAvgUsers and also by using a Poisson distribution to randomly vary the number in a realistic manner. The private inner class ResponseHandler is responsible for

accepting responses to the requests sent by UserBase. Therefore the traffic generation pattern is independent of receiving the responses to the requests.

#### **WorldGeometry.java**

This class defines the world geometry and it divides the world into six regions. The regions are indexed as 0, 1, 2, 3, 4, and 5 corresponding to six continents of the world.

### **Package: cloudsim.ext.datacenter**

#### **Class / Interface**

#### **Brief Description**

##### **ActiveVmLoadBalancer.java**

ActiveVmLoadBalancer class performs the task of load balancing among Virtual Machines (VMs). It balances the load among available VMs in such a way that all the VMs has same number of active tasks at any given time.

##### **DatacenterController.java**

DataCenterController class is used for managing a single data center. It inherits the DatacenterBroker class and overrides some of the default features of the inherited class for managing a single data center.

##### **RoundRobinVmLoadBalancer.java**

This class performs the task of load balancing. It balances the load among available VMs in a round robin fashion.

##### **ThrottledVmLoadBalancer.java**

The ThrottledVmLoadBalancer class balances the load among available VMs. Each VM is allocated a single task at a time and can be allocated another task only when it finishes the execution of the current task. This class does not implement queuing of tasks, but it returns valid VM only if available. The DatacenterController should implement task queuing locally.

##### **VirtualMachineState.java**

This enumerator holds the state of VMs. It has two values BUSY indicating VM is busy and AVAILABLE indicating VM is available for task execution.

##### **VmLoadBalancer.java**

All load balancing algorithms in CloudAnalyst simulator should inherit VmLoadBalancer class. This class provides the basic functionality of load balancing technique. It includes methods like getNextAvailableVm() and allocatedVM(int currVm) that defines the behavior of virtual machine load balancer.

## Package: **cloudsim.ext.event**

Class / Interface	Brief Description
<b>BaseCloudSimObservable.java</b>	The BaseCloudSimObservable class is an abstract class and it implements the CloudsimObservable interface. This class is used to observe the various events generated by different entities during the simulation process.
<b>CloudSimEvent.java</b>	The CloudSimEvent class implements serializable interface. This class provides the information about the events being generated in the simulation world.
<b>CloudSimEventListener.java</b>	CloudSimEventListener interface defines the contract for a listener object to the CloudSimEvent framework .
<b>CloudSimEvents.java</b>	The CloudSimEvents interface defines some static and constants which are used during the event handling mechanism process.
<b>CloudsimObservable.java</b>	The CloudSimObservable interface includes methods that are to add, remove and fire CloudSimEventListeners.

## Package: **cloudsim.ext.gui**

Class / Interface	Brief Description
<b>DataCenterUIElement.java</b>	The DataCenterUIElement class inherits SimulationUIElement class and implements serializable interface. This class holds the data pertaining to a specific data center in the simulation world.
<b>GuiMain.java</b>	The GuiMain class displays the welcome screen of the simulator. It includes buttons to run and cancel simulation etc.
<b>MachineUIElement.java</b>	The MachineUIElement class implements serializable interface. It holds the data about a specific physical machine in the simulation world.
<b>SimulationUIElement.java</b>	The SimulationUIElement is an abstract class and it implements serializable interface. This class acts as a base class for DataCenterUIElement and UserBaseUIElement class.
<b>UserBaseUIElement.java</b>	This class extends SimulationUIElement class and implements

serializable interface. This class encapsulates UserBase and holds the data about a specific UserBase. This information is used to construct a UserBase while initializing the simulation.

#### **VmAllocationUIElement.java**

The VmAllocationUIElement class implements serializable interface and holds the data specific to a data center VM allocation

### **Package: cloudsim.ext.gui.screens**

#### **Class / Interface**

#### **Brief Description**

#### **ConfigureSimulationPanel.java**

This class displays the configuration panel of the simulator. It displays a tabbed panel where the user can configure the various aspects of the simulator.

#### **InternetCharacteristicsScreen.java**

The InternetCharacteristicsScreen class is used to configure the various characteristics of the internet entity. This configuration setting controls the behavior of the internet. The user can configure the bandwidth and latency among different regions of the simulation world.

#### **ResultsScreen.java**

The ResultsScreen class displays the results of the simulation. The results are displayed both in tabular and graphical form.

#### **SimulationPanel.java**

This class displays the main simulation panel and it displays the world divided into six regions.

### **Package: cloudsim.ext.gui.utils**

#### **Class / Interface**

#### **Brief Description**

#### **AbstractListTableModel.java**

This class creates a table model. It gets the data from list data type and displays in the form of the table

#### **CommunicationPath.java**

This utility class is used by the GUI to represent a communication path.

#### **MultilineTableHeaderRender.java**

This class is used to display multi-line table headers.

#### **SimMeasure.java**

The SimMeasure class is used to transfer the statistical measures gathered from the simulation onto the screens.

### **SimGraph.java**

The SimGraph class is used to create a simple area graph based on the statistical data gathered during the simulation. It displays the graph on the screen.

### **SimpleTableModel.java**

This class extends the AbstractListTableModel class. It is used to create a simple table model to be used in the screen.

## **Package: cloudsim.ext.servicebroker**

### **Class / Interface**

### **Brief Description**

#### **CloudAppServiceBroker.java**

CloudAppServiceBroker interface defines the contract for a service broker. It defines a single method getDestination(GeoLocatable inquirer) which returns the name of selected datacenter based on service brokerage policy used. This interface needs to be implemented by all the service broker policies.

#### **ServiceProximityServiceBroker.java**

The ServiceProximityServiceBroker class implements closest data center policy. This broker policy selects the datacentre which is closest to the user base generating the request. In case two or more data centers are available in the selected region, it selects the data center randomly.

#### **BestResponseTimeServiceBroker.java**

This service broker policy class selects the datacenter with optimize response time i.e. the data center with best response time. First, it inquires about the data centers which are close to the request generating area. It then finds out the best performing data center among rest of the datacenters. It then compares the response time of closest data center and already found best datacenter. Based on this, it selects the data center with best response time.

#### **DynamicServiceBroker.java**

The DynamicServiceBroker class extends ServiceProximity-ServiceBroker and implements CloudAppServiceBroker interface. This class is not fully implemented in the simulator. The basic idea is it increases or decreases the VMs depending upon the load of the data center.

## **Package: cloudsim.ext.stat**

### **Class / Interface**

### **Brief Description**

**HourlyEventCounter.java**

This class records a list of events occurring during the day. The list is obtained after grouping hourly events.

**HourlyStat.java**

This class records the statistics occurring during the day. The statistics is obtained after grouping hourly statistics measures.

**Package: cloudsim.ext.util**

**Class / Interface**

**Brief Description**

**CommPath.java**

This class provides the information about the communication delays among the various regions of the simulation world.

**InternetEntityRegistry.java**

The InternetEntityRegistry maintains a listing of all the entities in the simulation world in several indexed forms for fast searching.

**IOUtil.java**

This class provides functionality for serializing objects as XML and deserializing XML as objects.

**MiscUtil.java**

This utility class contains miscellaneous methods used during the simulation process.

**ObservableList.java**

The ObservableList class implements a list whose additions and removals can be monitored.

**PdfExporter.java**

The PdfExporter class is used to export the results in the PDF format.

**IV. EXPERIMENTAL USAGE OF CLOUDANALYST SIMULATOR**

The CloudAnalyst simulation tool is simple and easy to use. This simulation tool is open source and can be easily modified as per the requirements of the users. In this section, we would collect real data from various sources and create a simulated cloud environment using CloudAnalyst simulation tool. As this tool, tests the behavior and performance of the large-scale internet application on the cloud, we would select social networking site like Facebook and gather the data pertaining to Facebook from various sources. According to socialbakers.com [11], the social networking giant Facebook has 1.09 billion Daily Active Users (DAU) and 1.65 billion Monthly Active Users (MAU). 90% of these users are mobile users i.e. they use mobile devices like smart-phones and gadgets to access the App of this social networking giant. Mobile technology and the growth of smart-phones has fuelled the growth of Facebook. On an average, a user spends 20 minutes per Facebook visit. The United States of America leads the list of countries in number of Facebook users. Facebook is an important asset for social marketers. The table 1 below lists the number of Facebook users in six continents of the world up to 15<sup>th</sup> of November, 2015. For

the purpose of simulation, we assume that during peak hour's number of simultaneous users are 5% of the number of users in the table given below and during off peak hours the users are taken as 1/10 of the number of users during peak hours.

Table 1. Continent wise Facebook users up to 15.11.2015

Continent	Facebook users up to 15.11.2015	Number of simultaneous Peak hour Users (5 % of Facebook User) (Assumption)	Number of simultaneous Off Peak hour Users (1/10 of Peak hour users) (Assumption)
<b>North America</b>	213,075,500 [12]	10653775	1065377
<b>South America</b>	210,874,200 [12]	10543710	1054371
<b>Europe</b>	309,576,660 [13]	15478833	1547883
<b>Asia</b>	503,708,200 [14]	25185410	2518541
<b>Africa</b>	124,568,500 [15]	6228425	622842
<b>Australia and Oceania</b>	18,239,110 [16]	911955	91195
<b>Total</b>	<b>1380,042,170</b>	<b>69002108</b>	<b>6900209</b>

Initially, Facebook outsources its datacenter needs and processes to its neighbors in the silicon valley. From 2011 onwards Facebook starts developing in-house data centers to cater the needs of its ever-increasing users. The location of various data centers of Facebook is as under:

Table 2. Location of Facebook Data Centers

Sr. No.	Location of Data Center [17]	Country	Continent
1.	Prineville, Oregon	United States of America	North America
2.	Altoona, Iowa	United States of America	North America
3.	Forest City, North Carolina	United States of America	North America
4.	Luleå, Sweden	Sweden	Europe
5.	Fort Worth, Texas	United States of America	North America
6.	Clonee, Ireland*	Ireland	Europe

\*Datacenter under construction

Network latency is the delay added by the network as the request travels from source to destination and vice versa. The network latency[10] matrix values (in milliseconds) is shown as under:

Table 3. Network Latency Matrix (in milliseconds)

Region Id / Region		0	1	2	3	4	5
		North America	South America	Europe	Asia	Africa	Ocenia
0	North America	25	100	150	250	250	100
1	South America	100	25	250	500	350	200
2	Europe	150	250	25	150	150	200

<b>3</b>	<b>Asia</b>	250	500	150	25	500	500
<b>4</b>	<b>Africa</b>	250	350	150	500	25	500
<b>5</b>	<b>Ocenia</b>	100	200	200	500	500	25

The bandwidth matrix provides the available bandwidth among various regions of the world. The bandwidth matrix [10] values (in Mbps) is shown below:

Table 4. Bandwidth Matrix (inMbps)

Region Id / Region		<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
		North America	South America	Europe	Asia	Africa	Ocenia
<b>0</b>	<b>North America</b>	2000	1000	1000	1000	1000	1000
<b>1</b>	<b>South America</b>	1000	800	1000	1000	1000	1000
<b>2</b>	<b>Europe</b>	1000	1000	2500	1000	1000	1000
<b>3</b>	<b>Asia</b>	1000	1000	1000	1500	1000	1000
<b>4</b>	<b>Africa</b>	1000	1000	1000	1000	500	1000
<b>5</b>	<b>Ocenia</b>	1000	1000	1000	1000	1000	2000

The other parameters [10] used to configure the simulation are shown in table below:

Table 5. Configuration parameters of CloudAnalyst Simulation Tool

Parameter	Value Assigned				
Cost per VM per hour (1024Mb, 100MIPS)	\$ 0.10				
Cost per memory	\$ 0.05				
Cost per storage	\$ 0.10				
Cost per 1Gb of data transfer (from/to Internet)	\$ 0.10				
Data Center – Number of Virtual Machines	DC1	DC2	DC3	DC4	DC5
	50	50	50	50	150
VM Image Size	10000				
VM Memory	1024 Mb				
VM Bandwidth	1000				
Data Center – Architecture	X86				
Data Center – OS	Linux				
Data Center – VMM	Xen				
Data Center – Number of Physical Machines	DC1	DC2	DC3	DC4	DC5
	20	20	20	20	60
Data Center – Memory per Machine	204800 Mb				

Data Center – Storage per machine	100000000 Mb
Data Center – Available BW per Machine	1000000
Data Center – Number of processors per machine	4
Data Center – Processor speed	10000 MIPS
Data Center – VM Policy	Time Shared
User Grouping Factor	1000
Request Grouping Factor	100
Executable Instruction Length	250

Using the information mentioned in the above tables, the simulation is configured. Various combinations of service broker policy and load balancing policy are worked and the results obtained are listed in the table below.

Table 6. Results of Simulated Cloud

Sr. No.	Service Broker Policy	Load Balancing Policy	Average Data Center Request Processing Time (ms)	Average Overall Response Time (ms)	Cost (in \$)
1	Closest Data Centre	Round Robin	257.62	622.89	\$ 111.33
2	Closest Data Centre	Equally spread current execution load	263.85	630.10	\$ 111.33
3	Closest Data Centre	Throttled	132.98	502.34	\$ 111.33
4	Optimise Response Time	Round Robin	260.06	628.26	\$ 111.33
5	Optimise Response Time	Equally spread current execution load	264.28	622.79	\$ 111.33
6	Optimise Response Time	Throttled	133.05	505.25	\$ 111.33

The following points can be concluded from the results obtained:

1. The combination of closest data center service broker policy and throttled virtual machine load balancing algorithm performs better than other combinations. In this scenario, average data center request processing time is 132.98 ms and average overall response time is 502.34 ms.
2. The Cost of application deployment is same in all the combinations.
3. Asia with region id 3 has a large number of users, if one or more data centers in North America is moved to Asia region, then it would produce better results.

## V. CONCLUSION AND FUTURE SCOPE

Cloud computing is an emerging area of research among academicians and industrialists. Doing research in cloud computing is difficult because real cloud environment is complex and dynamic. Study and usage of cloud simulation tools are required to simulate real cloud environment and perform testing in a controlled and monitored way.

In this article, we try to address this problem. First, we explore CloudAnalyst, a simulation tool to simulate cloud environment. We explore the various packages of the tool and give a brief description of the various classes within these packages. Finally, we showed the usage of the simulation by collecting data of real cloud environment. Data pertaining to Facebook is collected and a virtual cloud is simulated by using the collected data.

In future, we would try to incorporate more performance parameters in CloudAnalyst open tool to analyze the performance of cloud. We would also try to design new service broker policy or virtual machine load balancing policy to improve the performance of the cloud.

## REFERENCES

- [1] <http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf> on 20.05. 2016.
- [2] R.N. Calheiros, R. Ranjan, A. Beloglazov, C.A.F.D. Rose, and R. Buyya , “CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms”, in Software- Practice and Experience, pp 41:23-50, 2011.
- [3] V. Korenkov, A. Nechaeskiy, G. Ososkov, D. Pryahina. V. Trofimov, A.Uzhiniskiy, and N. Balashov, “ Web- Service Development of the Grid-Cloud Simulation Tools”, in Procedia Computer Science, volume 66, pp 533-539, 2015.
- [4] K. Fatema, V.C. Emeakaroha, P.D. Healy, J.P. Morrison, and T. Lynn, “A Survey of Cloud monitoring tools: Taxonomy, capabilities and objectives”, in Journal of parallel and Distributed Computing, pp 2918-2933, 2015.
- [5] D. Kliazovich, P. Bouvry, and S. U. Khan, "A Packet-level Simulator of Energy- aware Cloud Computing Data Centers," Journal of Supercomputing, vol. 62, no. 3, pp. 1263-1283, 2012.
- [6] W. Zhao, Y. Peng, F. Xie, and Z. Dai, "Modeling and simulation of cloud computing: A review", in Cloud Computing Congress (APCloudCC), pp20-24, 2012.
- [7] U. Sinha, and M. Shekhar, "Comparison of Various Cloud Simulation Tools available in Cloud Computing", in International Journal of Advanced Research in Computer and Communication Engineering, Volume4, Issue 3, pp171-176, 2015.
- [8] S.K. Garg, and R. Buyya, "An environment for modeling and simulation of message-passing parallel applications for cloud computing", in Software-Practice and Experience, 2012.
- [9] B. Wickremasinghe. R.N. Calheiros, and R. Buyya, "CloudAnalyst: A CloudSim Based Visual Modeller for Analysing Cloud Computing Environments and Applications", in the proceedings of 24<sup>th</sup> IEEE International Conference on Advanced Information Networking and Applications, pp 446-452, 2010.
- [10] B. Wickremasinghe, “CloudAnalyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale Cloud Computing Environments”, MEDC Project Report, University of Melbourne, 2009.
- [11] <http://www.socialbakers.com/statistics/facebook/> on 06.07.2016.
- [12] <http://www.internetworldstats.com/stats2.htm> on 06.07.2016.
- [13] <http://www.internetworldstats.com/stats4.htm> on 06.07.2016.
- [14] <http://www.internetworldstats.com/stats3.htm> on 06.07.2016.
- [15] <http://www.internetworldstats.com/stats1.htm> on 06.07.2016.
- [16] <http://www.internetworldstats.com/stats6.htm> on 06.07.2016.
- [17] <http://www.zdnet.com/pictures/facebook-data-centers-worldwide-by-the-numbers-and-in-pictures> on 07.07.2016.

# Goal Modeling Techniques for Requirements Engineering

Nagy Ramadan Darwish<sup>1</sup>, Bassem S.M. Zohdy<sup>2</sup>

<sup>1</sup>*Information Systems Department, Cairo University,  
Cairo.*

<sup>2</sup>*BTEC Department, Canadian International College  
Cairo*

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

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

## I. INTRODUCTION

Recently the term goal is used in requirements engineering techniques, goals promoted into requirements engineering for a lot of causes, this is due to different and enormous activities and objectives of RE [14]. Goals in some RE are playing a fundamental role, researchers in this domain made a lot of efforts to prove the importance, significance and usefulness of this concept [2]. As RE pays attention to real life observation, in order to translate these observations to mathematical specification language [1]. The importance of goal could be very powerful in; requirements acquisition, as to how relate the theses goal to business situation as it gives a detailed description of this business and how it is operate, also it helps to clarify requirements, dealing with conflicts, aiding software design [13]. This research is organized as follows; first section introduce what is a goal and differences between goals and requirements, in the second section a detailed description of types of goals and the source of extracting goals, in the third part; elaboration of goal modeling techniques, which implies the birth of goal modeling techniques, the reason of using goal modeling, the benefits of using goal modeling, also when to use goal modeling, and goal oriented approaches, in fourth part an illustration of early and late requirements goal modeling techniques.

## II. GOALS, REQUIREMENTS AND GOALS ACTIVITIES

Any developed system can be judged or evaluated due to the degree of meeting the purpose it was done for, so in the first place identifying the purpose of what the system it was done for considered one of the main activities in software development. Goal oriented approaches had been risen to meet multi-stakeholders criteria and handling the consistency between their actions, in recent years the popularity of goal oriented approaches has increased due to; Inadequacy of traditional systems with more and more complex systems; most techniques focus on modeling and specifications of software alone; also non function requirements are left outside requirements specification mostly, in addition, the traditional techniques don't allow alternative system configurations. While goal oriented approaches can handle the previously specified reasons.

In order to distinguish between goals and requirements, first identifying the goal, according to van Lamsweerde [15, 16], goal is an objective achieved by the system through cooperation of agents in the software to be, which implies that goal is purpose of the system or software to be, another definition for goal by Anton, goal is high level objectives of the business, organization or system. Also this definition implies that goal is related to business or any organization stakeholders, as it is also implies that goals are purposes of system to be.

As the definitions of goal illustrated, this could lead to differentiate between goals and requirements, as goal is a high level objective for organization or business, and objective achieved by system through cooperation of agents in a software to be, as requirements is an objective achieved under the responsibility of one agent [15, 16], this could imply that goal is broader term than requirements, and this is the reason of rise the techniques of goal modeling techniques.

Goal oriented requirement engineering contains the following activities [14], first is goal elicitation, as stated in literature identifying goals is not an easy task, as goals could be stated by different means like stakeholders or any other organizational sources of information and then these goals could be passed to software engineers, most frequently goals are implicit and therefore the elicitation process must take place. A preliminary analysis of the current system/organization is an important source of goal identification. This analysis can result in a list of problems and deficiencies that can be precisely formulated. as mentioned goals could be elicited from many sources of information, these sources could be multiple kinds of documents. it is noted that stakeholders tend to express their requirements in terms of operations or actions, rather than goals. So, it makes sense to look for action words such as “schedule” or “reserve” when gathering requirements for a meeting scheduler system. Also another activity is goal refinement; once goals have been identified the aim is usually to refine them into progressively simpler goals until these goals can be easily operationalized and implemented. This process is usually done by asking the HOW questions and refining goals through AND/OR refinements. Many GORE approaches stress that when determining how a high-level goal can be refined, one need to consider alternative ways of refining it to make sure that as many options as possible are explicitly represented in goal models and analyzed with respect to high-level criteria. After goal refinement here comes a various types of goal analysis or sometimes called obstacle analysis, Potts identifying obstacles for specific goals by asking certain questions, Goal-Based Requirements Analysis Method (GBRAM) uses same method in managing obstacles, once obstacle is identified, a set of events must be built, and these events identifies why goal could be fail, so those built events could be incarnation of obstacles for each goal [9]. it is noteworthy that more specific goal leads in more preventing the obstacles to be. As obstacles can be formally identified as follows. Given a formal specification for a goal G, calculate the preconditions for obtaining the negation of G from the domain theory. Each obtained precondition is an obstacle [15]. Another activity is assigning goals to agents, this in KAOS approach, and agents are assigned leaf-level goals based on their abilities. The process is quite like GBRAM. KAOS permits requirements engineers to analyze substitution configurations of the boundary between the system-to-be and its environment through the use of OR responsibility links. That is why it is possible to compare several system configurations. In GBRAM several agents could be responsible for the same goal at different times.

### III. TYPES OF GOALS

Goals could be categorized in many different ways according to literature, according to [13] goals can be classified as functional and non-functional goals, functional goals means services that expected to be output of the system, as non-functional goals means the qualities which system will deliver, such as security, customizability, and flexibility as an example, according to [dar93], goals could be classified as satisfaction goals and information goals, information goals can be defined as functional goals involved in keeping agents informed with objects states, as satisfaction goals are concerned with replying to agents questions. Goals also could be classified as Performance goals, which are specialized into time and space performance goals, the former being specialized into response time and throughput goals [15]. Security goals are specialized into confidentiality, integrity and availability goals [15]; the latter can be specialized in turn until reaching domain-specific security goals.

Another classification made is between soft goals, whose satisfaction cannot be established in a clear-cut sense [192], and hard goals whose satisfaction can be established through verification techniques [15]. Soft goals are especially useful for comparing alternative goal refinements and choosing one that contributes the most to them.

Another classification axis is based on types of temporal behavior prescribed by the goal. [15]. Achieve (resp. cease) goals generate system behaviors, in that they require some target property to be eventually satisfied in some future state (resp. denied); maintain (resp. avoid) goals restrict behaviors, in that they require some target property to be permanently satisfied in every future state (resp. denied) unless some other property holds. Optimize goals compare behaviors to favor those, which better ensure some soft target property.

Goal types and taxonomies are used to define heuristics for goal acquisition, goal refinement, requirements derivation, and semi-formal consistency/completeness checking [15], or to retrieve goal specifications in the context of specification reuse [15].

#### IV. GOAL MODELING TECHNIQUES

Goal oriented approaches have been linked to the four RE activities as shown in the below table (Table 1) linking to them goal oriented approaches that can handle requirements within each activity.

<i>RE Activity</i>	<i>Goal Analysis Contribution</i>	<i>Goal-Oriented Approach</i>
• requirements elicitation	1. understanding the current organisational situation,	GOMS, Goal-based Workflow, <i>i*</i> , EKD
	2. understanding the need for change	ISAC, F <sup>3</sup>
• requirements negotiation	3. providing the deliberation context of the RE process	SIBYL, REMAP, The reasoning loop model
• requirements specification	4. relating business goals to functional and non-functional system components	KAOS, GBRAM , the NFR framework, the Goal-scenario coupling frame work
• requirements validation	5. validating system specifications against stakeholders' goals	GSN, GQM

**Table (1)** Goal oriented approaches' link with RE activities [14]

One of the most popular examples for late requirements goal modeling techniques is the UML. UML was originally based on the object oriented modeling technique whose aim is to provide a standard way to visualize the design of a system. UML has two types of views: systematic (or structural) view and dynamic (or behavioral) view. In spite of being one of the most known techniques for modeling late requirements, UML suffers many disadvantages such as: it is so time consuming, we can't identify exactly who benefits from UML, the UML diagram might be overcomplicated for the customer to understand and the emphasis most of the times is mainly on the design which annoys the developer in his work a lot.

In order to model early requirements properly, agent oriented approach has been defined to analyze such requirements and select the best alternatives that results later in achieving their goals. From the popular early requirements goal models are:

#### I. *i\** framework

*i\** framework has been proposed by Eric Yu [17] which later was discovered for many drawbacks that lead to the appearance of variants of *i\** which are Tropos and GRL [3]. The *i\** framework has two ways of representation: the graphical and the formal representations. As for the graphical representation, it has two main modeling components; the Strategic Dependency (SD) model and the Strategic Rationale (SR) model. The SD model is concerned with describing mainly the dependency relationship between actors within the organization and as a result, it helps in understanding how a certain goal is embedded within the organization. Therefore, it always answers the "Whys" questions. On the other hand, the SR model is concerned with describing stakeholders' interests and how they might be addressed through various system configurations, and stakeholders' evaluation of various alternatives respecting their interests [4]. *i\** actors can be categorized into agents, positions, and roles. An agent occupies a position, a position covers a role. Consequently an agent plays role covered by a certain position. Actors and their categorization can also be decomposed into another actor through the is-part-of relationship type. Actors have some intentional properties such as goal, belief, attributes, and commitment. *i\** framework provides a number of analyzing levels in terms of ability, workability, viability, and believability. *i\** models and analyzes security and privacy requirement using secure *i\** (SI\*) [18].

The *i\** basic elements (shown in figure 1) are: The intentional elements [3]: which are goals, soft goals, tasks, resources of an actor boundary. Links: which connect the *i\** model elements together through means-end, task-decomposition or contribution links (positive (Make/Help links), negative (Hurt/ Break links) and the unknown. The satisfaction levels. Reasoning elements: they are represented through routines, rules and beliefs.

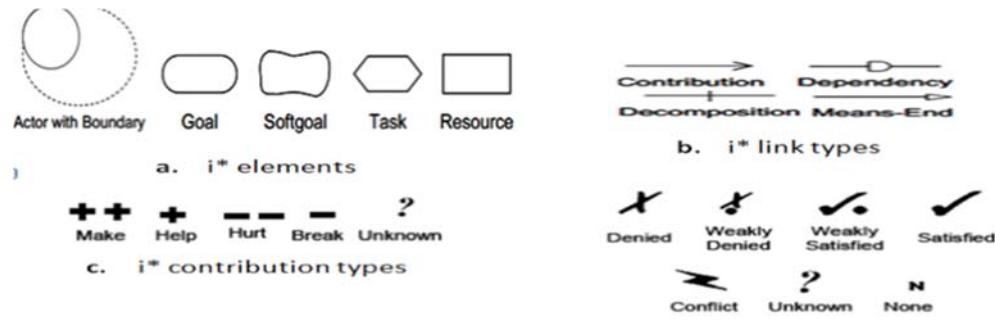


FIG. (1) THE I\* BASIC ELEMENTS

For large and complex organizations, the  $i^*$  graphical representation is not suitable as a result,  $i^*$  formal notation is used. The formal notation uses the First Order Logic (FOL). Six predicates are used indicating the possible six analysis labels  $v = \{FS(), PS(), FD(), PD(), C(), U()\}$  where:  
 $FS()$ : represents the fully satisfied,  $PS()$ : stands for the partially satisfied,  $FD()$ : is the fully denied label,  $PD()$ : represents the partially denied,  $C()$ : stands for the conflict label and the  $U()$ : is the unknown label. The predicate holds when the label applies. Examples for propagation rules for labels are found in fig. 2.

Link Type		Original Rule
Dependency		$(v \in V) v(i_s) \Rightarrow v(i_d)$
Decomposition		$(\bigwedge_{j=1}^n FS(i_j)) \Rightarrow FS(i_d)$ $\dots$ $(\bigvee_{j=1}^n FD(i_j)) \Rightarrow FD(i_d)$
Contribution		$(c = Make) FS(i_s) \Rightarrow FS(i_d)$ $(c = Help) FS(i_s) \Rightarrow PS(i_d)$ $\dots$ $(c = Unk, v \in V) v(i_s) \Rightarrow U(i_d)$ $(c \in \{Make, Help, \dots\}) U(i_s) \Rightarrow U(i_d)$

FIG. 2. I\* FORWARD PROPAGATION RULES

**i\* tools.** In spite of the existence of many tools supporting  $i^*$  language such as OME and REDEPEND, it always suffered from having some strange situations which mainly appears due to the incompleteness of the formalization of  $i^*$ . The reason for this incompleteness is the existence of many confused situations, e.g., the “is-a” relation is used profusely, however, it is not defined as a  $i^*$  constructor. Moreover, many incomplete definitions exist, e.g., no indication about the type and the number of roots in the internal decomposition of an actor. In addition to the presences of some ambiguous definitions, e.g., the dependency link must have two different importance degrees each implies one of the involved actor (depender, dependee).

## II. Tropos

Tropos is an agent-oriented software development methodology. It adopts  $i^*$  model. Tropos supports four phases from the software development, which are early requirements, late requirements, architectural design, and detailed design. The early requirement phase is intended to understand the organizational context that the system-to-be will be built on. The Late requirement phase is concerned with defining the system-to-be's functional and non-functional requirements. The architectural design is concerned with defining the system global architecture, whereas the detailed design concerned with defining the behavior of each software component in more details. Tropos can represent organizational goals either graphically or formally [8]. Tropos models and analyzes security and privacy requirement using secure Tropos [19].

**Tropos graphical representation** has two diagrams for modeling and analysis of organizational requirements and goals, which are the actor diagram and the rationale diagram. The actor diagram, similar to SD model in  $i^*$  while the rationale diagram is similar to SR model in  $i^*$ . Tropos basic elements are the same as those of  $i^*$  differing only in the links types, where Tropos

provides AND/OR decomposition link instead of the task- decomposition of i\*. Tropos Formal Representation is preferred to model large and complex organizations requirements. It uses the FOL of that of i\*. Tropos forward propagation rules are found in figure 3. Backward propagation rules are found in figure 4[12].

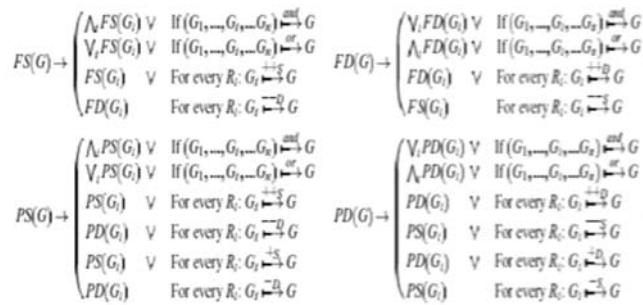
Goal	Invariant Axioms
$G :$	$FS(G) \rightarrow PS(G), FD(G) \rightarrow PD(G)$
<b>Goal relation</b>	<b>Relation Axioms</b>
$(G_1, \dots, G_i, \dots, G_n) \xrightarrow{\text{and}} G :$	$(\bigwedge_i FS(G_i)) \rightarrow FS(G), (\bigwedge_i PS(G_i)) \rightarrow PS(G)$
	$\bigwedge_i (FD(G_i) \rightarrow FD(G)), \bigwedge_i (PD(G_i) \rightarrow PD(G))$
$(G_1, \dots, G_i, \dots, G_n) \xrightarrow{\text{or}} G :$	$(\bigwedge_i FD(G_i)) \rightarrow FD(G), (\bigwedge_i PD(G_i)) \rightarrow PD(G)$
	$\bigwedge_i (FS(G_i) \rightarrow FS(G)), \bigwedge_i (PS(G_i) \rightarrow PS(G))$
$G_2 \xleftarrow{\pm S} G_1 :$	$PS(G_2) \rightarrow PS(G_1)$
$G_2 \xleftarrow{\pm D} G_1 :$	$PS(G_2) \rightarrow PD(G_1)$
$G_2 \xleftarrow{\pm S} G_1 :$	$FS(G_2) \rightarrow FS(G_1), PS(G_2) \rightarrow PS(G_1)$
$G_2 \xleftarrow{\pm D} G_1 :$	$FS(G_2) \rightarrow FD(G_1), PS(G_2) \rightarrow PD(G_1)$
$G_2 \xleftarrow{\pm D} G_1 :$	$PD(G_2) \rightarrow PD(G_1)$
$G_2 \xleftarrow{\pm D} G_1 :$	$PD(G_2) \rightarrow PS(G_1)$
$G_2 \xleftarrow{\pm Q} G_1 :$	$FD(G_2) \rightarrow FD(G_1), PD(G_2) \rightarrow PD(G_1)$
$G_2 \xleftarrow{\pm Q} G_1 :$	$FD(G_2) \rightarrow FS(G_1), PD(G_2) \rightarrow PS(G_1)$

**Fig. (3) TROPOS FORWARD PROPAGATION RULES**

Tropos Tools Tropos forward and backward propagation is supported by the Goal Reasoning Tool (Gr-Tool<sup>1</sup>). The Gr-Tool is a graphical tool for representing goal models and applying the required analysis algorithm. GOALSOLVE and GOALMINSOLVE tools are implemented to support the backward propagation first and second approach.

### III. Goal-oriented Requirement Language (GRL)

GRL is an agent-oriented and goal-oriented modeling language that supports reasoning about non-functional requirements and quality attributes. It is influenced by both the i\* and the NFR frameworks for specifying non-functional requirements. Recently, the International Telecommunication Union (ITU-T) as a part of the User Requirements Notation [5] standardizes GRL.



**Fig. 4. Tropos backward propagation rule**

User Requirements Notation (URN) is used to support all the RE phases. It allows requirement engineers to elicit and specify requirements then analyze such requirements to ensure its completeness and correctness. URN consists of two complementary languages which are Goal-oriented Requirement Language (GRL) and Use Case Map (UCM) [6].

The benefit from using GRL is that it can be integrated with a scenario notation and can define a clear separation between its elements and their graphical representation. Moreover, it enables a scalable and consistent representation for multiple diagrams for the same goal model.

It has been influenced by i\* language in: GRL goal model has three basic concepts which are actors, intentional elements and links, GRL links types are decomposition, contribution, dependencies, The main differences respecting i\* are that GRL offer a new link type namely, correlation link. Correlation link describes the side effects of an element on another element. Also, GRL offers constructors for enabling relationships with external elements. Moreover, the use of URN links and metadata for enabling linking GRL with UCM elements. However, GRL supports only one type of actor, whereas i\* support the notations of role, agent, and position.

There exist many GRL tools; some of them are jUCMNav and OME. JUCMNav tool is an Eclipse plug-in for the creation, analysis, and transformation of URN models.

Amyot presented a tool providing a lightweight profile for GRL that enables creating a goal model in i\* style. Such profile supplements GRL with i\* missing concepts using the advantage of URN links and metadata. Moreover, it restricts the usage of GRL to i\* through the usage of UML's Object Constraint Language (OCL). The tool is implemented in the jUCMNav tool.

The goal satisfaction analysis procedure is applied on a goal model in order to select alternatives that aim to satisfy the desired goal. Goal satisfaction is either forward or backward propagation whether it is qualitative or quantitative [7].

**Forward propagation.** The forward propagation starts by initializing a set of alternative with a satisfaction value, and then propagates such values upward iteratively through links and forward propagation rules until reaching the top goals [20].

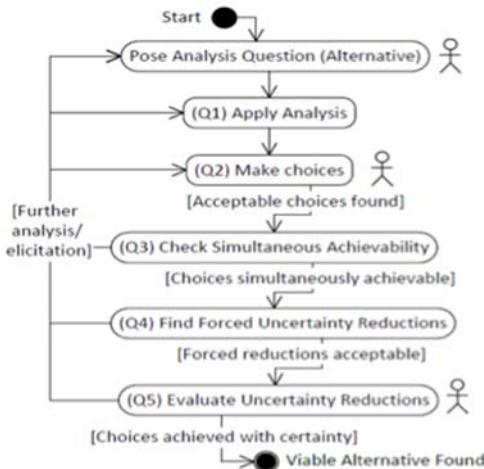
**Backward propagation.** The backward propagation starts by initializing the top desired goals with satisfaction values, and then propagates such values downward iteratively through links and backward propagation rules. i\* framework analysis supports both forward and backward propagation analysis on both qualitative and quantitative.

Tropos analysis supports both forward and backward propagation analysis on both qualitative and quantitative. It solves the back propagation problem using two approaches. In the first approach, it reduces the problem of the backward propagation to that of the propositional satisfiability (SAT). In the second approach, Tropos tries to find the set of alternatives with minimum cost that achieve the desired top goals using the Minimum-Weight Propositional Satisfiability (MW-SAT). GRL analysis supports both forward and backward propagation analysis on qualitative, quantitative and hybrid analysis.

In real world applications, most of the goal models suffer from uncertainty and this is due to having gaps in the knowledge domain, disagreement between stakeholders or the presence of uncertainty over requirement details. Therefore, it is important to handle uncertainty since ignoring uncertainty may lead to selecting alternatives that may not be sufficient to achieve the desired goals or eliminating viable alternatives.

Horkoff proposed a semi-automated tool to handle uncertainty in early requirements represented in i\* using MAVO framework in addition to her proposed formal analysis formula based on the forward qualitative analysis [10]. Horkoff goal model analysis methodology is depicted in fig. 5 below. Initially, it constructs a set of all possible concretizations where each of them represents a concrete goal model results from resolving a specific uncertainty requirement. Then, for each concretization she applies forward qualitative analysis. Afterwards the user is allowed to select her choices for each top goal, which is then checked for simultaneous availability in any of the possible set of concretization. Reduced uncertainty respecting the user acceptability and the domain consistency is evaluated for the concretization results in the simultaneous achievable choices. Any appearing changes will lead to further analysis.

The results for contextual goal modeling and reasoning framework defined that more than one alternative can satisfy the top desired goal. The decision of selecting among them depends on the applied context. Thus, it is important to enrich goal models with context. Consequently, Raian proposed a goal-oriented requirement engineering modeling and reasoning framework for systems operating under various contexts [11]. The framework relates context and goals using Tropos. Then, it analyzes all contexts to identify ways for verifying them. In order to derive requirements reflecting a certain context automatically, the framework proposes two reasoning techniques, namely, design time and runtime reasoning techniques. The runtime reasoning technique concerns deriving goal model variants reflecting context and user priorities. The design time reasoning technique concerns deriving requirements for the system-to-be with minimum cost and valid in all considered contexts.



**Fig. 5.** Horkoff goal model analysis methodology

## V. CONCLUSION

It has been realized that goal-modeling techniques are being preferred than traditional methods such as object oriented approaches. This research showed that requirements are considered to be a subset of goals and in return goals have been the major umbrella that concerns everyone for modeling. Moreover, this research found that late requirements modeling haven't got so much interest such as that of early requirements modeling. An illustration of three well known early requirements models which are; i\* framework, Tropos and GRL. Also, the research addressed in brief UML, which is considered the most popular late requirements modeling technique.

## REFERENCES

1. P. Save, ``Classification of Research Efforts in Requirements Engineering," ACM Computing Surveys, vol. 29, no. 4, 1997, pp. 315-321.
  2. Yu ESK, Mylopoulos J , 1998 Why goal-oriented requirements engineering. In: Dubois E, Opdahl AL, Pohl K (eds) Proceedings of the 4th international workshop on requirements engineering: foundation for software quality (RESFQ 1998). Presses Universitaires de Namur, Namur
  3. C. P. Ayala , C. Cares , J. P. Carvallo , G. Grau , M. Haya , G. Salazar , X. Franch , E. Mayol , C. Quer,"A Comparative Analysis of i\*-Based Agent-Oriented Modeling Languages," 7<sup>th</sup> International Conference on Software Engineering and Knowledge Engineering, pp. 43-50, 2005.
  4. D. Quartel, W. Engelsman, H. Jonkers, M. van Sinderen, "A goal-oriented requirements modeling language for enterprise architecture," Enterprise Distributed Object Computing Conference, pp. 3-13, EDOC '09, IEEE International, 2009.
  5. D. Amyot, J. Horkoff, D. Gross, G. Mussbacher, "A Lightweight GRL Profile for i\* Modeling," ER 2009 Workshops on Advances in Conceptual Modeling, Lecture Notes, pp. 254-264, 2009
  6. D. Amyot and G. Mussbacher, "Development of Telecommunications Standards and Services with the User Requirements Notation".
  7. S. Ghanavati, D. Amyot, L. Peyton, "Compliance Analysis Based on a Goal-oriented Requirement Language Evaluation Methodology," Requirements Engineering Conference, RE '09. 17th IEEE International, 2009.
  8. P. Giorgini, J. Mylopoulos, R. Sebastiani, "Goal-Oriented Requirements Analysis and Reasoning in the Tropos Methodology," International Journal Engineering Applications of Artificial Intelligence, Volume 18, issue 2, pp. 159-171, 2005.
  9. J. Horkoff, E. Yu "Analyzing Goal Models – Different Approaches and How to Choose Among Them," ACM Symposium on Applied Computing, pp. 675-682, 2011.
  10. J. Horkoff , R. Salay, M. Chechik, D Sandro, "Supporting Early Decision-Making in the Presence of Uncertainty," Requirements Engineering Conference (RE), 2014 IEEE 22nd International , pp. 33-42, 2014.
  11. A. Raian, F. Dalpiaz, P. Giorgini, "A Goal-based Framework for Contextual Requirements Modeling and Analysis," Requirements Engineering Journal, Volume 15, Issue 4, November 2010, pp. 439-458, 2010.
  12. J. Horkoff and E. Yu, "Finding Solutions in Goal Models: An Interactive Backward Reasoning Approach"
  13. Torkar, R., Gorschek, T., Feldt, R., Raja, U. A., & Kamran, K., 2009. Requirements traceability state-of-the-art: A systematic review and industry case study . *IST Journal* .

14. Kavakli, E. a., 2003. Goal driven requirements engineering: evaluation of current methods. *Proceedings of the 8th CAiSE/IFIP8.*, (p. 16).
15. VAn Lamsweerde, A., 2001. Goal-Oriented Requirements Engineering: A Guided Tour. *Requirements Engineering, Proceedings of teh 5th IEEE international Symposium on IEEE* (pp. 249-262). IEEE.
16. Van Lamsweerde, A., 2000. Requirements engineering in the year 00: A research perspective. *Proceedings of the 22nd international conference on Software engineering*. (pp. 5-19). ACM.
17. Yu, E. S., 1997. Towards modeling and reasoning support for early-phase requirements engineering. In *Requirements Engineering Proceedings of the Third IEEE International Symposium* (pp. 226-235). IEEE.
18. Kiyavitskaya, N., & Zannone, N., 2008. Requirements model generation to support requirements elicitation: the Secure Tropos experience. *Automated Software Engineering* (pp. 149-173). Springer.
19. Massaccia, F., & Zannone, M. P. 2005. Using a Security Requirements Engineering Methodology in Practice: the compliance with the Italian Data Protection Legislation. *Computer Standards & Interfaces* (pp. 445-455). Elsevier.
20. Yu, J. H., 2010. Finding Solutions in Goal Models: An Interactive Backward Reasoning Approach. In *Conceptual Modeling-ER* (pp. 59-75). Springer Berlin Heidelberg.

# Using Multiple Criteria Decision Making Approaches to Assess the Quality of Web Sites

Rim Rekik

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

Ilhem Kallel

REGIM-Lab: Research Groups in Intelligent Machines, University of Sfax, Tunisia  
ISIMS: Higher Institute of Computer Science and Multimedia of Sfax, University of Sfax, Tunisia

Jorge Casillas

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

Adel M. Alimi

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

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

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

## I. INTRODUCTION

Multiple Criteria Decision Making (MCDM) is a largely used discipline to solve complex decision problems involving more than one criterion [1, 2]. MCDM also is continuously growing in fields of Mathematics, Decision Sciences, Business, Management and Accounting, Medicine, Social Sciences, Environmental Science, Economics, Econometrics and Finance, etc. Its evolution throughout the years is interesting; some limitations of the use of MCDM include modeling human judgment in a clear way (as uncertainty, imprecision), let the appearance of hybrid methods. Fig.1 shows the spread of MCDM techniques areas. It is clear that Engineering and Computer science areas have the most important part in utilizing MCDM for solving decision problems according to Scopus database.

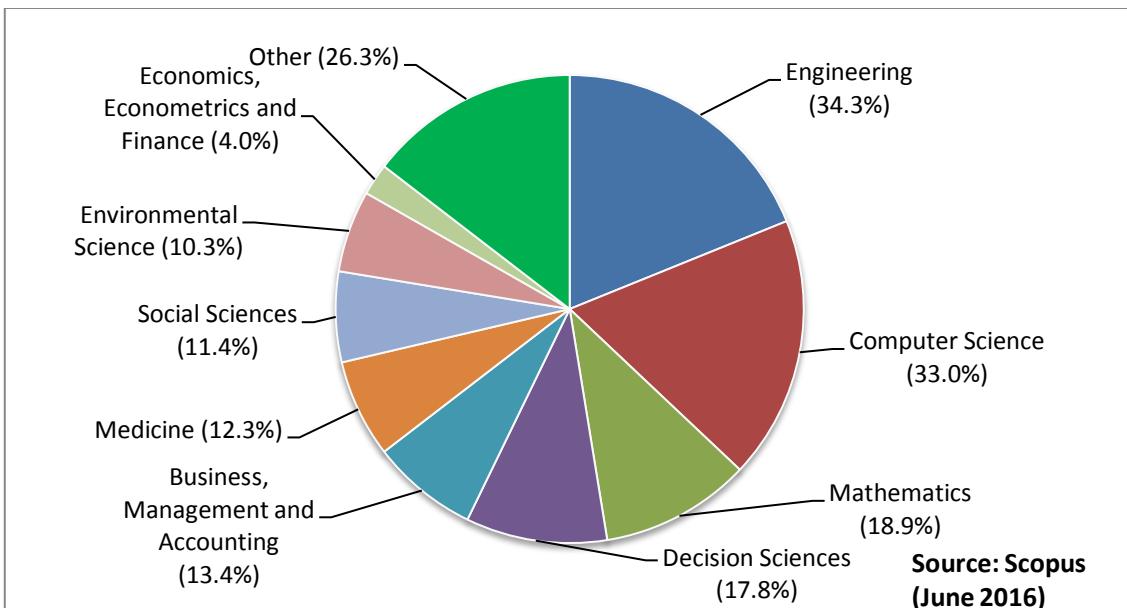


Figure 1. The spread of MCDM techniques areas

MCDM concerns the problems that need the views' point of decision makers facing multiple conflicting criteria. Unfortunately, human judgments' preferences are often unclear to express by exact numerical values. In the classical MCDM problems, certainty is required to evaluate criteria weights and ratings by crisp values. Consequently, MCDM is supported by soft computing techniques such as fuzzy sets, neural networks and genetic algorithms due to the imprecision and vagueness of decisions. Indeed, according to the proponents of fuzzy logic, it is more natural to express judgments by fuzzy numbers instead of crisp values.

The field of MCDM assessment evolves in parallel to hybrid MCDM and fuzzy MCDM to choose the best method or the most useful hybrid methods. Fig. 2 shows an evolution in the last decade related to MCDM, by number of publications according to Scopus database. Recently, development in fuzzy MCDM is becoming increasingly important. Mardani et al. [3] explored fuzzy MCDM techniques and applications between 1994 and 2014. The study presents the developments of fuzzy models of multiple criteria decision analysis.

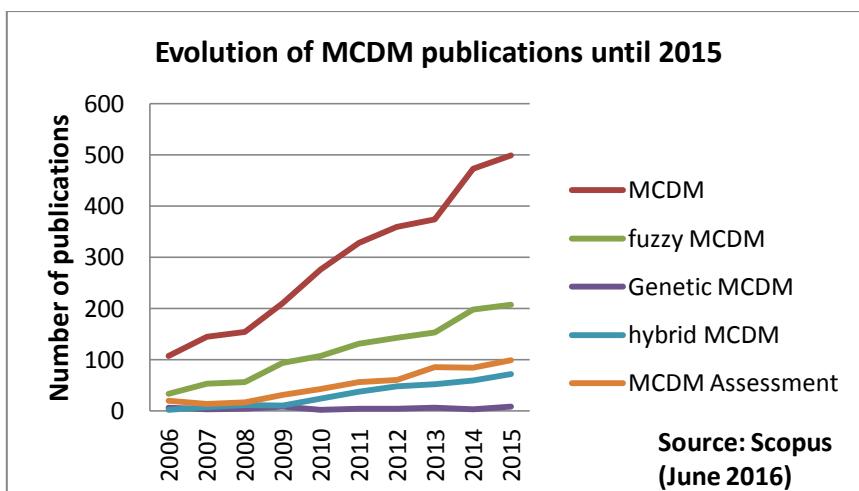


Figure 2. Evolution of publications' number related to MCDM

Section 2 presents a background in fuzzy sets theory related to MCDM. Section 3 reports the followed method. Section 4 answers the research questions. A discussion of them and the strength and limitation of the study are presented in Section 5 and conclusions in Section 6.

## II. FUZZY SETS IN MCDM

Originally, Zadeh [4] proposed the fuzzy set theory which Bellman and Zadeh [5] combined with the MCDM to solve problems that encountered limitations using conventional MCDM techniques. Usually, due to the subjectivity and imprecision, solutions given by decision makers for rating criteria can't be clearly presented by crisp data. Thus, the fuzzy MCDM approaches evolved to deal with uncertain decisions.

In crisp set theory, the membership function  $\mu_A$  (characteristic function of  $x$  in  $A$ ) values are 0 or 1, but nothing in between, expressed by (1).

$$\mu_A = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases} \quad (1)$$

Where  $X$  represents a universal set. A fuzzy set  $A$  in  $X$  is expressed by its membership function  $\mu_A: X \rightarrow [0-1]$ . Here,  $\mu_A(x)$  characterized by the degree of membership of element  $x$  in fuzzy set  $A$  for each  $x \in X$ . Fig. 3 presents a crisp and a fuzzy set.

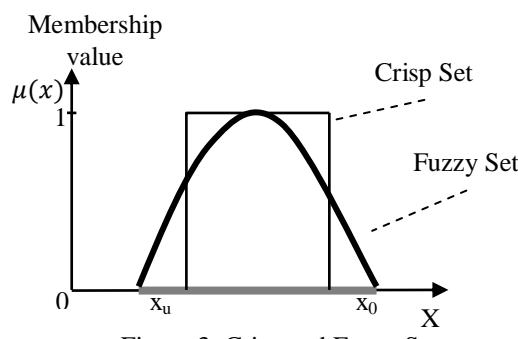


Figure 3. Crisp and Fuzzy Set

### A. Fuzzy numbers

Fuzzy numbers could be used in presenting the preferences of decision makers expressed by linguistic variables. Triangular Fuzzy Number (TFN) is defined by its simplicity in both concept and computation. It is the most popular technique among different shapes used to define fuzzy number memberships.

According to the definition of van Laarhoven and Pedrycz [6], a TFN is represented with a triplet  $(L, M, U)$  for *Lower, Medium and Upper* numbers. A membership value of a triangular fuzzy number  $A$  is defined by (2).

$$\mu_A(x) = \begin{cases} \frac{x-L}{M-L}, & L \leq x < M \\ \frac{x-L}{M-L}, & L \leq x \leq M \\ 0, & \text{Otherwise} \end{cases} \quad (2)$$

Trapezoidal Fuzzy Number (TrFN) is another category of fuzzy numbers. It is applied in the multi-criteria decision making field [7]. Some discussions on TFNs and TrFNs are stated by Chen and Ku [8].

A new aggregation method to solve multi-attribute group decision using TFNs and TrFNs was proposed by Xu et al. [9]. The method is used to aggregate heterogeneous decision information into Atanassov's intuitionistic fuzzy numbers (AIFNs) for more useful information.

### B. Fuzzy linguistic approach

When making decisions, experts' judgment can be expressed using linguistic terms in order to reduce vagueness, imprecision of point of views and to enhance reliability and clarity of preferences. Fuzzy sets are used to deal with the qualitative aspects of linguistic values using variables. The fuzzy linguistic approach employs an interpretation of linguistic variables as fuzzy numbers. These are defined as membership functions. Consequently, the preferences of decision makers are expressed by fuzzy numbers. In MCDM, a problem of ranking fuzzy numbers is usually treated.

## III. RESEARCH METHODOLOGY

Kitchenham and Charters [10] provide a Systematic Literature Review (SLR) methodology. This process is composed of three phases: planning, conducting and reporting the review. Here, Section III.A involves "planning the review" phase for developing a clear review protocol. It specifies the objective, the main raised research questions, the adopted search strategy and a set of established inclusion and exclusion criteria to select publication. Next, Section III.B treats "conducting the review" phase for executing the review protocol. A dissemination of results is presented in Section IV from extracted data of selected papers that deal with MCDM in assessing web sites quality. Some developments and purposes in this domain are revealed.

### A. Planning the review

A phase of planning the review is dedicated to define the objective of the SLR which is exploring different MCDM approaches in assessment of web sites quality existing in literature and the research questions. It is also devoted to develop a clear review protocol composed by search process strategy and a definition of inclusion/exclusion criteria.

#### 1) Research questions

According to the objective of this review, a set of three research questions include:

**RQ1:** What are the developments in MCDM in the field of assessment of web sites quality?

**RQ2:** What is the objective of recent research using MCDM in this field?

**RQ3:** What are the features and limitations of MCDM methods?

#### 2) Search process and inclusion/ exclusion criteria

It is important to follow a search strategy in order to insure a convincing review conducting in the phase 2. This produces exploring scientific publications within the domain. The search focuses on related journals and conferences using Elsevier's Scopus database. It is necessary to define search criteria based on key concepts as selection words. Indeed, some words are considered such as "MCDM", "multiple criteria", "multi-criteria", "decision making", "assessment", "assessing", "evaluation", "quality" and "web sites". A combination of these terms should be done to enlarge the scope of searching for better results.

Since all collected papers cannot be included in the SLR, some inclusion and exclusion criteria are introduced. In order to select the most relevant ones, the criteria that specify whether a study will be included or excluded are determined. The first inclusion criterion based on terms appeared in the titles, abstracts and keywords in studies by browsing computer science discipline; an identification of relevant ones was established. However, papers published before 2009 and languages other than English written studies were excluded. In

addition, some sub-disciplines as “Web services” not related exactly to the topic of assessment were excluded. After obtaining a large set of papers, a step of eliminating short ones (up to 4 pages) was performed.

#### B. Conducting the review

An initial search on Scopus database returned 4321 document. That figure was reduced to 1041 by limiting the search to subject area of “Computer Science”, (written in English) and published between 2009-2015. This was further refined to 758 by excluding keywords that do not respond to the topic such “Web services”. From this 578 accessible documents rescanned to consider other inclusion/exclusion criteria, such as short papers and articles that deal with MCDM techniques. Finally, 18 papers were reviewed to contribute to the research. Some of them exist in the other scientific e-sources. Fig. 4 illustrated the SLR process.

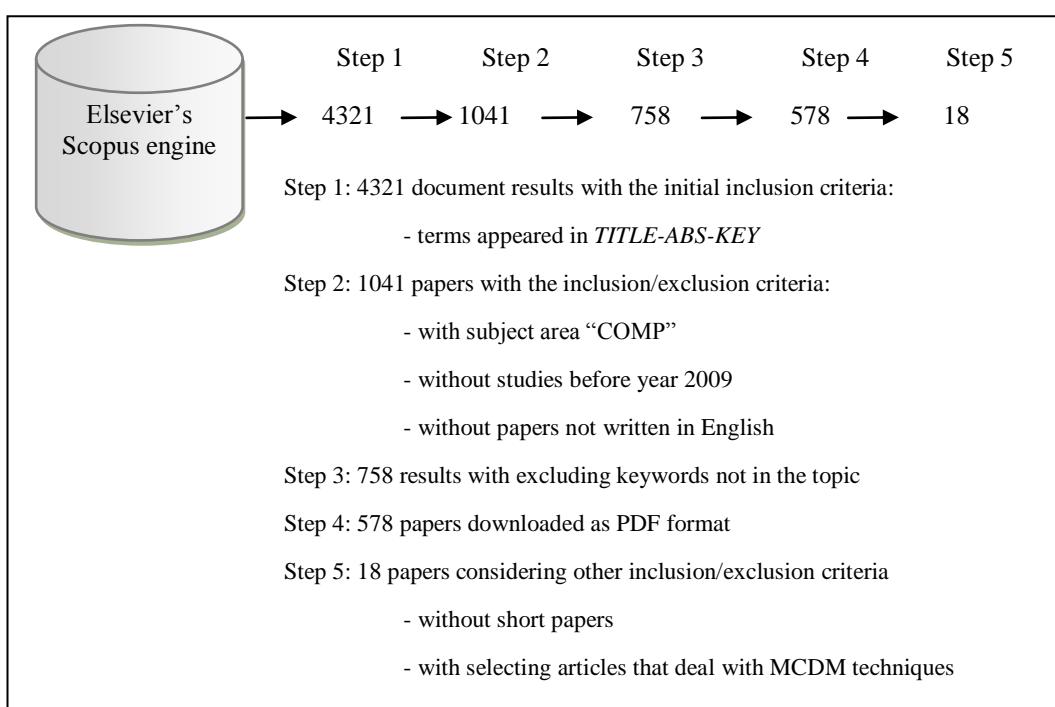


Figure 4. Paper selection process

## IV. EXISTING RESEARCH

#### A. RQ1: What are the developments in MCDM in the field of assessment of web sites quality?

MCDM methods are increasingly being used in the last decade. In order to evaluate web site quality, any process goes through a step of specifying certain criteria. The distribution of criteria importance and ranking web sites are generally solved by MCDM techniques. There are many methods as AHP, ANP, TOPSIS, VIKOR, PROMETHEE, ELECTRE; choosing the best method is, it-self, a multi-criteria decision making problem. They are cited below in case of comparing and selecting criteria for the evaluation of quality of websites. Rekik et al. [11] provided an overview of MCDM assessment methods in the field of assessment of web sites quality.

##### 1) Crisp hybrid and non-hybrid MCDM methods

A web site is evaluated according to a set of criteria but they do not have the same importance to highlight the quality. They differ also according to the web site category. Weighing criteria using crisp MCDM are used as Analytic Hierarchy Process (AHP) developed by Saaty [12] and applied by Akincilar and Dagdeviren [13] in

evaluating the quality of hotel web sites. Moreover, among recent studies, Cebi [14] presents an integrated MCDM method to find out the web site design parameters with Delphi method. The study focuses on determining the importance degrees of those parameters and their interactions according to the type of web site with the Decision-Making Trial and Evaluation Laboratory (DEMATEL) method. They have discussed those parameters to help the designers building a good web site in order to satisfy users' needs. DEMATEL method is generally used to determine interrelations among criteria as in [15].

Ranking alternatives also can be solved by Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) presented by Brans and Vincke [16]. Akincilar and Dagdeviren [13] apply it to rank web sites. The study considered so hybrid MCDM for weighing criteria and ranking web sites to provide results for customers and enterprises. Kostoglou et al. [17] assess the quality of universities web sites using AHP method for ranking, despite its limitation.

## 2) Fuzzy hybrid and non-hybrid MCDM methods

Hu [18] proposes a genetic algorithm based learning method for distributing automatically the degrees of importance of criteria. The evaluation concerns electronic Service Quality (e-SQ) of travel web sites characterized by multiple criteria. It deals with the respondents' subjective decision represented by fuzzy numbers. The decision problem is modeled by a hierarchical structure. The motivation is to find critical criteria concerned by customers to attract them for making transactions on the web site. Hu and Liao [19] based on the same last learning method of weighing criteria evaluate internet banking services' quality.

Chou and Cheng [20] use a hybrid fuzzy MCDM approach which is composed by Fuzzy Analytic Network process (FANP) to weight criteria and Fuzzy VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) to rank professional accounting firms web sites in a linguistic terms. Basically, ANP developed by Saaty [21] to deal with interrelationships and feedback among criteria. While VIKOR introduced by Opricovic and Tzeng [22] to make a compromise ranking from a set of alternatives.

Kaya [23] evaluates E-business category using an integrated fuzzy AHP-TOPSIS method. Fuzzy AHP is applied to distribute weights for criteria in order to select highest ones and fuzzy TOPSIS to rank web sites. Originally, Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) proposed by Hwang and Yoon [24] to solve MCDM problems.

Dey et al. [25], Yu et al. [26] interested by B2C E-commerce web sites based on AHP to structure the problem as a hierarchy and distribute weights for criteria. Then, fuzzy TOPSIS is used to rank alternatives. In the same category of evaluation, Aydin and Kahraman [27] propose a fuzzy AHP process for weighing alternatives including quantitative and qualitative factors allowing multiple decision makers view points. They compare the results with fuzzy VIKOR. Sun and Lin [28] suggest fuzzy TOPSIS to determine weights for criteria in order to know the most important factors of shopping web sites which are security and trust.

Hsu et al. [29] treat the multiplicity of criteria for e-SQ of travel web sites by applying consistent fuzzy preference relations method as a first phase to distribute weights (without considering interdependence perspective). They propose ANP technique for the degree of influence for criteria and sub-criteria as a second phase and to determine their weights in a third phase considering interdependence perspective. Based on the furnished results, they confirm that criteria and sub-criteria relevant to e-SQ are interdependent.

Carrasco et al. [30] define an implementation of a Linguistic Multi-Criteria Decision Making model (LMCDM), using IBM SPSS Modeler to obtain a scale for attributes by feedback of guests that makes in

evidence service quality in tourism. Del Vasto-Terrientes et al. [31] propose a method ELECTRE-III-H for ranking web sites of tourist destination brands based on the outranking relation interpreted as a fuzzy relation. The method based in decomposing the problem into multi-level sub-problems. It helps the decision maker to analyze hierarchically structured criteria by allowing the propagation of results upwards in the tree. Ip et al. [32] suggest a fuzzy AHP approach for evaluating hotel web site functionality to avoid decision makers' uncertainty in establishing the relative weights of each criterion. Its purpose is to know the most important criterion which is according to the results "Reservation Information" to inform hotel industry about users' preferences for criteria.

Büyüközkan et al. [33] apply Fuzzy Axiomatic Design (FAD) for the evaluation of e-learning web sites. They compare its results of weighing and ranking criteria using experts' opinion with fuzzy TOPSIS.

Kabak and Burmaoglu [34] use a hybrid MCDM methodology that combines the Fuzzy Decision Making Trial and Evaluation Laboratory (FDEMATEL) to make relations between criteria, the fuzzy ANP to determine weights of criteria. The evaluation concerns the performance of e-procurement web sites.

Markaki et al. [35] treat the assessment of quality of e-government web sites with fuzzy AHP as a Multi-Attribute Decision Making (MADM) method to know the service levels that should be provided by sites and to inform government administrators with it.

#### *B. RQ2: What is the objective of recent research using MCDM in this field?*

The methods described above are based on fuzzy decision making for weighing and ranking criteria or web sites. Some others focused on structuring the problem as a hierarchy or making interrelations between criteria. In Table 1, a classification of the above works that consider the use of MCDM techniques according to the purpose of evaluation is established.

TABLE 1.  
SUMMARY OF LITERATURE ON MCDM METHODS TO ASSESS WEB SITES QUALITY

Method Objective	Method	References
Ranking web sites	Fuzzy TOPSIS	Kaya, 2010 [23] Yu et al., 2011 [26] Dey et al., 2015 [25]
	Fuzzy VIKOR	Chou and Cheng, 2012 [20]
	AHP	Kostoglou et al., 2014 [17]
	PROMETHEE	Akincilar and Dagdeviren, 2014 [13]
	ELECTRE-III-H	Del Vasto-Terrientes et al., 2015 [31]
Weighing criteria	Fuzzy TOPSIS	Sun and Lin, 2009 [28] Büyüközkan et al., 2010 [33]
	Genetic Algorithm based learning method	Hu, 2009 [18] Hu and Liao, 2011 [19]
	Fuzzy AD	Büyüközkan et al., 2010 [33]
	AHP	Kaya, 2010 [23] Yu et al., 2011 [26] Akincilar and Dagdeviren, 2014 [13] Dey et al., 2015 [25]
	Fuzzy AHP	Büyüközkan et al., 2010 [33] Markaki et al., 2010 [35] Ip et al., 2012 [32] Aydin and Kahraman, 2012 [27]
	Fuzzy ANP	Chou and Cheng, 2012 [20] Kabak and Burmaoglu, 2013 [34]

	Consistent fuzzy preference relations method	Hsu et al., 2012 [29]
	Fuzzy VIKOR	Aydin and Kahraman, 2012 [27]
	DEMATEL	Cebi, 2013 [14]
	Linguistic MCDM	Carrasco et al., 2014 [30]
Structuring the problem as a hierarchy	Fuzzy MCDM	Hu, 2009 [18] Hu and Liao, 2011 [19]
	Fuzzy AHP	Markaki et al., 2010 [35] Ip et al., 2012 [32] Aydin and Kahraman, 2012 [27]
	AHP	Yu et al., 2011 [26] Akincilar and Dagdeviren, 2014 [13] Dey et al., 2015 [25]
Making relations between criteria	ANP	Hsu et al., 2012 [29]
	Fuzzy ANP	Chou and Cheng, 2012 [20]
	FDEMATEL	Kabak and Burmaoglu, 2013 [34]
	DEMATEL	Cebi, 2013 [14]

### C. RQ3: What are the features and limitations of MCDM methods?

According to previous analysis of RQs, a classification of MCDM methods is proposed into four categories which are scoring methods, compromising methods, outranking methods and other methods. This classification can be useful for future works to select appropriate method for solving MCDM according to the nature of the problem.

#### 1) Scoring methods

Major approaches include scoring (or utility) theory. Its fundamental feature is to express the decision maker's preference via a score or utility. Popular methods in this category are the AHP [12], ANP [21] is also a generic form of AHP that treats dependencies between criteria. Indeed, AHP structures the problem as a hierarchy while ANP structures it into a network. These methods have the ability to check inconsistencies of decision makers. Their disadvantage is that the decision maker finds difficulty to express his opinion using the 9 point scale [12]. This limitation can be solved by using fuzzy numbers as presented in Table 2. In the following, fuzzy AHP is described by its fundamental steps.

First, the decision problem is modeled as a hierarchy illustrated by Fig. 5. It is partitioned into a high level criteria  $\{\text{criterion}_1.. \text{criterion}_n\}$ , a level for sub-criteria  $\{(C_{11}, C_{12})..(C_{n1}, C_{n2})\}$ , and a level for alternatives  $\{A_1..A_m\}$ .

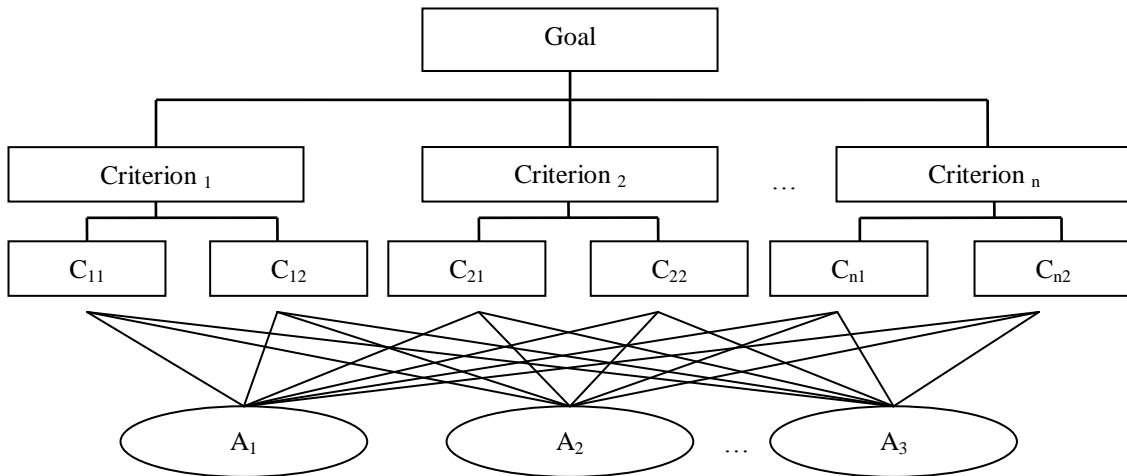


Figure 5. Hierarchy Decision Structure

The next step is to have subjective opinions of decisions makers to set up priorities for criteria, or degrees of preference, expressed by linguistic terms such as criterion<sub>1</sub> has *very strong importance* than criterion<sub>2</sub> for the evaluation. Then, a pair wise comparison matrix is constructed between each criterion with conversion linguistic terms to fuzzy numbers using a scale defined in Table 2.

TABLE 2. A scale for comparison [36]

Crisp Scale	Fuzzy Number	Degree of Preference	Membership Function
1	$\tilde{1}$	Equal importance	(1,1,2)
3	$\tilde{3}$	Moderate importance	(2,3,4)
5	$\tilde{5}$	Strong or essential importance	(4,5,6)
7	$\tilde{7}$	Very strong importance	(6,7,8)
9	$\tilde{9}$	Extreme importance	(8,9,10)

To be sure that the original preference ratings are consistent, a check of a Consistency Ratio (CR) should be done after calculating it. The comparisons are supposed to be acceptable, if the CR is less than 0.10. Then, for each decision alternative, the weighted average rating is calculated. Thus, it is obvious to determine the alternative with the highest score.

## 2) Compromising methods

The other category of approaches in MCDM utilizes a compromising method that finds a feasible solution for ranking problems closest to the ideal (the best compromise) and helps decision makers to reach a final solution(s). The TOPSIS and VIKOR methods are among this category. These two methods are based in different aggregating functions for ranking and combined with fuzzy sets theory for better results. Fuzzy TOPSIS and fuzzy VIKOR are widely applied as MCDM techniques. They are detailed below and compared to each other.

### a) Fuzzy TOPSIS:

Step 1: Construction of the fuzzy decision matrix  $\tilde{D}$ , defined by (3), based on linguistic terms and assigned to criteria  $C$  and alternatives  $A$ .

$$\tilde{D} = \begin{matrix} C_1 & C_2 & \dots & C_n \\ A_1 & \left( \begin{matrix} \tilde{x}_{11} & \tilde{x}_{12} & \dots & \tilde{x}_{1n} \\ \vdots & & & \vdots \\ \tilde{x}_{1m} & \dots & \dots & \tilde{x}_{mn} \end{matrix} \right) \\ A_2 \\ A_3 \end{matrix} \quad (3)$$

Step 2: Aggregation of decision makers' judgments  $\tilde{x}_{ij}$  according to (4).

$$\tilde{x}_{ij} = \frac{1}{k} (\tilde{x}_{ij}^1 + \tilde{x}_{ij}^2 + \dots + \tilde{x}_{ij}^k) \quad (4)$$

where k is the decision maker number and  $\tilde{x}_{ij}^k$  his fuzzy judgment.

Step 3: Normalization of the fuzzy decision matrix

Step 4: Computation of the normalized fuzzy decision matrix

Step 5: According to triangular fuzzy numbers (TFNs), a definition of positive and negative ideal reference points ( $FPI^+$  and  $FPI^-$ ).

Step 6: Calculation of the distance from  $FPI^+$  and  $FPI^-$  for each alternative

Step 7: Calculation of similarities to ideal solution and ranking alternatives.

b) *Fuzzy VIKOR:*

Fuzzy VIKOR considers in the presence of conflicting criteria two distance measurements, based on an aggregating function in the compromising programming method in order to measure "closeness" to the "ideal" solution.

Step 1: Identifying the objective of the decision making process, determining the best values of criteria or alternatives and structuring a hierarchy.

Step2: Identification of linguistic scale for evaluation of criteria

Step3: Rating alternatives respecting judgments of decision makers as expressed by (4).

Step 4: Expression of the fuzzy decision matrix  $\tilde{D}$  defined by (3).

Step 5: Defuzzifying  $\tilde{D}$  and fuzzy weight of each criterion into crisp values. Defuzzification methods as center of area (COA) can be used to determine the best fuzzy value ( $BFV, f_j^*$ ) and worst fuzzy value ( $WFV, f_j^-$ ).

Step6: Computing values of  $S_i$  and  $R_i$  ( $i = 1, \dots, m$ ), defined by (5) and (6).

$$S_i = \sum_{j=1}^n w_j (f_j^* - f_{ij}) / (f_j^* - f_j^-) \quad (5)$$

where  $S_i$  the separation measure of  $A_i$  from the ( $BFV, f_j^*$ )

$$R_i = \max_{j=1,\dots,n} [w_j(f_j^* - f_{ij}) / (f_j^* - f_j^-)] \quad (6)$$

where  $R_i$  the separation measure of  $A_i$  from the ( $WFV, f_j^-$ )

Step 7: Computing the values  $Q_i$  that represent a weight for the strategy of maximum group utility

Step 8: Ranking and selecting best alternatives as a compromise solution

Table 3 summarizes the steps of fuzzy TOPSIS and fuzzy VIKOR methods and highlights the features of each of them. Both of them belong to compromising methods as it is detailed in answering RQ3.

TABLE 3. Comparison of fuzzy TOPSIS and fuzzy VIKOR methods

Fuzzy TOPSIS	Fuzzy VIKOR
Steps	

The same first steps: - Expression of the fuzzy decision matrix $\tilde{D}$ - Aggregation of decision makers' judgments $\tilde{x}_{ij}$ - Computation of normalized fuzzy decision matrix	
Determination of positive and negative ideal reference points ( $FPI^+$ and $FPI^-$ ).	Determination of the best fuzzy value (BFV, $f_j^*$ ) and worst fuzzy value (WFV, $f_j^-$ )
Calculation of the distance from $FPI^+$ and $FPI^-$ for each alternative	Computing values of $S_i$ and $R_i$ ( $i = 1, \dots, m$ ), where $S_i$ the separation measure of $A_i$ from the (BFV, $f_j^*$ ) and $R_i$ the separation measure of $A_i$ from the (WFV, $f_j^-$ )
	Computing the values $Q_i$ that represent a weight for the strategy of maximum group utility
Ranking and determination of compromise solutions which are the closeness to the ideal alternatives	
<i>Features</i>	
Express subjective opinions of decisions makers into linguistic terms	
Used for solving ranking problems by similarity to ideal solutions	Used for multiple conflicting criteria optimization of complex systems and the obtained compromised solution can be accepted because it has maximum group utility (the majority)

### 3) Outranking methods

The ‘ELimination and Et Choice Translating Reality’ (ELECTRE) methods developed by Roy [37] and PROMETHEE for decision aid are well known belong to outranking methods [1]. ELECTRE and PROMETHEE family methods solve complex and uncertain choice problems with multiple decision makers and multiple criteria. They also deal with ranking problems in the presence of incomparability between alternatives. They consist of a preference relation called an outranking relation among alternatives based on several attributes.

### 4) Other methods

In the field of artificial intelligence, genetic algorithm [18, 19] in the presence of multiple criteria can identify critical criteria for the evaluation by determining degrees of criteria’ importance automatically.

## V. DISCUSSION OF FINDINGS

In this paper, an overview of MCDM methods for assessing web sites quality is presented by following a systematic review strategy on articles written between 2009 and 2015. In order to reveal the most important findings, a discussion about interdependencies between criteria is treated below. This purpose is not well studied in literature (see Table 1). However, criteria and their importance differ according to the web site categories. So, distributing weights for criteria needs an MCDM method as presented in Table 1 which is based on expert decision and evaluation.

### A. Discussion of research questions

By disseminating results of RQ1, crisp MCDM methods are not largely used. The trend is to make hybridizations among them or a combination with fuzzy sets theory. Some of them are analyzed and compared.

The most important purposes of MCDM defined in RQ2 are weighing criteria and ranking web sites (alternatives) while the purpose of studying relations between criteria is not well explored in literature. DEMATEL and ANP are among conventional MCDM methods that deal with interactions between criteria.

The problematic of evaluating web sites implies a multiple criteria decision making due to the multiple conflicting criteria for assessment. A recent study Rekik et al. [38] used association rules mining to find

interrelations between criteria. In addition, a determination of relevant criteria was done. Web site category is also implied because criteria depend on it. A web site of E-commerce for example that supports payment mode will be not assessed as an educational one. Rekik et al. [39] studied the feedback between a set of criteria related to E-commerce category based on fuzzy ANP method. They determined weights for criteria to know the best ones. Machine learning techniques can be useful also on Knowledge Discovery in Databases (KDD) to obtain novel information, for example in marketing problems. Finding relationship between items in marketing databases is proposed by Orriols-Puig et al. [40].

The use of MCDM methods has been emerged recently and it is evolving continuously. Their application attempts to consider multiple criteria selected or extracted by experts for the evaluation. In Table 4, a summary of collected criteria from selected studies for the SLR according to the web site's category. The finding is to show the important task and step of choosing the criteria according to the web site's type in the process of evaluation.

TABLE 4. Criteria variation according to the web site's category and experts' selection

Web site category	Criteria	Reference
Institutional	Completeness, coverage, objectivity, research, web services	[17]
E-commerce	Product, design, technology, service quality, logistics	[26]
	Design and usability, product, security, service quality, fulfillment	[25]
	Ease of use, product, security, customer relationship, fulfillment	[27]
	Efficiency, practical, ease use, time-saving, communication, confident, security, trust, familiar, past experience, knowledgeable	[28]
Professional accounting firms	Accessibility, navigability, usability, privacy, relevance, understandability, richness, currency, responsiveness, reliability, assurance, empathy	[20]
Hotel	Customer oriented, technology oriented, marketing oriented, security oriented, other factors	[13]
	Hotel description, hotel facility information, reservation information, surrounding area information, user generated information	[32]
	Tangibles, reliability, responsiveness, assurance, empathy	[30]
Tourism	Usability and accessibility, visibility, brand treatment	[31]
E-learning	Right and understandable content, complete content, personalization, security, navigation, interactivity, user interface	[33]
Travel	Design, security, customer relationships, enjoyment	[29]
	Efficiency, fulfillment, system availability, security/privacy, responsiveness, compensation, contact, benefit, customization/personalization, tangibility, assurance/trust, continuous improvement	[18]
Internet banking	Efficiency, system availability, responsiveness, compensation, contact, tangibility, privacy, reliability, reputation, continue improvement, personalization, benefit	[19]
E-government	Usability, content, site quality, e-services, e-democracy features	[35]
	Navigability, speed, Standardisation, links, accuracy, richness, attractiveness, reliability, personalization, responsiveness	[34]
Any type	Usability, visual aspects, technical adequacy, content, security, communication, prestige	[14]
E-business	Relevance and currency, Understandability, Reliability, Empathy, Response speed, Personalization, Security, Price savings, Awareness	[23]

Different categories are concerned by the evaluation such as E-commerce, E-government, Hotel and Institutional. On the basis of a large set of criteria, an outranking hierarchical approach can be applied as in [31] to help the decision maker designing the decision model in a structured way and analyzing the preference relations in the set of alternatives.

The classification presented in RQ3 expounds the most important methods found in general and in the assessment of web sites quality area especially. The approaches used hybridizations of MCDM methods or combined with fuzzy sets theory as detailed in reporting results of RQ1.

According to the reporting results of RQ2 and RQ3, the most utilized methods for weighing criteria purpose are scoring methods. As summarized in Table 1, there are studies that deal with AHP [13, 23, 25, 26], some others with fuzzy AHP [27, 32, 33, 35] and with fuzzy ANP [20, 34]. While the most utilized methods for solving ranking web sites purpose are compromising methods as the use of fuzzy TOPSIS [23, 25, 26] and fuzzy VIKOR [20] and outranking methods as PROMETHEE [13] and ELECTRE-III-H [31]. Ranking web sites can be solved also using fuzzy sets theory as in Rekik and Kallel [41, 42] by choosing a set of criteria for the evaluation.

#### B. Study strength, limitation and future work

The aim of this SLR is to help research community having a scope in existing research and to derive future developments. It is limited to expose MCDM methods and developments in assessing web sites quality.

As future work, a proposition to enlarge it by expanding the list of studies not only involved in solving the problem by MCDM but also studies that used other methods implied in different steps of the evaluation's process. Another ambitious goal is to know the most important categories of web sites considered for the assessment in literature and the corresponding list of criteria.

#### VI. CONCLUSION

This study identifies the main purposes of using MCDM in the field of web sites assessment. In addition, the features of some of them have been revealed. This review classifies the most important methods into different categories and exposes hybridizations and integration between them. Some of them are combined with fuzzy sets theory to reduce subjectivity and uncertainty of complex decision problems. MCDM methods are based on expert judgment. The decisions are expressed by linguistic variables and then presented by fuzzy numbers.

#### ACKNOWLEDGEMENT

The authors express thanks to the Erasmus Mundus Al-Idrisi II for funding the research reported under the Grant Agreement number 2013–2401/001–001–EMA2. The authors acknowledge, too, the financial support of this work by grants from Tunisian General Direction of Scientific Research (DGRST) under the ARUB program. This work also was partially supported by the Spanish Ministry of Science and Innovation (grant no. TIN2014-57251-P).

#### REFERENCES

- [1] B. Roy, "Decision-Aiding Today: What Should We Expect?," in *Multicriteria Decision Making*, vol. 21, T. Gal, T. J. Stewart, and T. Hanne, Eds. Boston, MA: Springer US, 1999, pp. 1–35.
- [2] T. Gal, T. J. Stewart, and T. Hanne, Eds., *Multicriteria Decision Making*, vol. 21. Boston, MA: Springer US, 1999.
- [3] A. Mardani, A. Jusoh, and E. K. Zavadskas, "Fuzzy multiple criteria decision-making techniques and applications – Two decades review from 1994 to 2014," *Expert Systems with Applications*, vol. 42, no. 8, pp. 4126–4148, May 2015.
- [4] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
- [5] R. E. Bellman and L. A. Zadeh, "Decision-Making in a Fuzzy Environment," *Management Science*, vol. 17, no. 4, p. B-141, Dec. 1970.
- [6] P. J. M. van Laarhoven and W. Pedrycz, "A fuzzy extension of Saaty's priority theory," *Fuzzy Sets and Systems*, vol. 11, no. 1, pp. 229–241, Jan. 1983.
- [7] W. Jianqiang and Z. Zhong, "Aggregation operators on intuitionistic trapezoidal fuzzy number and its application to multi-criteria decision making problems," *Systems Engineering and Electronics, Journal of*, vol. 20, no. 2, pp. 321–326, Apr. 2009.
- [8] T.-Y. Chen and T.-C. Ku, "Importance-assessing method with fuzzy number-valued fuzzy measures and discussions on TFNs and TrFNs," *International Journal of Fuzzy Systems*, vol. 10, no. 2, p. 92, 2008.

- [9] J. Xu, S.-P. Wan, and J.-Y. Dong, "Aggregating decision information into Atanassov's intuitionistic fuzzy numbers for heterogeneous multi-attribute group decision making," *Applied Soft Computing*, 2016.
- [10] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.
- [11] R. Rekik, I. Kallel, and A. M. Alimi, "Quality Evaluation of Web Sites: A Comparative Study of some Multiple Criteria Decision Making Methods," in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2015, pp. 585–590.
- [12] T. L. Saaty, "The analytic hierarchy process: planning, priority setting, resources allocation," *New York: McGraw*, 1980.
- [13] A. Akincilar and M. Dagdeviren, "A hybrid multi-criteria decision making model to evaluate hotel websites," *International Journal of Hospitality Management*, vol. 36, pp. 263–271, Jan. 2014.
- [14] S. Cebi, "Determining importance degrees of website design parameters based on interactions and types of websites," *Decision Support Systems*, vol. 54, no. 2, pp. 1030–1043, Jan. 2013.
- [15] G.-H. Tzeng, C.-H. Chiang, and C.-W. Li, "Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1028–1044, May 2007.
- [16] J.-P. Brans and P. Vincke, "Note—A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making)," *Management science*, vol. 31, no. 6, pp. 647–656, 1985.
- [17] Kostoglou V., Papathanasiou J., Moschidis O., and Ladopoulou P., "A comparative analysis of Greek universities presence on the World Wide Web using an analytical MCDM methodology," *International Journal of Multicriteria Decision Making*, vol. 4, no. 3, pp. 279–298, 2014.
- [18] Y.-C. Hu, "Fuzzy multiple-criteria decision making in the determination of critical criteria for assessing service quality of travel websites," *Expert Systems with Applications*, vol. 36, no. 3 PART 2, pp. 6439–6445, 2009.
- [19] Y.-C. Hu and P.-C. Liao, "Finding critical criteria of evaluating electronic service quality of Internet banking using fuzzy multiple-criteria decision making," *Applied Soft Computing Journal*, vol. 11, no. 4, pp. 3764–3770, 2011.
- [20] W.-C. Chou and Y.-P. Cheng, "A hybrid fuzzy MCDM approach for evaluating website quality of professional accounting firms," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2783–2793, 2012.
- [21] T. L. Saaty, "Decision making with dependence and feedback: The analytic network process," *RWS Publication*, 1996.
- [22] S. Opricovic and G.-H. Tzeng, "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS," *European journal of operational research*, vol. 156, no. 2, pp. 445–455, 2004.
- [23] T. Kaya, "Multi-attribute Evaluation of Website Quality in E-business Using an Integrated Fuzzy AHPTOPSIS Methodology," *International Journal of Computational Intelligence Systems*, vol. 3, no. 3, pp. 301–314, Sep. 2010.
- [24] C.-L. Hwang and K. Yoon, *Multiple Attribute Decision Making*, vol. 186. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981.
- [25] Dey S., Jana B., Gourisaria M.K., Mohanty S.N., and Chatterjee R., "Evaluation of Indian B2C E-shopping websites under multi criteria decision-making using fuzzy hybrid technique," *International Journal of Applied Engineering Research*, vol. 10, no. 9, pp. 24551–24580, 2015.
- [26] X. Yu, S. Guo, J. Guo, and X. Huang, "Rank B2C e-commerce websites in e-alliance based on AHP and fuzzy TOPSIS," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3550–3557, 2011.
- [27] S. Aydin and C. Kahraman, "Evaluation of E-commerce website quality using fuzzy multi-criteria decision making approach," *IAENG International Journal of Computer Science*, vol. 39, no. 1, pp. 64–70, 2012.
- [28] C.-C. Sun and G. T. R. Lin, "Using fuzzy TOPSIS method for evaluating the competitive advantages of shopping websites," *Expert Systems with Applications*, vol. 36, no. 9, pp. 11764–11771, 2009.
- [29] T.-H. Hsu, L.-C. Hung, and J.-W. Tang, "The multiple criteria and sub-criteria for electronic service quality evaluation: An interdependence perspective," *Online Information Review*, vol. 36, no. 2, pp. 241–260, 2012.
- [30] R. A. Carrasco, M. F. Blasco, and E. Herrera-Viedma, *An implementation of a linguistic multi-criteria decision making model: An application to tourism*, vol. 8536 LNAI. 2014.
- [31] L. Del Vasto-Terriente, A. Valls, R. Slowinski, and P. Zielińiewicz, "ELECTRE-III-H: An outranking-based decision aiding method for hierarchically structured criteria," *Expert Systems with Applications*, vol. 42, no. 11, pp. 4910–4926, Jul. 2015.
- [32] C. Ip, R. Law, and H. "Andy" Lee, "The Evaluation of Hotel Website Functionality by Fuzzy Analytic Hierarchy Process," *Journal of Travel & Tourism Marketing*, vol. 29, no. 3, pp. 263–278, Apr. 2012.
- [33] G. Büyüközkan, J. Arsenyan, and G. Ertek, "Evaluation of e-learning web sites using fuzzy axiomatic design based approach," *International Journal of Computational Intelligence Systems*, vol. 3, no. 1, pp. 28–42, 2010.
- [34] Kabak M. and Burmaoğlu S., "A holistic evaluation of the e-procurement website by using a hybrid MCDM methodology," *Electronic Government*, vol. 10, no. 2, pp. 125–150, 2013.
- [35] O. I. Markaki, D. E. Charilas, and D. Askounis, "Application of Fuzzy Analytic Hierarchy Process to Evaluate the Quality of E-Government Web Sites," 2010, pp. 219–224.
- [36] T. L. Saaty, "Decision Making, Scaling, and Number Crunching," *Decision Sciences*, vol. 20, no. 2, pp. 404–409, Jun. 1989.
- [37] B. Roy, "The outranking approach and the foundations of electre methods," *Theory and Decision*, vol. 31, no. 1, pp. 49–73, 1991.

- [38] R. Rekik, I. Kallel, and A. M. Alimi, "Extraction of association rules used for assessing web sites' quality from a set of criteria," in *2014 14th International Conference on Hybrid Intelligent Systems (HIS)*, 2014, pp. 291–296.
- [39] R. Rekik, I. Kallel, and A. M. Alimi, "Ranking Criteria based on Fuzzy ANP for Assessing E-commerce Web Sites," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016)*, 2016, p. in press.
- [40] A. Orriols-Puig, F. J. Martínez-López, J. Casillas, and N. Lee, "A soft-computing-based method for the automatic discovery of fuzzy rules in databases: Uses for academic research and management support in marketing," *Journal of Business Research*, vol. 66, no. 9, pp. 1332–1337, 2013.
- [41] R. Rekik and I. Kallel, "Fuzz-Web: A Methodology Based on Fuzzy Logic for Assessing Web Sites," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 5, pp. 126–136, 2013.
- [42] R. Rekik and I. Kallel, "Fuzzy reduced method for evaluating the quality of institutional web sites," in *2011 7th International Conference on Next Generation Web Services Practices (NWeSP)*, 2011, pp. 296–301.

# Analysis of Rank Aggregation techniques for Metasearch: A Case study

Parneet Kaur<sup>1,\*</sup>, Manpreet Singh<sup>2</sup>, Gurpreet Singh Josan<sup>3</sup>

<sup>1</sup>Research Scholar, Department of Computer Science and Engineering, I.K.G Punjab Technical University, Jalandhar, Punjab, India. Email:kaur.parneet@gmail.com

\* Asst. Professor, CSE Deptt., Ambala College of Engineering and Applied Research, Haryana, India

<sup>2</sup>Department of IT and Computer Science and Engineering, GNDEC, Ludhiana, Punjab, India

<sup>3</sup>Department of IT and Computer Science, Punjabi University, Patiala, Punjab, India

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

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

## I. INTRODUCTION

Rank aggregation is the technique used for merging the web results based list into single confined list. The area of rank aggregation has been used earlier for social choice and primarily on voting methods. But now days, rank aggregation is applied on various optimization based problems related to web like spam fighting, search engine and meta-search engine designing and many more. The information for aggregating the ranks is available easily on web. Hence, web has become a universal repository of human knowledge and culture, which allow the sharing of ideas and information on large scale. The growth of web is unstoppable because of its huge user dependency. Users nowadays don't limit their search to search engines but also make use of purposeful tools. To identify relevant information, Meta-search engines (MSE's) are used on large scale. MSE's make use of relevant query merging techniques whose back bone is rank aggregation methods. The problem of ranking started back at late 13<sup>th</sup> century in the area Social Choice. The milestone of ranking was kept in the paper called as Anatomy of Search Engine [5] which introduces a unique and new method based on Markov Chain to rank web pages [6]. With the passage of time these methods are well adopted by researchers and scholars to improve performance and designing of search engines. Each ranking method is blended with some speciality as they are based on matrix, optimization, statistics, game theory etc.

The very first glimpse of ranking method was found in the Greece system of voting called as Simpler Plural Voting [29] which was based on technique in which every voter submit ordered list of their candidates of their

choice and aggregation of this ordered list was done to find the winner else in other case re-voting was done. Due to this reason modified and improved methods were proposed by [6],[23]. With the onset of ranking methods various rating methods are also developed side by side because both are interdependent on each other.

## II. EARLIER WORK DONE

The main work done in Rank Aggregation in web was proposed in [10]. The author had taken the problem of the Rank Aggregation and emphasised the technique in relation of web designing of Metasearch along with improvement of search precision through associations, choice of documents based on multiple criteria's. The first method explored in the field of aggregation was the Kemeny method [1]. This method was used to provide rank in the area of Social Choice Theory. The work done by Diaconis [8] developed two new methods of rank metrics based on distances and called as the Spearman distance and the Kendall-tau distance metric. The Optimal method [26],[27] was also based on these two distance measures and satisfy the Condorcet Criteria of neutrality and consistency. Dwork [10] explained the new concept of Markov Chain apart from ranking methods. Another contribution was done in the process of Rank function in the paper [5] which uses indexing where a query was considered for performing operations based on aggregation. It basically uses two ways as a Rank function. The first one is based on providing a score value to a document as per its importance or similarity with the query. In second way, at least set of two documents is taken and ordering is done as per the query.

The paper by Beg [2],[3] proposed a new concept based on soft computing techniques like Genetic and Fuzzy methods with rank aggregation techniques. The author obtained the optimized results using Spearman Footrule distance called as Footrule optimal aggregation (FOA) for full lists. Due to NP hard nature of FOA for partial lists Genetic Algorithm was used to refine the output for the PFOA. The author used the Borda Score method for aggregating the list. The work done by Shimura proposed the Shimura technique [25] for ranking and rating uses the concept of fuzzy ordering. This method fits best for the non transitive ranking. Another soft computing method discussed was the Entropy minimization technique which generates complex results unsuited for producing refined list. In the paper by Beg [4] focuses on the parameters like search quality, user feedbacks and biased rank aggregation methods. In order to measure the user search quality, the '*satisfaction*' of user was taken top most for getting desired results. The final results by Beg paper helps to improve the quality of search engine by satisfaction rating provided by the user.

The methods used by Montague [20],[21] for aggregation were based on the decomposition of the normalize scores. The relevance score normalization method discussed in [12], [18] was used for non retrieved documents. Finally the relevance score generated are combined into single value for each document. Some of the normalized algorithms are the CombMIN , CombMED , CombMAX etc.

Renda and Straccia paper [22] explored the aggregation method with new techniques of rank fusion for solving the problems related to metasearch. Paper by [24] covered the aggregating rank method for similar items which helps in resolving the problem of providing correct rank. Various similarity functions were used in the paper like pair-wise similarity which normalized the function called as Snorm. Also work done by Meng [19]

focused on the various result merging strategies. The results were collected from various search engines and then given to various MSEs which were joined together to form a single ranked list. Similarly the D-WISE method by Ding [9] designed a new technique using the local rank of a document provided by the search engine and the ranking score.

The usage of various open source tools like Weka, Rapid Miner etc are also applied in context of web for obtaining the ranks. In the work of Kaur [13] the Rapid Miner tool was used for obtaining the frequency of similar keywords related to query which is based on operation for calculation of similarity among various URLs. In this paper top 3 URLs were obtained, relevant to the user query using Rapid Miner Similarity attribute, which helps in reducing the user efforts and saves time spent on all non relevant URLs. Also in Kaur [14] the Weka tool was applied to calculate the ranks obtained by students in education sector.

### III. RANKING METHODS FOR WEB

Purpose of aggregating rank is to design a new, better and powerful ranking list by filtering and merging various rank lists. The various rank aggregation methods are being developed which can directly applied in the area of web. Some of these methods are as follows:

#### A. Borda Method

This method was developed by Jean Charles de Borda in 1770 [6]. The method is based on the scores in which each candidate is allotted a score that is equal to the number of candidates who outrank each other. This method was applied first in the area of Social Choice. In this method rank is represented by R and each item with n. The method consists of ranking L provided to n items. The list of item say  $lk \in L$ . The  $BC_{lk}(a)$  is equivalent to items b in  $lk$  such that  $lk(b) > lk(a)$ . Hence, the formula for Borda score is as follows:

$$BC_{lk}(a) = \sum_{l \in L} BC_l(a) \quad (1)$$

Order of ranking starts from highest to lowest sorting scores  $BC_l$ . The Borda score for a list  $lk$  where  $|lk| = |U| - d$ , calculates the  $BC_{lk}(a)$  for all items  $a \in lk$ , where  $b \neq lk$  and the Borda Score is given as:

$$BC_{lk}(b) = \frac{(d+1)^2 + (d+1)}{2d} \quad (2)$$

Borda Methods are further modified into Borda Fuse which is based on fusion which takes the “consensus” ranking. This method computes a single ranking list which is preferred over several lists.

#### B. Markov Method

This method of ranking is designed by Markov in 1906 and its various methods based on chains are proposed by[10]. Markov is a stochastic process consisting of the chain for a set of states having items. A non negative stochastic matrix M of  $n \times n$  item is also formed in the Markov chain. This method works by initially moving from a particular state to other state. The Markov chains based method uses different heuristics rules to construct the transition probability matrix. The states are basically web pages rank by rank engines. The four heuristics rules are MC1, MC2, MC3 and MC4.

### C. CombMNZ Method

The CombMNZ method was proposed by [11,18] which works both on score as well as on ranks. This algorithm calculates the Borda rank by multiplying each value of the page with the value of total number of occurrences of that page. The descending order gives the new improved Borda score as calculated in [23]. The equation 3 CombMNZ method is given as:

$$DB(R^A) = DB(R^1) \cup \dots \cup DB(R^S) \quad (3)$$

Where  $S$  is the rankers given as  $\{R^1, \dots, R^S\}$ .

### D. Mean By Variance Method (MBV)

The MBV method is based on the soft computing technique of Fuzzy Logic and used for computing the aggregating the ranks for a given set of lists. The scores are obtained using the ranks of the lists. These scores are arranged in the decreasing order and largest value is provided with minimum rank. The formula for the Mean By Variance technique is given in equation 4.

$$mbv(i) = \left( \frac{\bar{m}_{di}}{\bar{\sigma}_{di}^2} \right) \quad (4)$$

Where  $\bar{m}_{di}$  is the mean of the documents and  $\bar{\sigma}_{di}^2$  is the variance of the position for documents  $di$ .

### E. Page Rank Method

Brin and Page [5] came up with the new method of ranking pages for the web called as PageRank method. This method takes under consideration the title tag, the anchor tag along with some ranking algorithms to rank a page [17]. The equation for PageRank is given as:

$$PR(u) = (1 - e) + e \sum_{v \in B(u)} \frac{PR(v)}{N_v} \quad (5)$$

Where  $PR(u)$  denotes the page rank for page  $(u)$  and  $N_v$  denotes the number of out-links of page  $(v)$ . The  $e$  denotes the dampening factor with 0.85 value.

### F. HITS method

HITS method was developed by Klienberg [16]. This method is comprised of hubs and authority similar to in-links and out-links. The author has taken the number of edges among hubs and authority which matters a lot. The hubs and authority are provided with weight which is used to calculate the output. The iterations are executed using hubs and authority which is computed using equation 6 as:

$$a_i = \sum_{j \in B(i)} h_j \text{ and } h_j = \sum_{i \in F(j)} a_i \quad (6)$$

With the iteration, the graph is formed which reflects the overall structure of the HITS algorithm.

#### IV. AGGREGATION METHODS BASED ON DISTANCE METRICS

##### A. Kendall-tau distance metrics

The Kendall distance finds the pair-wise disagreements between two rating list say  $a$  and  $b$  and also gives the ranking. The formula for Kendall's distance metrics [15] is given in equation 7 as:

$$K(a,b) = \sum_{\{m,n\} \in R} KT^* a, b(\alpha, \beta) \quad (7)$$

Where  $R$  is the set of unordered pairs of distinct items in  $Y$  and  $KT^* a, b(\alpha, \beta)$  is given as:

$$KT^* a, b(\alpha, \beta) = \begin{cases} 0 & \text{if } x_i, x_j \text{ are in the same order in } \alpha \text{ and } \beta \\ 1 & \text{if } x_i, x_j \text{ are in the inverse order in } \alpha \text{ and } \beta \end{cases} \quad (8)$$

The normalization of Kendall-tau rank metrics is done by dividing its maximum value [7] with  $\binom{m}{2}$ . This method is computed in  $m \log m$  time.

##### B. Spearman's distance

Spearman Rank metrics method compute the sum of the absolute difference between the score values of the rankings. The formula for Spearman distance metrics is given in equation 9 as:

$$FT(\alpha, \beta) = \sum_{a=1}^m |\alpha(a) - \beta(b)| \quad (9)$$

where  $\alpha$  and  $\beta$  are total ranking. The normalized footrule distance metrics  $FT$  [7] is obtained by dividing with its maximum value as  $\left(\frac{m^2}{2}\right)$ . Also the  $FT$  value which is equivalent to 1 stands for distinct ranking and 0 stands for similar ranking. The  $FT$  can also be computed in linear time.

#### V. EXPERIMENTAL RESULTS

In this paper, experiment is carried out using rank aggregation methods which are Borda, Mean By Variance and CombMNZ methods with a web based query on four popular search engines which are Google<sup>1</sup>, Altavista<sup>2</sup>, Excite<sup>3</sup> and Deeperweb<sup>4</sup>. The query chosen for experimental analysis is *Alcoholism* which is taken among 37 queries from [10] and applied on the four mentioned search engines. The top 40 results are collected from all the search engines in order to generate Universal list. Then top 15 URLs are selected for list formation through filtration i.e. selection of all common URLs are chosen from the Universal list. The Spearman's Footrule distance and Kendall-Tau distance measures are used to calculate the positions between the URLs obtained from the query *Alcoholism* for score generation. The normalized Spearman's Footrule and Kendall-Tau distance measure is also obtained using equations 7 and 8. The top 15 results obtained for the query *Alcoholism* is given in Table 1 as full list.

TABLE 1

FULL LIST OF TOP 15 RESULTS FOR THE QUERY ALCOHOLISM FROM UNIVERSAL LIST

1. <http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm>
2. <http://en.wikipedia.org/wiki/Alcoholism>
3. <http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm>
4. <http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-use-disorders/genetics-alcohol-use-disorders>
5. <http://www.webmd.com/mental-health/addiction/alcohol-abuse-and-dependence-topic-overview>
6. <http://emedicine.medscape.com/article/285913-overview>
7. <http://alcoholism.about.com/>
8. <http://www.merriam-webster.com/dictionary/alcoholism>
9. <https://www.psychologytoday.com/conditions/alcohol-abuse>
10. <http://dictionary.reference.com/browse/alcoholism>
11. <http://www.healthline.com/health-slideshow/best-alcoholism-blogs>
12. <http://www.learn-about-alcoholism.com/>
13. <http://www.britannica.com/EBchecked/topic/13448/alcoholism/251753/Causes-of-alcoholism>
14. <http://www.timberlineknolls.com/alcohol-addiction/signs-effects>
15. <http://umm.edu/health/medical/altmed/condition/alcoholism>

The search engines used are as follows:

1. [www.google.com](http://www.google.com)
2. [www.altavista.com](http://www.altavista.com)
3. [www.excite.com](http://www.excite.com)
4. [www.deeperweb](http://www.deeperweb)

To evaluate the search results of Table 1 some definitions for aggregating the lists [2] used are given as:

*A. Definition 1*

Suppose a set of elements in Universe  $U$  and  $P \in U$  is given, also  $lt$  represents an ordered list with respect to  $U$ . Hence, the elements in the list are  $lt = \{lt_1, lt_2, \dots, lt_k\}$ , where  $lt_i \in P$  and  $lt_1 > lt_2 > \dots > lt_{|lt|}$  where  $>$  is an relation of order. The position of element in a list is  $lt(a)$  is further  $a \in U \cap a \in lt$ . For each element a unique value is assigned such as  $U = \{1, 2, 3, \dots, U\}$ .

*B. Definition 2*

Full or Complete List: A List having all the elements is known as Full or Complete List of a Universe U.

### C. Definition 3

Suppose the set of elements in a Full or Complete list is denoted by  $FL = \{fl_1, fl_2, \dots, fl_k\}$ , then the normalized Kendall-tau distance metrics is represented with equation 10 for full and Complete list as:

$$KT(lt, FL) = \frac{\sum_{i=1}^z KT(lt, FL)}{z} \quad (10)$$

Also the normalized aggregated Spearman's Footrule distance for complete or full list is given in equation 11 as:

$$FT(lt, FL) = \frac{\sum_{i=1}^z FT(lt, FL)}{z} \quad (11)$$

### D. Definition 4

Suppose the aggregated list used for obtaining a new Full or Complete List is denoted by a set lists  $FL = \{fl_1, fl_2, \dots, fl_n\}$  and the final aggregated list obtained always provides a normalized Kendall and Spearman's distance metrics values for the given list.

### E. Definition 5

A Partial or Incomplete list supposes  $lt_p$  is one which contains those elements in a subset of Universe U which are part of Universal Set say U.

For the experiment analysis, the dataset for the Universal list is taken from search engines which are Google<sup>1</sup>, AltaVista<sup>2</sup>, Excite<sup>3</sup> and DeeperWeb<sup>4</sup> for the query *Alcoholism*. Total 120 results are taken from the Universal List. As per the *Definition 2*, only top 15 elements are selected to form a Full list  $l$  from the given Universal list and filtration is done manually by removing those elements which are not common in order to form full list for three search engines as list  $l_1, l_2, l_3$  as shown in Table 2.

TABLE 2  
LIST 1, LIST 2, LIST 3 FROM SEARCH ENGINES

Lists with Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
List L1	1	7	2	6	5	11	3	10	14	12	9	4	15	8	13
List L2	2	1	7	6	3	15	8	11	5	13	12	14	10	9	4
List L3	2	1	3	5	4	6	8	10	15	9	12	7	11	13	14

The dataset generated from the query *Alcoholism* in Table 2 is applied to the rank aggregation methods such as Borda Method, Mean By Variance Method and CombMNZ methods to generate the aggregated list.

As per *Definition 3 and Definition 4*, the Kendall tau distance and the Spearman's Footrule distance is calculated between each List 1, 2, 3 corresponding to the Full List as shown in the Table 3.

TABLE 3  
DISTANCE MEASURES FOR THREE LIST L1, L2 AND L3.

Lists	F(L1,L2)	F(L2,L3)	F(L2,L3)
Kendall Tau distance	0.295	0.3714	0.533
Spearman Foot-rule distance	0.533	0.355	0.4177

It is observed from the Table 3, that for List 1 and List 2, the Kendall Tau distance came out to be very less as compare to Spearman's Footrule. As per *Definition 5*, the Normalized Aggregated Spearman's Footrule and Normalized Aggregated Kendall Tau distance is calculated and shown in the Table 4.

TABLE 4  
AGGREGATED DISTANCE MEASURE FOR THE LIST L1, L2 AND L3

Aggregated Normalized Distance for Kendall and Spearman Foot rule distances for List 1,2, 3	Normalized Aggregated Result
$F(l, L) = \frac{\sum_{l=1}^k F(l, L)}{k}$	0.3614
$K(l, L) = \frac{\sum_{l=1}^k K(l, L)}{k}$	0.2634

It is observed from the Table 4 that the normalized results obtain from Kendall tau is again less than the normalized Spearman's Footrule distance. Hence, it is concluded that performance of Kendall Tau for given Query 1 is producing better results as distance is minimized between various items for the List 1,2,3.

The Aggregated list generated by using Borda method, Mean By Variance method as well as using CombMNZ methods are applied using Kendall Tau and Spearman's Foot rule distances to further calculates the distance value which is shown in the Table 5 .The calculations are carried out using Matlab.

TABLE 5  
FINAL RESULT TABLE FOR BORDA, MEAN BY VARIANCE AND COMBMNZ METHODS

Methods Used	Spearman Footrule Distance(SFD)	Kendall Tau Distance (KTD)
Borda	0.283	0.200
Mean By Variance	0.746	0.542
CombMNZ	0.533	0.400

From the Table 5, it is observed that the Borda method with Kendall Tau distance gives better results than with Spearman's Footrule distance. Also the results generated for Mean By Variance and CombMNZ normalized methods with Kendall Tau produces less distances which helps in showing the performance of the methods. It has been observed from the study that the Rank Aggregation techniques when derived using distance measure helps in improving the performance of search engines since distances between lists are minimized. These techniques are further helpful in comparing the results between various Search engines and also the better aggregated results are used for designing of Meta-search engines.

The Borda method derives the best aggregated List with top 15 results in relevant order for the query *Alcoholism* as shown in the Table 6. The results obtained are further evaluated using Open source tool called as Rapid Miner ([www.rapidminer.com](http://www.rapidminer.com)) which shows the occurrence of keyword Alcoholism, Alcohol, Alcoholic obtained from main keyword which is Alcoholism to further check the relevancy of URLs obtained after rank aggregation. The URLs with maximum keyword occurrences are arranged in the increasing order and matched with the list obtained from various rank aggregation methods as shown in the Table 7.

TABLE 6  
AGGREGATED LIST FOR THE QUERY ALCOHOLISM

The Aggregated Refined List for the Query Alcoholism	New Rank Using Borda Score
<a href="http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm">http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm</a>	1
<a href="http://en.wikipedia.org/wiki/Alcoholism">http://en.wikipedia.org/wiki/Alcoholism</a>	2
<a href="http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-use-disorders/genetics-alcohol-use-disorders">http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-use-disorders/genetics-alcohol-use-disorders</a>	4
<a href="http://www.healthline.com/health-slideshow/best-alcoholism-blogs">http://www.healthline.com/health-slideshow/best-alcoholism-blogs</a>	11
<a href="http://emedicine.medscape.com/article/285913-overview">http://emedicine.medscape.com/article/285913-overview</a>	6
<a href="http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm">http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm</a>	3
<a href="http://www.webmd.com/mental-health/addiction/alcohol-abuse-and-dependence-topic-overview">http://www.webmd.com/mental-health/addiction/alcohol-abuse-and-dependence-topic-overview</a>	5
<a href="https://www.psychologytoday.com/conditions/alcohol-abuse">https://www.psychologytoday.com/conditions/alcohol-abuse</a>	9

<a href="http://www.britannica.com/EBchecked/topic/13448/alcoholism/251753/Cau-ses-of-alcoholism">http://www.britannica.com/EBchecked/topic/13448/alcoholism/251753/Cau-ses-of-alcoholism</a>	13
<a href="http://dictionary.reference.com/browse/alcoholism">http://dictionary.reference.com/browse/alcoholism</a>	10
<a href="http://alcoholism.about.com/">http://alcoholism.about.com/</a>	7
<a href="http://www.learn-about-alcoholism.com/">http://www.learn-about-alcoholism.com/</a>	12
<a href="http://umm.edu/health/medical/altmed/condition/alcoholism">http://umm.edu/health/medical/altmed/condition/alcoholism</a>	15
<a href="http://www.timberlineknolls.com/alcohol-addiction/signs-effects">http://www.timberlineknolls.com/alcohol-addiction/signs-effects</a>	14
<a href="http://www.merriam-webster.com/dictionary/alcoholism">http://www.merriam-webster.com/dictionary/alcoholism</a>	8

TABLE 7  
RESULTS OF TOP 15 WEBSITES WITH OCCURRENCE OF KEYWORDS

S.No.	Website or URLs	Alcoholism	Alcohol	Alcoholic	Rank
1.	<a href="http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm">http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm</a>	0.068	0.007	0.003	1
2.	<a href="http://en.wikipedia.org/wiki/Alcoholism">http://en.wikipedia.org/wiki/Alcoholism</a>	0.010	0.005	0.000	4
3.	<a href="http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm">http://www.nlm.nih.gov/medlineplus/ency/article/000944.htm</a>	0.009	0.003	0.003	5
4.	<a href="http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-use-disorders/genetics-alcohol-use-disorders">http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-use-disorders/genetics-alcohol-use-disorders</a>	0.000	0.000	0.000	14
5.	<a href="http://www.webmd.com/mental-health/addiction/alcohol-abuse-and-dependence-topic-overview">http://www.webmd.com/mental-health/addiction/alcohol-abuse-and-dependence-topic-overview</a>	0.002	0.016	0.001	10
6.	<a href="http://emedicine.medscape.com/article/285913-overview">http://emedicine.medscape.com/article/285913-overview</a>	0.002	0.005	0.001	11
7.	<a href="http://alcoholism.about.com/">http://alcoholism.about.com/</a>	0.004	0.001	0.001	9
8.	<a href="http://www.merriam-webster.com/dictionary/alcoholism">http://www.merriam-webster.com/dictionary/alcoholism</a>	0.007	0.004	0.005	6
9.	<a href="https://www.psychologytoday.com/conditions/alcohol-abuse">https://www.psychologytoday.com/conditions/alcohol-abuse</a>	0.002	0.007	0.001	12
10.	<a href="http://dictionary.reference.com/browse/alcoholism">http://dictionary.reference.com/browse/alcoholism</a>	0.015	0.004	0.007	3
11.	<a href="http://www.healthline.com/health-slideshow/best-alcoholism-blogs">http://www.healthline.com/health-slideshow/best-alcoholism-blogs</a>	0.059	0.009	0.000	2
12.	<a href="http://www.learn-about-alcoholism.com/">http://www.learn-about-alcoholism.com/</a>	0.007	0.021	0.002	7
13.	<a href="http://www.britannica.com/EBchecked/topic/13448/alcoholism/251753/Causes-of-alcoholism">http://www.britannica.com/EBchecked/topic/13448/alcoholism/251753/Cau-ses-of-alcoholism</a>	0.000	0.004	0.007	15

14.	<a href="http://www.timberlineknolls.com/alcohol-addiction/signs-effects">http://www.timberlineknolls.com/alcohol-addiction/signs-effects</a>	0.002	0.006	0.006	13
15.	http://umm.edu/health/medical/altmed/condition/alcoholism	0.005	0.019	0.002	8

The Table 7 also shows the frequency of occurrence of keywords in various websites. Further it has been observed that the best aggregated list obtained from Borda Score method matches up to some extent with the results obtained from the Open source tool Rapid Miner as this software generates the result in form of the terms with maximum occurrences of keyword. Also the query *Alcoholism* as applied in the Rapid Miner tool shown in Figure 1 gives the final list which contains maximum occurrence of the keyword which matches with the list obtained from Borda method to a large extent. Hence the performance of Borda method comes out to be best in terms of keyword occurrence.

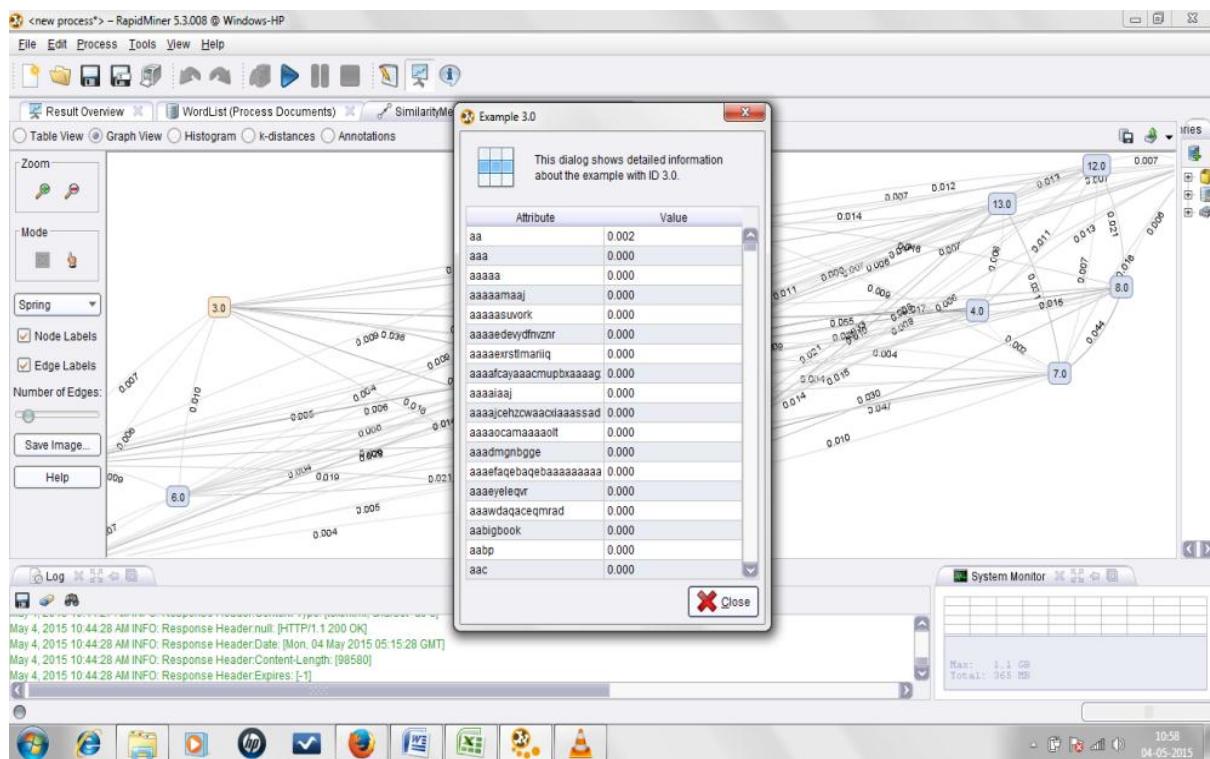


Figure 1. Frequency of keywords with occurrence among various websites.

The Figure 1 indicates the occurrence of various keywords individually in top 10 websites selected in Table 1. The snapshot is given in Figure 1 from the Rapid Miner tool ([www.rapidminer.com](http://www.rapidminer.com)) for keyword *Alcoholism*.

## VI. CONCLUSION AND FUTURE SCOPE

In this paper, rank aggregation methods are discussed, investigated and experimental implementation are done to check the performance of search engines on the basis of scores obtained through the distance measures. Also, the relevancy check is carried out on a dataset of top 15 search items obtained from four famous search engines

Google<sup>1</sup>, AltaVista<sup>2</sup>, Excite<sup>3</sup> and Deeperweb<sup>4</sup>. It is observed from this study that minimum distance values obtained from rank aggregation improves the performance of search engine and also the quality of metasearch. Aim of this work is to apply various methods and techniques of ranking based on scores and to check the frequency of occurrence of keywords for some specific query. The future scope of this work is to apply other rank aggregation methods from single document to multiple queries for feature extraction, selection of important lists for ranking and many more. To optimize the list, the rank aggregation methods can also be moulded with some soft computing techniques which will help in producing excellent results in terms of ranking and rating. Also, the soft computing techniques will further help in improving the performance and quality of search for web list.

## REFERENCES

- [1] J.J. Bartholdi, C.A. Tovey, M.A.Trick, “Voting schemes for which it can be difficult to tell who won the election”, *Social Choice and Welfare*, 6(2), 1989 157-165.
- [2] M.M.S. Beg, N. Ahmad, “Soft computing techniques for rank aggregation on the World Wide Web”, *Journal of World Wide Web*, 6 (1), 2003, pp.5-22.
- [3] M.M.S. Beg, “Parallel Rank Aggregation for the World Wide Web. World Wide Web”, Kluwer Academic Publishers, 2004, pp.5-22.
- [4] M.M.S. Beg, N.Ahmad, “Web search enhancements by mining user actions”, *The Journal of Information Sciences*, 177,2006, pp.5203-5218.
- [5] S. Brin, L. Page, “The anatomy of a large-scale hypertextual Web search engine”, In Proceedings of Seventh International World Wide Web Conference, Computer Networks, 30(1-7),1998, pp.107-117.
- [6] J.C. Borda, “ Mémoire sur les élections au scrutin. Histoire de l”, Académie Royale des Sciences, 1781.
- [7] P. Diaconis, R. Graham, “Spearman's footrule as a measure of disarray”, *Journal of the Royal Statistical Society, Series B*, 39(2), 1979, pp.262-268.
- [8] P. Diaconis, “Group Representation in Probability and Statistics”, IMS Lecture Series 11, Institute of Mathematical Statistics. Hayward, CA, 1988.
- [9] Y. Ding, “Applying Weighted Page Rank to author Citation Networks”, In Proceedings of JASIST ,2011, pp.1-19.
- [10] C. Dwork, R.Kumar, M. Noar, D. Sivakumar, “Rank aggregation methods for the Web”, In Proceedings of ACM Conference on World Wide Web, 2001, pp. 613–622.
- [11] D.E. Goldberg, “Genetic Algorithms in Search, Optimization, and Machine Learning”, Book by Addison Wesley, 1989.
- [12] K.L. Fox, O. Frieder, M. Knepper, E. Snowberg, “SENTINEL: A multiple engine information retrieval and visualization system”, *Journal of the ASIS*, 50(7), 1999.
- [13] P. Kaur, G. Singh. S. Khurmi, “Analysis of Similar Documents Using Rapid Miner Tool”, In the Proceedings of International Conference on issues and challenges in Intelligent Computing Techniques by IEEE,Delhi Section, 2014, pp. 470-475.
- [14] P. Kaur, M. Singh, G.S. Josan, “Classification and prediction based data mining algorithms to predict slow learners in education sector”, In Proceedings of 3rd International Conference on Recent Trends in Computing, Procedia Computer Science 57, Elsevier, 2015, pp. 500 – 508.

- [15] M.G. Kendall, “A New Measure of Rank Correlation”, *Biometrika*, 30:1/2, 1938, pp.81–93.
- [16] J. Kleinberg, “Authoritative sources in a hyperlinked environment”, *Journal of the ACM* 46, 1999.
- [17] A.N. Langville, C.D. Meyer, “Google’s PageRank and Beyond: The Science of Search Engine Rankings”, Book. Princeton University Press, 2006.
- [18] A.N. Langville, C.D. Meyer, “Book of Who’s #1?: The Science of Rating and Ranking”, Princeton University Press, Princeton, NJ, USA, 2012.
- [19] K. Massey, “Statistical Models Applied to the Rating of Sports Teams”. Bachelor’s thesis, Bluefield College, 1997.
- [20] W. Meng, Y. Lu, L. Shu, C. Yu, K. Liu, “Evaluation of result merging strategies for metasearch engines”, In Proceedings of 6th International Conference on Web Information Systems Engineering (WISE Conference), New York, CZ2005.
- [21] M.H. Montague, J.A. Aslam, “Relevance score normalization for Meta search”, In Proceeding of Conference on Information and Knowledge Management, Atlanta, GA, 2001, pp. 427-433.
- [22] M.E. Renda, U. Straccia, “Web metasearch: Rank vs. score based rank aggregation methods”, In Proceedings of ACM SAC , 2003, pp. 841–846.
- [23] D.J. Saari, “The mathematics of voting: democratic symmetry”, *Economist*. 83, 2000, Available at <http://vidal.ece.sc.edu/822/papers/econ-voting.html>.
- [24] D. Sculley, “Rank aggregation for similar items”, In Proceedings of the SIAM Conference on Data Mining (SDM), Springer, 2006.
- [25] M. Shimura, “Fuzzy set concept in Rank Ordering Objects”, *Journal of Mathematical Analysis and Applications*, 43, 1973, pp.717-733.
- [26] X. Wenpu. A. Ghorbani, “Weighted pagerank algorithm”, In Proceedings of 2<sup>nd</sup> Annual Conference on IEEE, Communication Networks and Services Research, 2004, pp. 305-314.
- [27] H.P. Young, A. Levenglick, “A Consistent Extension of Condorcet’s Election Principle”, *SIAM Journal on Applied Mathematics*, 35(2), 1978, pp.285-300.
- [28] H.P. Young, “Condorcet’s theory of Voting”, *American Political Science Review*, 82, 1988, pp.1231-1244.
- [29] Single Plural Voting, [https://en.wikipedia.org/wiki/Plurality\\_%28voting%29](https://en.wikipedia.org/wiki/Plurality_%28voting%29). Dated-21-03-2016.

# Construction of a Jacobi matrix by given n eigenpairs

Seyed Abolfazl Shahzadeh Fazeli<sup>1,\*</sup>, Soodeh Kakuei Nejad<sup>2</sup>, Maryam Babaie<sup>3</sup>

<sup>1</sup>Parallel Processing Laboratory, Faculty of Mathematics, Yazd University, Yazd, Iran

fazeli@yazd.ac.ir

<sup>2</sup>Parallel Processing Laboratory, Faculty of Mathematics, Yazd University, Yazd, Iran

kakuei\_nejad90@yahoo.com

<sup>3</sup>Parallel Processing Laboratory, Faculty of Mathematics, Yazd University, Yazd, Iran

m.babaie193@gmail.com

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

**Keywords:** Inverse problem, Jacobi matrix, Eigenpair

## I. INTRODUCTION

A Jacobi matrix  $J$  is a matrix of the following form:

$$J = \begin{vmatrix} a_1 & b_1 & & & & \\ b_1 & a_2 & b_2 & & & \\ & b_2 & a_3 & b_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & b_{n-2} & a_{n-1} & b_{n-1} \\ & & & & b_{n-1} & a_n \end{vmatrix} \quad b_i > 0$$

An inverse eigenvalue problem for a Jacobi matrix is how to determine the elements of this matrix from given eigen data. Some applications of this problem are including vibration theory and structural design, for instant, the vibrating rod model [1, 2]. Some new results have been obtained on the construction of a Jacobi matrix [3, 4]. The problem of constructing a Jacobi matrix from its four or five eigenpairs was discussed in [5]. The problem of constructing a Jacobi matrix from its six or seven eigenpairs is the subject of our study.

*Problem1:* Let  $\lambda_1, \lambda_2, \dots, \lambda_{2n}$  be  $2n$  distinct real scalars (with  $\lambda_1 > \lambda_2 > \dots > \lambda_{2n}$ ) and  $x_1 = [x_{1,1}, x_{1,2}, \dots, x_{1,2n}]^T$ ,  $x_2 = [x_{2,1}, x_{2,2}, \dots, x_{2,2n}]^T$ , ...,  $x_{2n} = [x_{2n,1}, x_{2n,2}, \dots, x_{2n,2n}]^T$  be  $2n$  real orthogonal vectors of size  $2n$ , finding a Jacobi matrix  $J$  of size  $2n$  such that  $(\lambda_1, x_1) \cdot (\lambda_2, x_2) \cdot \dots \cdot (\lambda_{2n}, x_{2n})$  are its  $2n$  eigenpairs.

*Problem2:* Let  $\lambda_1, \lambda_2, \dots, \lambda_{2n+1}$  be  $2n+1$  distinct real scalars (with  $\lambda_1 > \lambda_2 > \dots > \lambda_{2n+1}$ ) and  $x_1 = [x_{1,1}, x_{1,2}, \dots, x_{1,2n+1}]^T$ ,  $x_2 = [x_{2,1}, x_{2,2}, \dots, x_{2,2n+1}]^T$ , ...,  $x_{2n+1} = [x_{2n+1,1}, x_{2n+1,2}, \dots, x_{2n+1,2n+1}]^T$  be  $2n+1$  real

orthogonal vectors of size  $2n+1$ , finding a Jacobi matrix  $J$  of size  $2n+1$  such that  $(\lambda_1, x_1) \cdot (\lambda_2, x_2) \cdot \dots \cdot (\lambda_{2n+1}, x_{2n+1})$  are its  $2n+1$  eigenpairs.

The remaining text of this paper is organized as follows: In Sections 2, the sufficient conditions for the existence and uniqueness of the solution of reconstruction of a Jacobi matrix by given  $2n$  or  $2n+1$  eigenvalues and corresponding eigenvectors are proposed. Numerical algorithms and some numerical examples are given in Section 3.

*Lemma 1.1.* See [6,7]. Let  $\lambda, \mu$  be two distinct real scalars (with  $\lambda > \mu$ ) and  $x = [x_1, x_2, \dots, x_n]^T, y = [y_1, y_2, \dots, y_n]^T$  be two real orthogonal vectors of size  $n$ , there is a unique Jacobi matrix  $J$  such that  $(\lambda, x)$  and  $(\mu, y)$  are its two eigenpairs if the next condition is satisfied:

$$\frac{d_k}{D_k} > 0 \quad , \quad (k = 1, 2, \dots, n - 1),$$

where

$$d_k = \sum_{i=1}^k x_i y_i, \quad (k = 1, 2, \dots, n),$$

$$D_k = \begin{vmatrix} x_k & x_{k+1} \\ y_k & y_{k+1} \end{vmatrix} \neq 0 \quad , \quad (k = 1, 2, \dots, n - 1).$$

and the elements of matrix  $J$  are

$$b_k = \frac{(\lambda - \mu)d_k}{D_k}, \quad (k = 1, 2, \dots, n - 1),$$

$$a_1 = \lambda - \frac{b_1 x_2}{x_1},$$

$$a_n = \lambda - \frac{b_{n-1} x_{n-1}}{x_n},$$

$$a_k = \begin{cases} \lambda - \frac{(b_{k-1} x_{k-1} + b_k x_{k+1})}{x_k}, & x_k \neq 0 \quad , \\ \mu - \frac{(b_{k-1} y_{k-1} + b_k y_{k+1})}{y_k}, & x_k = 0 \quad , \end{cases} \quad k = 2, 3, \dots, n - 1).$$

From Lemma 1.1 under some conditions, two eigenpairs can determine a unique Jacobi matrix. Subsequently, we have the Theorem 1.1.

*Lemma 1.2* (see [8]). Let  $\lambda, \mu$  and  $\nu$  be three distinct real scalars (with  $\lambda > \mu > \nu$ ) and  $x = [x_1, x_2, \dots, x_n]^T, y = [y_1, y_2, \dots, y_n]^T, z = [z_1, z_2, \dots, z_n]^T$  be three real orthogonal vectors of size  $n$ , there is a unique Jacobi matrix  $J$  such that  $(\lambda, x) \cdot (\mu, y)$  and  $(\nu, z)$  are its three eigenpairs if the following conditions are satisfied:

$$(\lambda - \mu)d_k^{(1)}/D_k^{(1)} = (\lambda - \nu)d_k^{(2)}/D_k^{(2)} > 0$$

If  $x_k = 0$  then

$$(\lambda - \mu)d_j^{(1)}/D_j^{(1)} = (\mu - \nu)d_j^{(3)}/D_j^{(3)}, \quad j = k, k - 1,$$

where

$$d_k^{(1)} = \sum_{i=1}^k x_i y_i, \quad d_k^{(2)} = \sum_{i=1}^k x_i z_i, \quad d_k^{(3)} = \sum_{i=1}^k y_i z_i, \\ D_k^{(1)} = \begin{vmatrix} y_k & y_{k+1} \\ x_k & x_{k+1} \end{vmatrix}, \quad D_k^{(2)} = \begin{vmatrix} z_k & z_{k+1} \\ x_k & x_{k+1} \end{vmatrix}, \quad D_k^{(3)} = \begin{vmatrix} z_k & z_{k+1} \\ y_k & y_{k+1} \end{vmatrix}, \quad (k = 1, 2, \dots, n-1),$$

and the elements of matrix J are

$$b_k = \frac{(\lambda - \mu)d_k^{(1)}}{D_k^{(1)}}, \quad (k = 1, 2, \dots, n-1),$$

$$a_1 = \lambda - \frac{b_1 x_2}{x_1}$$

$$a_n = \lambda - \frac{b_{n-1} x_{n-1}}{x_n}$$

$$a_k = \begin{cases} \lambda - \frac{(b_{k-1} x_{k-1} + b_k x_{k+1})}{x_k}, & x_k \neq 0, \\ \mu - \frac{(b_{k-1} y_{k-1} + b_k y_{k+1})}{y_k}, & x_k = 0, \end{cases} \quad k = 2, 3, \dots, n-1.$$

From Lemma 1.2, under some conditions three eigenpairs can determine a unique Jacobi matrix. Therefore, we have the Theorem 1.1.

*Theorem 1.1.* Let  $\lambda, \mu, \xi, \eta$  be four distinct real scalars (with  $\lambda > \mu > \xi > \eta$ ) and  $x = [x_1, x_2, \dots, x_n]^T, y = [y_1, y_2, \dots, y_n]^T, m = [m_1, m_2, \dots, m_n]^T, r = [r_1, r_2, \dots, r_n]^T$  be four real orthogonal vectors of size  $n$ , there is a Jacobi matrix  $J$  of size  $n$  such that  $(\lambda, x)$ ,  $(\mu, y)$ ,  $(\xi, m)$  and  $(\eta, r)$  are its four eigenpairs if the following conditions are satisfied:

$$(\lambda - \mu)d_k^{(1)}/D_k^{(1)} = (\lambda - \xi)d_k^{(2)}/D_k^{(2)} = (\lambda - \eta)d_k^{(3)}/D_k^{(3)} > 0$$

If  $x_k = 0$  then

$$(\lambda - \mu)d_j^{(1)}/D_j^{(1)} = (\mu - \xi)d_j^{(4)}/D_j^{(4)} = (\mu - \eta)d_j^{(5)}/D_j^{(5)}, \quad j = k, k-1$$

where

$$d_k^{(1)} = \sum_{i=1}^k x_i y_i, \quad d_k^{(2)} = \sum_{i=1}^k x_i m_i, \quad d_k^{(3)} = \sum_{i=1}^k x_i r_i, \\ d_k^{(4)} = \sum_{i=1}^k y_i m_i, \quad d_k^{(5)} = \sum_{i=1}^k y_i r_i, \quad d_k^{(6)} = \sum_{i=1}^k m_i r_i, \\ D_k^{(1)} = \begin{vmatrix} y_k & y_{k+1} \\ x_k & x_{k+1} \end{vmatrix}, \quad D_k^{(2)} = \begin{vmatrix} m_k & m_{k+1} \\ x_k & x_{k+1} \end{vmatrix}, \quad D_k^{(3)} = \begin{vmatrix} r_k & r_{k+1} \\ x_k & x_{k+1} \end{vmatrix}, \\ D_k^{(4)} = \begin{vmatrix} m_k & m_{k+1} \\ y_k & y_{k+1} \end{vmatrix}, \quad D_k^{(5)} = \begin{vmatrix} r_k & r_{k+1} \\ y_k & y_{k+1} \end{vmatrix}, \quad D_k^{(6)} = \begin{vmatrix} r_k & r_{k+1} \\ m_k & m_{k+1} \end{vmatrix}, \quad (k = 1, 2, \dots, n-1).$$

Proof. (see [5])

*Theorem 1.2.* Let  $\lambda, \mu, v, \xi$  and  $\eta$  be five distinct real scalars (with  $\lambda > \mu > v > \xi > \eta$ ) and  $x = [x_1, x_2, \dots, x_n]^T$ ,  $y = [y_1, y_2, \dots, y_n]^T$ ,  $z = [z_1, z_2, \dots, z_n]^T$ ,  $m = [m_1, m_2, \dots, m_n]^T$  and  $r = [r_1, r_2, \dots, r_n]^T$  be five real orthogonal vectors of size  $n$ , there is a Jacobi matrix  $J$  of size  $n$  such that  $(\lambda, x) \prec (\mu, y) \prec (v, z) \prec (\xi, m)$  and  $(\eta, r)$  are its five eigenpairs if the following conditions are satisfied:

$$i). (\lambda - \mu)d_k^{(1)}/D_k^{(1)} = (\lambda - v)d_k^{(2)}/D_k^{(2)} = (\lambda - \xi)d_k^{(3)}/D_k^{(3)} = (\lambda - \eta)d_k^{(4)}/D_k^{(4)} > 0$$

ii). If  $x_k = 0$  then

$$(\lambda - \mu)d_j^{(1)}/D_j^{(1)} = (\mu - v)d_j^{(5)}/D_j^{(5)} = (\mu - \xi)d_j^{(6)}/D_j^{(6)} = (\mu - \eta)d_j^{(7)}/D_j^{(7)}, j = k, k-1$$

where

$$\begin{aligned} d_k^{(1)} &= \sum_{i=1}^k x_i y_i, & d_k^{(2)} &= \sum_{i=1}^k x_i z_i, & d_k^{(3)} &= \sum_{i=1}^k x_i m_i, \\ d_k^{(4)} &= \sum_{i=1}^k x_i r_i, & d_k^{(5)} &= \sum_{i=1}^k y_i z_i, & d_k^{(6)} &= \sum_{i=1}^k y_i m_i, \\ d_k^{(7)} &= \sum_{i=1}^k y_i r_i, & d_k^{(8)} &= \sum_{i=1}^k z_i m_i, & d_k^{(9)} &= \sum_{i=1}^k z_i r_i, \end{aligned}$$

$$d_k^{(10)} = \sum_{i=1}^k m_i r_i, \quad (k = 1, 2, \dots, n)$$

$$\begin{aligned} D_k^{(1)} &= \begin{vmatrix} y_k & y_{k+1} \\ x_k & x_{k+1} \end{vmatrix}, & D_k^{(2)} &= \begin{vmatrix} z_k & z_{k+1} \\ x_k & x_{k+1} \end{vmatrix}, & D_k^{(3)} &= \begin{vmatrix} m_k & m_{k+1} \\ x_k & x_{k+1} \end{vmatrix}, \\ D_k^{(4)} &= \begin{vmatrix} r_k & r_{k+1} \\ y_k & y_{k+1} \end{vmatrix}, & D_k^{(5)} &= \begin{vmatrix} z_k & z_{k+1} \\ y_k & y_{k+1} \end{vmatrix}, & D_k^{(6)} &= \begin{vmatrix} m_k & m_{k+1} \\ y_k & y_{k+1} \end{vmatrix}, \\ D_k^{(7)} &= \begin{vmatrix} r_k & r_{k+1} \\ y_k & y_{k+1} \end{vmatrix}, & D_k^{(8)} &= \begin{vmatrix} m_k & m_{k+1} \\ z_k & z_{k+1} \end{vmatrix}, & D_k^{(9)} &= \begin{vmatrix} r_k & r_{k+1} \\ z_k & z_{k+1} \end{vmatrix}, \\ D_k^{(10)} &= \begin{vmatrix} r_k & r_{k+1} \\ m_k & m_{k+1} \end{vmatrix}, \quad (k = 1, 2, \dots, n-1). \end{aligned}$$

Proof. see[5].

## II. THE SOLVABILITY CONDITIONS OF CONSTRUCTION OF A JACOBI MATRIX BY GIVEN 2N OR 2N+1 EIGENPAIRS

In this section, we present following theorems to establish a constructive algorithm for construction of a Jacobi matrix with  $2n$  or  $2n+1$  predefined eigenpairs.

*Theorem 2.1.* Problem 1 has a unique solution if the following conditions are satisfied:

$$i) (\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = (\lambda_1 - \lambda_3)d_k^{(2)}/D_k^{(2)} = (\lambda_1 - \lambda_4)d_k^{(3)}/D_k^{(3)} \dots = (\lambda_1 - \lambda_{2n})d_k^{(2n-1)}/D_k^{(2n-1)} > 0$$

ii) If  $x_{1k} = 0$  then

$$\begin{aligned} (\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} &= (\lambda_2 - \lambda_3)d_j^{(2n)}/D_j^{(2n)} = (\lambda_2 - \lambda_4)d_j^{(2n+1)}/D_j^{(2n+1)} \\ &\dots = (\lambda_2 - \lambda_n)d_j^{(4n-3)}/D_j^{(4n-3)}, \quad j = k, k-1 \end{aligned}$$

where

$$d_k^{(1)} = \sum_{i=1}^k x_{1i}x_{2i}, d_k^{(2)} = \sum_{i=1}^k x_{1i}x_{3i}, \dots, d_k^{(n(2n-1))} = \sum_{i=1}^k x_{2n-1i}x_{2ni}$$

where ( $k = 1, 2, \dots, 2n$ ) and

$$D_k^{(1)} = \begin{vmatrix} x_{2k} & x_{2k+1} \\ x_{1k} & x_{1k+1} \end{vmatrix}, D_k^{(2)} = \begin{vmatrix} x_{3k} & x_{3k+1} \\ x_{1k} & x_{k+1} \end{vmatrix}, \dots, D_k^{(n(2n-1))} = \begin{vmatrix} x_{2n\ k} & x_{2n\ k+1} \\ x_{2n-1\ k} & x_{2n-1\ k+1} \end{vmatrix}$$

where ( $k = 1, 2, \dots, 2n - 1$ ).

Proof: By induction on the number of eigenpairs  $2n$ .

base case:  $n = 2$  and according to Lemma 1.1, under certain conditions we can determine one unique Jacobi matrix with 2 eigenpairs.

Induction hypothesis : with having  $2n-1$  eigenpairs under certain conditions we can determine one unique Jacobi matrix.

According to induction hypothesis under certain conditions we can determine one unique Jacobi matrix with  $2n-2$  eigenpairs, denoted by  $J'', J', J$ , respectively.

$$\begin{aligned} & (\lambda_1, x_1), (\lambda_2, x_2), \dots, (\lambda_{2n-3}, x_{2n-3}), (\lambda_{2n-2}, x_{2n-2}) \\ & (\lambda_1, x_1), (\lambda_2, x_2), \dots, (\lambda_{2n-3}, x_{2n-3}), (\lambda_{2n-1}, x_{2n-1}) \\ & (\lambda_1, x_1), (\lambda_2, x_2), \dots, (\lambda_{2n-3}, x_{2n-3}), (\lambda_{2n}, x_{2n}) \end{aligned}$$

.For  $J$  we have

$$i) (\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = \dots (\lambda_1 - \lambda_{2n-3})d_k^{(2n-4)}/D_k^{(2n-4)} = (\lambda_1 - \lambda_{2n-2})d_k^{(2n-3)}/D_k^{(2n-3)}$$

ii) If  $x_{1k} = 0$  then

$$(\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} = \dots (\lambda_2 - \lambda_{2n-3})d_j^{(4n-6)}/D_j^{(4n-6)} = (\lambda_2 - \lambda_{2n-2})d_j^{(4n-5)}/D_j^{(4n-5)}$$

For  $J'$  we have

$$i) (\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = \dots (\lambda_1 - \lambda_{2n-3})d_k^{(2n-4)}/D_k^{(2n-4)} = (\lambda_1 - \lambda_{2n-1})d_k^{(2n-2)}/D_k^{(2n-2)}$$

ii) If  $x_{1k} = 0$  then

$$(\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} = \dots (\lambda_2 - \lambda_{2n-3})d_j^{(4n-6)}/D_j^{(4n-6)} = (\lambda_2 - \lambda_{2n-1})d_j^{(4n-4)}/D_j^{(4n-4)}$$

For  $J''$  we have

$$i) (\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = \dots (\lambda_1 - \lambda_{2n-3})d_k^{(2n-4)}/D_k^{(2n-4)} = (\lambda_1 - \lambda_{2n})d_k^{(2n-1)}/D_k^{(2n-1)}$$

ii) If  $x_{1k} = 0$  then

$$(\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} = \dots (\lambda_2 - \lambda_{2n-3})d_j^{(4n-6)}/D_j^{(4n-6)} = (\lambda_2 - \lambda_{2n})d_j^{(4n-3)}/D_j^{(4n-3)}$$

Consequently, we have the conditions of theorem and  $J = J' = J''$  such that their elements are

$$\begin{aligned} b_k &= \frac{(\lambda_1 - \lambda_2)d_k^{(1)}}{D_k^{(1)}}, \quad (k = 1, 2, \dots, 2n - 1), \\ a_1 &= \lambda_1 - \frac{b_1 x_{12}}{x_{11}}, \\ a_{2n} &= \lambda_1 - \frac{b_{2n-1} x_{12n-1}}{x_{12n}} \end{aligned}$$

$$a_k = \begin{cases} \lambda_1 - \frac{(b_{k-1}x_{1 k-1} + b_k x_{1 k+1})}{x_{1 k}}, & x_{1 k} \neq 0 \\ \lambda_2 - \frac{(b_{k-1}x_{2 k-1} + b_k x_{2 k+1})}{x_{2 k}}, & x_{1 k} = 0 \end{cases}, \quad k = (2, 3, \dots, 2n-1).$$

*Theorem 2.2.* Problem 2 has a unique solution if the following conditions are satisfied:

- i)  $(\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = (\lambda_1 - \lambda_3)d_k^{(2)}/D_k^{(2)} = (\lambda_1 - \lambda_4)d_k^{(3)}/D_k^{(3)} \dots = (\lambda_1 - \lambda_{2n+1})d_k^{(2n)}/D_k^{(2n)} > 0$
- ii) If  $x_{1 k} = 0$  then

$$\begin{aligned} (\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} &= (\lambda_2 - \lambda_3)d_j^{(2n+1)}/D_j^{(2n+1)} = (\lambda_2 - \lambda_4)d_j^{(2n+2)}/D_j^{(2n+2)} \\ &\dots = (\lambda_2 - \lambda_{2n+1})d_j^{(4n-1)}/D_j^{(4n-1)}, \quad j = k, k-1 \end{aligned}$$

where

$$d_k^{(1)} = \sum_{i=1}^k x_{1 i} x_{2 i}, \quad d_k^{(2)} = \sum_{i=1}^k x_{1 i} x_{3 i}, \dots, \quad d_k^{(n(2n+1))} = \sum_{i=1}^k x_{2n i} x_{2n+1 i}$$

where ( $k = 1, 2, \dots, 2n+1$ ) and

$$D_k^{(1)} = \begin{vmatrix} x_{2 k} & x_{2 k+1} \\ x_{1 k} & x_{1 k+1} \end{vmatrix}, \quad D_k^{(2)} = \begin{vmatrix} x_{3 k} & x_{3 k+1} \\ x_{1 k} & x_{1 k+1} \end{vmatrix}, \dots, \quad D_k^{(n(2n+1))} = \begin{vmatrix} x_{2n+1 k} & x_{2n+1 k+1} \\ x_{2n k} & x_{2n k+1} \end{vmatrix}$$

where ( $k = 1, 2, \dots, 2n$ ).

Proof: By induction on the number of eigenpairs  $n$ .

base case:  $n = 3$  and according to Lemma 1.2, under certain conditions we can determine one unique Jacobi matrix with 2 eigenpairs.

Induction hypothesis : with having  $2n$  eigenpairs under certain conditions we can determine one unique Jacobi matrix.

According to induction hypothesis under certain conditions we can determine one unique Jacobi matrix with  $2n-1$  eigenpairs, denoted by  $J'', J', J$ , respectively.

$$(\lambda_1, x_1), (\lambda_2, x_2), \dots, (\lambda_{2n-2}, x_{2n-2}), (\lambda_{2n-1}, x_{2n-1})$$

$$(\lambda_1, x_1), (\lambda_2, x_2), \dots, (\lambda_{2n-2}, x_{2n-2}), (\lambda_{2n}, x_{2n})$$

$$(\lambda_1, x_1), (\lambda_2, x_2), \dots, (\lambda_{2n-2}, x_{2n-2}), (\lambda_{2n+1}, x_{2n+1})$$

.For  $J$  we have

$$i) (\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = \dots (\lambda_1 - \lambda_{2n-2})d_k^{(2n-3)}/D_k^{(2n-3)} = (\lambda_1 - \lambda_{2n-1})d_k^{(2n-2)}/D_k^{(2n-2)}$$

- ii) If  $x_{1 k} = 0$  then

$$(\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} = \dots (\lambda_2 - \lambda_{2n-2})d_j^{(4n-4)}/D_j^{(4n-4)} = (\lambda_2 - \lambda_{2n-1})d_j^{(4n-3)}/D_j^{(4n-3)}$$

For  $J'$  we have

$$i) (\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = \dots (\lambda_1 - \lambda_{2n-2})d_k^{(2n-3)}/D_k^{(2n-3)} = (\lambda_1 - \lambda_{2n})d_k^{(2n-1)}/D_k^{(2n-1)}$$

- ii) If  $x_{1 k} = 0$  then

$$(\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} = \dots (\lambda_2 - \lambda_{2n-2})d_j^{(4n-4)}/D_j^{(4n-4)} = (\lambda_2 - \lambda_{2n})d_j^{(4n-2)}/D_j^{(4n-2)}$$

For  $J''$  we have

$$i) (\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = \dots (\lambda_1 - \lambda_{2n-2})d_k^{(2n-3)}/D_k^{(2n-3)} = (\lambda_1 - \lambda_{2n+1})d_k^{(2n)}/D_k^{(2n)}$$

ii) If  $x_{1k} = 0$  then

$$(\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} = \dots (\lambda_2 - \lambda_{2n-2})d_j^{(4n-4)}/D_j^{(4n-4)} = (\lambda_2 - \lambda_{2n+1})d_j^{(4n-1)}/D_j^{(4n-1)}$$

Consequently, we have the conditions of theorem and  $J = J' = J''$  such that their elements are

$$\begin{aligned} b_k &= \frac{(\lambda_1 - \lambda_2)d_k^{(1)}}{D_k^{(1)}}, \quad (k = 1, 2, \dots, 2n), \\ a_1 &= \lambda_1 - \frac{b_1 x_{12}}{x_{11}}, \\ a_{2n+1} &= \lambda_1 - \frac{b_{2n} x_{12n}}{x_{12n+1}}, \\ a_k &= \begin{cases} \lambda_1 - \frac{(b_{k-1} x_{1k-1} + b_k x_{1k+1})}{x_{1k}}, & x_{1k} \neq 0, \\ \lambda_2 - \frac{(b_{k-1} x_{2k-1} + b_k x_{2k+1})}{x_{2k}}, & x_{2k} = 0, \end{cases} \quad k = (2, 3, \dots, 2n). \end{aligned}$$

According to the previous sections, for determination of solvability conditions for each problem that was expressed, former theorems are considered. Lemma 1.1, asserts that under certain conditions an unique Jacobi matrix can be constructed with two eigenpairs, while Theorem 1.1 asserts that determination of solvability conditions of problem for constructing Jacobi matrix with four eigenpairs. In the same way, Lemma 1.2 was used for proof of Theorem 1.2.

By repeating the above process we can obtain solvability of problem 1 and 2. In the next section we express the algorithms for obtaining the solution of problem 1 and 2.

### III. NUMERICAL ALGORITHMS AND EXAMPLES

From Theorem 2.1, we propose a numerical algorithm for finding the unique solution of problem 1 as follows.

#### Algorithm1

Input: The real numbers  $\lambda_1 > \lambda_2 > \dots > \lambda_{2n}$  and mutually orthogonal vectors  $x_1, x_2, \dots, x_{2n}$

Output: The symmetric Jacobi matrix having the eigenpairs  $(\lambda_1, x_1)$ ,  $(\lambda_2, x_2)$ , ...,  $(\lambda_{2n}, x_{2n})$

compute  $d_k^{(i)}$  and  $D_k^{(i)}$ ,  $i = 1 \dots n(2n - 1)$

if any one of  $D_k^{(i)} = 0$ ,  $i = 1 \dots n(2n - 1)$  the Problem1 can not be solved by this method

for  $k = 1:2n-1$

when  $x_{1k} = 0$

if

$$(\lambda_1 - \lambda_2)d_j^{(1)}/D_j^{(1)} = (\lambda_2 - \lambda_3)d_j^{(2n)}/D_j^{(2n)} = (\lambda_2 - \lambda_4)d_j^{(2n+1)}/D_j^{(2n+1)}$$

$$\dots = (\lambda_2 - \lambda_n) d_j^{(4n-3)} / D_j^{(4n-3)}, \quad j = k, k-1$$

Then

$$a_k = \lambda_2 - \frac{(b_{k-1}x_{2k-1} + b_kx_{2k+1})}{x_{2k}}$$

$$b_k = \frac{(\lambda_1 - \lambda_2)d_k^{(1)}}{D_k^{(1)}}$$

Else

“problem 1 has no solution”.

when  $x_{1k} \neq 0$

$$if (\lambda_1 - \lambda_2)d_k^{(1)} / D_k^{(1)} = (\lambda_1 - \lambda_3)d_k^{(2)} / D_k^{(2)} = (\lambda_1 - \lambda_4)d_k^{(3)} / D_k^{(3)} \dots = (\lambda_1 - \lambda_{2n})d_k^{(2n-1)} / D_k^{(2n-1)} > 0$$

Then

$$a_k = \lambda_1 - \frac{(b_{k-1}x_{1k-1} + b_kx_{1k+1})}{x_{1k}}$$

$$b_k = \frac{(\lambda_1 - \lambda_2)d_k^{(1)}}{D_k^{(1)}}$$

Else

“problem 1 has no solution”.

$$a_{2n} = \lambda_1 - \frac{b_{2n-1}x_{12n-1}}{x_{12n}}$$


---

From Theorem 2.2, we propose a numerical algorithm for finding the unique solution of problem 2 as follows.

---

### Algorithm2

---

*Input:* The real numbers  $\lambda_1 > \lambda_2 > \dots > \lambda_{2n+1}$  and mutually orthogonal vectors  $x_1, x_2, \dots, x_{2n+1}$

*Output:* The symmetric Jacobi matrix having the eigenpairs  $(\lambda_1, x_1)$ ,  $(\lambda_2, x_2)$ , ...,  $(\lambda_{2n+1}, x_{2n+1})$

compute  $d_k^{(i)}$  and  $D_k^{(i)}$ ,  $i = 1 \dots n(2n+1)$

if any one of  $D_k^{(i)} = 0$ ,  $i = 1 \dots n(2n+1)$  the Problem2 can not be solved by this method

for  $k = 1:2n$

when  $x_{1k} = 0$

if

$$(\lambda_1 - \lambda_2)d_j^{(1)} / D_j^{(1)} = (\lambda_2 - \lambda_3)d_j^{(2n+1)} / D_j^{(2n+1)} = (\lambda_2 - \lambda_4)d_j^{(2n+2)} / D_j^{(2n+2)}$$

$$\dots = (\lambda_2 - \lambda_{2n+1})d_j^{(4n-1)} / D_j^{(4n-1)}, \quad j = k, k-1 \quad \text{Then}$$

$$a_k = \lambda_2 - \frac{(b_{k-1}x_{2k-1} + b_kx_{2k+1})}{x_{2k}}$$

$$b_k = \frac{(\lambda_1 - \lambda_2)d_k^{(1)}}{D_k^{(1)}}$$

Else

“problem 2 has no solution”.

when  $x_{1k} \neq 0$

$$\text{if } (\lambda_1 - \lambda_2)d_k^{(1)}/D_k^{(1)} = (\lambda_1 - \lambda_3)d_k^{(2)}/D_k^{(2)} = (\lambda_1 - \lambda_4)d_k^{(3)}/D_k^{(3)} \dots = (\lambda_1 - \lambda_{2n+1})d_k^{(2n)}/D_k^{(2n)} > 0$$

Then

$$a_k = \lambda_1 - \frac{(b_{k-1}x_{1k-1} + b_kx_{1k+1})}{x_{1k}}$$

$$b_k = \frac{(\lambda_1 - \lambda_2)d_k^{(1)}}{D_k^{(1)}}$$

Else

“problem 1 has no solution”.

$$a_{2n+1} = \lambda_1 - \frac{b_{2n}x_{12n}}{x_{12n+1}}$$


---

In what follows an example is presented to show the validity of the algorithms.

*Example1:* Let  $\lambda_1 = 1.9603, \lambda_2 = 1.7967, \lambda_3 = 1.5255, \lambda_4 = 1.1834$

$\lambda_5 = 0.8166, \lambda_6 = 0.4745, \lambda_7 = 0.2033, \lambda_8 = 0.0397$  be eight real numbers and the eight vectors are:

$$x_1 = [0.2250 \quad 0.3742 \quad 0.4443 \quad 0.4604 \quad 0.4301 \quad 0.3603 \quad 0.2588 \quad 0.1351]$$

$$x_2 = [-0.3335 \quad -0.4601 \quad -0.3370 \quad -0.0609 \quad 0.2428 \quad 0.4460 \quad 0.4650 \quad 0.2926]$$

$$x_3 = [0.3960 \quad 0.3605 \quad -0.0759 \quad -0.4458 \quad -0.3885 \quad 0.0408 \quad 0.4318 \quad 0.4119]$$

$$x_4 = [0.4258 \quad 0.1353 \quad -0.4280 \quad -0.2926 \quad 0.3242 \quad 0.4118 \quad -0.1742 \quad -0.4759]$$

$$x_5 = [0.4258 \quad -0.1353 \quad -0.4280 \quad 0.2926 \quad 0.3242 \quad -0.4118 \quad -0.1742 \quad 0.4759]$$

$$x_6 = [0.3960 \quad -0.3605 \quad -0.0759 \quad 0.4458 \quad -0.3885 \quad -0.0408 \quad 0.4318 \quad -0.4119]$$

$$x_7 = [-0.3335 \quad 0.4601 \quad -0.3370 \quad 0.0609 \quad 0.2428 \quad -0.4460 \quad 0.4650 \quad -0.2926]$$

$$x_8 = [0.2250 \quad -0.3742 \quad 0.4443 \quad -0.4604 \quad 0.4301 \quad -0.3603 \quad 0.2588 \quad -0.1351]$$

It is easy to verify that these given numbers can satisfy the conditions of the Theorem 2.2. After calculating we have a Jacobi matrix:

$$\begin{pmatrix} 1 & .5774 & & & & & & \\ .5774 & 1 & .5164 & & & & & \\ & .5164 & 1 & .5071 & & & & \\ & & .5071 & 1 & .5040 & & & \\ & & & .5040 & 1 & .5025 & & \\ & & & & .5025 & 1 & .5017 & \\ & & & & & .5017 & 1 & .5013 \\ & & & & & & .5013 & 1 \end{pmatrix}$$

*Example2:* Let  $\lambda_1 = 1.9682, \lambda_2 = 1.8360, \lambda_3 = 1.6134, \lambda_4 = 10.3243, \lambda_5 = 1,$

$\lambda_6 = 0.6757, \lambda_7 = 0.3866, \lambda_8 = 0.1640, \lambda_9 = 0.0318$  be nine real numbers and the nine vectors are:

```

x1=[ 0.2016  0.3380  0.4084  0.4355  0.4257  0.3834  0.3134  0.2216  0.1147]
x2=[-0.3005 -0.4352 -0.3686 -0.1644  0.0980  0.3280  0.4484  0.4195  0.2514]
x3=[ 0.3610  0.3835  0.0519 -0.3277 -0.4511 -0.2220  0.1804  0.4430  0.3618]
x4=[ 0.3952  0.2219 -0.3025 -0.4194  0.0345  0.4429  0.2517 -0.2805 -0.4334]
x5=[ 0.4063  0.0000 -0.4543 -0.0000  0.4571 -0.0000 -0.4578 -0.0000  0.4581]
x6=[ 0.3952 -0.2219 -0.3025  0.4194  0.0345 -0.4429  0.2517  0.2805 -0.4334]
x7=[ 0.3610 -0.3835  0.0519  0.3277 -0.4511  0.2220  0.1804 -0.4430  0.3618]
x8=[ 0.3005 -0.4352  0.3686 -0.1644 -0.0980  0.3280 -0.4484  0.4195 -0.2514]
x9=[-0.2016  0.3380 -0.4084  0.4355 -0.4257  0.3834 -0.3134  0.2216 -0.1147]

```

It is easy to verify that these given numbers can satisfy the conditions of the Theorem 2.2. After calculating we have a Jacobi matrix:

$$\begin{pmatrix} 1 & .5774 & & & & & & & \\ .5774 & 1 & .5164 & & & & & & \\ & .5164 & 1 & .5071 & & & & & \\ & & .5071 & 1 & .5040 & & & & \\ & & & .5040 & 1 & .5025 & & & \\ & & & & .5025 & 1 & .5017 & & \\ & & & & & .5017 & 1 & .5013 & \\ & & & & & & .5013 & 1 & .5010 \\ & & & & & & & .5010 & 1 \end{pmatrix}$$

#### IV. CONCLUSION

As a summary, we have presented some sufficient conditions, as well as simple methods to construct a Jacobi matrix from its  $2n$  or  $2n+1$  eigenpairs. Numerical examples have been given to illustrate the effectiveness of our results and the proposed method. The idea in this paper may provide deep insights for other banded matrix inverse eigenvalue problems.

#### REFERENCES

- [1] P. Nylen and F. Uhlig, "Inverse eigenvalue problem: existence of special spring-mass systems," *Inverse Problems*, vol. 13, no. 4, pp. 1071–1081, 1997.
- [2] N. Radwan, "An inverse eigenvalue problem for symmetric and normal matrices," *Linear Algebra and Its Applications*, vol. 248, pp. 101–109, 1996.
- [3] Z. S. Wang, "Inverse eigenvalue problem for real symmetric five-diagonal matrix," *Numerical Mathematics*, vol. 24, no. 4, pp. 366–376, 2002.
- [4] G. M. L. Gladwell, *Inverse Problems in Vibration*, vol. 119 of Solid Mechanics and Its Applications, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2nd edition, 2004.
- [5] Z. Wang and B. zhong , An Inverse Eigenvalue Problem for Jacobi Matrices , Hindawi Publishing Corporation , Mathematical Problems in Engineering 1-11 , 2011.
- [6] H. Dai, "Inverse eigenvalue problems for Jacobi matrices and symmetric tridiagonal matrices," *Numerical Mathematics*, vol. 12, no. 1, pp. 1–13, 1990.
- [7] A. P. Liao, L. Zhang, and X. Y. Hu, "Conditions for the existence of a unique solution for inverse eigenproblems of tridiagonal symmetric matrices," *Journal on Numerical Methods and Computer Applications*, vol. 21, no. 2, pp. 102–111, 2000.
- [8] X. Y. Hu and X. Z. Zhou, "Inverse eigenvalue problems for symmetric tridiagonal matrices," *Journal on Numerical Methods and Computer Applications*, vol. 17, pp. 150–156, 1996.

# Enhancing Genetic Algorithms using Multi Mutations: Experimental Results on the Travelling Salesman Problem

Ahmad B. A. Hassanat <sup>1\*</sup>, Esra'a Alkafaween <sup>2</sup>, Nedal A. Al-Nawaiseh <sup>3</sup>, Mohammad A. Abbadi <sup>4</sup>, Mouhammd Alkasassbeh <sup>5</sup>, and Mahmoud B. Alhasanat <sup>6</sup>

<sup>1,2,4,5</sup> IT Department, Mutah University, Mutah, Karak, Jordan.

<sup>3</sup> Department of Public Health and Community Medicine, Mutah University, Mutah, Karak, Jordan.

<sup>6</sup> Department of Civil Engineering, Al-Hussein Bin Talal University, Maan, Maan, Jordan.

\* Corresponding Author: Ahmad B. A. Hassanat, Mutah Street, Mutah, Karak, 61711, Jordan, Email address:

**Abstract**—Mutation is one of the most important stages of genetic algorithms because of its impact on the exploration of the search space, and in overcoming premature convergence. Since there are many types of mutations one common problem lies in selecting the appropriate type. The decision then becomes more difficult and needs more trial and error to find the best mutation to be used.

This paper investigates the use of more than one mutation operator to enhance the performance of genetic algorithms. New mutation operators are proposed, in addition to two election strategies for the mutation operators. One is based on selecting the best mutation operator and the other randomly selects any operator.

Several experiments were conducted on the Travelling Salesman Problem (TSP) to evaluate the proposed methods. These were compared to the well-known exchange mutation and rearrangement mutation. The results show the importance of some of the proposed methods, in addition to the significant enhancement of the genetic algorithms' performance, particularly when using more than one mutation operator.

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

## I. INTRODUCTION

GENETIC algorithms (GA) are adaptive heuristic random search techniques [1], and are a sub-family of evolutionary algorithms that mimic the theory of evolution and natural selection. The basic principles of genetic algorithm were presented by John Holland in the 1970s [2]. The effectiveness of genetic algorithms has been proven by solving many optimization problems [3], [4] and [5].

There are many applications of genetic algorithms in various areas, such as image processing [6], software engineering [7], computer networks [8], robotics [9], and speech recognition [10].

Genetic algorithms are concerned, in general, with how to produce new chromosomes (individuals) that possess certain features through recombination (crossover) and mutation operators. Therefore, individuals with appropriate characteristics have the strongest chance of survival, while individuals with inappropriate characteristics are less likely to survive. This simulates Darwin's theory of evolution by natural selection, colloquially described as survival of the fittest [11], [12] and [13].

Typically, GAs start with a number of random solutions (initial population). These solutions are encoded according to the current problem, forming a chromosome for each individual (solution). The quality of each individual is then evaluated using a fitness function, after which the current population changes to a new population by applying three basic operators: selection, crossover and mutation. The efficiency of a genetic algorithm is based on the appropriate choice of these operators and strategy parameters [14] associated with ratios, such as crossover ratio and mutation ratio [15]. Many researchers have shown the effect of the two operators—crossover and mutation—on the success of the GA, and where success lies in both, whether crossover is used alone or mutation alone or both, as in [16] and [17].

One of the common issues with genetic algorithms is premature convergence [18] which is directly related to the loss of diversity [19]. Achieving population diversity is a desired goal, as the search space becomes better (diverse) accordingly, and also avoids a suboptimal solution. According to Holland, mutation is considered an important mechanism to maintain diversity [20]. Researchers [21], explored new areas in the search space, thus avoiding the convergence of the local optimum [22]. The need for mutation is to prevent loss of genetic material where the crossover does not guarantee access to new parts of the search space [23]. Therefore, random changes in the gene through mutation helps provide diversity in the population [15].

Genetic algorithms have evolved from what was prevalent in the era of Holland [24]. Classical mutation (bit-flip mutation) developed by Holland with different encoding problems no longer fits for TSP because it is difficult to encode a TSP as a binary string that does not have ordering dependencies [25]. Therefore, several types of mutation of various types of encoding have been proposed, including Exchange Mutation [26], Displacement Mutation [27], Uniform Mutation and Creep Mutation [28], Inversion Mutation [29], etc. One problem lies in our selection of which type(s) to use to solve a specific problem, which increases the difficulty in our decision and requiring more trial and error to find the best mutation to be used. To overcome this problem, several researchers have developed new types of GA that use more than one mutation operator at the same time [30], [31] and [32]. This paper contributes to previous work to overcome the problem of determining which mutation to use.

The contribution of this paper is two-fold: (1) proposals of new mutation operators for TSP, and (2) investigations into the effect of using more than one of these mutations on the performance of the GA.

The rest of this paper presents some of the related previous work and the proposed methods. This paper also discusses the experimental results, which were designed to evaluate the proposed methods. Conclusions and future work are presented at the end of the paper.

## II. RELATED WORK

To increase the effectiveness of the algorithm in tackling a problem, researchers have focused on improving the performance of Genetic Algorithms to overcome premature convergence.

Soni and Kumar studied many types of mutations that provide approximate solutions to the TSP [28]. Larrañaga et al. presented a review of how to represent travelling salesman problems and the advantages and disadvantages of different crossover and mutation operators [25]. Louis and Tang proposed a new mutation called greedy-swap mutation, so that two cities are chosen randomly in the same chromosome, and switching between them if the length of the new tour obtained is shorter than the previous ones [33].

Hong et al. proposed an algorithm called the Dynamic Genetic Algorithm (DGA) to simultaneously apply more than one crossover and mutation operator. This algorithm automatically selects the appropriate crossover and appropriate mutation, and automatically adjusts the crossover and mutation ratios, based on the evaluation results of the respective offspring in the next generation. In comparing this algorithm with the simple genetic algorithm that commonly uses one crossover process and one process of mutation, the results showed the success of the proposed algorithm in performance [30].

Deep and Mebrahtu proposed an Inverted Exchange mutation and Inverted Displacement mutation, which combine inverted mutation with exchange mutation and combines inverted mutation with displacement mutation. The experiment was performed on the TSP problem and the results were compared with several existing operators [23].

Hong et al. proposed a Dynamic Mutation Genetic Algorithm (DMGA) to simultaneously apply more than one mutation to generate the next generation. The mutation ratio is also dynamically adjusted according to the progress value that depends on the fitness of the individual. This decreases the ratio of mutation if the mutation operator is inappropriate, and vice versa, increases the ratio of mutation if the operator is appropriate [34] [31]. Dynamically adjusting the mutation ratio was studied and used later by several researchers [ [35] and [36]].

Hilding and Ward proposed an Automated Operator Selection (AOS) technique which eliminated the difficulties that appear when choosing crossover or mutation operators for any problem. In this technique, they allowed the genetic algorithm to use more than one crossover and mutation operators; taking advantage of the most effective operators to solve problems. The operators were automatically chosen based on their performance, and thereby reducing the time spent choosing the most suitable operator. The experiments were performed on the 01-knapsack problem. This approach was more effective as compared to the traditional genetic algorithm [32].

Dong and Wu proposed a dynamic mutation probability, which calculates the mutation rate by the ratio between the fitness of the individual and the most fit in the population. This ratio helps the algorithm to avoid local optima and also leads to the population's diversification [37]. Patil and Bhende presented a study of the various mutation-based operators in terms of performance, improvement and quality of solution. A comparison was made between Dynamic Mutation Algorithm, Schema Mutation Genetic Algorithm, Compound Mutation Algorithm, Clustered-based Adaptive Mutation Algorithm, and Hyper Mutation-Based Dynamic Algorithm [38].

## III. METHODS

Many researchers have attempted to prevent local convergence in different ways. Since mutation is a key operation in the search process, we found several mutation methods in the literature. The question is: what is the best method to use? To answer this question, and in the hope of avoiding local optima and increasing the diversification of the population, we have proposed and implemented 10 types of mutations to be compared with two of the well-known types, namely, Exchange mutation and Rearrangement mutation [39].

In the following we describe each operator. It is important to note that mutation methods described next subsections were

designed specifically for the TSP problem. However, they can be customized to fit other problems, such as the knapsack problem with special treatment that goes with the definition of the problem and the designed chromosome.

#### A. Worst gene with random gene mutation (WGWRGM)

To perform this mutation, we need to search for the "worst" gene in the chromosome from index 0 to L-1, where L is the length of the chromosome. The worst gene varies depending on the definition of the worst for each problem and each method. Basically, the worst gene is the point in a specific chromosome that contributes the maximum to increase the cost of that chromosome (solution).

In this method, the worst gene in the TSP's chromosome is the city with the maximum distance from its left neighbor, while the worst gene in the knapsack problem is the point with the lowest value-to-weight ratio, and so on. The worst gene is defined based on the definition of the problem.

After identifying the worst gene for a TSP chromosome, another gene is randomly selected, and then both genes are swapped, as in the Exchange mutation. In the knapsack problem, however, the worst gene is not swapped with a random gene but removed from the solution (converted to zero in the binary string), and another random (zero) gene is converted to one, to hopefully create a better offspring. Figure 1 shows an example of WGWRGM.

The worst gene (WG) can be calculated for a minimization problem such as TSP using:

$$WG = \underset{1 \leq i < n}{\operatorname{argmax}} Distance(C[i], C[i + 1]) \quad (1)$$

and for the maximization problem, such as the knapsack problem using:

$$WG = \underset{0 \leq i < n}{\operatorname{argmin}} \left( \frac{Value(C[i])}{weight(C[i])} \right) \quad (2)$$

where C represents the chromosome,  $i$  is the index of a gene within a chromosome, and the distance function for the TSP can be calculated using either Euclidian distance or the distances table between cities. In the case of TSP, searching for the WG starts at index 1, assuming that the route-starting city is located at index 0, while this is not the case for other problems such as the knapsack problem (Equation 2).

The previous equations are used for the chromosome, and the worst gene of this chromosome that exhibits the maximum distance is used for the mutation operation.

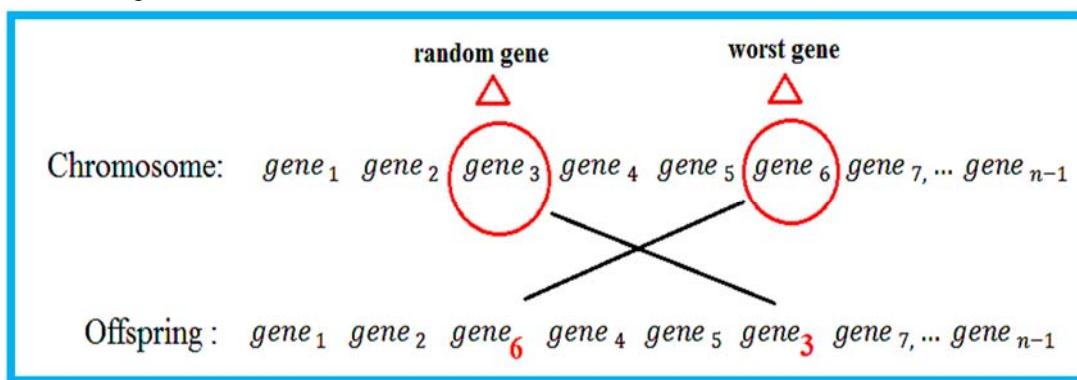


Fig. 1. Example of WGWRGM

**Example 1.** Example of applying WGWRGM to a specific chromosome of a particular TSP, the measurements are in centimeters (cm) to demonstrate the example on a printed paper. Suppose that the chromosome chosen for mutation is:

CHR1: A→B→E→D→C→A, as depicted in Figure 2(a).

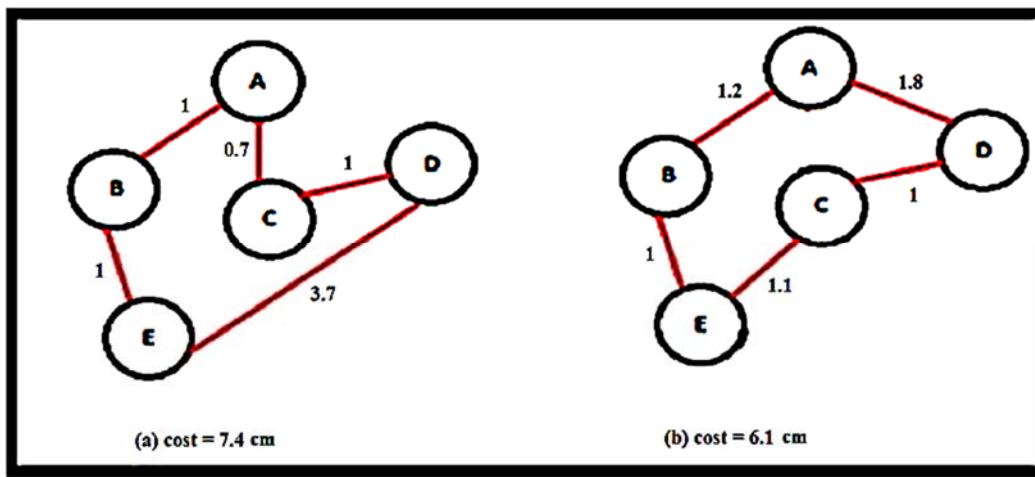


Fig. 1. Example of applying WGWRGM to a specific chromosome of a particular TSP

To apply WGWRGM:

Step 1: Find the worst gene in the parent. According to Figure 2, the worst gene is (D).

Step 2: Suppose that the city which has been selected at random is (C).

Step 3: Apply the Exchange mutation in this chromosome by swapping the positions of the two cities (see Figure 2(b)).

The output offspring becomes: A→B→E→C→D→A.

#### B. Worst gene with worst gene mutation (WGWWGM)

Although this type is similar to the WGWRGM, the difference is searching for the two worst genes then exchange positions of both the selected genes with each other. Finding both worst genes is similar to finding the two maximum values in a single array, if the problem being dealt with is a minimization problem. For the maximization problem, the algorithm of finding the two minimum values is used. The definition of the worst gene concept is different from one problem to another. For example, the two worst genes in the knapsack problem can be found by applying Equation (2) twice. Instead of swapping, both become zeros and two random (zeros) genes become ones. Figure (3) shows a TSP example of the WGWWGM.

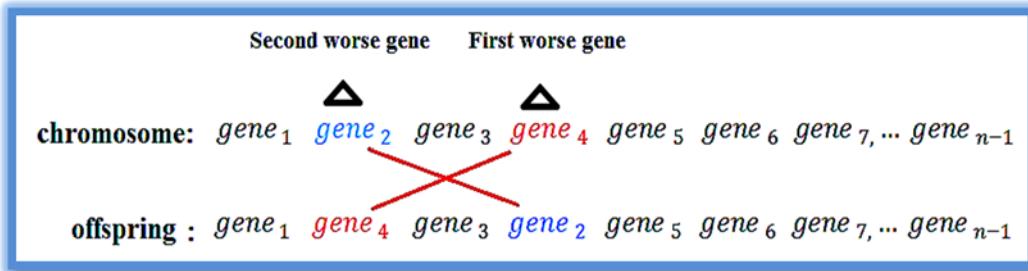


Fig.2. Example of WGWWGM

#### C. Worst left and right gene with random gene mutation (WLRGWRGM)

This method is also similar to the WGWRGM but the difference is that the worst gene is the one with the maximum total distance between that gene and both of its neighbors—the left and the right neighbors. Considering both distances (left and right) might be more informative than considering only one distance from left or right.

The worst gene (WLRgene) can be calculated for the TSP using:

$$W_{LRgene} = \underset{1 \leq i < n-2}{\operatorname{argmax}}(Distance(C[i], C[i - 1]) + Distance(C[i], C[i + 1])) \quad (3)$$

and if it is a maximization problem using:

$$W_{LRgene} = \underset{1 \leq i < n-2}{\operatorname{argmin}}(Distance(C[i], C[i - 1]) + Distance(C[i], C[i + 1])) \quad (4)$$

Equation (3) can be used for minimization problems, and Equation (4) for maximization problems, e.g. finding the maximum route in TSP. The extreme genes, the first and last ones in a chromosome, can be handled in a circular way, i.e. the left of the first

gene is the last gene.

The worst gene for minimization problems is the one that the sum of the distances with its left and right neighbors is the maximum among all genes within a chromosome; and vice versa for Maximization problems. In this mutation, the position of the worst gene is altered with the position of another gene chosen randomly.

This mutation is not defined for the knapsack problem, as the distance is not defined for such a problem.

**Example 2.** Example of applying WLRGWRGM to a specific chromosome of a particular TSP, Figure 4(a) represents the chromosome chosen for mutation, which is: A→B→E→H→F→D→C→A.

According to Figure 4 (a), the WLRgene is city D because the total distance from city D to city F and from city D to city C is the maximum distance (4.5 cm). If randomly choosing city H to swap with the WLRgene, the output offspring after applying WLRGWRGM mutation is A→B→E→D→F→H→C→A (see Figure 4(b)).

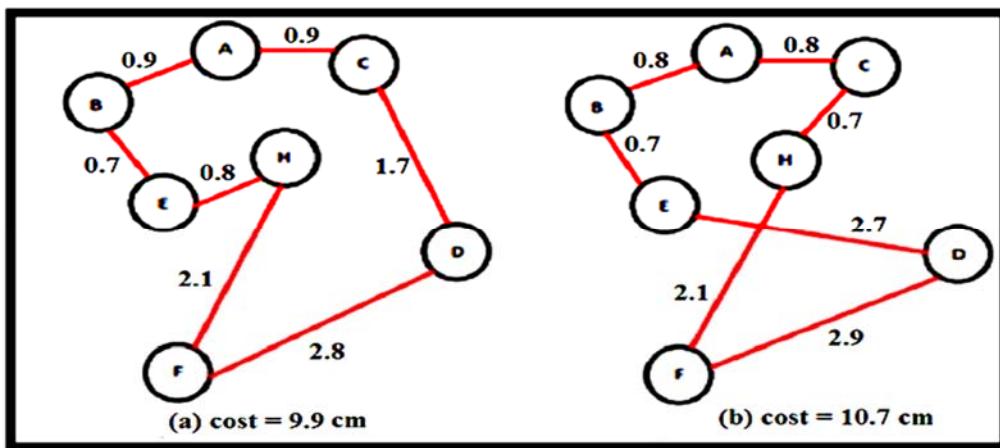


Fig.3. Example of applying WLRGWRGM on a specific chromosome of particular TSP

As can be seen from Figure 4, the new offspring does not provide a better solution which is true for many mutations. Due to randomness, there is no guarantee for better offspring all the time.

#### D. Worst gene with nearest neighbor mutation (WGNNM)

This method uses the idea of the nearest neighbor city, which provides an heuristic search process for mutation. Basically, the worst gene is swapped with one of the neighbors of its nearest city. The WGNNM is performed as follows:

*Step 1:* Search for the gene (city) in a tour characterized by the worst with its left and right neighbors (WLRgene) as in WLRGWRGM mutation. This city is called the worst city.

*Step 2:* Find the nearest city to the worst city (from the graph) and call it Ncity. Then search for the index of that city in the chromosome and call it Ni.

We need to replace the worst city with another one around the Ncity other than the Ncity itself. The term around is defined by a predefined range, centered at the Ncity. To give the algorithm some degree of randomness, the algorithm arbitrarily used ( $Ni \pm 5$ ) as a range around the index of the Ncity. The out-of-range problem with the extreme points is solved by dealing with the chromosome as a circular structure.

*Step 3:* Select a random index within the range. The city at that index is called random city.

*Step 4:* Swap between the worst city and the random city.

Example 3. Example of applying WGNNM to a specific chromosome of a particular TSP, suppose that the chromosome chosen for mutation is: A→B→F→D→E→C→H→A, as depicted in Figure 5(a).

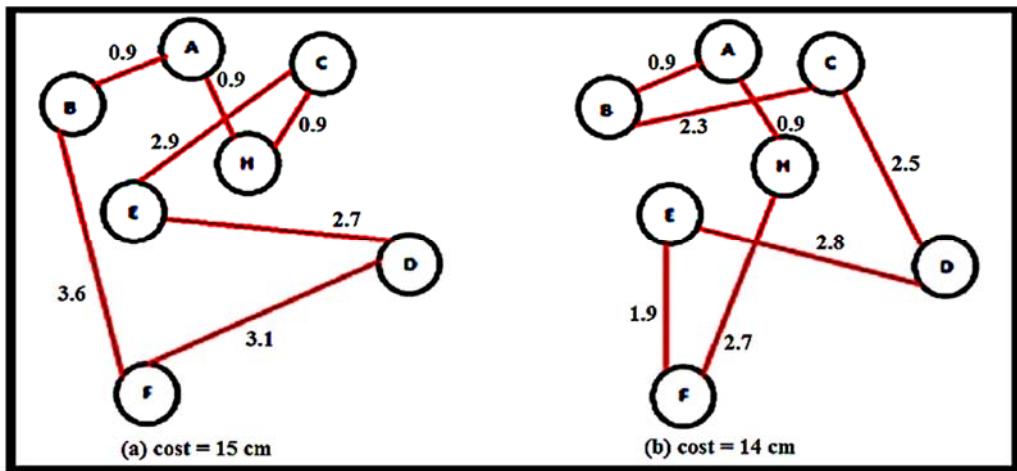


Fig. 4. Example of applying WGWNNM to a specific chromosome of particular TSP

By applying WGWNNM:

*Step 1:* Find the WLRgene in the chromosome. According to the graph, the worst city is F (6.7 cm).

*Step 2:* Find the nearest city to the worst city, which is E according to the distance table. This city is called Ncity.

*Step 3:* Search for a city around Ncity at random in the range  $\pm 5$ . Suppose we choose city C.

*Step 4:* Apply the Exchange mutation in this chromosome by swapping the position of the two cities F and C (see Figure 5(b)). The output offspring is A→B→C→D→E→F→H→A.

This mutation cannot be defined for the knapsack problem, as the nearest neighbor approach is not defined for such a problem.

#### E. Worst gene with the worst around the nearest neighbor mutation (WGWWNNM)

This mutation is similar to the WGWNNM but the only difference is in the selection of the swapped city. The swapped city is not randomly selected around the nearest city as in WGWNNM, but rather is chosen based on its distance from the nearest city. By considering the furthest city from the nearest city to be swapped with the worst city, this brings nearest cities together, and sends furthest cities far away.

This mutation will hopefully provide better offspring. However, there is no guarantee, as the swapped furthest city might be allocated in a place neighboring very far away cities, which creates a new offspring with longer TSP route.

The WGWWNNM is also cannot be defined for the knapsack problem, as the distance is not defined for such a problem neither the nearest neighbor approach.

#### F. Worst gene inserted beside nearest neighbor mutation (WGIBNNM)

This type of mutation is similar to the WGWNNM, after finding the indices of the worst city and its nearest city. The worst city is moved to be a neighbor to its nearest city, and the rest of the cities are then shifted either left or right depending on the locations of the worst city and its nearest city.

In other words, if the worst city was found to the right of its nearest city, the worst city is moved to the left of its nearest city, and the other cities are shifted to the right of the location of the worst city. If the worst city was found to the left of its nearest neighbor, the worst city is moved to the location prior to the location of its nearest city, and the rest of the cities between this location and the previous location of the worst city are shifted to the right of that location, and vice versa.

**Example 4.** Example of applying WGIBNNM to a specific chromosome of a particular TSP, suppose that the chromosome chosen for mutation is: A→B→F→D→E→C→H→A, as depicted in Figure 5(a). By applying WGIBNNM:

*Step 1:* Find the WLRgene in the chromosome. According to the graph, the worst city is F (6.7 cm).

*Step 2:* Find the nearest city to the worst city, which is E according to the distance table. This city is called Ncity.

*Step 3:* Now F is moved prior to E, and (A and B) are shifted right to get a new chromosome A→B→D→F→E→C→H→A.

#### G. Random gene inserted beside nearest neighbor mutation (RGIBNNM)

This mutation is almost the same as the WGIBNNM, except that the worst city is selected randomly, i.e. the worst city concept here is not defined, it is just a random city, and is not based on its negative contribution to the fitness of the chromosome. We propose the RGIBNNM to enhance the performance of the WGIBNNM by enforcing some randomness to increase diversity in the search space.

The RGIBNNM is also cannot be defined for the knapsack problem, as the distance is not defined for such a problem neither

the nearest neighbor approach.

#### H. Swap worst gene locally mutation (SWGLM)

This mutation is based on finding the worst gene using WLRGWRGM, then it swaps related genes locally, either the left neighbors are swapped, or the worst gene is swapped with its right neighbor. The best resulting offspring decides which genes will be swapped. This mutation is summarized as follows:

*Step 1:* Search for the worst gene, the same as for WLRGWRGM.

*Step 2:* Swap the left neighbor of the worst gene with its left neighbor, and calculate the fitness (C1) of the new offspring (F1).

*Step 3:* Swap the worst gene with its right neighbor, and calculate the fitness (C2) of the new offspring (F2).

*Step 4:* If C1 > C2, then return F2 as the legitimate offspring and delete F1, otherwise return F1 as the legitimate offspring and delete F2 (see Figure 6).

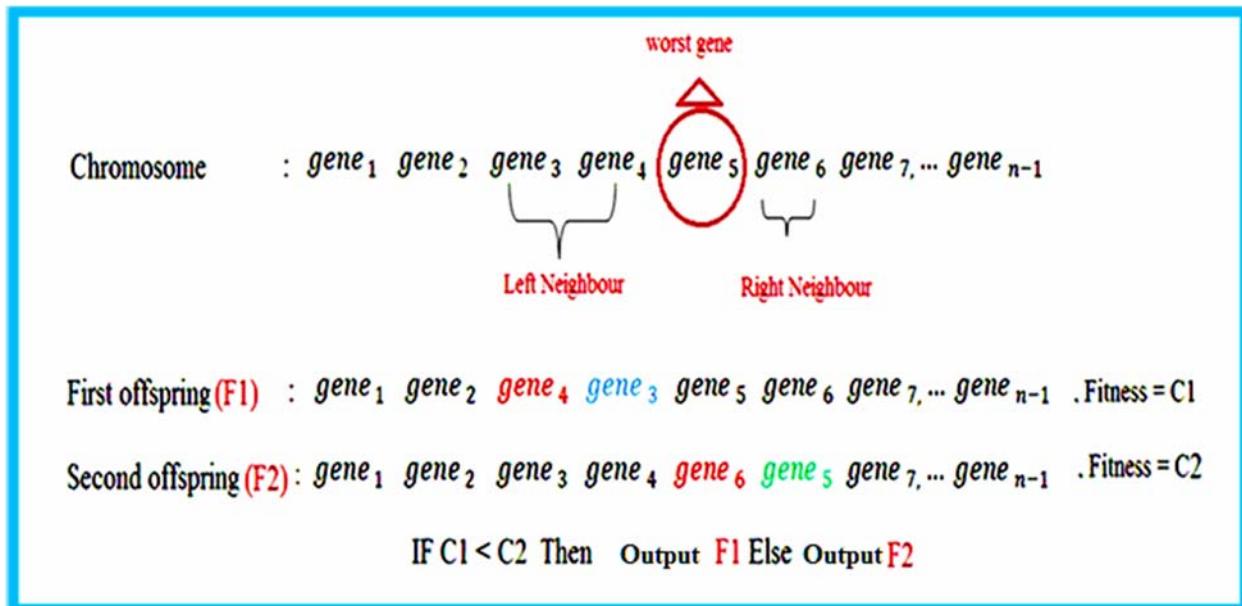


Fig. 5. Example of SWGLM

Example 5. Example of applying SWGLM to a specific chromosome of a particular TSP, suppose that the chromosome chosen for mutation is: A→B→F→E→H→D→C→A, as depicted in Figure 7(a). To apply SWGLM:

*Step 1:* Find the worst gene in the chromosome. According to the graph, the worst city is E (6.2 cm).

*Step 2:* Swap the two left neighbors of E, which are B and F. The first offspring become A→F→B→E→H→D→C→A, and the cost of this offspring is C1 (15 cm) (see Figure 7(b)).

*Step 3:* Swap between worst city E and its right neighbor H. The second offspring become A→B→F→H→E→D→C→A. The cost of this offspring is C2 (10.2 cm) (see Figure 7(c)).

*Step 4:* Compare the cost (C1, C2) and the least among them is the output offspring.

Based on the graph the output offspring is A→B→F→H→E→D→C→A (Figure 7(b)).

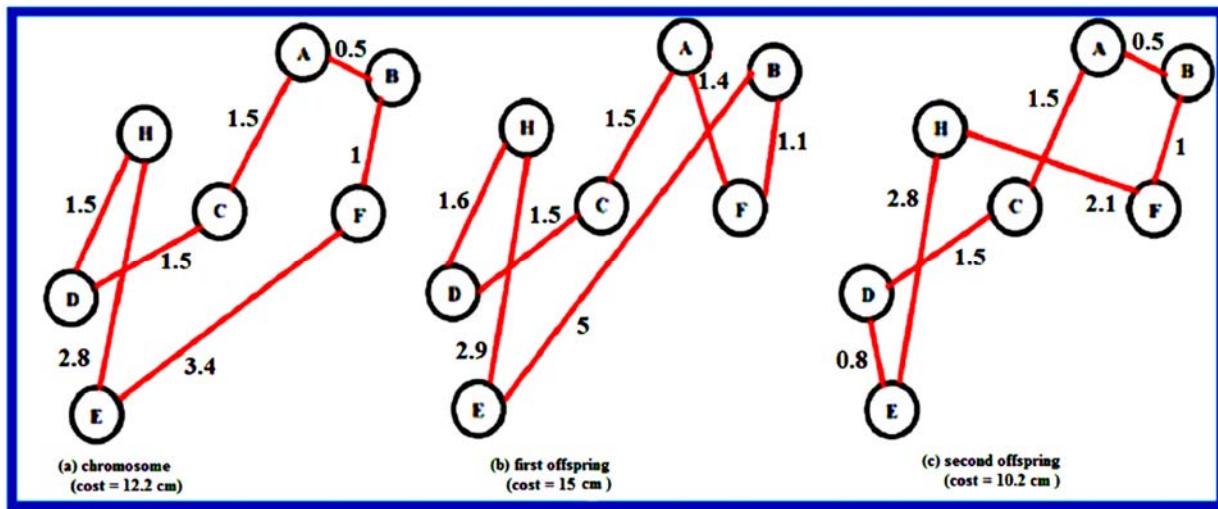


Fig. 6. Example of applying SWGLM to a specific chromosome of particular TSP

#### I. Insert best random gene before worst gene mutation (IBRGBWGM)

This method is based on finding the worst gene, as in WGWRGM, which is the city with the maximum distance from its left neighbor. Choose a random number of cities, insert the one with the minimum distance to both the worst city and its left neighbor between them. This mutation is summarized as follows:

*Step 1:* Search for the city that is characterized by the worst city as in WGWRGM and find the index of its previous city.

*Step 2:* Select a certain number of random cities. In this work we chose five random cities arbitrarily excluding the worst city and its previous neighbor (PN).

*Step 3:* For each random city calculate the distance to the worst city (D1) and the distance to PN (D2).

*Step 4:* Find the best city from the random cities, which is the one with the minimum distance (D1+D2).

*Step 5:* Move the best city and insert it between the worst city and PN.

*Step 6:* Shift cities which are located between the old and the new location of the best city to legitimize the chromosome.

**Example 6.** Example of applying IBRGBWGM to a specific chromosome of a particular TSP, Figure 8(a) represents the chromosome chosen for mutation, which is: Chromosome: A→B→E→D→C→A.

According to Figure 8(a), the worst gene is city E because the distance to its left equals four centimeters. According to the graph, the best city is C—distance (C, E) + distance (C, B) is the minimum. The output offspring after applying the IBRGBWGM mutation is A→B→C→E→D→A (see Figure 8(b)).

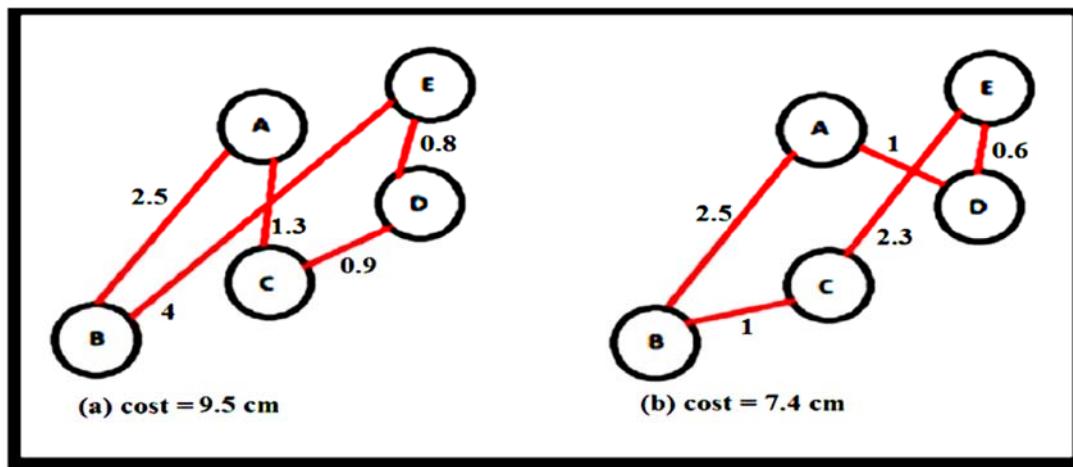


Fig. 7. Example of applying IBRGBWGM to a specific chromosome of particular TSP

#### J. Insert best random gene before random gene mutation (IBRGBRMG)

Sometimes the worst gene is located in the best possible location, thus swapping it with another gene might yield weak offspring. Therefore, it is important to have another mutation which does not depend on finding the worst gene but instead uses a random

gene. This mutation is similar to IBRGBWGM, however, the difference is that the worst city is not chosen based on any distance but is instead chosen randomly to impose some diversity among the new offspring.

Another important motivation for proposing IBRGBRGM is the computation time. As with finding the worst gene, enforce a linear computation time along the chromosome— $O(n)$  where  $n$  is the length of the chromosome. Finding the nearest neighbor approach also exhibits  $O(n)$  time complexity, while choosing a random gene takes only  $O(1)$ . Finding the nearest gene from a constant ( $k$ ) number of randomly selected genes takes  $O(k)$ , which is approximate to  $O(1)$  when  $n$  (number of the cities in a TSP instance) is very large.

#### K. Multi Mutation Operators Algorithms

A traditional genetic algorithm normally uses just one mutation operator. We propose using more than one mutation operator. Those different mutations are supposed to lead to different directions in the search space, thus increasing diversity in the population, and therefore improving the performance of the genetic algorithm. To do this we opted for two selection approaches: the best mutation, and a randomly chosen mutation.

##### 1) Select the best mutation algorithm (SBM)

This algorithm simultaneously applies multiple mutation operators to the same chromosome. To prevent duplication, it only considers the best offspring that is not found in the population to add to the population.

In this work, we defined 10 mutations to apply. The SBM implements the entire aforementioned methods—WGWRGM, WGWWGM, WLRGWRGM, WGWNM, WGWWNNM, WGIBNNM, RGIBNNM, SWGLM, IBRGBWGM and IBRGBRGM—one after the other with each mutation producing one offspring. The best offspring that does not already exist in the population is added. In TSP the best offspring is the one with the minimum TSP route.

Using such a diverse collection of mutations anticipates that such processes encourage diversity in the population, thus avoids convergence to local optima and provides better final solutions.

##### 2) Select any mutation algorithm (SAM)

This algorithm tries to apply a mutation each time, which is selected from a collection of operators. The selection strategy is random. Each operator has the same probability to be chosen. The algorithm randomly chooses one of the aforementioned mutations each time it is called by the GA. Therefore, in each generation different mutations are chosen. This means that there is a different direction of the search space which is what we are aiming for; increasing diversity and attempting to enhance the performance of the genetic algorithm.

## IV. EXPERIMENTS AND DISCUSSIONS

To evaluate the proposed methods, we conducted two sets of experiments on different TSP problems. The aim of the first set of experiments was to examine convergence to a minimum value of each method separately. The second set of experiments was designed to examine the efficiency of the SBM and SAM algorithms and compare their performance with the proposed mutation operators—WGWRGM, WGWWGM, WLRGWRGM, WGWNM, WGWWNNM, WGIBNNM, RGIBNNM, SWGLM, IBRGBWGM and IBRGBRGM—using the TSPLIB, a collection of travelling salesman problem datasets maintained by Gerhard Reinelt at <http://comopt.ifii.uni-heidelberg.de/software/TSPLIB95/>. The results of these experiments were compared with two existing mutations: Exchange mutation [26], and Rearrangement mutation [39].

In the first set of experiments, the mutation operators were tested using three test instances taken from TSPLIB [40], consisting of berlin52, ch130 and a280, each consisting of 52, 130, and 280 cities respectively.

The genetic algorithm parameters used are as follows:

- 1) Population size = 100.
- 2) The probability of crossover = 0.0.
- 3) Mutation's probability = 1.0.
- 4) The selection strategy is based on keeping the best  $k$  solutions, whether they are old parents or offspring resulted from the mutation operator(s), where  $k$  is the constant size of the population.
- 5) The termination criterion is based on a fixed number of generations reached. In our experiments the maximum number of generations = 1,600.
- 6) The chromosome used is a string of random sequence of cities' numbers, thus, the chromosome length is associated with the problem size  $n$ , which is the number of cities for each TSP problem.

The GA was applied 10 times using each of the proposed mutation, the average of the best solutions from the 10 runs, for each generation, for each method, for each TSP instance was recorded, starting from generation 1 up to generation 1,600.

Results from the first test indicate that the best performance was recorded by the SBM, followed by the SAM. This compared well with the rest of the mutations because it showed good convergence to a minimum value.

The efficiency of each of the 14 mutations (10 proposed, 2 from the literature, and 2 selection strategies) is shown in Figures 9-11. A closer look at these figures reveals that the SBM and SAM algorithms outperform all other methods in the speed of convergence.

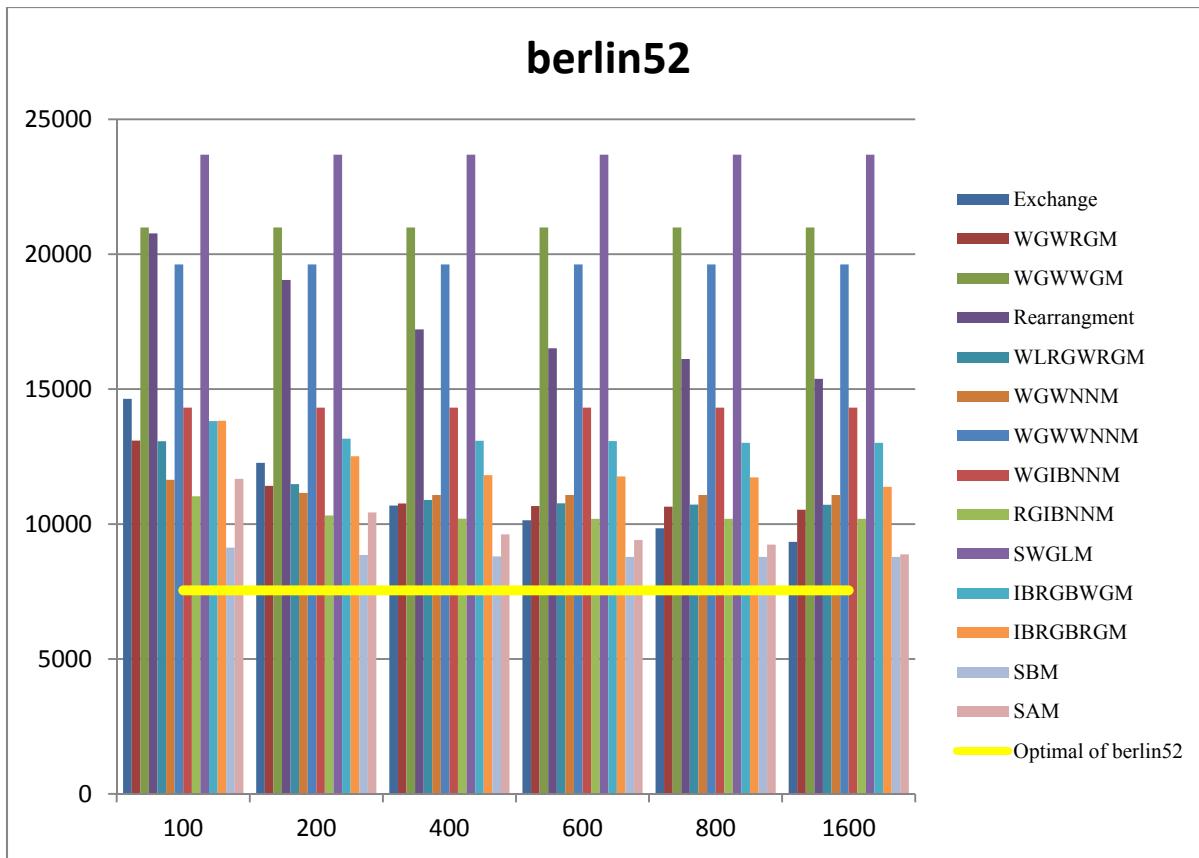


Fig. 8. Mutation's convergence to the minimum value, TSP (berlin52)

As seen in Figure 9, the results indicate the efficiency of the SBM and SAM algorithms, where the speed of convergence of a near optimal solution with the progress of the generations is faster than the use of a certain type of mutation alone. The Exchange mutation followed by RGIBNNM also showed the extent of their influence on the quality of the solution.

One result in Figure 10 indicates that the SBM algorithm showed faster convergence to the minimum value followed by SAM, and these algorithms showed better performance than the remaining mutations. At the level of mutation alone, the WLRGWRGM mutation followed by WGWRGM showed a better performance than the other mutations.

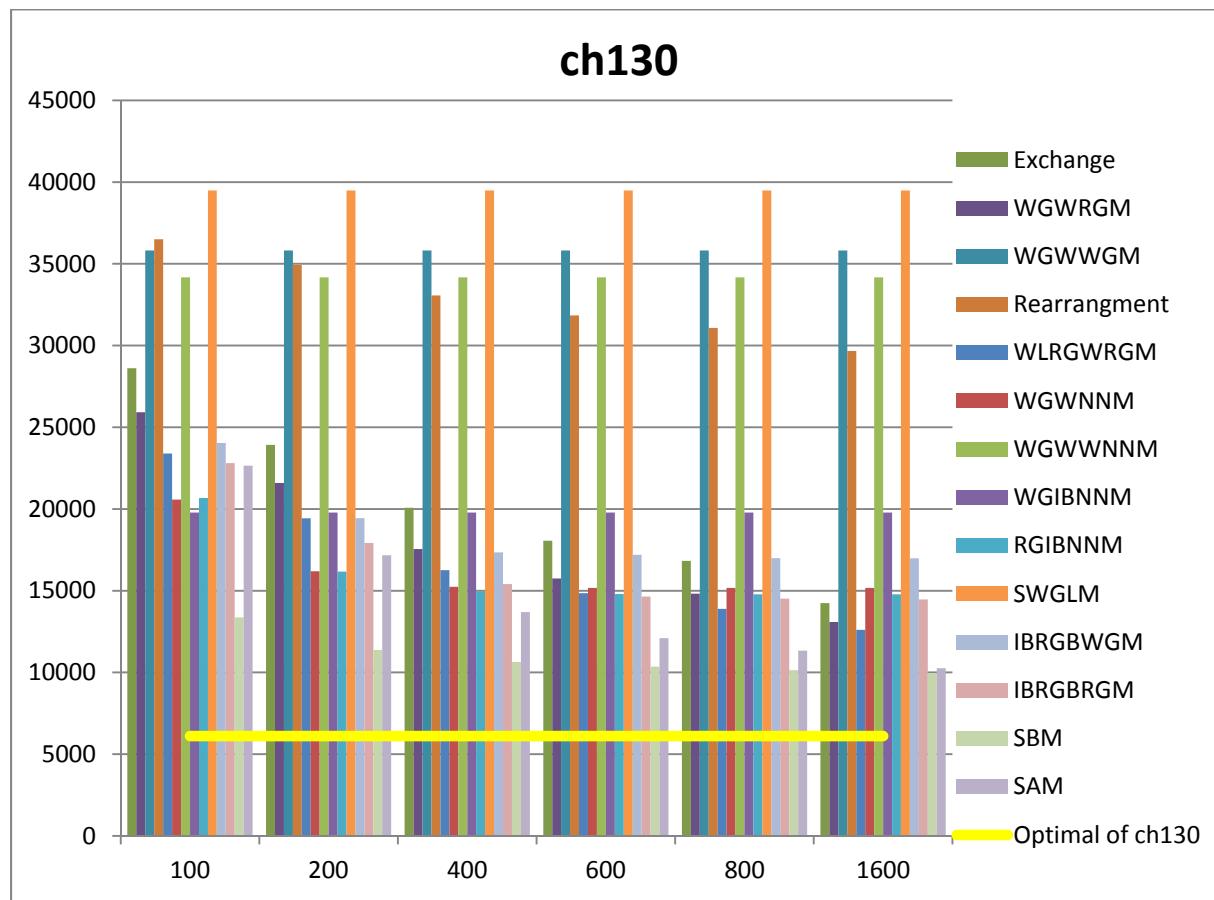


Fig. 9. Mutation's convergence to the minimum value, TSP (ch130)

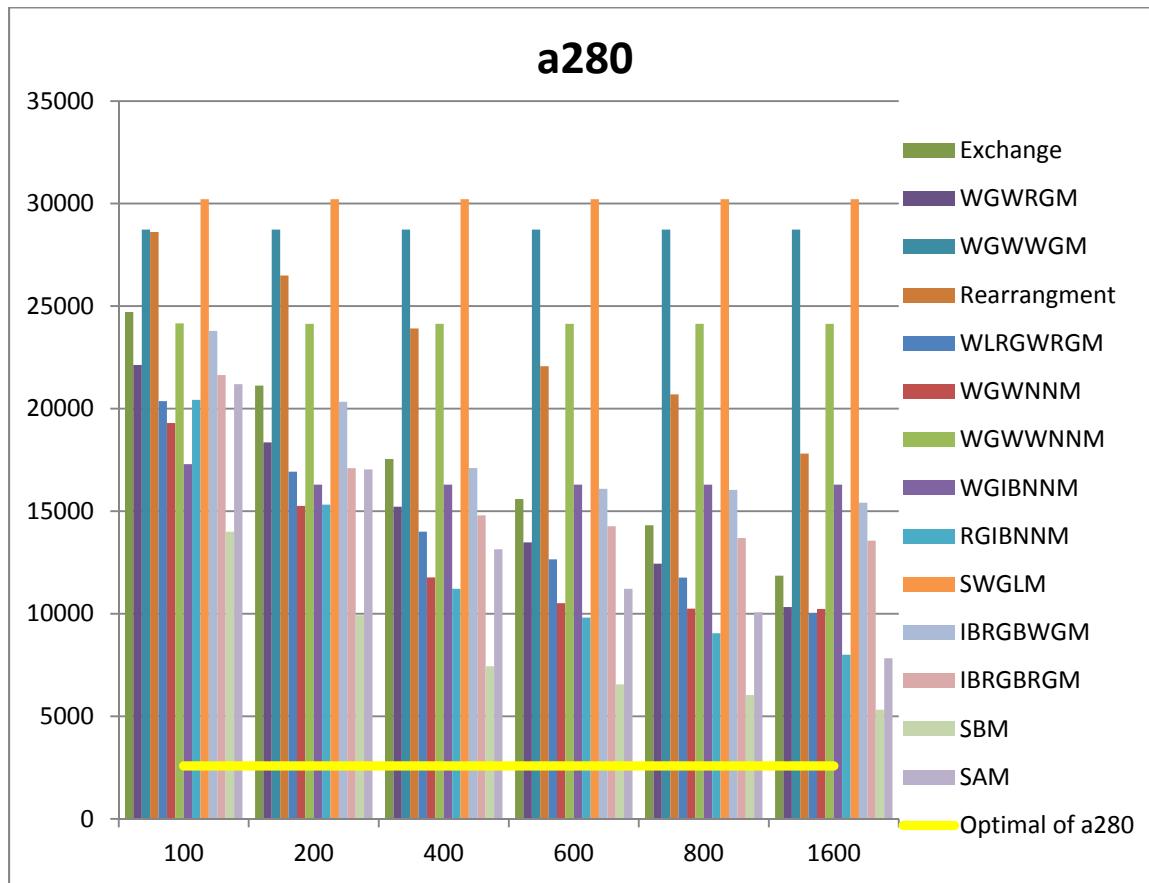


Fig. 10. Mutation's convergence to the minimum value, TSP(a280)

As can be seen from Figure 11, the best performance was recorded by the SBM algorithm. This showed faster convergence to the minimum value than any other mutation, followed by the SAM algorithm. At the level of mutations alone, RGIBNNM, followed by WLRGWRGM and WGWNNM in addition to WGWRGM mutations showed a better performance than the rest of the mutations. Because of the slow convergence of the SWGLM and WGGWGM mutations, they achieved the worst result.

The reason behind the good performance of the SBM is that it tries several mutations and chooses the best among them; however, this comes at the cost of time consumed. Although the SBM outperformed the SAM, SAM is still better than SBM in terms of time spent because SBM tries all mutations available and chooses the best, while SAM selects any one randomly. Moreover, the difference between the two results is sometimes not significant. The good performance of the SAM is due to using a different mutation each time, and this leads to an increase in the diversity of the solutions, and thus enhances the overall performance of the GA.

The second set of experiments attempted to measure the effectiveness of the SBM and SAM comparing to optimal solutions. These methods and all the proposed operators, in addition to the Exchange mutation and Rearrangement mutation, were tested using 13 TSP instances taken from the TSPLIB. They include a280, att48, berlin52, bier127, ch130, eil51, kroA100, pr76, pr144, u159, rat783, brd14051, and usa13509.

The genetic algorithm parameters that were selected were the same as in the first test; however, the recorded results were the average of the solutions at the last generation (1,600) after executing the algorithm 10 times (see Table 1).

TABLE I  
RESULTS OF 13 TSP INSTANCES OBTAINED BY 14 MUTATION OPERATORS AFTER 1,600 GENERATIONS

Mutation	a280	att48	berlin52	bier127	ch130	eil51	kroA100	pr76	pr144	u159	rat783	brd14051	usa13509
Exchange	11860	41749.4	9338.4	217739	13923	514.8	44815	169713	219250	133616	83155	36964078	1878070618
Rearrangement	17810	73119	15381	377025	29671	802.1	78546	272815	373603	208038	116095	35411256	1788855536
WGWRGM	10325	42221.8	10529	252213	13084	503.1	42259	168850	190946	122144	71748	41534181	2117784066
WGGWGM	28734	93108	20994	528898	35817	1050	119607	420047	660178	339365	165796	39752677	2035635792

<b>WLRGWRGM</b>	10043	43225.6	10714	262604	12606	524	44158	167912	200323	116924	68705	33441004	1681692076
<b>WGWNM</b>	10233	46517.3	11075	338476	15172	589.9	50393	199048	234684	129658	58338	32788677	1613016352
<b>WGWWNNM</b>	24139	89746.5	19625	543930	34178	1073	107043	408988	557415	301068	143057	39139603	2065593522
<b>WGIBNNM</b>	16300	62576	14314	446290	19781	657.7	67283	234865	310768	199013	104155	30505628	1549822430
<b>RGIBNNM</b>	8000.2	49855	10193	225990	14777	551	47938	194527	213205	116383	56263	34597287	1735470678
<b>SWGLM</b>	30212	120925	23689	559770	39487	1275	139929	467464	696683	386194	166447	41361128	2126239629
<b>IBRGBWGM</b>	15416	66912.4	13009	328296	16987	659.7	66358	228258	321485	180738	101146	36218274	1853569535
<b>IBRGRGM</b>	13562	45749.6	11378	256321	14465	583.4	48408	214855	261076	164734	68005	36058022	1822032402
<b>SBM</b>	<b>5316.1</b>	<b>37575.8</b>	<b>8782.9</b>	<b>190978</b>	<b>9958.4</b>	<b>459.1</b>	<b>35063</b>	<b>147595</b>	<b>137256</b>	<b>78225</b>	<b>34777</b>	<b>27638514</b>	<b>1377597129</b>
<b>SAM</b>	7830.7	38612.8	8875.3	201895	10262	469.9	<b>33145</b>	<b>147369</b>	142124	88452	59216	34314633	1708749204
<b>Optimal</b>	<b>2579</b>	<b>10628</b>	<b>7542</b>	<b>118282</b>	<b>6110</b>	<b>426</b>	<b>21282</b>	<b>108159</b>	<b>58537</b>	<b>42080</b>	<b>8806</b>	<b>469385</b>	<b>19982859</b>

TABLE II  
RANKS OF MUTATION OPERATORS AFTER 1,600 GENERATIONS

Mutation	a280	att48	berlin52	bier127	ch130	eil51	kroA100	pr76	pr144	u159	rat783	brd14051	usa13509	Average
<b>Exchange</b>	8	4	4	4	6	5	6	6	7	8	9	11	11	7
<b>Rearrangement</b>	12	12	12	11	12	12	12	12	12	12	12	13	12	12
<b>WGWRGM</b>	7	5	6	6	5	4	4	5	4	6	8	8	8	6
<b>WGWWGM</b>	14	14	14	13	14	13	14	14	14	14	14	15	14	14
<b>WLRGWRGM</b>	5	6	7	8	4	6	5	4	5	5	7	5	5	6
<b>WGWNM</b>	6	8	8	10	9	9	9	8	8	7	4	4	4	7
<b>WGWWNNM</b>	13	13	13	14	13	14	13	13	13	13	13	12	13	13
<b>WGIBNNM</b>	11	10	11	12	11	10	11	11	10	11	11	3	3	10
<b>RGIBNNM</b>	4	9	5	5	8	7	7	7	6	4	3	7	7	6
<b>SWGLM</b>	15	15	15	15	15	15	15	15	15	15	15	14	15	15
<b>IBRGBWGM</b>	10	11	10	9	10	11	10	10	11	10	10	10	10	10
<b>IBRGRGM</b>	9	7	9	7	7	8	8	9	9	9	6	9	9	8
<b>SBM</b>	2	2	2	2	2	2	3	3	2	2	2	2	2	2
<b>SAM</b>	3	3	3	3	3	3	2	2	3	3	5	6	6	3
<b>Optimal</b>	1	1	1	1	1	1	1	1	1	1	1	1	1	1

As can be seen in Table 1, results indicate the efficiency of the SBM algorithm in most of the problems, such as a280, rat87, berlins52, bier127, ch130, att48, pr144, u159, and eil51. It converges to the optimal faster than the exchange method, and the rest of the test data (instances) were outperformed by the SAM algorithm, such as pr76 and kroA100.

Considering methods that use one mutation only, the WGWRGM, WLRGWRGM and RGIBNNM performed better than other methods (see Table 2). The WGWRGM mutation was the best in three problems, eil51, kroA100 and pr144, and the RGIBNNM mutation was the best in three problems, a280, rat783 and u159. WLRGWRGM also showed convergence in the rest of the instances better than other methods. This method was the best in two problems. The Exchange mutation was the best in three problems, att48, berlins52 and bier127.

In these experiments, SWGLM showed weak performance, followed by WGWWGM which showed slow convergence to a minimum value. However, the importance of these operators has emerged in the diversity of the population, where both helped to achieve new areas for searching to be used by SAM and SBM.

The good performance of SBM was expected and not surprising, because SBM uses a number of mutations and chooses the best among them. Figure 12 shows the average selection probability for each mutation.

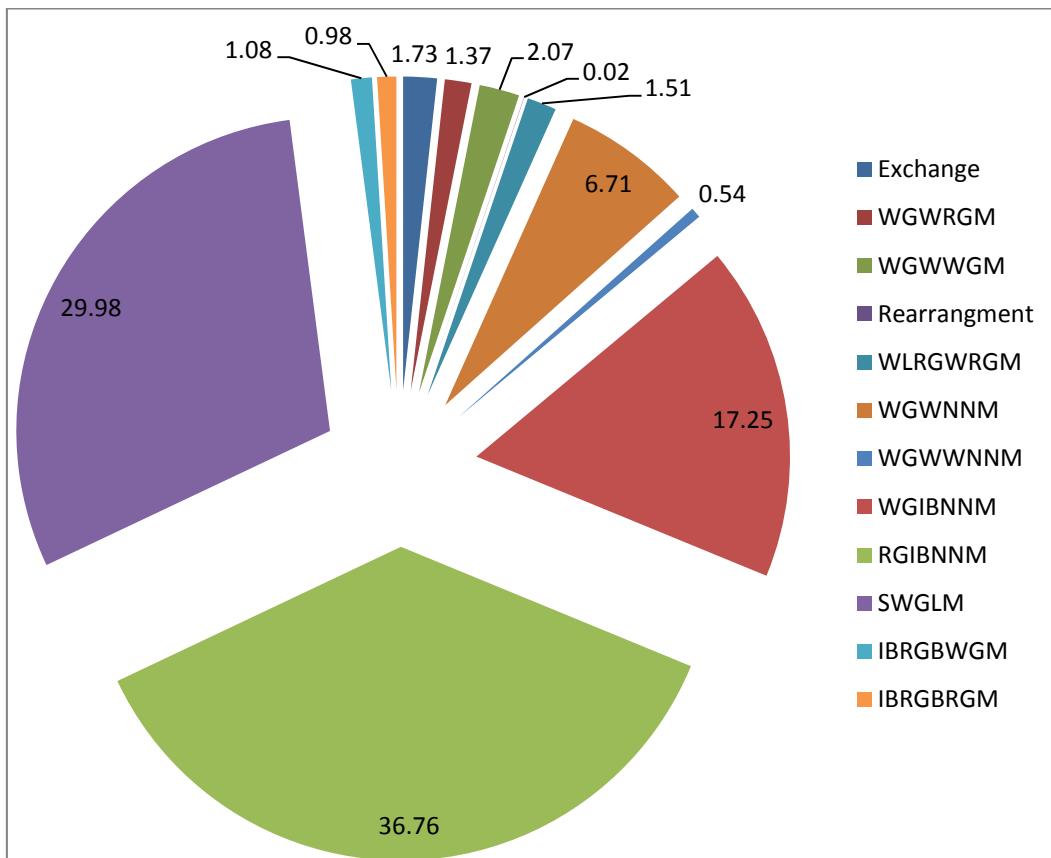


Fig. 11. The average selection probabilities for mutations used (all numbers are in percent).

As can be seen from Figure 12, the most selected mutation is the RGIBNNM, with an average probability of 36.76%. This is not surprising as this mutation performed, on average, better than most of the other methods (see Tables 1 and 2). Moreover, the least selected mutations were Rearrangement and WGWWNNM with 0.02% and 0.54% respectively. This was also not surprising as both mutations performed the weakest comparing to the other mutations. What is surprising is to have the SWGLM—the mutation of the weakest performance—selected by SBM with a probability of 29.98% ranked second. Perhaps the SWGLM contributes well to diversity, which increases the performance of the SBM.

It is interesting to note that the gap between SBM and SAM decreases as the number of generation increases (see Figures 9, 10 and 11), and sometimes the difference is not significant as in Figures 9 and 10 at generation 1,600. This shows that SAM is better than SBM in terms of time and accuracy if we used large number of generations. But if we want to use a small number of generations, SBM would be a better choice, as it converges to better solutions faster. Table 3 shows the average time consumed for each method for each TSP instance using single 3.06 GHz Pentium 4 CPU.

TABLE III  
AVERAGE TIME (IN MILLISECONDS) CONSUMED BY EACH MUTATION AFTER 1,600 GENERATIONS

Mutation	a280	att48	berlin52	bier127	ch130	eil51	kroA100	pr76	pr144	u159	rat783	brd14051	usa13509
Exchange	<b>19843</b>	<b>14172</b>	14008	15934	16562	13724	15825	<b>14476</b>	<b>16749</b>	17049	40598	<b>514352</b>	544840
WGWRGM	31481	21060	18984	21799	18424	15103	20179	20029	22077	25978	56772	666619	713398
WGWGWM	26840	18510	17732	19676	19662	14981	17647	14879	17573	21644	49813	630079	666275
Rearrangement	33122	15552	18872	20245	20756	18369	21096	17188	23967	22321	69531	1130893	1060503
WLRGWRGM	30126	17603	17969	22654	22736	17173	21103	19669	22784	22478	61324	1026570	741261
WGNNM	37008	19360	16800	17980	30933	18177	21804	23597	28001	23240	78486	845265	809090
WGWWNNM	28452	17347	14415	17544	18470	13645	16967	15142	20581	20589	53550	792461	784913
WGIBNNM	28668	16937	<b>13844</b>	17994	20854	13900	18434	19831	21182	24417	65065	1005676	897441
RGIBNNM	24139	16860	17442	21564	20274	14448	19486	16354	20062	22487	44072	642389	498415
SWGLM	28409	17343	14772	17208	21388	16581	17434	18373	22412	23081	57737	800739	774224
IBRGBWGM	27982	17350	13944	16224	18153	13586	17234	15132	18850	24867	52849	668979	601137
IBRGBRGM	27633	19907	15241	<b>15457</b>	<b>16195</b>	<b>13266</b>	<b>14912</b>	17308	18921	<b>15980</b>	<b>38180</b>	613229	<b>465612</b>
SBM	100975	30443	29452	53541	58931	30706	48506	40462	65555	66766	260287	8430944	7957288
SAM	27337	16214	15003	16746	18462	16975	19389	16155	20636	21133	58314	1048040	1445483
SBM/SAM	4	2	2	3	3	2	3	3	3	3	4	8	6

As can be seen from Table 3, the consumed time by the individual mutations, the first 12 mutations, is not significantly different. However, we find that the IBRGBRGM and the Exchange mutations consumed slightly less time than the others. This is because they do not need to do any special treatments to the mutated chromosome, such as finding the worst gene, which justifies the increase in the time consumed by the other mutations.

In addition, it is expected that SBM would consume more time than the others. Compared to the SAM, the SBM consumes at least double the time consumed by the SAM (see the last row in Table 3). This triggers the question: do the results of the SMB justify its high consumption of time? The answer depends mainly on the TSP instance itself, as can be seen from Table 3. The larger the number of instances the higher the time consumed by the SBM, and vice versa. Having the SBM converge faster than any other method for a better solution, with little time consumed when applied in small TSP instances makes the SBM a better choice. In the case of large TSP instances (greater than 200 cities), we do not recommend the use of SBM but instead recommend the use of SAM, as the results of SBM are not significantly better than the SAM, and the consumed time is much higher.

Although the aim of this paper is not to find the optimal solution for TSP instances, the solutions of the proposed algorithms were close to optimal solutions in some cases (e.g. columns 4 and 7 of Table 1), and none could achieve an optimal solution. Perhaps employing crossover operators and increasing the number of generations would enhance the solutions of the proposed methods. This shows the importance of using appropriate parameters along with mutation (such as population size, crossover ratio, number of generations, etc.), due to the effective impact of their convergence to an optimal or near optimal solution. Therefore, it is unfair to compare our findings to state-of-the-art GAs, as we just investigated the efficiency of the proposed mutations, and whether of to use a single mutation or more at each generation.

## V. CONCLUSIONS

We have proposed several mutation methods—WGRGM, WGWGM, WLRGWRGM, WGWNM, WGWWNNM, WGIBNNM, RGIBNNM, SWGLM, IBRGBWGM and IBRGBRGM—to enhance the performance of GA while searching for near optimal solutions for the TSP, in addition to proposing two selection approaches—SBM and SAM. Several experiments were conducted to evaluate those methods on several TSP problems, which showed the efficiency of some of the proposed methods over the well-known Exchange mutation and Rearrangement mutations. Some of the proposed mutations can be used for other problems with some modifications and not only oriented to the TSP problem, such as the knapsack problem. Here the concept of the worst gene is defined by its value-over-weight ratio, except for those which uses the distance and the nearest neighbor approaches.

The results of the experiments conducted for this study also suggest that using more than one mutation method in the GA is preferable, because it allows the GA to avoid local optima; the proposed SBM and SAM strategies enhance the performance of the GA. This approach, using more than one mutation for GA, is supported [30], [31] and [34].

For the use of each mutation alone, some mutations showed better performance than others, and this does not mean that the rest of the mutations had been proven to fail. Even those with the weakest performance can be effective in dealing with other problems because every problem has a different search space. In this work, we found them effective in SBM and SAM, where they encouraged diversity and hence increased the efficiency of both algorithms.

Our future work will include the development of some types of new crossovers, using the same approaches, i.e. trying more than one crossover each time to support the proposed approaches [41], and attempting to further enhance the performance of GA. Additionally we will apply the proposed methods to different problems using different benchmark data and different distance metrics such as [42].

## ACKNOWLEDGMENT

The first author made part of his contribution to the work presented in this paper while he was a Visiting Research Fellow at the Sarajevo School of Science and Technology ([www.ssst.edu.ba](http://www.ssst.edu.ba)). He thanks the University for hosting his visit and for all the support and assistance afforded to him during the visit.

## REFERENCES

- [1] A. Singh and R. Singh, "Exploring Travelling Salesman Problem using Genetic Algorithm," *International Journal of Engineering Research & Technology (IJERT)*, vol. 3, no. 2, 2014.
- [2] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, Cambridge, MA: MIT Press, 1975.
- [3] D. E. Goldberg, "Genetic algorithms in search, optimization, and machine learning," *Addison Wesley*, 1989.
- [4] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing*, vol. 4, no. 2, pp. 65-85, 1994.
- [5] P. W. Tsang and A. T. Au, "A genetic algorithm for projective invariant object recognition," in *TENCON'96. Proceedings., 1996 IEEE TENCON. Digital Signal Processing Applications*, 1996.

- [6] M. Paulinas and A. Ušinskas, "A survey of genetic algorithms applications for image enhancement and segmentation," *Information Technology and control*, vol. 36, no. 3, 2015.
- [7] P. R. Srivastava and T. H. Kim, "Application of genetic algorithm in software testing," *International Journal of software Engineering and its Applications*, vol. 3, no. 4, pp. 87-96, 2009.
- [8] A. A. Mohammed and G. Nagib, "Optimal Routing In Ad-Hoc Network Using Genetic Algorithm," *International Journal of Advanced Networking and Applications*, vol. 3, no. 05, pp. 1323-1328, 2012.
- [9] H. V. H. Ayala and L. dos Santos Coelho, "Tuning of PID controller based on a multiobjective genetic algorithm applied to a robotic manipulator," *Expert Systems with Applications*, vol. 39, no. 10, pp. 8968-8974, 2012.
- [10] H. Gupta and D. S. Wadhwa, "Speech feature extraction and recognition using genetic algorithm," *International Journal of Emerging*, vol. 4, no. 1, 2014.
- [11] J. Zhong, X. Hu, M. Gu and J. Zhang, "Comparison of performance between different selection strategies on simple genetic algorithms," in *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference*, 2005.
- [12] W. Mustafa, "Optimization of Production Systems Using Genetic Algorithms," *International Journal of Computational Intelligence and Applications*, vol. 3, no. 3, pp. 233-248, 2003.
- [13] A. E. Eiben and J. E. Smith, *Introduction to evolutionary computing*, Springer Science & Business Media, 2003.
- [14] A. E. Eiben, Z. Michalewicz, M. Schoenauer and J. E. Smith, "Parameter control in evolutionary algorithms," in *Parameter setting in evolutionary algorithms*, pp. 19-46, 2007.
- [15] S. Yang, "Adaptive non-uniform mutation based on statistics for genetic algorithms," in *Proceedings of the 2002 Genetic and Evolutionary Computation, GECCO'02, Part II*, Berlin, 2002.
- [16] W. M. Spears, "Crossover or mutation," *Foundations of genetic algorithms*, vol. 2, pp. 221-237, 1992.
- [17] K. Deb and S. Agrawal, "Understanding interactions among genetic algorithm parameters," *Foundations of Genetic Algorithms*, pp. 265-286, 1999.
- [18] E. S. Nicoară, "Mechanisms to avoid the premature convergence of genetic algorithms," *Petroleum-Gas University of Ploiești Bulletin, Math.-Info.-Phys.*, pp. 87-96, 2009.
- [19] J. Y. Suh and D. Van Gucht, "Incorporating heuristic Information into Genetic Search," in *Proceedings of the Second International Conference on Genetic Algorithms*, 1987.
- [20] K. Deb and D. Deb, "Analysing mutation schemes for real-parameter genetic algorithms," *International Journal of Artificial Intelligence and Soft Computing*, vol. 4, no. 1, pp. 1-28, 2014.
- [21] S. Wagner, M. Affenzeller, A. Beham, G. K. Kronberger and S. M. Winkler, "Mutation effects in genetic algorithms with offspring selection applied to combinatorial optimization problems," in *In Proceeding of 22nd European modeling and simulation symposium EMSS*, 2010.
- [22] I. Korejo, S. Yang, K. Brohi and Z. U. Khuhro, "Multi-Population Methods with Adaptive Mutation for Multi-Modal Optimization Problems," *International Journal on Soft Computing, Artificial Intelligence and Applications (IJSACI)*, vol. 2, no. 2, 2013.
- [23] K. Deep and H. Mebrahtu, "Combined mutation operators of genetic algorithm for the travelling salesman problem," *International Journal of Combinatorial Optimization Problems and Informatics*, vol. 2, no. 3, pp. 1-23, 2011.
- [24] T. Bäck and H. P. Schwefel, "An overview of evolutionary algorithms for parameter optimization," *Evolutionary computation*, vol. 1, no. 1, pp. 1-23, 1993.
- [25] P. Larrañaga, C. M. Kuijpers, R. H. Murga, I. Inza and S. Dizdarevic, "Genetic algorithms for the travelling salesman problem: A review of representations and operators," *Artificial Intelligence Review*, vol. 13, no. 2, pp. 129-170, 1999.
- [26] W. Banzhaf, "The ‘molecular’ traveling salesman," *Biological Cybernetics*, vol. 64, no. 1, pp. 7-14, 1990.
- [27] M. Z. T I, *Genetic Algorithms+ data Structures= Evolutionary Programs*, Berlin: Springer, 1992.
- [28] N. Soni and T. Kumar, "Study of Various Mutation Operators in Genetic Algorithms," *(IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, pp. 4519-4521, 2014.
- [29] D. A. Fogel, "A parallel processing approach to a multiple travelling salesman problem using evolutionary programming," in *Proceedings of the Fourth annual Symposium on Parallel Processing*, 1990.
- [30] T. P. Hong, H. S. Wang, W. Y. Lin and W. Y. Lee, "Evolution of appropriate crossover and mutation operators in a genetic process," *Applied Intelligence*, vol. 16, no. 1, pp. 7-17, 2002.
- [31] T. P. Hong, H. S. Wang and W. C. Chen, "Simultaneously applying multiple mutation operators in genetic algorithms," *Journal of heuristics*, vol. 6, no. 4, pp. 439-455, 2000.
- [32] F. Hilding and K. Ward, "Automated Crossover and Mutation Operator Selection on Genetic Algorithms," in *Proceedings of the 9th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, Melbourne, 2005.
- [33] S. J. Louis and R. Tang, "Interactive genetic algorithms for the traveling salesman problem," in *Genetic and Evolutionary Computation Conf.(ICGA-99)*, 1999.
- [34] T. P. Hong and H. S. Wang, "A Dynamic Mutation Genetic Algorithm," in *Systems, Man, and Cybernetics, 1996., IEEE International Conference*, 1996.
- [35] J. Clune, D. Misevic, C. Ofria, R. E. Lenski, S. F. Elena and R. Sanju, "Natural Selection Fails to Optimize Mutation Rates for Long-Term Adaptation on Rugged Fitness Landscapes," vol. 4, no. 9, 2008.
- [36] B. Wang, X. Wei, J. Dong and Q. Zhang, "Correction: Improved Lower Bounds of DNA Tags Based on a Modified Genetic Algorithm," vol. 10, no. 6, 2015.
- [37] M. Dong and Y. Wu, "Dynamic Crossover and Mutation Genetic Algorithm Based on Expansion Sampling," *In Artificial Intelligence and Computational Intelligence*, pp. 141-149, 2009.
- [38] S. Patil and M. Bhende, "Comparison and Analysis of Different Mutation Strategies to improve the Performance of Genetic Algorithm," *(IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, pp. 4669-4673, 2014.
- [39] O. M. Sallabi and Y. El-Haddad, "An improved genetic algorithm to solve the traveling salesman problem," *World Academy of Science, Engineering and Technology*, vol. 52, no. 3, pp. 471-474, 2009.

- [40] Reinelt and Gerhard, "TSPLIB. University of Heidelberg," 1996. [Online]. Available: <http://comopt.ifii.uni-heidelberg.de/software/TSPLIB95/>. [Accessed 17 9 2015].
- [41] A. B. A. Hassanat and E. Alkafaween, "On Enhancing Genetic Algorithms Using New Crossovers," *International Journal. of Computer Applications in Technology (IJCAT)*, 2016.
- [42] A. B. A. Hassanat, "Dimensionality Invariant Similarity Measure," *Journal of American Science*, vol. 10, no. 8, pp. 221-226, 2014.

# Hybrid Metaheuristic Optimization based on ACO and Standard PSO applied to Traveling Salesman Problem

Sonia Kefi <sup>#1</sup>, Nizar Rokbani <sup>\*\*2</sup>, Adel Mohamed Alimi <sup>#3</sup>

<sup>#</sup> REGIM-Lab: Research Groups in Intelligent Machines, University of Sfax

ENIS, Sfax, Tunisia

<sup>1</sup> sonia.kefi@ieee.org

<sup>3</sup> adel.alimi@ieee.org

<sup>\*</sup> High Institute of Applied Sciences and Technology of Sousse, University of Sousse

Sousse, Tunisia

<sup>2</sup> nizar.rokbani@ieee.org

## Abstract

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

## I. INTRODUCTION

Combinatorial optimization occupies a very important place in operations research, discrete mathematics and computer science. Firstly, its importance justified by the great difficulty of optimization problems and secondly by many practical applications can be formulated as a combinatorial optimization problem. Although the problems combinatorial optimization are often easy to define, they are generally difficult to solve. Indeed, most of these problems belong to the class *NP-hard* problems, but until now, it has not an effective algorithmic solution for all information.

Several problems are considered as combinatorial optimization problems, including the routing problems as one of the most popular and difficult to solve. The most known problems and the most used of routing problems are the Traveling Salesman Problem (TSP) [1] and the Vehicle Routing Problem (VRP) [2], which are the subject of several literature researches [3], [4].

Regarding to the importance of these problems, many methods of resolution were developed in operating research and intelligence artificial. These methods can be classified into two major categories: *exact methods*(complete) that ensure the completeness of the resolution and *approximate methods* (incomplete) which neglect the completeness to increase efficiency.

Therefore, several researchers often use *approximate methods* that provide close optimal solutions with a reasonable execution time. The approximation methods are called *Metaheuristics* [5]. Metaheuristic algorithms form a class of approximate problem-solving methods that try to combine various heuristic approaches to efficiently explore the solution space of a problem. On high level, they are successful optimization strategies and aim at reusability for different problems and in different scenarios.

Metaheuristics are adaptable and applicable to a wide class of problems, so, several metaheuristics methods are developed and based on solution of the problem as neighborhood search methods like Tabu Search [6], Swarm Intelligence methode like Particle Swarm Optimization [7] and Ant Colony optimization [8], etc. Fig. 1 presents a proposed taxonomy of metaheuristics methods. This taxonomy presents differents class of Metaheuristics methods existing in litterature and used to solve combinatorial optimization problem. In recent years, the interest in hybrid metaheuristic increases continuously in operations research and artificial intelligence.

For this reason, this paper presents a hybrid metaheuristic which was proposed, called Ant Supervised by particle swarm optimization (AS-PSO). This method is based on ACO and PSO algorithms to solve routing problem especially TSP. In [9], [10], the authors have developed several variants of Ant supervised by PSO (AS-PSO) like the simplified AS-PSO because they have used the simplified variant of PSO. The objective of their methods, the PSO is applied to optimize the parameters  $(\alpha, \beta)$  of ACO, an extended work of AS-PSO, presented in [11], can optimize the four ACO parameters  $(\alpha, \beta, \tau, \rho)$ . A close proposal appears in [12] and [13]. The classical AS-PSO schema to which a 2-Opt algorithm instead of the 3-Opt algorithm in [13]. The AS-PSO-2Opt [14] method can optimize the three ACO parameters  $(\alpha, \beta, \rho)$  by PSO algorithm and decrease the possibility of sinking in local minimum with the 2-Opt algorithm. The first main advantage of this hybrid AS-PSO-2Opt [14] metaheuristic method is the capacity to achieve the best global solution than others variants of ACO, AS-PSO [9], [10] and Extended AS-PSO [11].

In this paper, we present a Standard Ant Supervised by PSO for solving routing problem as TSP (SAS-PSO). The main objective of this work is to demonstrate that the SAS-PSO is a judicious hybrid metaheuristic method to solve the TSP. Therefore, we have compared the results obtained by our hybrid proposed method with the others methods from literature. This comparison is performed with several TSP benchmarks from TSPLIB [15]. The realized comparison is based on statistics such as results average, standard deviation, pourcentage error, etc.

This paper is structured as follows. In Section 2, a brief literature related to the metaheuristics methods is

presented. In Section 3, an overview of AS-PSO and its different variants are detailed. Then, our proposed hybrid metaheuristic method is described in Section 4. Additionally, the experimentation carried out and a brief description of the TSP are made in Section 5. Finally, conclusions and future work are explained in Section 6.

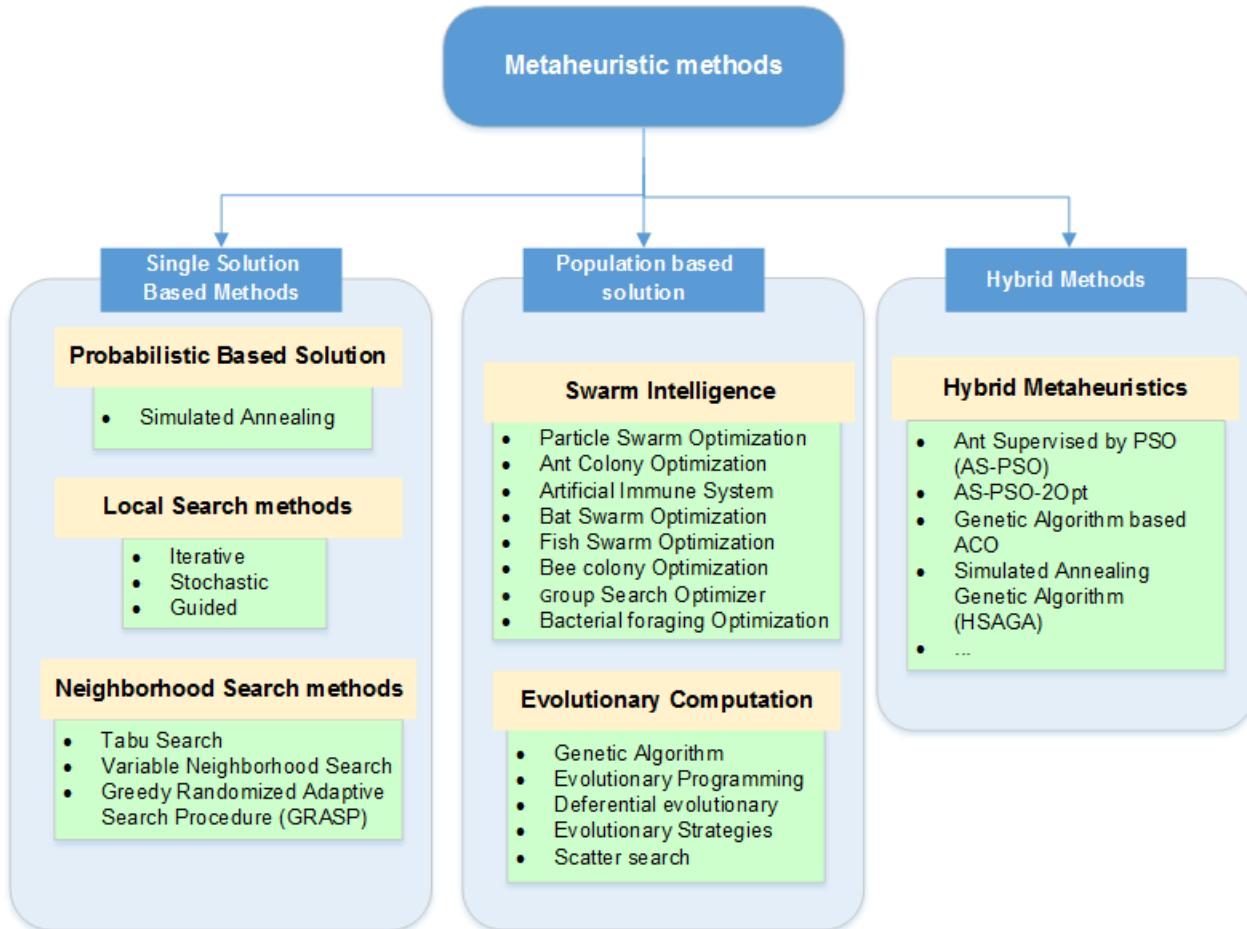


Fig. 1. Proposed taxonomy of metaheuristics methods

## II. RELATED WORKS

The Traveling Salesman Problem (TSP) [16] is an interesting optimization problem in the literature and it has attracted several researchers. The TSP has also been applied as a benchmark problem for several metaheuristics methods proposed in the last decade. There are several reasons for the choice of the TSP as the problem to explain our hybrid metaheuristic, that it is an important *NP-hard* optimization problem and the availability of several benchmarks that we allow to perform experiments and tests and compare our method with other methods. This section focus especially on different metaheuristics methods applied to TSP which based on single solution, population solution and hybrid methods.

#### A. Single Solution Based Metaheuristics Methods

The Single Solution based metaheuristics methods, also called *trajectory methods*. They are based on the notion of neighborhood to find solution in the search space of the problem. The most popular examples of single solution based metaheuristics methods are Tabu Search (TS) [6], Simulated Annealing (SA) [17], Iterated Local Search (ILS) [18], Guided Local Search (GLS) [19], Variable Neighborhood Search (VNS) [20], [21], Greedy Randomized Adaptive Search Procedure (GRASP) [22]. Table I presents a global overview of several method of single solution based metaheuristics like Tabu Search, Simulated Annealing and Local Search methods as 2-Opt and 3-Opt algorithms and their principles.

#### B. Population Solution Based Metaheuristics Methods

Population Solution based metaheuristics methods begin from an initial population of solutions, this is the main difference between them and the single solution based metaheuristics methods which start from a single solution. In opposition to local research methods, the population solution methods improve, progressively, a population of solutions. The advantage of these methods is to use the population as a diversity factor. Most of the population based metaheuristics methods are nature-inspired methods. It has two major classes of population solution methods based on optimization algorithms are evolutionary algorithms [23] and swarm intelligence based algorithms [24]. The most popular of evolutionary algorithms are Genetic Algorithm (GA)[25], Genetic Programming (GP) [26], evolutionary programming (EP) [27], Differential Evolutionary (DE) [28]. Although GA is the most widely used one in the literature. In recent years, Swarm Intelligence has been used by many research scientists which is inspired by the collective behaviour of social insect colonies and other animal societies. Although Ant Colony Optimization (ACO) [8], particle Swarm Optimization (PSO) [7], Bat Algorithm (BA) [29] and Artificial Bee Colony (ABC) [30] are popular swarm intelligence algorithms optimization. Table II describes some population based metaheuristics.

#### C. Hybrid Metaheuristics methods

Hybrid metaheuristics have emerged along the paradigm itself, but most researchers accorded little interest [5]. Now, These methods are very used by the researchers in their study because the best results for several combinatorial optimization problems have been obtained with hybrid algorithms. Hybrid methods can be divided into two groups: the hybrid metaheuristics that combine several metaheuristics and hybrid methods that combine an exact method and a metaheuristic. In this section we describe the first group *hybrid methaheuristics* which will used to design our method. Swarm Intelligence, Local search methods and evolutionary algorithms have been successfully combined in many applications. We can classify these hybridizations according to the taxonomy proposed by Talbi [31]. This classification allows to compare qualitatively the hybrid metaheuristics. The taxonomy has two aspects. the first one, a *hierarchical classification* allows to identify the hybridization structure and the second one, a *fat (general) classification* specifies the details of the algorithms involved in hybridization as in Fig.2.

The hierarchical classification of hybrid metaheuristics is divided into two classes: the *low-level* hybridization and the *high-level* hybridization.

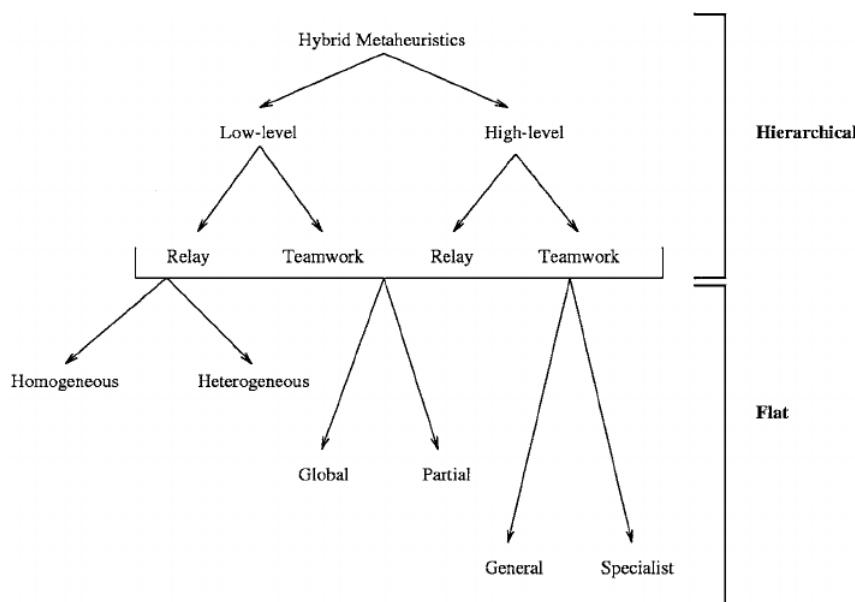


Fig. 2. Classification of hybrid metaheuristics [31]

- **Low-level hybridization** where according to a metaheuristic is replaced by another metaheuristic.
- **High-level hybridization** is obtained when two metaheuristics are hybridized without their internal functioning is related.

The previous two hybridization classes is divided into two subclasses: relay and teamwork.

- **Relay hybridization** is applied when metaheuristics are executed sequentially, one using the result as the previous entry.
- **Teamwork (co-evolutionary) hybridization** occurs when agents cooperate in parallel to explore the solution space.

The combination of all the above named classes gives four different classes:

- **low-level relay hybridization (LRH)** which represents algorithms in which a metaheuristic is incorporated into another metaheuristic single solution.
- **Low-level teamwork hybridization (LTH)** which consists to incorporate an algorithm local search operated with population based metaheuristic and focused on exploration.
- **high-level relay (HRH)** consists to use a set of metaheuristics which are executed in a sequence. Therefore, one can improve the solution of a population algorithm by using it as a starting local search solution algorithm.
- **high-level teamwork hybridization (HTH)** involves a set complete metaheuristics working in parallel and cooperate to find the optimal solution of a problem.

Table III presents more detailed description of each hybridization classes with some examples of work performed.

TABLE I. RELEVANT RELATED WORKS ON TSP BASED ON SINGLE SOLUTION METAHEURISTICS

Ref. work	Methods	Methods Features	Strength	Unhandled issues
<ul style="list-style-type: none"> <li>● Knox (1994) [32]</li> <li>● He et al. (2005) [33]</li> <li>● Zhong et al. (2008) [34]</li> </ul>	Tabu search	It can continue the search even when a local optimum solution is reached using the memory principle to avoid flashbacks	<p>This is an adaptive memory method:</p> <ul style="list-style-type: none"> <li>● Short-term memory: diversification.</li> <li>● Long-term memory: intensification.</li> </ul> <p>This method is generally better in execution time.</p>	<ul style="list-style-type: none"> <li>● Few intuitive settings.</li> <li>● Require significant resources if the taboo list is too large.</li> <li>● No demonstration of convergence.</li> </ul>
<ul style="list-style-type: none"> <li>● Song et al. (2003) [35]</li> <li>● Chen et al. (2006) [36]</li> <li>● Shakouri et al. (2009) [37]</li> </ul>	Simulated annealing	Used to find the hamiltonian path with minimum cost and to minimize the total length of the path and modify the order of crossed cities.	<ul style="list-style-type: none"> <li>● Easy to implement</li> <li>● Gives generally good solutions compared to algorithms Classic research</li> <li>● Can be used in most optimization problems</li> <li>● It converges to a global optimum (when the number of iterations tends to infinity)</li> </ul>	<ul style="list-style-type: none"> <li>● In a local minimum, it is impossible to stop.</li> <li>● The inability to know if the solution found is optimal;</li> <li>● Degradation of problem performance where there is little local minima.</li> </ul>
<ul style="list-style-type: none"> <li>● Mavroidis et al. (2007) [38]</li> <li>● Engels and Manthey (2009) [39]</li> <li>● Wu et al. (2015) [1]</li> </ul>	Lin-Kernighan (LK) local search methods (2-Opt, 3-Opt algorithm)	In each iteration, it randomly switches the value of k components of the vector solution. It was initially used to remove the crosses in a path (a possible solution) made by the TSP.	<ul style="list-style-type: none"> <li>● Avoid the falling in the local minimal and improve the final solution.</li> </ul>	<ul style="list-style-type: none"> <li>● High execution time.</li> </ul>

TABLE II. RELEVANT RELATED WORKS ON TSP BASED ON POPULATION SOLUTION METAHEURISTICS

Ref. work	Methods	Methods Features	Strength	Unhandled issues
<ul style="list-style-type: none"> <li>Wang et al. (2007) [40]</li> <li>Lui and Zeng (2009) [41]</li> <li>Kondo and Watanabe (2011) [42]</li> </ul>	Genetic Algorithm (GA)	It is based on the evolution of a population. From a population of N solutions, operators are applied to simulate the manipulation of the genome as the crossover or the mutation to reach a population of solutions using a cost function.	<ul style="list-style-type: none"> <li>Very useful for complex problem.</li> <li>Use the cut method between cities for TSP.</li> </ul>	<ul style="list-style-type: none"> <li>Very high time calculation because GA handle several solutions simultaneously.</li> <li>Convergence to a local optimum.</li> <li>Does not guarantee that the solution found is the best after a number of generations.</li> </ul>
<ul style="list-style-type: none"> <li>Islam et al. (2006) [43]</li> <li>Ugur and Aydin (2009) [44]</li> <li>Zhou (2009) [45]</li> <li>Yang et al. (2010) [46]</li> </ul>	Ant Colony Optimization (ACO)	An artificial ants colony cooperate to find good solutions to difficult discrete optimization problems using pheromone trail.	<ul style="list-style-type: none"> <li>Robust algorithm, high adaptability.</li> <li>Perfect for problems based on graphs.</li> </ul>	<ul style="list-style-type: none"> <li>A blocking condition can happen.</li> <li>Can lead a rapid emergence of a stagnation situation.</li> <li>Does not apply to all types of problems.</li> </ul>
<ul style="list-style-type: none"> <li>Zhi et al. (2004) [47]</li> <li>Zhong et al. (2007) [48]</li> <li>Shi et al. (2007) [49]</li> <li>He et al. (2009) [50]</li> </ul>	Particle Swarm Optimization (PSO)	PSO is an stochastic iterative optimization method inspired from optimal behavior of animals groups. It is applied to both problems of continuous variables as problems of discrete variables	<ul style="list-style-type: none"> <li>PSO converges quickly to an approximate solution with good quality.</li> <li>Simple and easy to implement.</li> </ul>	<ul style="list-style-type: none"> <li>It is very difficult to find good settings for the structure of the problem.</li> </ul>

TABLE III. RELEVANT RELATED WORKS ON TSP BASED ON HYBRID METAHEURISTICS METHODS

Ref. work	Hybrid Methods	Design	Methods Features	Strength	Unhandled issues
Chen and Chien (2011) [51]	Genetic SA and ACO with PSO	LRH(GA(SA)+ACO+PSO)	The simulated annealing based on genetic algorithm and ant colony system with particle swarm optimization	<ul style="list-style-type: none"> <li>Method outperformed traditional local search algorithms.</li> </ul>	<ul style="list-style-type: none"> <li>Does not apply to different types of problems.</li> </ul>
Geng et al. (2011) [52]	Simulated Annealing algorithm and Greedy Search (GS)	LTH(SA(GS))	The simulated annealing improved by GS to solve TSP	<ul style="list-style-type: none"> <li>The experimental results show that the proposed algorithm provides better compromise between computational time and accuracy among some recent algorithms for the TSP.</li> </ul>	<ul style="list-style-type: none"> <li>Can fall into a local minimum.</li> </ul>
Shi (2013) [53]	Artificial Bee Colony (ABC) Algorithm and Simulated Annealing (SA)	HRH(ABC(GA)+SA)	Hybrid ABC-SA adopts natural number coding to improve search efficiency. The heuristic crossover operator and a mutation operator are adopted to local search food source.	<ul style="list-style-type: none"> <li>Can avoid prematurity and advance stringency for TSP</li> <li>To improve the global search ability of ABC, SA is used to choose food source.</li> </ul>	<ul style="list-style-type: none"> <li>No performance for large scale problem.</li> </ul>
Kefi et al. (2016) [12]	ACO supervised by PSO and 2-Opt	HRH(PSO+ACO+2Opt)	Based on hybridization between population based solution methods, ACO and standard Particle swarm optimization (SPSO) and single solution method 2-Opt local search to solve TSP	<ul style="list-style-type: none"> <li>Cost path is improved by 2-Opt algorithm.</li> <li>2-Opt can reduce the probability of falling into a local minimum</li> </ul>	<ul style="list-style-type: none"> <li>Is not efficient for large scale test benches.</li> </ul>
Eloumi et al. (2014) [12]	PSO modified by ACO	HTH(PSO+ACO)	ACO algorithm is used to modify the parameters of PSO to improve the performance using pheromone trail heuristic information.	<ul style="list-style-type: none"> <li>Results prove the completion time and the best length.</li> </ul>	<ul style="list-style-type: none"> <li>The tour is not improved, it contains several connections.</li> </ul>

#### D. Synthesis

From the related works presented in the previous subsections (Tables I-III), the hybrid metaheuristics methods, based on population solution like swarm intelligence methods and single solution methods, are in progressive state during the last decade. The current trends of hybridization increase the intelligence of systems thus it can improve the result. For this reason, we have thought to develop a hybrid metaheuristic method to solve the routing problem and combines the swarm intelligence method like ACO and PSO and the local search meth as 2-Opt algorithm to improve the result.

### III. MATERIALS AND METHODS

We present two optimization methods which are used to develop our proposed method.

#### A. Particle Swarm Optimization (PSO)

PSO was introduced by Eberhart and Kennedy in 1995 [7]. It is inspired by bird flocks' behavior when they fly to look for their foods. Each particle, modeled as a potential solution to the optimization problem, traverses the search space in search of the global optimum. The displacement of a particle is influenced by three components: (1) a *physical component*: the particle tends to follow its current direction ; (2) a *cognitive component*: the particle tends to move towards the best site in which it has already passed; (3) a *social component*: the particle tends to head for the best position already achieved by its neighbors.

PSO is based of three steps. At the beginning of the algorithm, we must define the number of particle that we need to solve the problem and all particles are dispersed randomly in space research. Then, the particles form a bench and explore the search space while maintaining cohesion between them and gathering around the optimum. Depending on the configuration of the algorithm, the particles end up in the same location, which highlights a tendendy to move towards the optimum. Finally, all particles should updated their velocities using the following 1.

$$v_{ij}(t+1) = wv_{ij}(t) + c_1r_1(p_{il}(t) - x_{ij}(t)) + c_2r_2(p_{ig}(t) - x_{ij}(t)) \quad (1)$$

Where  $c_1$  and  $c_2$  are constants,  $r_1$  and  $r_2$  are random variables in  $[0, 1]$ .  $p_{il}(t)$  is the best local solution of the  $i^{th}$  particle and the  $p_{ig}(t)$  is the best global solution of all particles in the  $(t+1)^{th}$  iteration.

The displacement of particle is made according to 2.

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad 1 \leq i, j \leq N \quad (2)$$

The best local and the best solutions are choosen according to a fitness function defined in the proposed problem. These iterative steps are stopped when the termination condition is reached.

### B. Ant Colony Optimization (ACO)

The ACO was proposed by Dorigo et al. [8]. It is a metaheuristic in which an artificial ants colony cooperate in finding good solutions to difficult discrete optimization problems. The ACO metaheuristic is based on a generic problem representation and the definition of the ants behavior. Given this formulation, the ants in ACO build solutions to the problem being solved by moving concurrently and asynchronously on an appropriately defined construction graph.

ACO can be applied to the traveling salesman problem [54] where the ants visits all the nodes without repeating any node in the feasible tour. These nodes are stocked in the transition matrix according the probabilistic action rules in 3.

$$\begin{cases} P_{ij}^n = \frac{((\tau_{ij}^{n-1})^\alpha)(\eta_{ij})^\beta}{\sum_{j \in \Omega_i} (\tau_{ij}^{n-1})^\alpha (\eta_{ij})^\beta}, & \text{if } j \in \Omega_i \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

Where  $\tau_{ij}$  represents the amount of pheromone and refers to the desirability of visiting city  $j$  directly after city  $i$ .  $\Omega_i$  is the feasible set of cities of ant  $k$  which it has not visited yet, in step  $t$ . The parameters  $\alpha$  and  $\beta$  give the influence of the pheromone and the heuristic information  $\eta_{ij}$ , which is defined by  $\eta_{ij} = \frac{1}{d_{ij}}$ , where  $d_{ij}$  is the distance between cities  $i$  and  $j$ .

The pheromone trails are updated when all ants construct optimal tours. Pheromone evaporation is determined by (4), where  $\rho$  is the evaporation rate,  $B_t$ , the best tour. The evaporation rate parameter  $\rho$  allows the algorithm to avoid the accumulation of pheromone trails.

$$\tau_{ij} = (1 - \rho)\tau_{ij} + \rho\Delta\tau_{ij}^{n-1}, \forall (i, j) \in B_t \quad (4)$$

The ant system (AS) is the first ACO algorithm which has been proposed in the literature [9]. The update of the pheromone trails is applied according 5.

$$\tau_{ij} = (1 - \rho)\tau_{ij} + \Delta\tau_{ij} \quad (5)$$

With  $\rho \in [0, 1]$  is called evaporation rate. After evaporation, all ants deposit pheromone on arcs that they have crossed in their tour by 6.

$$\tau_{ij} = \tau_{ij} + \sum_{k=1}^m \Delta\tau_{ij}^k, \forall (i, j) \in B_t \quad (6)$$

Where  $\Delta\tau_{ij}^k$  defines the amount of pheromone, laid on edge  $(i, j)$  by the  $k$ th ant which it has visited; it is given by 7:

$$\Delta\tau_{ij}^k = \begin{cases} \frac{1}{C^k} & \text{if } arc(i, j) \in B_t \\ 0 & \text{Otherwise} \end{cases} \quad (7)$$

Where  $C^k$  is the tour length of the  $k$ th ant,  $B_t$  is the best tour and it forms a set of shortest path found within  $t$  iterations.

#### IV. OVERVIEW OF AS-PSO, ANT SUPERVISED BY PSO

A swarm is a large number of homogenous agent interacting locally among themselves and their environment in order to emerge. In this section, we are interested to present a hybridization approach combined ACO and PSO.

##### A. Motivation of AS-PSO

Meta-heuristics have emerged along with the paradigm itself, they are gaining popularity because they have worked well in some hard optimization problems such as TSP, VRP or hamiltonian path.

The hybridization between ACO and PSO is called *combinatorial optimization*. First, a general framework of the nodes representing different functions and all different paths that connects them. Second, the purpose of studying the performance and the number of cycles set time. Finally, our goal is to minimize the distance of different paths and the execution time. In fact, the main idea of this hybrid metaheuristic consists to present the robustness of the hierachic metaheuristic design for any routing problems. This hybridation aptes to minimize the distance of different paths and the execution time. In general and using several algorithms, the first one is used as a solver while the others are used to optimize the solver. For AS-PSO, it is proposed within this principle, it combines the strengths of both, PSO and ACO, the ACO is the solver and PSO is used to optimize ACO parameters.

##### B. The AS-PSO proposal adjusting $[\alpha, \beta]$

In several studies, PSO and ACO have been used to solve several optimization problems. However, using these algorithms separately showed the effectiveness of these methods compared to hybrid methods. Why don't we use both to solve optimization problems? In [9], [10], [55], the authors allow PSO to optimize parameters  $[\alpha, \beta]$  of ACO, knowing that ACO allows to optimize discrete problems and PSO, generally, optimizes for continuous problems. The Fig. 3 shows roughly the organization of AS-PSO. The heuristic, ACO acts directly with the optimization problem to get information, while the metaheuristic, PSO, adjusts the parameters  $[\alpha, \beta]$  of ACO. Therefore the heuristic, ACO is supervised by the metaheuristic, PSO. From this concept, the authors of [9] and [55] have proposed the method which is called Ant Supervised by PSO, AS-PSO and they have developed several variants of AS-PSO.

The running of the classical ACO is based on parameters that are often defined in the algorithm. Thus, to find parameters that are appropriate for a problem, the user needs to perform many tests.

With AS-PSO, the user does not need to define ACO parameters by trial and error. In fact, the system provides all parameters using PSO. Algorithm 1 describes principal steps of AS-PSO.

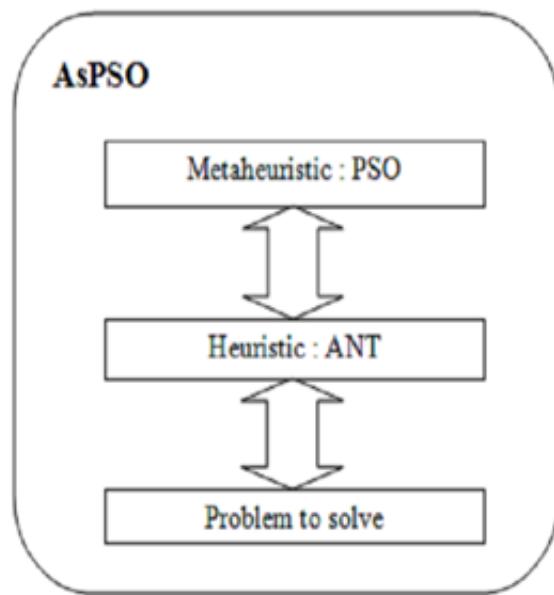


Fig. 3. AS-PSO architectural overview [9]

---

**Algorithm 1** Ant Supervised by PSO (AS-PSO) [9]

- 
- 1: Initialize the parameters of each PSO particle.
  - 2: For step index  $n = 1 : N$ 
    - For each particle index  $k = 1 : K$   
Consecutively run ACO parameters
    - Update all the parameters
  - 3: Save the best parameters
- 

Classically in ACO, the constants  $\alpha$  and  $\beta$  represent the influence of pheromone information and heuristic information. During the initialization, the functional parameters of the PSO algorithm are enabled. Each particle produces a couple of parameters  $[\alpha, \beta]$ . ACO retrieves these parameters and runs in order to find a solution from the parameters proposed by each particle.

Once the research phase of the solution is complete, each parameter is updated according to 1 of PSO and the best solution are saved. Therefore, the PSO helps the ACO to find its best parameters. Finally, in AS-PSO, there are no fixed parameters but an optimization process can generate suitable ACO parameters in which PSO is applied.

*C. The Extended AS-PSO  $[\alpha, \beta, \rho, \tau]$*

In [11], the authors applied the principle of AS-PSO previously described, so a new variant of AS-PSO is presented where the optimized ACO parameters are extended to  $[\alpha, \beta, \rho, \tau]$ . Generally, these parameters are constant, while

are determined according to the experimental results. The main difference of the extended AS-PSO consists in the determination of the pheromone trails  $\tau$  which are assigned to all edges and evaporation rate  $\rho$  which is used to update the pheromone trails. For the parameters  $\alpha$  and  $\beta$  in 3 are obtained according to 1 and 2 of PSO.

Extended AS-PSO algorithm will not be complete without the evaporation process of pheromone trails. Indeed, to avoid being trapped in sub-optimal solutions, it is necessary to allow the system to forget the bad solutions. Therefore, it is necessary to change the weight of the edges by a decay constant values of the edges at each iteration. Consequently, we update the pheromone trails  $\tau$  according to 5.

#### D. Hybrid AS-PSO $[\alpha, \beta, \rho]$ and 2-Opt Algorithm

1) *Definition of 2-Opt Algorithm:* The 2-Opt algorithm is a local search method which is proposed by Croes [56] to solve the TSP by improving an initial solution. Its principle is very simple: at each iteration it randomly switches the value of two components of the vector solution. It was initially used to remove the crosses in a path (a possible solution) made by the traveling salesman problem. This method was later extended 3-Opt. The principle is then randomly remove 3 connections in the path of the traveling salesman, and then reconnect the network of all possible ways. The best solution is retained and the process is repeated for others subsets contained three connections. This method was later generalized k-Opt, that is to say, by seeking to swap k variables in each iteration. But in the general case, the 3-Opt is rarely exceeded. It has also led to the method of Lin-Kernighan [56], which is one of the most efficient algorithm to solve the TSP.

In 2-Opt method, there are arcs in the graph where they overlap in the tour. This overlap can increase the tour length. Then, it must replace the two arcs in the specified nodes that gives the best tour length. To decrease the best tour length, several heuristic algorithms such as GA, PSO and ACO can fall into local minimum, which leads to obtain the best tour length. To fall the minimum local situations and improve the result, we can use an algorithm k-Opt and especially 2-Opt algorithm.

2) *Presentation of Hybrid AS-PSO-2Opt  $[\alpha, \beta, \rho]$ :* In the AS-PSO-2Opt method, a 2-Opt algorithm is added to AS-PSO method [9], [11]. Firstly, the authors have retained the same principle of AS-PSO. The initial step, ants are randomly dispersed in cities. Then, pheromone trails  $\tau$  are affected to all arcs. All ants achieve their first tour taking into account distances from the city. The ACO parameters  $\alpha$  and  $\beta$  are calculated using the two equations of PSO. For the fitness function of PSO, this hybrid used the same function of AS-PSO which is determined by the Euclidean distance of the tour length and the final solution of our method is considered as the shortest tour. When all the ants finish the construction of their tours, the pheromone trails are updated and the parameter  $\rho$ , pheromone evaporation rate, is determined, also, by the two equations of PSO.

The principal difference in this hybrid method compared to others previous variant of AS-PSO consists to the use of an improvement algorithm which can improve the initial solution using 2-Opt algorithm. This improvement concerns the initial solution, where two or more nodes can be changed. In this method, the 2-Opt algorithm is applied to the best ant solution obtained from AS algorithm, in order to eliminate the connection in the final path and to avoid the falling in local optimum.

## V. STANDARD AS-PSO $[\alpha, \beta, \rho]$ AND 2-OPT ALGORITHM

### A. Motivation of Standard PSO

Since 2006, three successive Standard PSO (SPSO) versions have been developed [57], namely SPSO 2006, 2007, and 2011. These versions have the same principles and they aims to improve the performance of all PSO variants.

Standard PSO is developed to overcome the problems of PSO. The principal problem of PSO is the choice of parameters which is very difficult to find good parameters for the structure of the problem. Also, in PSO, particles can exceed the boundaries of the search space so all particles can fall in the local mimimum. To solve this problem, standard PSO fixed a lower bound and an upper bound of variables for velocity as in 8 and particle position. The velocity and particle position are updated according to 9 and 10, respectively.

$$-v_{max} \leq v \leq v_{max} \quad (8)$$

$$v_{ij}(t+1) = wv_{ij}(t) + U(0, c)(p_{il}(t) - x_{ij}(t)) + U(0, c)(p_{ig}(t) - x_{ij}(t)) \quad (9)$$

where  $U(0, c)$  a random number in  $[0, c]$  given according to a uniform distribution.

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad 1 \leq i, j \leq N \quad (10)$$

The parameters values are fixed as:

- $w = \frac{1}{2\ln(2)} \simeq 0.721$
- $c = \frac{1}{2} + \ln(2) = 1.193$

### B. Our Proposed hybrid method, Standard AS-PSO-2Opt

The hybridization classification of Talbi [31] is more generally. Included in this classification the same classes of hybrids such as the integration of an algorithm in another algorithm (the low-level hybridization) and the execution in series (high-level relay hybridization) and parallel algorithms (the high-level teamwork hybridization).

For our method we have opted to a *High-level teamwork hybridization (HTH)*. The HTH scheme involves three metaheuristics which are ACO and standard PSO selfcontained algorithms performing a search in parallel, and cooperating to find an optimum. Also, we have used use the local serach algorithm 2-Opt to improve the obtained result. Intuitively, HTH will ultimately perform at least as well as one algorithm alone, more often perform better, each algorithm providing information to the others to help them. Since our problem to solve is the TSP, so, we have used the global hybrid because all the algorithms search in the whole research space. The goal is here to explore the space more thoroughly. The architectural overview of our Standard AS-PSO-2Opt is presented in Fig. 4 and the flowchart presents the pseudocode of our proposed method is described in Fig. 5.

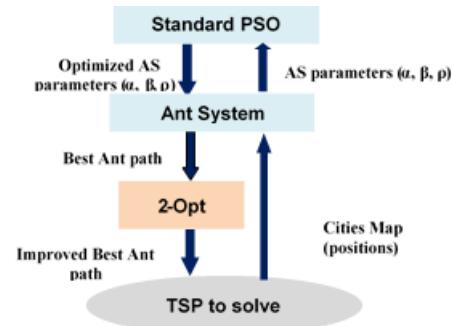


Fig. 4. Architecture of Standard AS-PSO-2Opt

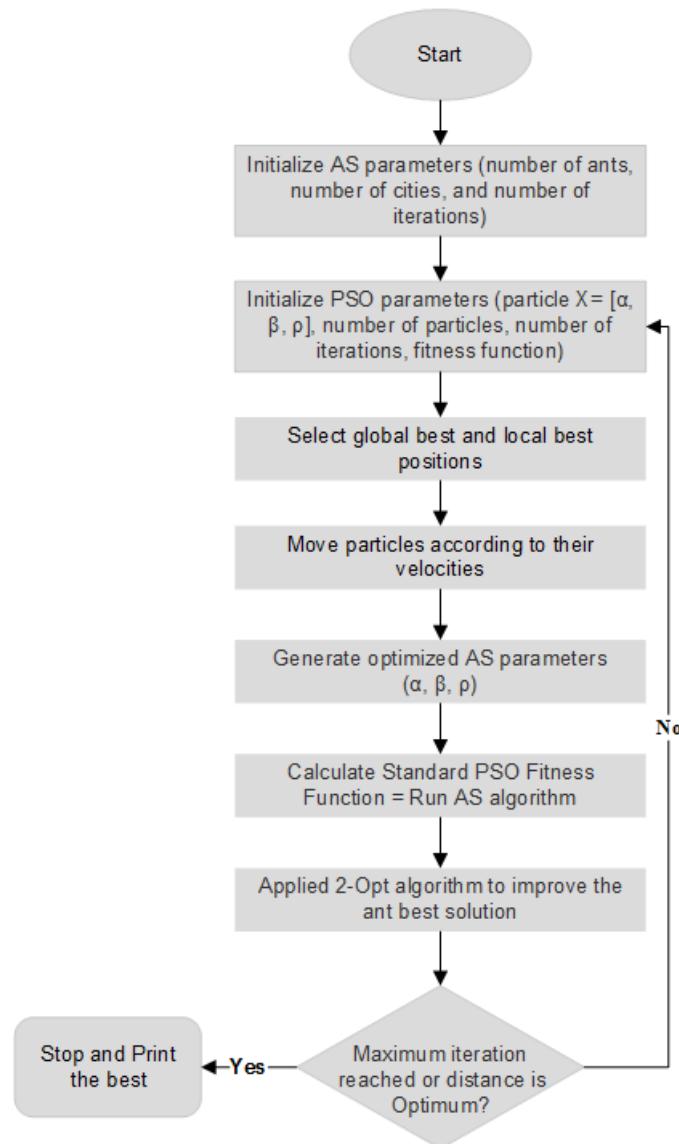


Fig. 5. The flowchart of the Standard AS-PSO-2Opt

## VI. SIMULATION, RESULTS, AND DISCUSSION

The proposed hybrid metaheuristic method is a statistical analysis was performed through 1000 executions using key test benches form TSPLIB [15]. The number of ants is equal to 20 in all experiments. For PSO algorithm, it is executed with 10 particles.

The selection of the domain of ACO parameters ( $\alpha, \beta, \rho$ ) is very important. In Ref. [8], Dorigo and Stutzle choose the values of parameters ACO as  $\alpha = 1$ ,  $2 \leq \beta \leq 5$  and  $\rho = 0.5$ . However, from the results obtained, we can conclude that these values are not the best choice. In Ref. [13], the authors select the values of parameters as  $0 \leq \alpha \leq \beta \leq 2$  and  $\rho = 0.1$ . From litterature, we can synthesis if  $\alpha = 0$ , the nearest cities are more probably selected: this gives a classic stochastic greedy algorithm (with several starting points because initially, ants are randomly dispersed on the cities). If  $\beta = 0$ , so only the amount of pheromone is adopted, without any heuristic information. This gives bad results because ants have not the visibility of total environment to choose the next closest cities and avoid the farest cities. Particularly, for values of  $\alpha > 2$  it leads to the quick emergence of a stagnation situation, which all ants follow the same path and build the same tour. This phenomenon gives a suboptimal solution.

For our system, these parameters are optimized by PSO. Therefore, PSO algorithm generated several values of each parameter for each ant and the best combination of parameters values that gives a better solution is taken by the ant. The domain of ACO parameters generated by PSO algorithm after execution of our method are shown in Table IV for several test benches as the eil51, berlin52, st70, eil76, rat99, kroA100, eil101, ch150 and kroA200.

After execution of Standard AS-PSO-2Opt algorithm, the optimum tour length and the variation of three parameters ( $\alpha, \beta, \rho$ ) of AS are shown in Figs. 6, 7, 8, 9, 10, respectively. Therefore, Fig. 6 presents the optimum course of berlin52 test bench, which is equal to 7542 as the Best Known Solution (BKS). Fig. 7 shows the optimum course of the eil76 test bench, which is equal to 538 as the BKS. Fig. 8 illustrates the optimum course of the rat99 test bench, which is equal to 21443. Fig. 9 illustrates the optimum course of the eil101 test bench, which is equal to 630. Finally, Fig. 10 shows the optimum course of the ch150 test bench, which is equal to 6559.

TABLE IV. Domain of Ant System parameters

AS Parameters	$\alpha$	$\beta$	$\rho$
Domain	[0.5, 2]	[1, 5]	[0.01, 0.5]

The best result SAS-PSO-2Opt with self-adapted parameters is presented in Table V for the eil51, berlin52, st70, eil76, rat99, kroA100, eil101, ch150, and kroA200 test benches respectively. The best results of our method are presented in bold. Comparative results of the proposed method with other studies in the literature are shown in Table V. We can conclude that the proposed method gives reasonable results compared to the optimal solution of all problems and better than the results of others studies. Moreover, we can see that for the problem berlin52 and eil76, our method gives results equal to the best known solution as 7542 and 538, respectively. Also, for the others problems, results are very close to best known solution. Finally, from the obtained results, we can synthesize that the Standard AS-PSO-2Opt method is very efficience for all test benches described in Table V.

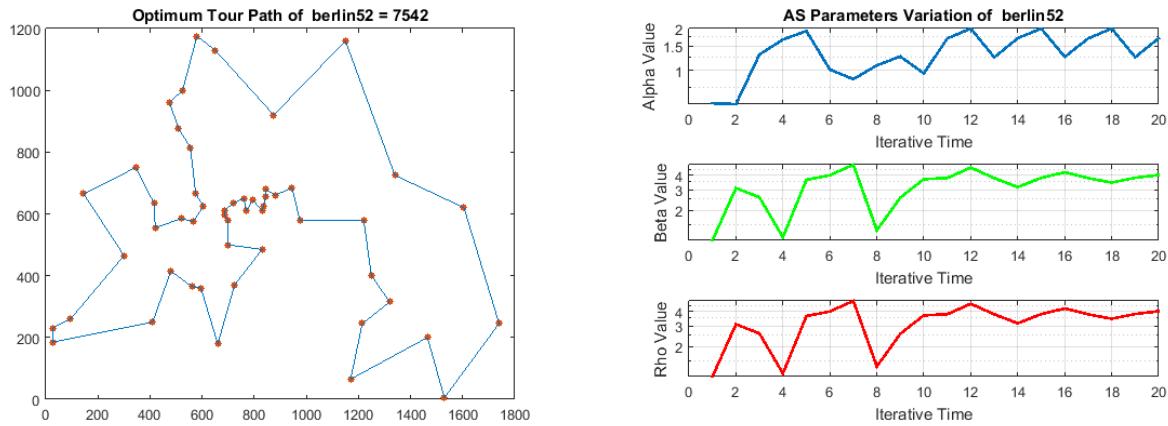


Fig. 6. Optimum tour of SAS-PSO-2Opt and AS parameters variation of berl52 test bench

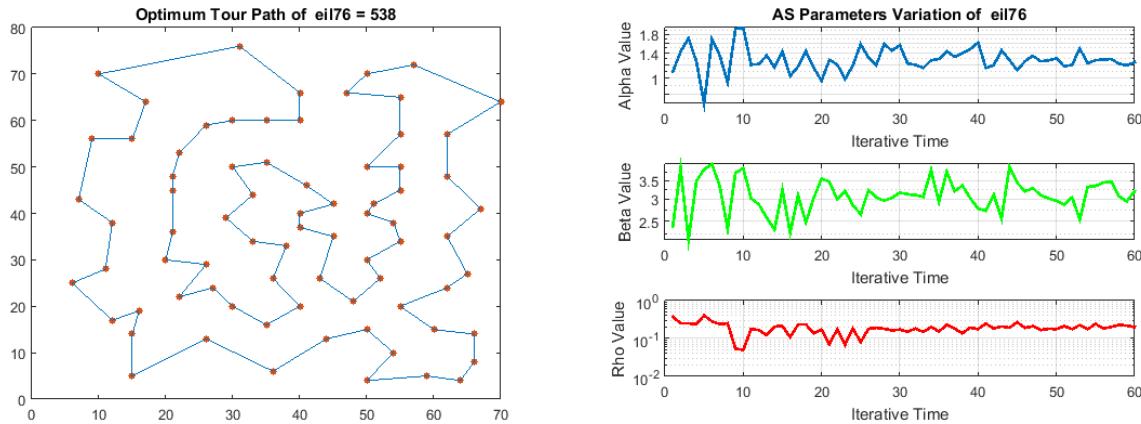


Fig. 7. Optimum tour of SAS-PSO-2Opt and AS parameters variation of eil76 test bench

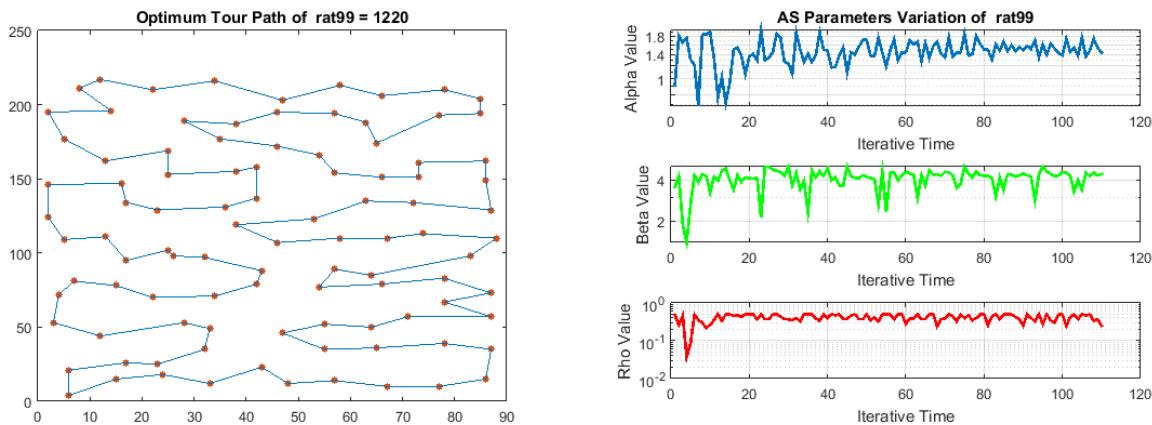


Fig. 8. Optimum tour of SAS-PSO-2Opt and AS parameters variation of rat99 test bench

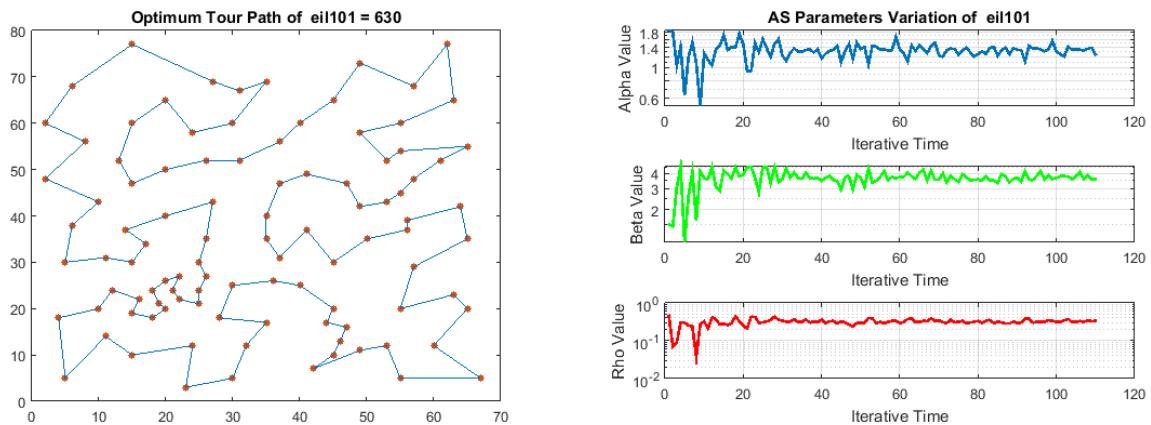


Fig. 9. Optimum tour of SAS-PSO-2Opt and AS parameters variation of eil101 test bench

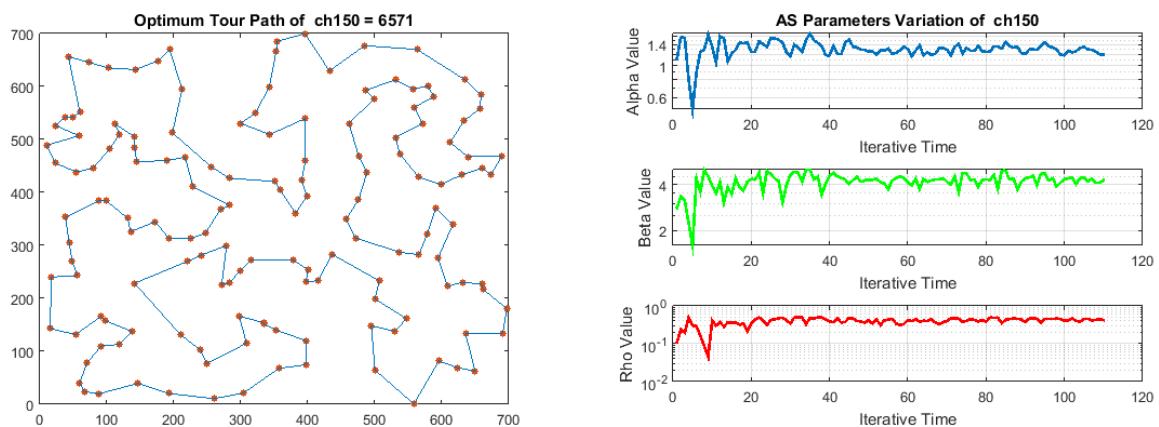


Fig. 10. Optimum tour of SAS-PSO-2Opt and AS parameters variation of ch150 test bench

TABLE V. THE COMPUTATIONAL RESULT OF THE PROPOSED METHOD AND OTHERS METHODS FROM THE LITTERATURE. AVG IS THE AVERAGE TOUR LENGTH; SD IS THE STANDARD DEVIATION; ERROR IS THE PERCENTAGE ERROR

Methods	Problems	eil51	berlin52	st70	eil76	rat99	kroA100	eil101	ch150	kroA200
	BKS	426	7542	675	538	1211	21282	629	6528	29368
ACOMAC (2004) [58]	AVG	430.68	-	-	555.70	-	21457.00	-	-	-
	SD	-			-	-	-			
	Error	1.10			3.29		0.82			
RABNET-TSP (2006) [59]	AVG	438.70	8073.97	-	556.10	-	21868.47	654.83	6753.20	30257.53
	SD	3.52	270.14		8.03		245.76	6.57	83.01	342.98
	Error	2.98	7.05		3.36		2.76	4.11	3.45	3.03
Modified RABNET-TSP (2009) [60]	AVG	437.47	7932.50	-	556.33	-	21522.73	648.63	6738.37	30190.27
	SD	4.20	277.25		5.30		93.34	3.85	76.14	273.38
	Error	2.69	5.18		3.41		1.13	3.12	3.22	2.80
IVRS-2opt (2012) [61]	AVG	439.25	7556.58	-	-	-	21498.61	648.67	-	-
	SD	-	-		-		-			
	Error	1.20	0.07				1.02	3.13		
ACO-2opt (2012) [61]	AVG	430.68	-	-	-	-	23441.80	672.37	-	-
	SD	-	-		-		-			
	Error	3.11	0.19				10.15	6.90		
Hybrid ACO (2012) [62]	AVG	431.20	7560.54	-	-	1241.33	-	-	-	-
	SD	2.00	67.48			9.60				
	Error	1.22	0.23			2.5				
GA-AntSystem (2012) [63]	AVG	-	7634.00	-	542.00	-	-	-	-	29946.00
	SD	-	-		-		-			-
	Error	1.22	0.74							1.97
ACO-Tagushi Method (2013) [64]	AVG	435.40	7635.40	-	565.50	-	21567.10	655.00	-	-
	SD	-	-		-		-			
	Error	2.21	1.24		5.11		1.34	4.13		
ACO-ABC (2014) [65]	AVG	443.39	7544.37	700.58	557.98	-	22435.31	683.39	6677.12	-
	SD	5.25	0.00	7.51	4.10		231.34	6.56	19.30	
	Error	4.08	0.03	3.79	3.71		5.42	8.65	2.28	
PSO-ACO-3Opt ( $\alpha, \beta$ ) (2015) [13]	AVG	<b>426.45</b>	7543.20	678.20	538.30	1227.40	21445.10	632.70	6563.95	<b>29646.05</b>
	SD	0.61	2.37	1.47	0.47	1.98	78.24	2.12	27.58	114.71
	Error	0.11	0.02	0.47	0.06	1.35	0.77	0.59	0.55	0.95
Standard AS-PSO-2Opt ( $\alpha, \beta, \rho$ )	AVG	427	<b>7542</b>	<b>676</b>	<b>538</b>	<b>1220</b>	<b>21443</b>	<b>630</b>	<b>6559</b>	29837
	SD	7.41	202.62	16.39	8.81	24.57	13.38	4.45	133.15	359.28
	Error	0.21	0.00	0.14	0.00	0.74	0.15	0.75	0.48	1.60

## VII. CONCLUSION

This paper presents a hybrid metaheuristic optimization method. This hybrid method is a combinatory optimization which combines several metaheuristics as Ant Colony Optimization, Particle Swarm Optimization and 2-Opt local search algorithm. The main objective of this study is to propose an efficient approach to solve the routing problem like the Travelling Salesman Problem. To evaluate our proposed hybrid approaches, we have used several standard test benches from TSPLIB and we have compared the results with other hybrid metaheuristics approaches from literature. Finally, we can conclude that in several test benches our method gives best results than the others hybrid methods. As perspective, we intend in the near future to test other variants of ACO to improve our result for large scale problems and reduce their standard deviation.

## ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

## REFERENCES

- [1] Y. Wu, T. Weise, and R. Chiong, "Local search for the traveling salesman problem: A comparative study," in *IEEE 14th International Conference on Cognitive Informatics Cognitive Computing*, July 2015, pp. 213–220.
- [2] N. Christofides, "The vehicle routing problem," *RAIRO - Operations Research - Recherche Opérationnelle*, vol. 10, pp. 55–70, 1976.
- [3] C. Wang, D. Mu, F. Zhao, and J. W. Sutherland, "A parallel simulated annealing method for the vehicle routing problem with simultaneous pickup/delivery and time windows," *Computers & Industrial Engineering*, vol. 83, pp. 111–122, 2015.
- [4] C. Groba, A. Sartal, and X. H. Vzquez, "Solving the dynamic traveling salesman problem using a genetic algorithm with trajectory prediction: An application to fish aggregating devices," *Computers & Operations Research*, vol. 56, pp. 22–32, 2015.
- [5] E. Alba, *Parallel Metaheuristics: A New Class of Algorithms*. Wiley-Interscience, 2005.
- [6] F. Glover, "Tabu search - part i," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.
- [7] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *IEEE International Conference on Neural Networks*, vol. 4, nov 1995, pp. 1942–1948.
- [8] M. Dorigo and T. Stützle, *Ant Colony Optimization*. Scituate, MA, USA: Bradford Company, 2004.
- [9] N. Rokbani, A. L. Momasso, and A. Alimi, "As-psos, ant supervised by pso meta-heuristic with application to tsp," in *International Conference on Control, Engineering & Information Technology, CEIT'2013*, 2013, pp. 148–152.
- [10] N. Rokbani, A. Abraham, and A. M. Alimi, "Fuzzy ant supervised by pso and simplified ant supervised pso applied to tsp," in *13th International Conference on Hybrid Intelligent Systems (HIS)*, Dec 2013, pp. 251–255.
- [11] S. Kefi, N. Rokbani, P. Kromer, and A. M. Alimi, *A New Ant Supervised-PSO Variant Applied to Traveling Salesman Problem*. Springer International Publishing, 2015, pp. 87–101.
- [12] W. Elloumi, H. E. Abed, A. Abraham, and A. M. Alimi, "A comparative study of the improvement of performance using a pso modified by aco applied to tsp," *Applied Soft Computing*, vol. 25, pp. 234 – 241, 2014.
- [13] M. Mahi, O. K. Baykan, and H. Kodaz, "A new hybrid method based on particle swarm optimization, ant colony optimization and 3-opt algorithms for traveling salesman problem," *Applied Soft Computing*, vol. 30, pp. 484–490, 2015.
- [14] S. Kefi, N. Rokbani, P. Kromer, and A. M. Alimi, "Ant supervised by pso and 2-opt algorithm, as-psos-2opt, applied to traveling salesman problem," in *IEEE International conference on System Man and Cybernetics SMC*, Hungary, oct 2016.
- [15] G. Reinelt, "Tsplib-a traveling salesman problem library," *ORSA Journal on Computing*, vol. 3, no. 4, pp. 376–384, 1991. [Online]. Available: <http://dx.doi.org/10.1287/ijoc.3.4.376>
- [16] D. L. Applegate, R. E. Bixby, V. Chvatal, and W. J. Cook, *The Traveling Salesman Problem: A Computational Study*. NJ, USA: Princeton University Press, 2007.

- [17] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
- [18] H. R. Lourenço, O. C. Martin, and T. Stützle, *Iterated Local Search*. Boston, MA: Springer US, 2003, pp. 320–353.
- [19] C. Voudouris and E. Tsang, "Guided local search," *Eur. J. Oper. Res.*, vol. 113, pp. 469–499, 1999.
- [20] N. Mladenovic, "A variable neighborhood algorithm - a new metaheuristic for combinatorial optimization," in *Abstracts of papers presented at Optimization Days*, Canada, May 1995, p. 112.
- [21] N. Mladenovic and P. Hansen, "Variable neighborhood search," *Computers and Operations Research*, vol. 24, p. 10971100, 1997.
- [22] T. Feo and M. Resende, "Greedy randomized adaptive search procedures," *Journal of Global Optim.*, vol. 6, pp. 109–133, 1995.
- [23] A. Eiben and J. Smith, *Introduction to Evolutionary Computing*. Berlin Heidelberg, 2003.
- [24] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*. New York, NY, USA: Oxford University Press, 1999.
- [25] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI, USA: University of Michigan Press, 1975.
- [26] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems)*. The MIT Press, December 1992.
- [27] L. Fogel, *Intelligence through Simulated Evolution : Forty Years of Evolutionary Programming*. John Wiley, 1999.
- [28] R. Storn and K. Price, "Differential evolution a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optim.*, vol. 11, pp. 341–359, 1997.
- [29] X.-S. Yang, *A New Metaheuristic Bat-Inspired Algorithm*. Springer Berlin Heidelberg, 2010, pp. 65–74.
- [30] D. Teodorović and M. DellOrco, "Bee colony optimization—a cooperative learning approach to complex transportation problems," in *Advanced OR and AI Methods in Transportation: Proceedings of 16th Mini-EURO Conference and 10th Meeting of EWGT - Poznan: Publishing House of the Polish Operational and System Research*, sep 2005, pp. 51–60.
- [31] E.-G. Talbi, "A taxonomy of hybrid metaheuristics," *Journal of Heuristics*, vol. 8, no. 5, pp. 541–564, 2002.
- [32] J. Knox, "Tabu search performance on the symmetric traveling salesman problem," *Computers & Operations Research*, vol. 21, no. 8, pp. 867–876, 1994.
- [33] Y. He, Y. Qiu, G. Liu, and K. Lei, "A parallel adaptive tabu search approach for traveling salesman problems," in *The International Conference on Natural Language Processing and Knowledge Engineering*, Oct 2005, pp. 796–801.
- [34] Y. Zhong, C. Wu, L. Li, and Z. Ning, "The study of neighborhood structure of tabu search algorithm for traveling salesman problem," in *The Fourth International Conference on Natural Computation*, vol. 1, Oct 2008, pp. 491–495.
- [35] C. Song, K. Lee, and W. D. Lee, "Extended simulated annealing for augmented tsp and multi-salesmen tsp," in *The International Joint Conference on Neural Networks*, vol. 3, July 2003, pp. 2340–2343.
- [36] Y. Chen and P. Zhang, "Optimized annealing of traveling salesman problem from the nth-nearest-neighbor distribution," *Physica A: Statistical Mechanics and its Applications*, vol. 371, no. 2, pp. 627–632, 2006.
- [37] H. S. G., K. Shojaei, and M. B. T., "Investigation on the choice of the initial temperature in the simulated annealing: A mushy state sa for tsp," in *17th Mediterranean Conference on Control and Automation (MED 09)*, June 2009, pp. 1050–1055.
- [38] I. Mavridis, I. Papaefstathiou, and D. Pnevmatikatos, "Hardware implementation of 2-opt local search algorithm for the traveling salesman problem," in *18th IEEE/IFIP International Workshop on Rapid System Prototyping (RSP)*, May 2007, pp. 41–47.
- [39] C. Engels and B. Manthey, "Average-case approximation ratio of the 2-opt algorithm for the tsp," *Operations Research Letters*, vol. 37, no. 2, pp. 83–84, 2009.
- [40] L. Y. Wang, J. Zhang, and H. Li, "An improved genetic algorithm for tsp," in *International Conference on Machine Learning and Cybernetics*, vol. 2, Aug 2007, pp. 925–928.
- [41] F. Liu and G. Zeng, "Study of genetic algorithm with reinforcement learning to solve the tsp," *Expert Systems with Applications*, vol. 36, no. 3, Part 2, pp. 6995–7001, 2009.
- [42] F. Kondo and T. Watanabe, "A study on distributed parameter free genetic algorithm for tsp problem," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2011, pp. 675–680.
- [43] M. M. M. Islam, M. W. H. Sadid, S. M. M. A. Rashid, and M. J. Kabir, "An implementation of aco system for solving np-complete problem, tsp," in *International Conference on Electrical and Computer Engineering*, Dec 2006, pp. 304–307.
- [44] A. Ugur and D. Aydin, "An interactive simulation and analysis software for solving tsp using ant colony optimization algorithms," *Advances in Engineering Software*, vol. 40, no. 5, pp. 341–349, 2009.

- [45] Y. Zhou, "Runtime analysis of an ant colony optimization algorithm for tsp instances," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 5, pp. 1083–1092, Oct 2009.
- [46] J. Yang, X. Shi, M. Marchese, and Y. Liang, "An ant colony optimization method for generalized tsp problem," *Progress in Natural Science*, vol. 18, no. 11, pp. 1417–1422, 2008.
- [47] X. H. Zhi, X. L. Xing, Q. X. Wang, L. H. Zhang, X. W. Yang, C. G. Zhou, and Y. C. Liang, "A discrete pso method for generalized tsp problem," in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*, vol. 4, Aug 2004, pp. 2378–2383.
- [48] W. Zhong, J. Zhang, and W. Chen, "A novel discrete particle swarm optimization to solve traveling salesman problem," in *IEEE Congress on Evolutionary Computation*, Sept 2007, pp. 3283–3287.
- [49] X. Shi, Y. Liang, H. Lee, C. Lu, and Q. Wang, "Particle swarm optimization-based algorithms for tsp and generalized tsp," *Information Processing Letters*, vol. 103, no. 5, pp. 169–176, 2007.
- [50] J. He, D. Shi, and L. Wang, "An adaptive discrete particle swarm optimization for tsp problem," in *Conference on Computational Intelligence and Industrial Applications, PACIIA 2009.*, vol. 2, Nov 2009, pp. 393–396.
- [51] S. Chen and C. Chien, "Solving the traveling salesman problem based on the genetic simulated annealing ant colony system with particle swarm optimization techniques," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14 439–14 450, 2011.
- [52] X. Geng, Z. Chen, W. Yang, D. Shi, and K. Zhao, "Solving the traveling salesman problem based on an adaptive simulated annealing algorithm with greedy search," *Applied Soft Computing*, vol. 11, no. 4, pp. 3680 – 3689, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1568494611000573>
- [53] A Hybrid Artificial Bee Colony Algorithm Combined with Simulated Annealing Algorithm for Traveling Salesman Problem, Dec 2013.
- [54] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," in *IEEE Transactions on Evolutionary Computation*, 1997.
- [55] W. Elloumi, N. Rokbani, and A. M. Alimi, "Ant supervised by pso," in *International Symposium on Computational Intelligence and Intelligent Informatics*, 2009, pp. 161–166.
- [56] G. A. Croes, "A method for solving traveling salesman problems," *Operations Research*, vol. 6, no. 6, pp. 791–812, 1958.
- [57] M. Clerc, "Standard particle swarm optimisation: From 2006 to 2011," in *Technical report*, 2012.
- [58] C.-F. Tsai, C.-W. Tsai, and C.-C. Tseng, "A new hybrid heuristic approach for solving large traveling salesman problem," *Information Sciences*, vol. 166, pp. 67–81, 2004.
- [59] R. Pasti and L. N. de Castro, "A neuro-immune network for solving the traveling salesman problem," in *IEEE International Joint Conference on Neural Network Proceedings*, 2006, pp. 3760–3766.
- [60] T. A. Masutti and L. N. de Castro, "A self-organizing neural network using ideas from the immune system to solve the traveling salesman problem," *Information Sciences*, vol. 179, no. 10, pp. 1454 – 1468, 2009, including Special Issue on Artificial Imune Systems. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0020025508005446>
- [61] K. Jun-man and Z. Yi, "Application of an improved ant colony optimization on generalized traveling salesman problem," *Energy Procedia*, vol. 17, pp. 319–325, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1876610212004377>
- [62] W. Junqiang and O. Aijia, "A hybrid algorithm of aco and delete-cross method for tsp," in *IEEE International Conference on Industrial Control and Electronics Engineering*, 2012, pp. 1694–1696.
- [63] D. Gaifang, G. W. W, and T. Kevin, "Solving the traveling salesman problem using cooperative genetic ant systems," *Expert Systems with Applications*, vol. 39, no. 5, pp. 5006–5011, 2012.
- [64] M. Peker, B. ŞEN, and P. Y. Kumru, "An efficient solving of the traveling salesman problem: the ant colony system having parameters optimized by the taguchi method," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 21, no. 1, pp. 2015–2036, 2013.
- [65] M. Gunduz, M. Kiran, and E. Ozceylan, "A hierachic approach based on swarm intelligence to solve traveling salesman problem," *Turkish Journal of Elec Eng & Computer Sciences*, 2015.

# Generative Software Development Techniques of User Interface: Survey and Open Issues

Thouraya SBOUI<sup>1</sup>, Mounir BEN AYED<sup>2</sup>

<sup>1</sup>*Department of Computer Science, National Engineering School of Sfax, Sfax, Tunisia*

<sup>2</sup>*Department of Computer Science and Communication, Faculty of Science of Sfax, Sfax, Tunisia*

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

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

## I. INTRODUCTION

The development of User Interface (UI) is a complex task. This complexity is related to the application complexity, task complexity and also context adaptation complexity. To resolve this complexity, software designers recommend the use of generative development process that advocates abstraction and reuse in order to facilitate development and to increase automation in program development. Model Driven Engineering (MDE) process is one of the processes. MDE aims to change the focus of software development from code to models while enabling the systematic use of models as primary engineering artifacts throughout the engineering lifecycle. The application of MDE to user interfaces gives birth to the Model Based User Interface Development (MBUID) approach [4] [5] [6]. MBUID strives to cope with the model driven challenges mentioned above. It concentrates on high-level models to analyze and specify a UI without giving importance to implementation details. MBUID supports some UI specifics like usability: the measure of an interface's potential to accomplish the goals of the user and adaptability, and the UI ability to change its layout and elements according to the context. In addition, Software Product Line (SPL) [1] [2] [12] is another paradigm to be considered. SPL is a process which creates a collection of similar software systems from a shared set of software assets. It defines two levels : a domain engineering level, which designs the software line, and an application engineering level which generates a specific product. The specificity of SPLs is the design of product commonalities and variabilities.

This paper presents an overview of generative approaches used to develop user interfaces. The paper structure is organized as follows: first this introduction, then the second section is a background on the evolution of UI development , the third section is a presentation of generative techniques used to develop

UIs, the fourth section conducts a brief Systematic Literature Review (SLR) of UI-SPL approaches, the fifth section highlights the automation aspect of UI-SPL approaches, the sixth section summarize the findings of section 4 in the form of a common process, in the seventh section, we validate the proposed schema by trying to tag UI-SPL proposals. The penultimate section presents two challenges for futures works. Finally, we conclude the paper.

## II. Background

User Interface development [9] [11] [13] is evolving just as is the case with general software development. Figure 1 describes the 4 outstanding generations: old live cycle, object oriented process, model driven process and software product line process.

The first generation is the generation of the classic life cycle [12] [13] like the waterfall model, V-Model and spiral model. The benefit of using these methods is the ease of development of complex software while their disadvantage is that they are reserved to generate and evaluate the functional side of applications (i.e. no specific attention is given for the UI).

The second generation is the generation of object oriented processes [22] (i.e OOSE Method, Boosh method and the Unified process). These processes are inspired from object-oriented programming [40] concepts and are crucial in software analysis and design because they address fundamental issues of software modifiability, adaptation, and evolution [12]. The benefit of using object oriented processes is that they are more adapted to UIs thanks to “UML profiles” [3]; a concept specific to Unified Modeling Language (UML). Profiles are used to adapt UML to any business domain.

Both the old life cycle and object oriented processes are single-software processes, meaning that if the same interface needs to be generated with other characteristics (implementation technology, new user preferences); the process has to be re-implemented.

This requirement was the impetus for software engineers to design and develop generative processes. Generative process is a family-software process that generates a variety of software products. Generative processes [31] are MDE process and SPL process. Both of them use models as a key concept for software development. For example Model Driven Architecture (MDA) [39] process (a specific MDE process proposed by the Object Management Group consortium) generates a family of platform-specific softwares while SPL process generates a family of variable software (The variety in SPL is larger, e.g. variety according to the platform, the environment of use, the user profile and so on). The focus of this paper is the use of generative processes for UI development.

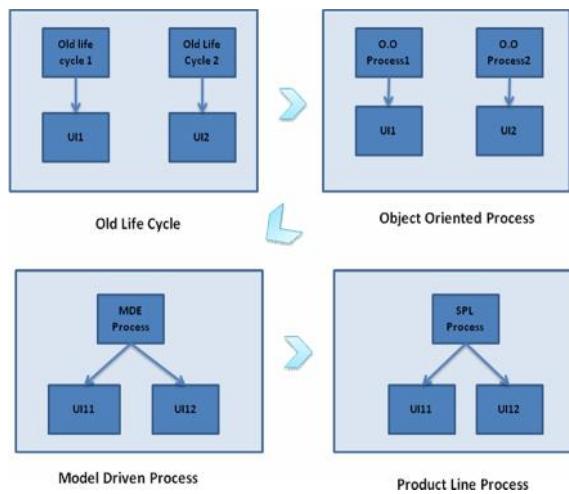


Figure1. UI development through the evolution of Software Development Processes

### III. Generative techniques for UIs development

#### A. Model Driven Approaches

Model Driven Engineering (MDE) [35] is a software development methodology which considers the model as a central artifact in software engineering leading to the design of the system and the generation of the desired software. One of the key points of MDE is the raise of abstractions [16]. MDE defines four levels of abstraction: the meta-meta-model, the meta-model, the model and the code. These levels are connected by means of transformations.

Within the field of UI development, much research has been published covering the state of the art of model driven approaches. The following is survey of some of the literature.

In [24], Akiki et al. present an overview and an evaluation of model driven user interface development systems. They classify the state-of-the-art systems under three categories: reference architectures, practical techniques, and supporting tools. Then, they evaluate them according to a set of criterias. Furthermore, authors of [24] propose the consideration of some other elements such as the empowerment of the end user, designer interventions and the management of the gap existing between adaptative techniques and supported tools.

Another important paradigm is the Model Based User Interface Development (MBUID) paradigm. This paradigm is a particular MDE approach which simulates the Model Driven Architecture process. The contribution of the MBUID, compared to MDA, is the definition of models specific to UIs. These models are arranged into four levels in the step-wise development life cycle.

Most MBUID approaches rely on the Cameleon Reference Framework (CRF) [6]. CRF was developed by Calvary et al. in 2002, and then it was revised and refined in 2003. It serves as a reference for the development of UIs that support multiple targets or multiple contexts of use.

The CRF is illustrated in figure2. Within this platform there are both task and domain models. The task model describes the logical activities that have to be carried out in order to reach the user's goals while the domain model is a conceptual model of the domain that incorporates both behavior and data. Both models are used as input to generate the abstract user interface (AUI) which expresses the UI in terms of Abstract

Interaction Units (AIU) (or Abstract Interaction Objects (AIO)), as well as the relationships among them. These AIUs are independent of any implementation technology or modality (e.g., graphical, vocal or gestural). The AUI may give rise to one or many Concrete User Interface (CUI). A CUI expresses the UI in terms of Concrete Interaction Units (CIU) (or Concrete Interaction Objects (CIOs)). These CIUs are modality-dependent but implementation technology independent. The CUI is used to derive the Final Interactive UI (FUI).

In this context, Meixner et al. [5] propose the classification of MBUID approaches into four generations. The first generation focuses on identifying and abstracting relevant aspects of UIs whereas the second generation focuses on the extension of UI models (the definition of task model, presentation model and dialog model). The third generation challenge was to develop a UI for several different devices and platforms and the actual MBUID generation focuses on the generation of context-sensitive UIs for a variety of different devices, platforms and modalities.

Jaouadi et al. [26] present another survey describing adaptive interactive systems. They identify three main groups of the adaptation approach: model-based approach, components-based approach and a combined (component and model)-based approach. The classification was realized according to the adaptation features (adaptation type, adaptation moment, context element, adaptation level etc.). In [26], authors recommend the use of a model-driven approach.

The common thread of those surveys [5] [24] [26] is the recognition of the importance of the context-sensitivity in the development of interactive UI. A context is defined in many works [4] [6] [10] [34] by means of three elements: the user (the software customer, e.g.: a visually impaired person), the platform (software and hardware platforms on which the application is installed, e.g.: a mobile device with an android OS) and the environment (the physical place, e.g.: a noisy and luminous place).

Another survey [25] in the same context focuses on designing tools, transformation tools and code derivation tools (e.g. User Interface Description Language (UIDL), a design tool). A UIDL is a Domain Specific Language and an XML-compliant markup language describing a user interface in terms of aspects (structural, behavioral) and at different abstraction levels. For example, User Interface Extensible Markup Language (UsiXML) [7] [8] is a UIDL which supports the different levels of the CRF schema except the FUI level). [27] is a comparative review of some UIDLs. This review was produced in order to analyze how UIDLs support the various stages of UI development life cycle and development goals.

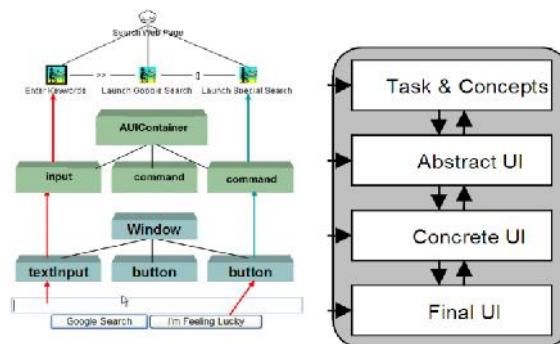


Figure 2. An instantiation of the CAMELEON Reference Framework [6] [4]

## B. Software Product Line Approaches

The second generative process is the Software Product Line Engineering (SPL) process [1] [2] [12]. SPL promises significant improvements in time-to-market, cost, and reliability through the system identification and the exploitation of commonalities and variations in software systems.

SPL is a two-level approach (see Figure 3): an abstract level called the domain engineering level and a concrete level called the application engineering level. The domain engineering level covers domain analysis (identification of common and variable features among the family members), domain design (development of common assets of all family products) and domain implementation (the implementation of the family-assets). The application engineering level covers application analysis (the configuration of the feature model), application design (instantiation of common assets in order to define the architecture of a specific product) and application implementation (the development of a specific product).

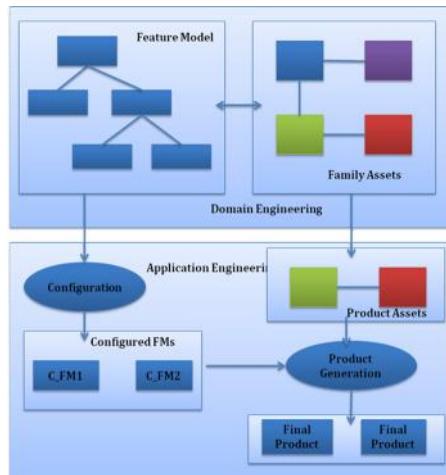


Figure3. Software Product Line process

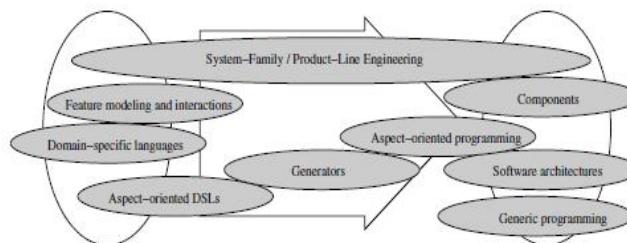


Figure4. Relationship between generative software development and other fields [31]

Both engineering levels define two spaces: a problem space and a solution space. The problem space is for expressing variability and product configuration while the solution space is for assets implementation and product generation. Figure4 highlights concepts and technologies used to implement the two spaces.

Feature modeling [36] is a key concept in product line engineering. Feature modeling by means of Feature Diagrams (FD) is a popular technique for capturing commonality and variability in Software Product Lines. A feature model is an intermediate step to move from problem space to solution space.

Unlike model-based approaches, there is no review covering the development of UIs using SPLE. The next section is a brief systematic literature review describing UI-SPL approaches.

#### IV. SPL- A Systematic Literature Review

A systematic review [37] is considered as the most valid and the most reliable method to identify and synthesize existing works. Therefore, it offers very strong results for decision making.

A systematic review has five main steps [37]: 1) framing the questions of the review, 2) definition of inclusion/exclusion criteria, 3) identifying relevant works, 4) assessing the quality of studies and 5) interpretation of the findings. The five steps are described below:

##### A. *Framing Research Questions*

Research questions reveal the problems to be addressed by this review. Five research questions are defined:

RQ1: What SPL methods exist to manage the variability of UIs or to derive usable UIs?

RQ2: What design elements are used for the implementation of those methods?

RQ3: What is the degree of automation of these methods?

RQ4: What motivates the adoption of existing methods?

RQ5: What are the limitations of existing methods?

##### B. *Search strategy and inclusion/exclusion criteria*

A manual search using specific conference proceedings (e.g. ACHI, SPLC) and specific journals (e.g. IJHCI, TCHI) was used. A search was conducted for all published paper from 2007 to 2016 in six conferences and workshops and three journals as shown in Table1. Given the limited number of works covering UI development using SPLE, both higher-quality sources and lower-quality sources were included. The quality of a source is defined by its rank and h-index; the higher the rank, the more selective the source is, the higher the h-index, the better the source is. However, it is possible to find sources which are considered as having an important venue in product line engineering but which do not have a rank (e.g. SPLC). Workshops were also considered as a research source (e.g. Variability Modeling of Software (VaMoS) Workshop and Advanced Visual Interfaces (AVI) Workshop).

The manual search was complemented by an automated search to ensure reliability of the studies rendered by the manual research and to ensure that no relevant approach escaped the manual search. For automatic research, academic databases were used (e.g. SpringerLink, IEEE Xplore, ACM digital library) and also the Google Scholar engine.

For an automatic search, both a simple search and an advanced search were employed. The advanced search used filters as a powerful search engine performs more precise searches by filtering them through a set of fields such as:

1. Search terms: Terms used were “software Product line”, “SPLE”, “User interface”, “variability management”, “UI”;
2. Publication period: As mentioned above, the chosen period was from 2007 to 2016 As SPL approaches addressing User interfaces began in 2007 by Garcés [23];
3. Publication venue: Journals, conferences and workshops were accessed.

For each rendered document (see Table 1), a set of inclusion and exclusion criteria were used to verify the relevance of the document:

Inclusion criterion1: Study is internal to SPLE. This study is only concerned with UI development;

Inclusion criterion 2: The study focuses on the variability management of UIs and the SPL process that generates interactive UIs;

Inclusion criterion3: Given the limited number of works which cover the second inclusion criteria, papers whose source is lower-quality were not exclude.

Exclusion criterion: Studies about general SPL process or about other types of systems (other than UI) were considered irrelevant.

TABLE I. Research Venues and Results

	<b>Acronym</b>	<b>Full Name</b>	<b>Rank/ H-index</b>	<b>Search results</b>
Conference/Workshop	GPCE	Generative Programming conference engineering	B/-	--
	SPLC	Software Product Line	-/-	[17] and [42]
	VaMos	Variability Modeling of software	-/5	[38] and [20]
	ACHI	Advances in Computer Human Interaction	C/8	[30]
	AVI	Advanced Visual Interfaces	B/24	--
Journal	IJHCI	International Journal of Human-Computer Interaction	A/38	--
	TCHI	ACM transactions on computer-human Interaction	B/46	--
	CEJCS	Central European Journal of Computer Science	-/-	[15]

### C. Relevant works

Some relevant works, which will be discussed in the following, are exposed in Table 1. For the rest, they were founded in the references of paper of Table 1 or using the Google Scholar engine.

Table 2 classifys relevant works according to a set of comparisons criteria. These criteria are used to analyze works, identify their limitations and define possible issues in future works.

Comparison criteria are defined as follows:

- Variability Management/Derivation Process: This criterion indicates the nature of the classified approach. In the literature review, two types of SPL approaches exist: those which manage the variability of UIs (definition of UI variability model and/or the configuration of UI variability model) and those proposed to derivate interactive user interfaces (going from feature model up to final UI);
- UI type: Indicates the type of the addressed UI. Most approaches are interested in Graphical User Interfaces. Such criterion can justify the choice of some design elements (the fourth comparison criteria);
- Approach type: This criterion indicates the concept used to implement the SPL approach. Such criterion can be a model, a component, an aspect or an architecture;
- Design elements: Indicates artifacts used to manage the UI variability (if the approach is a UI variability management approach) and/or others artifacts used to develop the derivation process (if the approach is a UI development process);
- Automation: This criterion indicates if the approach is automated or semi-automated.

#### D. Assessment and evaluation of findings

In [23], Garcés et al. propose a semi-automatic MD-SPL approach (MDA concepts are combined with SPL concepts) in order to develop a graphical user interfaces. Authors define a layered approach; each layer is related to a specific domain (business, architectural or technological) and for each domain, they define the metamodel, the correspondent model and the feature model. For each level, models are instantiated from metamodels and feature models are connected to models by means of weaving models. The approach layers are connected by means of transformations. In [23], generated UI was specific only to Java platform (no validation for the variability insured by feature models).

In [20], Pleuss et al. combine SPL and MBUID concepts in order to develop a semi-automated UI development approach. They define two engineering levels. In the domain engineering level, the authors define models which design the whole product family such as the feature model, the task model and the abstract user interface model. To design a particular product, authors use configured FM, AUI and task models related to particular user requirements. The relation between models is defined using mappings models. To customize the UI, authors opted for the clustering of AUI into presentations units. From the clustered AUI, the CUI is generated. The CUI can, in turn, be customized. What is lacking in this approach is the derivation of the final UI.

In [18], Muller proposes to combine the approaches of both Pleuss [20] and Lutteroth [19] (a constraint-based approach which defines a UI in terms of widgets layout). The Muller approach is too theoretical; the author proposes neither a schema to explain it nor a prototype to implement it.

In [17], Quinton et al. propose an automatic model driven approach that generates UIs for mobile devices by merging FM assets. To bridge the gap between application FD and the device FD, authors propose a pruning process which creates a reduced application metamodel. The role of this metamodel is to check if the product being derived can be executed in a given hardware. In [17], authors can extend their proposal by maximizing the consideration of targeted end-users (an essential element of the context of use).

In [38], Boucher et al. mention that direct configuration of FMs is not suitable and propose the decoupling of FM and UI configuration by combining the separation of concerns and generative techniques. Then, the CUI is generated from the AUI, with the later is being generated from the FM. The FM takes as input, views of the configuration UI, feature configuration workflow and the property sheet. The weakness of the Boucher proposal is its limitation to complex interfaces.

In [15], Bacikova and Poruban propose an automatic method for domain analysis of graphical user interfaces. The role of the proposed prototype is the extraction of useful information from existing GUIs. It takes as input a software system with a graphical user interface constructed of components and its output is a semi-formal domain model. The definition of the prototype is based on defined GUI stereotypes.

In [29], Arbodela et al. use a model driven approach based on a decision model to derive a specific product. The decision model takes as input the transformation model (defining the relationship between the feature model, the domain concepts metamodel and the architecture metamodel) and the feature model. The decision model is used with the product model and the configured feature model to generate the final product. Arboleda's approach is not dedicated to user interface; UI was just a use case to validate the approach.

In [14], Logre et al. support the development of a monitoring dashboards. In their approach, authors define a metamodel which allow the modelisation of dashboards relative to different requirements and a variability models which manage the technological variability of dashboard libraries. The presented prototype implements the link between the metamodel and the feature model and provides a semi-automated support for the approach (the user intervention is needed to choose a concrete widget). The weakness of Logre approach is that it is specific to dashboards; then the proposed metamodel can not be used for the generation of other types of UIs.

In [30], Gabilon et al. propose an automatic Dynamic Software Product Line (DSPL) process that generates a UI able to adapt its behavior when the context changes during the runtime. To generate an adaptive UI, authors used the configured feature model, the current context of use and components (features implementation). The weakness of this proposal is the use of components; not enough components are reusable even when it comes to the development of complex UIs.

In [28], Sottet et al. define an MD-SPL approach to manage UI variability and derive a variety of usable UIs. The variability management is performed using multiple FMs which define different variabilities (i.e. domain variability, interactive variability) allowing the separation of concerns. For the configuration, the authors adopt a partial and a staged configuration. The configuration is a user centered process presented by a UI generated (as is the case of the final UI) using an MD-SPL process. The problem

with the partial configuration is the inconsistency which may occur when designers want to produce a final product configuration (the aggregation of all configuration parts).

## V. Automation in SPL process

This section examines the overall spectrum of proposals presented in Table 2 from the viewpoint of automation in development. SPL approaches dedicated to the development of UIs are divided into automatic, semi-automatic (needing user intervention) and theoretical approaches (not yet developed).

In Garcés et al. [23], authors use a suite of eclipse tools (plug-ins): Eclipse Modeling Framework (EMF) to design meta-models and models, Graphical Modeling Framework (GMF) to create meta-models and instantiate correspondent models, ATL Transformation Language (ATL) as a transformation language, Atlas Model Weaver (AMW) to weave models (models and feature models) and Acceleo to generate the source code.

In Pleuss et al. [20], authors define a semi-automatic approach. The definition of feature model, task model, feature mappings (mapping between feature model and task model), and the adjustment of the AUI model is done manually. The definition of AUI model, product derivation (definition of task and AUI model relative to the product), the transformation AUI-CUI and the CUI implementation are performed automatically. For the remainder, feature configuration and task-AUI clustering, they are tool-supported.

In [18], the Muller approach is still at the theoretical stage. The author neither implements his approach nor describes tools that he proposes to use.

In Quito et al. [17], authors use a model-driven framework (AppliDE platform) which relies on Eclipse Modeling Framework (EMF). They use the AppliDE metamodel to build feature assets. To merge resulting assets which correspond to a particular product, they use a merge algorithm. From the merged assets, authors use model-transformation techniques to generate the application (for which they use Acceleo generators). For managing the gap between device FM and application FM, authors use the AppliDE metamodel (they prune it) and verify for which device the generated product is to conform.

In Boucher et al. [38], the approach is still in the theory (a vision paper). Authors propose an approach which generates a configuration UI. They merely quote tools that they plan to use. They use Text-based Variability Language (TVL) for a feature modeling, property sheet description and views description, workflows for feature configuration, and UIDLs for AUI and CUI description.

In [15], Bacikova and Poruban describe a user-assisted algorithm; DEAL (Domain Extraction Algorithm) which extracts domain information from GUIs. The algorithm takes as input a window or a dialog root component to traverse and as output, it returns the domain model (a feature model) in the form of a root feature.

In Arboleda et al. [29], the authors focus on the implementation of the decision model considered as a workflow. The adaptation mechanism (adaptation of transformation programs to the configuration of the feature model) of decision model is implemented using Aspect-Oriented Programming. Then, this model is used by the open Architecture Ware (oAW) engine to derivate the product.

In Logre et al. [14], authors propose a semi-automatic approach for the development of a composed dashboard. The user intervenes to define dashboard composition using the metamodel of concerns and the tool search for an equivalent concrete widget (using user interaction). Then the tool automatically generates the corresponding code of dashboards in HTML.

For Sottet et al. [28], the proposed approach is tool-supported. They use their own tool the AME platform. The designer defines domain model, IFML model and the connection between them. Then they generate the code.

Following this presentation of all SPL proposals designed for the development of user interfaces, the next section recommends the use of a model driven software product line process. This recommendation is inspired from proposals [43] which confirm the complementarity of MDE concepts and SPL concepts. The use of models, instead of components, aspects, templates and other technologies to implement an SPL approach ensures more reusability, best maintainability and more reliable SPL process [42].

TABLE II. Summary table of SPL approaches for UI development

Approach	Variability Management /Derivation Process	UI Type	Approach Type	Design Elements	Automation
Garcés et al. [23] 2007	Derivation Process Variability management	GUI	Model based	<b>Metamodels:</b> Business, architecture and java metamodels <b>Models:</b> business model, architecture model, Java model, architecture feature model and Java feature model	Semi-automatic
Pleuss et al. [20] 2010	Derivation Process	Not Specified	Model based	<b>Models:</b> task, AUI, CUI, UI feature, Clustered AUI and mappings models (describe connection between models)	Semi-automatic
Muller [18] 2011	Derivation process	GUI	Model-based	<b>Models:</b> MBUID models +ALM model (model describing the layout of GUI)	Not Specified
Quinton et al. [17] 2011	Derivation process	GUI	Model based	<b>Models:</b> Application FM, Device FM <b>Metamodels:</b> for managing gap between application FM and device FM	Automatic
Boucher et al. [38] 2012	Variability management (feature configuration)	GUI-Workflow based Interfaces	Model based	<b>Models:</b> UI feature model, feature configuration workflows model, AUI, CUI	Not specified
Bacikova and Poruban [15] 2012	Variability management (reverse Engineering)	GUI	Component based	<b>Models:</b> UI feature model	Automatic
Arboleda et al. [29] 2013	Derivation process	GUI	Model-based	<b>Metamodels:</b> problem space metamodel (domain concepts), solution space metamodel (product line architecture) <b>Models:</b> feature model, problem	Automatic

Approach	Variability Management /Derivation Process	UI Type	Approach Type	Design Elements	Automation
				space model (particular product model), transformation model (rules of product generation), decision model (product derivation)	
Logre et al. [14] 2014	Derivation Process	GUI-Dashboards	Model based	<b>Metamodels:</b> metamodel capturing dashboard concepts <b>Models:</b> visualization libraries FM	semi-automatic
Gabillon et al. [30] 2015	Derivation Process	GUI-Dashboards	Component based	<b>Models:</b> UI Feature model (UI features, context features) <b>Components:</b> FM implementation (assets)	Automatic
Sottet et al. [28] 2015	Variability management Derivation Process	Not specified	Model -based	<b>Models:</b> IFM (Interaction Flow Model), Domain model, CUI model, ISM (Implementation Specific Model), <b>Variability models:</b> models FM, mappings FM and transformations model	Semi-Automatic

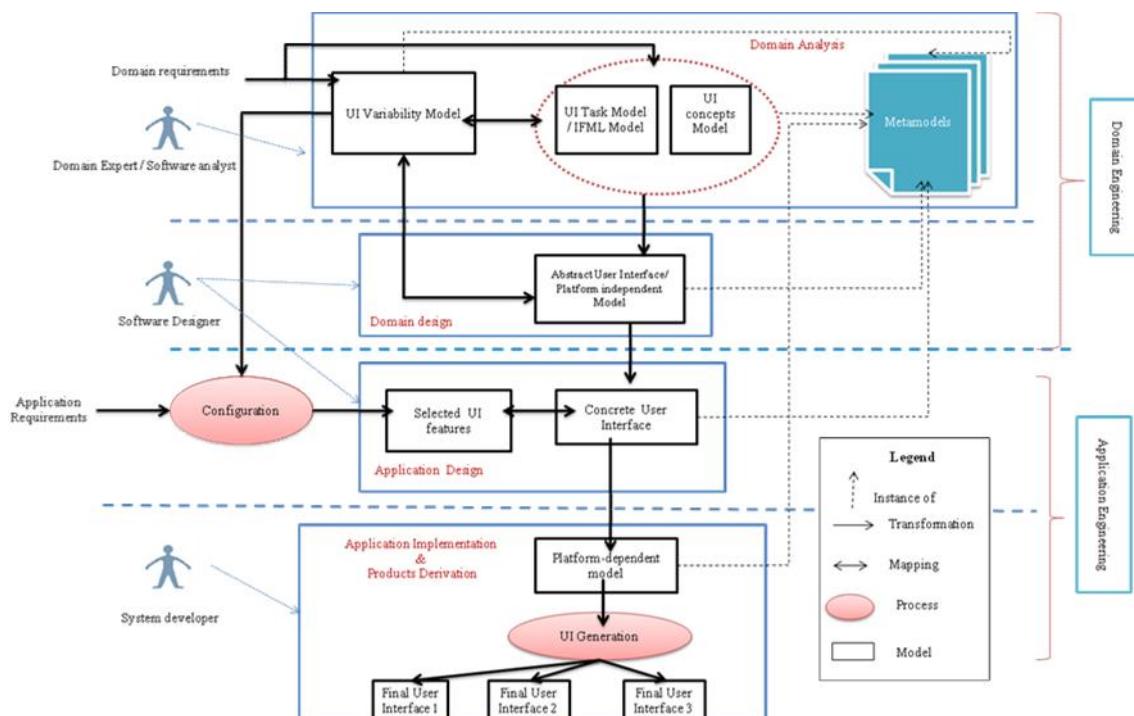


Figure5. General MD-SPL process for UI development

## VI. MD-SPL approaches- the Common Process

This section summarizes proposals described in Table2 in the form of a common schema (Figure 5). The schema combines the abstraction capability of MDE and the variability management capability of SPLE. In the following we describe the schema levels, common design elements and transformations connecting them.

#### A. A two-level MD-SPL process

In Figure 5, a process which highlights the traceability between MDE, MBUID and SPL paradigms was proposed. To make the summary schema easier for reading and editing, a two-layered process is defined; the domain and application engineering levels of the SPL process are employed. Each stage involves the common (common to most approaches) and appropriate design elements (e.g. MBUID, MDE or an SPL concept) which were logically distributed throughout the levels. Levels as well as design elements are connected through transformations.

The Domain Engineering level defines two sublevels; one reserved for the UI-family analysis and the other for UI-family design. For the analysis level, the following are presented:

- UI concepts using the “UI concepts model”: This model results from domain/requirements engineering. In UML, this model is represented by a class diagram grouping UI concepts;
- User interactions using the “task model” or the Interaction Flow Modeling Language (IFML) model. The task model describes the interactive tasks as viewed by the end user interacting with the UI. The Concur Task Trees (CTT) [33] is a visual notation for describing task models. The IFML [32] model is another type of model describing user interaction, and control behavior of the UI. This model has its own graphical notation described in the IFML specification document [32];
- Commonalities and variabilities of UI variants using the variability model. In most approaches the used variability model was the FM. Another variability model may be used; the decision model. A decision model serves to 1) document and plan variability in the development phase 2) guide users and automate product configuration during derivation phase. The decision model was used in Arboleda et al. [29] for product derivation;
- Meta-models: These define meta-data describing all models of the two engineering levels.

For the domain design, this sub-level presents one model:

- The AUI model (in MBUID-based approach) or the Platform Independent Model (in the MDE-based approach). Those models may describe the structure of the UI (i.e. the composition of a UI in terms of widgets and the relation between them) and also may represent the family-system architecture (the implementation of the variability model). Those models may be expressed using DSLs [31] such as UsiXML used to describe the AUI;

For the application engineering level, this stage is reserved for the design, development and generation of a specific UI. The application engineering level defines two sub-levels: the application design sub-level and the application implementation/product derivation sub-level. For the application design level, the following are defined:

- “Selected UI features”: This design element is an input data flow rather than a conceptual model. This flow is the set of selected features describing a specific UI;
- “CUI model”: This is a concrete AUI model relative to a particular UI modality. The CUI represents, also, the product architecture; meaning assets which correspond to selected features;

- The platform dependent model: At the lowest level, the CUI is implemented according to a specific implementation technology ;
- The final user interface: This is the interactive UI (the generated code).

#### B. Actors

The common process supports four types of actors, they include:

- Domain Expert: He has a deep knowledge of the domain; after expertise training, he delivers the glossary of UI terms to the software analyst;
- Software Analyst: After knowledge acquisition from the domain expert, the analyst defines functional and non-functional specifications and prepares conceptual models (UI variability model, UI concepts model...) which correspond to the specifications;
- Software Designer: This individual elaborates software architecture for the family-system (AUI/PIM) and for a the application (CUI);
- Software developer: This person is concerned with the implementation of product architecture for a specific technology and the generation of the final UI.
- Final Users: these are the people who have a stake or interest in the use of derived user interfaces. They are invited by the analyst to specify their requirements. In figure 5, final users are represented by their requirements (domain requirements/application requirements).

#### C. Transformations

Defined levels as well as employed design elements are connected by means of transformations. Transformation is an MDE concept which transforms, incrementally, models at a high level of abstraction into Platform Specific Models. A transformation is defined by rules which are written in a transformation language. Transformation may be automated (performed by the computer autonomously), semi-automated (requiring human intervention) or manually (performed by humans).

The different transformations of the common process are described in the following:

- 1) Requirements → domain analysis models/configuration: In the domain analysis level, domain experts use the stakeholder's "domain requirements" to design the feature model, the "UI concepts" model and also the "task model" or the "IFML model". In the application design level, "Application requirements" are used to configure the variability model. This transformation is done manually even if requirements are presented in an informal way (interview with final user and a tender specification);
- 2) <UI concepts model, task model/IFML model> → AUI: As was defined in the CRF platform, the AUI is generated using the "UI task model" and "UI concepts". However this transformation is not fixed, otherwise, the task model may be replaced by the "IFML model" as is the case in Sottet et al. [28] or the AUI can be generated only using the task model [20]. This transformation may be realized automatically using the model transformation techniques [45];
- 3) Feature model ↔ (task model/IFML model, UI concepts model): This transformation is a mapping [41] which highlights the correspondence feature task and feature UI concept.

Mapping is expressed using a “mapping model” [20]. This mapping ensures the coherence of the AUI;

- 4) “UI variability model”  $\leftrightarrow$  “AUI model/platform independent model”: this mapping associates for each property of the variability model and its correspondent in the abstract models;
- 5) “UI variability model”  $\rightarrow$  “selected UI features”: This transformation consists in selecting the right set of features and dropping the rest. This transformation is supported by many tools however it is a semi-automated transformation because it requires user intervention to select the appropriate features;
- 6) <selected UI features, AUI>  $\rightarrow$  CUI: To design the application, the application architecture has to be elaborated. This is defined by instantiating the AUI (or the platform independent model) model according to selected features;
- 7) CUI  $\rightarrow$  platform-dependent model: This transformation adapts the CUI model to the implementation technologies as is the case in Sottet et al. [28]
- 8) CUI  $\rightarrow$  FUI (product generation): A process which leads to final UI generation or simply a transformation process that may have many inputs and outputs of various types.

## VII. Validation (Figure 6 and 7)

In this section, the common process described above is used to position the proposals of Table 1. This positioning is performed by selecting design elements of the common process describing each approach. The positioning of SPL proposals is done using two schemas (in need of clarity); in Figure6, MB-SPL approaches are tagged (approaches that use MBUID core assets to implement the SPL process) and in Figure7, MD-SPL approaches are tagged (approaches using MDE core assets to implement the SPL process). All proposals described in Table 2 were tagged except the [30] proposal because it is a component-based proposal and we are only interested in model-based proposals.

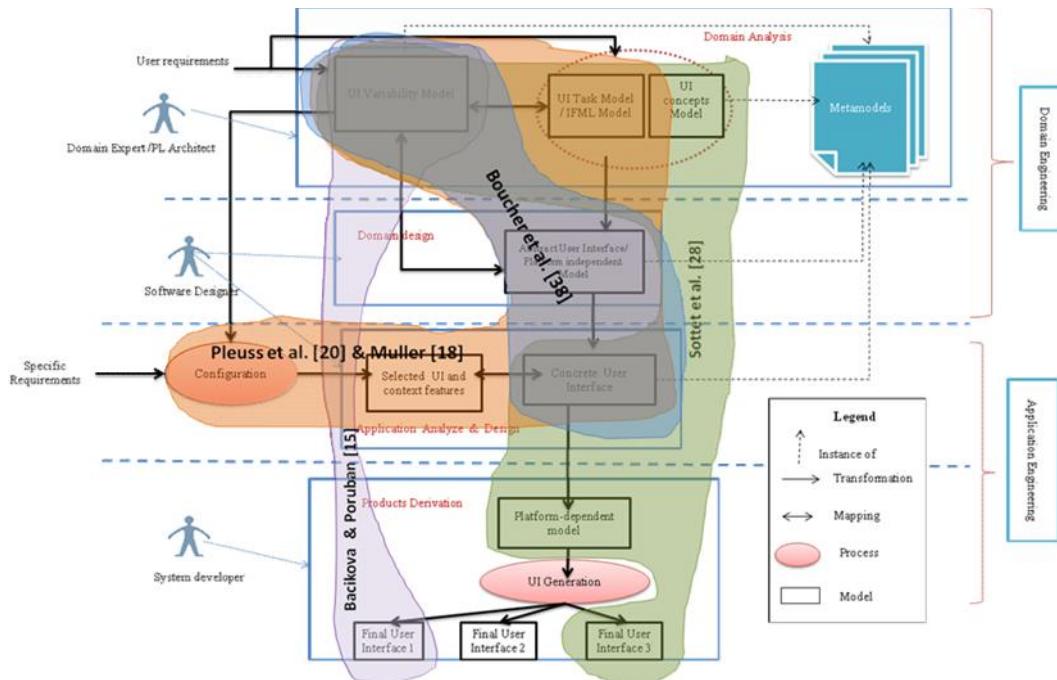


Figure 4: Tagging of MB-SPL proposals

#### A. Tagging of model based approaches

In Figure 6, Sottet et al. [28] use MBUID core assets for which they apply an update. So, they use the “domain model” which corresponds to the “UI concepts model” (in Figure 6), the “IFM model” instead of the task model. Then, they use the CUI model to define the pages of the application, their widgets as well as the navigation between pages of and the ISM (Implementation Specific Model); an additional model that implement the CUI according to the specifications of implementation technologies and to facilitate the code generation (the derivation of the final UI). In figure 6, the ISM corresponds to the platform specific platform. Concerning variability managing, authors use a multiple feature models (to describe the various facets of UI variability).

Boucher et al. [38], proposed an MBUID process to manage variability and generate a configuration UI. To reach this goal, they use the variability model to manage the variability of interfaces; then, they use the AUI model to implement the feature model. And from the AUI, they generate the CUI model.

For Pleuss et al. [20] and Muller [18], they use most MBUID concepts to implement the SPL approach; they use the feature model, the task model and the AUI model to define the domain engineering level. Then, they instantiates these models to define the product architecture which corresponds to CUI model in figure 5. The adding of Muller (compared to Pleuss et al.) is the consideration for UI layout aspect.

The Bacikova and Poruban [15] proposal is not a model base approach, however it concerns variability management. Then, their proposal include the “UI variability model”.

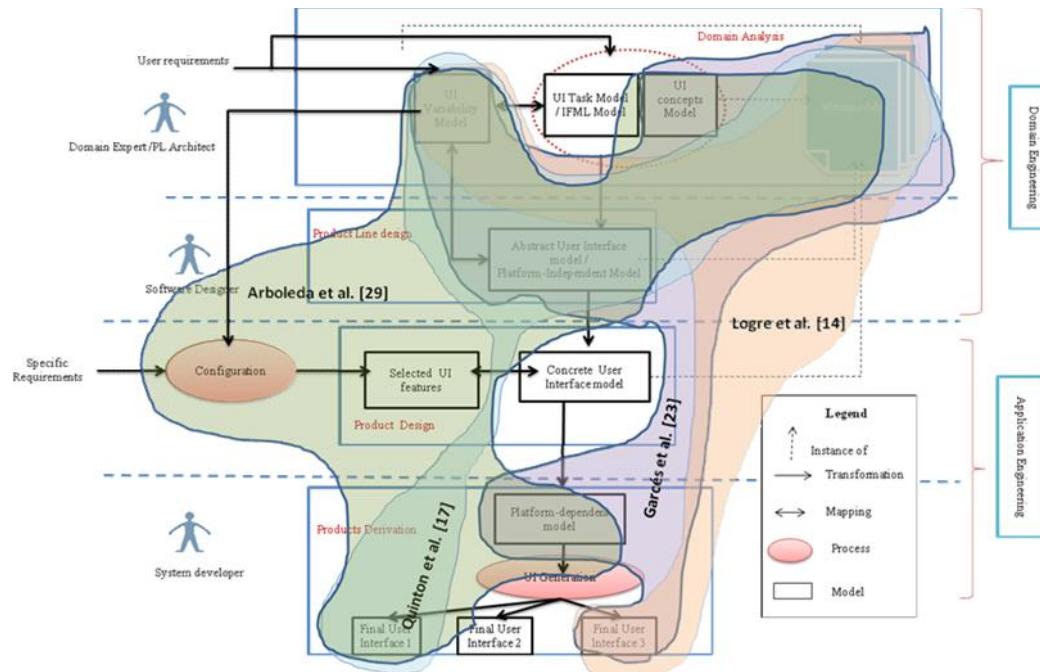


Figure 4: Tagging of MD-SPL proposals

#### B. Tagging of model driven approaches

In Figure 7, Logre et al. [14] use a meta-model to define concepts of dashboards, from this metamodel, they generate the feature model that depicts the technological variants of dashboards. Then, from the FM, dashboards are generated.

For Garcés et al. [23], they use metamodels to define business and architecture aspects of GUI and feature models to describe the variability of GUI from the same aspects. The code is generated from the java model, the result of the mapping between the architecture feature model and the architecture model. Arboleda et al. [29], uses MDE core assets (metamodel, models and transformation). Metamodels to define concepts of problem and solution space, models to implement the application level and transformations to derivate the final product. The variability model is also used to manage the variability of the system and to derivate the final product, authors use the decision model.

In Quinton et al. [17], authors use the feature model to manage the functional and the technological variabilities, they use model to define features assets and to manage the gap between the functional variability and the technological variability, they use a metamodel (which describes device concepts), for which they verify if the generated interface is conform to the targeted device.

#### VIII. Future Challenges

The evaluation of SLR findings and the positioning of approaches on the common schema were very useful for highlighting the shortcomings of existing proposals. In this section we describe the challenges that we will take up in future works.

**Challenge 1 - A context sensitive MD-SPL process:** we note that all approaches have not considered the context of use during the UI development process. The consideration of the context during the design time and/or the run time has become an important fact that ensures UI usability [9][11] and adaptability [4] [6] seen the increase of the number of devices, the development platforms and the user profiles. Pleuss et al. [21] took into account the context in their proposal. In [21], authors considered context as additional models used in the transformation of UI models. So, the user influences the AUI and the environment and the platform influence the CUI. However and into the case of an SPL approach, it is necessary to consider the context of use when configuring the product [44]. For that, the first challenge which is not yet met within a UI MD-SPL proposal is the consideration of the context using feature modeling;

**Challenge 2- Industrial Product:** we note that existing UI MD-SPL approaches are still in the stage of prototyping. However, we think that packaging these approaches as general-purpose products could increase their usefulness for real-life projects in the same way that existing commercial tools are useful for developing traditional UIs. The literature already offers several approaches and prototypes. Therefore, it might be the appropriate time for UI MD-SPL approaches for a joint venture between academics working on UIs MD-SPL and industrial partners with a real interest in adopting this approach for developing commercial applications.

## IX. Conclusion

In this paper, the generative approaches used to develop UIs have been reviewed. These include MDE-based approaches, the MBUID-based approaches and the SPL approaches. For MDE and MBUID (a particular MDE approach inspired from MDA process and dedicated for the development of UIs) approaches, the most important surveys that deal with the development of UIs were mentioned. After that, the focus was given to SPL approaches and a short SLR which identifies UI-SPL proposals was conducted. The findings of SLR were compared according to a set of criteria. Then, we proposed a summary schema on which we have tried to tag all approaches to validate it. In the last section, two future challenges were identified to manage the gap of existing approaches.

## REFERENCES

- [1] K. Pohl, G. Boeckle, and F. van der Linden, "Software Product Line Engineering". New York, NY: Springer, 2005
- [2] P. Clements, L. Northrop, "Software Product Lines". Addison-Wesley Boston, Boston, 2002
- [3] J. Van den Bergh and K. Coninx, "Using UML 2.0 and profiles for modeling context-sensitive User Interfaces", 2005.
- [4] J. M. González Calleros et al., "Model-Based UI XG Final Report". Available at: <http://www.w3.org/2005/Incubator/model-based-ui/XGR-mbui-20100504/>, 2010
- [5] G. MEIXNER, F. PATERNÒ, AND J. VANDERDONCKT, "Past, Present, and Future of Model-Based User Interface Development". i-com, 10(3), pp.2–11, 2011
- [6] G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, J. Vanderdonckt, "A Unifying Reference Framework for Multi-Target User Interfaces". Interacting with Computers, Vol. 15, num. 3, 289–308, 2003
- [7] Q. Limbourg, J. Vanderdonckt, B. Michotte, L. Bouillon, and V. López-Jaquero, "USIXML: a language supporting multi-path development of user interfaces". In International Workshop on Design, Specification, and Verification of Interactive Systems pp. 200-220. Springer Berlin Heidelberg, July 2004

- [8] J. Vanderdonckt, Q. Limbourg, B. Michotte, L. Bouillon, D. Trevisan, F. Mureille, "USIXML: a User Interface Description Language for Specifying Multimodal User Interfaces." W3C Workshop on Multimodal Interaction. Sophia Antipolis, 2004
- [9] J. Nielsen, "Usability Engineering". Academic Press, Boston, 1993
- [10] <https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/context-of-use>
- [11] T. Baccino, C. Bellino, T. Colombi, "Mesure de l'utilisabilité des Interfaces", Hermès Science Publisher: Paris 2005.
- [12] H. Gomaa, "Software Modeling and Design", Cambridge University Press, 2011
- [13] B. Kovas, F. Gaunet, X. Briffault, « les techniques d'analyse de l'activité pour l'IHM », Lavoisier,
- [14] I. Logre, S. Mosser, P. Collet, M. Riveilli, "Sensor data visualisation : a composition-based approach to support domain variability", In European Conference on Modelling Foundations and Applications (pp. 101-116). Springer International Publishing. 2014.
- [15] M. Bacikova, J. Poruban, Analyzing stereotypes of creating graphical user interfaces, Central European Journal of Computer Science vol. 2 num. 3, pp. 300-315, 2012
- [16] R. Silveira, Abstraction in Software Development. Available at:<http://legacy.rodrigo-silveira.com/abstraction-software-development/>, 2014
- [17] C. Quinton, S. Mosser, C. Parra, L. Duchien, Using Multiple Feature Models to Design Applications for Mobile Phones. MAPLE / SCALE workshop, colocated with SPLC'11, Aug 2011, Munich, Germany. pp.1-8, 2011
- [18] J. Muller, Generating graphical user interfaces for software product lines: A constraint-based approach. In 15. Interuniversita Doktorandenseminar Wirtschaftsinformatik, 2011
- [19] Ch. Lutteroth, G. Weber, Modular Specification of GUI Layout Using Constraints. In: Proceedings of the 19th Australian Conference on Software Engineering. Washington, DC, USA : IEEE Comp. Soc., 2008, pp. 300–309
- [20] A. Pleuss, G. Botterweck, D. Dhungana, Integrating automated product derivation and individual user interface design. In VAMOS'10, 2010.
- [21] A. Pleuss, S. Wollny and G. Botterweck, Model-driven development and evolution of customized user interfaces. In Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive computing systems, EICS '13, New York, NY, USA, 2013
- [22] J. Bézivin, and P. A Muller, UML: The birth and rise of a standard modeling notation. In International Conference on the Unified Modeling Language, pp. 1-8, Springer Berlin Heidelberg, June 1998
- [23] K. Garces, C. Parra, H. Arboleda, A. Yie and R. Casallas, "Variability management in a model-driven software product line," Rev. Av. en Sist. e Informatica, vol.4,no.2,2007.
- [24] P. A Akiki, A. K Bandara and Yu, Y. (20. Adaptive model-driven user interface development systems. ACM Computing Surveys, 47(1), 2015.
- [25] J. L PÉREZ-MEDINA, S. DUPUY-CHESSA, AND A. FRONT, "A Survey of Model Driven Engineering Tools for User Interface Design". In M. Winckler, H. Johnson, and P. Palanque, eds. Task Models and Diagrams for User Interface Design. Springer, pp. 84–97, 2007.
- [26] I. Jaouadi, R. Ben Djemaa and H. Ben-Abdallah, "Interactive Systems Adaptation Approaches: A Survey". In Proceedings of the 7th International Conference on Advances in Computer-Human Interactions ACHI (pp. 127-131), 2014.
- [27] "State of the art of User Interface Description Languages", Thales group. Available: <https://itea3.org/project/workpackage/document/download/468/08026-UsiXML-WP-1-D11v2-StateoftheArtofUIDL.doc>, 2011.
- [28] J. S Sotter, A. Vagner and A. G. Frey. "Model Transformation Configuration and Variability Management for User Interface Design". In International Conference on Model-Driven Engineering and Software Development (pp. 390-404). Springer International Publishing, (2015, February).
- [29] H. Arboleda, A. Romero, R. Casallas and J. C. Royer, "Product Derivation in a Model-Driven Software Product Line using Decision Models". In CIBSE (pp. 59-72). (2009, April).
- [30] Y. Gabillon, N. Biri and B. Otjacques, "Designing an adaptive user interface according to software product line engineering". Proc. ACHI, 15, 86-91. 2015.
- [31] K. Czarnecki, "Overview of generative software development", In Unconventional Programming Paradigms (pp. 326-341). Springer Berlin Heidelberg. 2005.

- [32] “IFML: The Interaction Flow Modeling Language”. Available at: <http://www.ifml.org>, 2013
- [33] F. Paternò, C. Santoro, L. D. “Spano, Concur Task Trees (CTT)”. Available at: <https://www.w3.org/2012/02/ctt/>, 2012
- [34] JS. Sotter, G. Calvary, JM. Favre, J. Coutaz, A. Demeure, L. Balme: Towards Model-Driven Engineering of Plastic User Interfaces. In Conference on Model Driven Engineering Languages and Systems (MoDELS’05) satellite proceedings, Springer LNCS, pp. 191—200, 2005
- [35] D. C. Schmidt, Model-driven engineering. COMPUTER-IEEE COMPUTER SOCIETY-, 39(2), 25, 2006
- [36] S. Apel, and C. Kästner, An Overview of Feature-Oriented Software Development. Journal of Object Technology, 8(5), 49-84, 2009
- [37] R. Landry, N. Becheikh, N. Amara, S. Ziam, O. Idrissi, and Y. Castonguay, La recherche, comment s'y retrouver? revue systématique des écrits sur le transfert de connaissances en éducation, 2008
- [38] Q. Boucher, G. Perrouin, and P. Heymans, Deriving configuration interfaces from feature models: A vision paper. In Proceedings of the Sixth International Workshop on Variability Modeling of Software-Intensive Systems (pp. 37-44). ACM, January 2012
- [39] MDA, Model Driven Architecture, <http://www.omg.org/mda>
- [40] T. Rentsch, Object oriented programming. ACM Sigplan Notices, 17(9), 51-57, 1982
- [41] K. Czarnecki, and M. Antkiewicz, Mapping features to models: A template approach based on superimposed variants. In International conference on generative programming and component engineering (pp. 422-437). Springer Berlin Heidelberg, September 2005
- [42] M. Schlee, and J. Vanderdonckt, Generative programming of graphical user interfaces. In Proceedings of the working conference on Advanced visual interfaces (pp. 403-406). ACM, May 2004
- [43] N. Anquetil, B. Grammel, I. Galvao Lourenco da Silva, J. A. R. Noppen, S. Shakil Khan, H. Arboleda, and A. Garcia, Traceability for model driven, software product line engineering, 2008
- [44] M. Acher, P. Collet, F. Fleurey, P. Lahire, S. Moisan and J.P Rigault, “Modeling context and dynamic adaptations with feature models”. In 4th International Workshop Models@ run. time at Models 2009 (MRT’09), p. 10, October 2009.
- [45] K. Czarnecki, and S. Helsen, Classification of model transformation approaches. In Proceedings of the 2nd OOPSLA Workshop on Generative Techniques in the Context of the Model Driven Architecture, Vol. 45, No. 3, pp. 1-17, October 2003

## Energy Aware Resource Management for Cloud Data Centers

Santosh Kumar, Manish Pandey  
Maulana Azad National Institute of Technology, Bhopal, India, 462003

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

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

### 1. INTRODUCTION

Cloud computing technology provides infrastructure, software, and platform as subscription based services in a pay-as-you-go model to consumers over the internet. The services provided by cloud computing are referred to as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). It allows organizations to focus on innovations rather than to set up a high cost infrastructure and their maintenance. The rapid increase in use of Cloud computing leads to establishment of large scale data centers, around the world containing thousands of computing nodes. However, Cloud data centers absorb huge amount of electrical energy resulting in high operating expenditure and carbon dioxide (CO<sub>2</sub>) emissions to the environment. Energy consumption by data centers worldwide has risen by 56% from 2005 to 2010, and in 2010 it was between 1.1% and 1.5% of the total electricity used [1]. Furthermore, carbon dioxide outflows by the ICT industry are recently estimated to be 2% of the global discharge of CO<sub>2</sub>, which is equivalent to the emissions of the aircraft industry [2] and significantly contributes to the greenhouse effect. Energy consumption by data centers is predicted to increase around 140 billion kilowatt-hours per year by 2020, which is equivalent to approximately production of 50 power plants per year [3]. One of the sources of energy waste is inefficient usage of resources of data centers. The problem of server under utilizations worsened by narrow power ranges of servers, even energy consumption of idle servers is up to 70% of their peak power [4]. The main concern of data centers deployments has been high performance and this demand has been achieved without giving careful attention to energy consumption. In view of increasing energy costs in data centers, data center resource management should focus not only on performance optimization but also on energy efficiency while fulfilling quality requirements.

Hence there is a need of improvement in existing resource allocation and management algorithm in cloud data centers as well as proposing new ones. One of the solutions of energy inefficiency in cloud data centers is virtualization technology [5]. Using virtualization technology cloud providers can create multiple instances of virtual machine (VMs) on a single physical machine which increases resource utilization. Switching idle nodes to sleep mode eliminate idle power consumption hence reduces energy consumption. Live migration can be used to dynamically consolidate VMs on the minimum number of physical nodes according to their current resource need

and hence putting more number of physical nodes in sleep mode. However this leads to a challenge of energy efficiency in cloud computing of minimize both resource usage and SLA violations. There are two extreme states of a host overutilization and under-utilization. Under-utilization is related to energy saving while over-utilization is related to SLA violation. Under-utilization of a host leads to waste of energy, dynamic VM consolidation is necessary when a host is underutilized to minimize energy waste. The goal of optimizing energy consumption can be achieved by placing VMs on the minimal number of hosts. Aggressive consolidation of VMs may however increase the possibility of over-utilization. Over-utilization is often caused by highly variable as well as aggressive VM consolidation. It can often cause performance degradation and lead to SLA violations. Live VM migration is often necessary when over-utilization occurs. Therefore, both over-utilization and under-utilization should be avoided and mitigated by employing live VM migration. However VM migration can also cause performance degradation [6]. It is therefore extremely challenging for cloud service providers to maintain a balance between energy efficiency and SLA requirements established between consumers and broker.

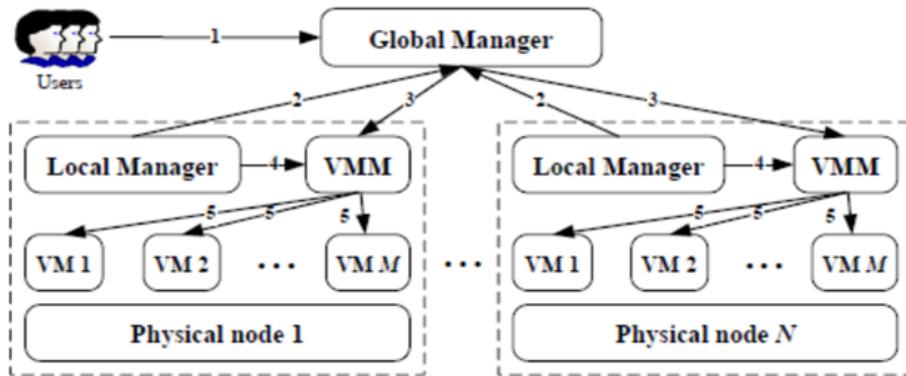


Figure 1.1 System model

This paper discusses system of an IaaS environment (Figure 1.1), which presents a large-scale data center consisting of  $N$  heterogeneous physical nodes. Characteristics of each host is defined by CPU performance in Millions Instructions Per Second (MIPS), amount of RAM and network bandwidth. The servers do not have local disks, the storage is provided as a Network Attached Storage (NAS) to enable live migration of VMs. The type of the environment implies no knowledge of application workloads and time for which VMs are provisioned. Different independent users demand for provisioning of  $M$  heterogeneous VMs according to their requirements to processing power defined in MIPS, amount of RAM and network bandwidth. The way that the VMs are overseen by independent users infers that the subsequent workload made because of joining different VMs on a single physical host is mixed. The mixed workload is formed by various types of applications, such as HPC and web-applications, which utilize the resources simultaneously. The users and resource providers establish a service level agreement to formalize the QoS delivered. In cases of SLA violations the service providers have to pay a penalty to the users.

The software layer of the system comprises of local and global managers. The local managers reside on each host as a module of the virtual machine manager (VMM). Local managers continuously monitor the host's CPU utilization, resizing the VMs based on their resource requirements, and deciding when and which VMs should be migrated from the host. The global manager resides on the master node and it maintains the overall view of the resource utilization by collecting information of all hosts from the local managers. The global manager issues commands to optimize the VM placement. Changes in power modes, resizing VMs and migration of VMs are the responsibility of VMM.

## 2. RELATED WORKS

One of the first works, in which energy efficiency has been considered in the context of virtualized data centers, has been proposed by Nathuji and Schwan [7]. They have proposed an infrastructure for cloud data center's resource management systems. The resource management has been divided into two parts local and global policies. At the local level local managers influences the guest OS's power management strategies. All the local managers give information of the current resource allocation to the global resource manager so that it can apply its policy to decide

whether there is a need of VM reallocation. However, they have not proposed a specific policy for adaptive resource management at the global level.

Kusic et al. [8] have characterized the problem of energy efficiency in virtualized heterogeneous data centers as a sequential optimization and acknowledged it with limited Lookahead control (LLC). The goal is to minimize both energy consumption and SLA violation and hence maximize profit of resources provider. Kalman filter is applied to estimate the number of future applications to predict the future state of the system and make the necessary reallocations. However, in contrast to approaches based on heuristics, the proposed model requires a regimen based on the simulation of specific application settings learning, which is very difficult to implement by the infrastructure as a service cloud provider, such as Amazon EC2.

Srikantaiah et al. [9] have exposed some concerns raised due to performing consolidation to optimize energy usage while satisfying required performance in heterogeneous cloud data centers. They have shown how resource utilization, energy usage and performance as multiple workloads with different resource usage are combined on common servers. They have found that the energy consumption per transaction results in a "U"-shaped curve, and it is possible to determine the optimal utilization point. To tackle the issue of optimization over multiple resources, they have proposed an algorithm to solve the multidimensional bin packing problem for the workload consolidation. However, the proposed algorithm is workload specific and application dependent.

Verma et al. [10] have defined the problem of energy-aware dynamic placement of applications in virtualized heterogeneous data centers as a continuous optimization. The VM placement is optimized in every time frame such that energy consumption can be minimized and performance can be maximized. As in [9], the authors have used a heuristic for the bin packing problem with variable bin sizes and costs.

Raghavendra et al. [11] have explored the issue of energy management in cloud data centers by joining and organizing five different force management approaches. The authors investigated the issue in terms of control theory and applied an input control loop to facilitate the controllers' activities. Like most of the previous works, the system deals only with the CPU management. The authors have found an important result of the experiments that depending on the workload the actual energy savings can vary but the benefits from coordination are subjectively comparative for all classes of workloads. However, the system does not support strict SLAs as well as variable SLAs for different kind of applications. This is suitable for enterprise environments, but not for Cloud computing providers, as more comprehensive support for SLAs is required in cloud data centers.

Cardosa et al. [12] have proposed a methodology for the issue of energy efficient allocation of VMs in virtualized heterogeneous cloud data centers. They have taken advantage of the inherent features min, max and share in virtualization technology these are minimum, maximum and proportion of the CPU allocated to VMs which share the same resource respectively. Their methodology does not support strict SLAs and to define the shares parameter it requires the knowledge of application priorities so it is suitable for only enterprise environments. Other limitations are that the allocation of VMs is static and only CPU is considered as resource during the VM reallocation.

Kumar et al. [13] have proposed a methodology for dynamic VM consolidation based on a prediction of stability, the likely hood that a proposed VM reallocation will remain effective for some time in the future. Predictions of future resource demands of applications are done using a time-varying probability density function. The problem is that the authors assume that the parameters of the distribution, such as the mean and standard deviation, are known a priori. They assume that these values can be obtained using offline profiling of applications and online calibration. However, offline profiling is unrealistic for IaaS environments. Moreover, the authors assume that the resource utilization follows a normal distribution, whereas numerous studies [14], [15] have shown that resource utilizations by applications are more complex and it is difficult to model using simple probability distributions.

Gandhi et al. [16] have focuses on distribution of available power among servers in a virtual heterogeneous data center to get high performance. To find the impact of different factors on mean response time, a queuing theoretic model has been proposed, which provides the prediction of the mean response time as a function of the power-to-frequency relationship, arrival rate, peak power budget, etc. The model is used to find the optimal power distribution for each configuration of these factors.

Beloglazov Buyya. [17] Proposed some novel techniques for the online detection of overloaded and underloaded host and adaptive adjustment of the utilization thresholds on the basis of statistical analysis of previously collected data of utilization of resources by VMs throughout their lifetime. Proposed techniques determine the value of upper utilization threshold based on strength of deviation of the CPU utilization. Higher deviation implies higher probability that the CPU utilization will reach 100% and cause an SLA violation. Few statistical methods are used to calculate the upper CPU utilization threshold. By combining these statistical methods to determine over-utilized and underutilized hosts with policies to select VMs to be migrated, various strategies can be formed. The destination hosts where VMs will be placed chosen in such a way that energy consumption is

minimized. Median Absolute Deviation (MAD), Local Regression (LR), Interquartile Range (IQR) and Local Regression Robust (LRR) are four techniques proposed by authors to determine adaptive upper threshold.

This paper extended the work of Beloglazov and Buyya [17] and proposes a technique for auto adjustment of the upper utilization threshold.

### 3. PROPOSED WORK

Dynamic consolidation of virtual machines can be divided into three parts. (1) Deciding when to consider a host is over-utilized requiring migration of one or more VMs. (2) Selecting VMs that should be migrated from an over-utilized host. (3) Finding a new placement of the VMs selected for migration from the over-utilized and under-utilized hosts. Algorithm for optimization of VM placement is given in [17] as Algorithm 1.

Table 3.1: VM placement optimization Algorithm

---

#### Algorithm 1

---

Input: HostList   Output: MigrationMap

1. For each Host in HostList do
  2.     If Host is overloaded then
  3.         Get VmsToMigrate from Overloaded Host
  4.         Get new VM placement for VmsToMigrate
  5.         Add new VM Placement in MigrationMap
  6.         Clear VmsToMigrate
  7.     For each Host in HostList do
  8.         If Host is underloaded then
  9.             Get VmList from underloaded Host
  10.          Get new VM placement for VmList
  11.          Add new VM placement in MigrationMap
  12.     Return MigrationMap
- 

#### 3.1. Host over utilization detection

Due to unpredictable and dynamic workloads a static value of utilization threshold is inappropriate. Therefore, Beloglazov and Buyya [17] proposed an auto adjustment technique of utilization threshold based on statistical analysis of previous data which was gathered throughout the lifetime of VMs. The main objective of his heuristic is to adjust the upper utilization threshold considering the deviation of CPU utilization. Higher the deviation, upper utilization threshold value will be lower, as the higher deviation implies more higher probability that the CPU utilization will reach 100% and cause an SLA violation. Four overload detection techniques proposed in [17] are discussed below.

- Median Absolute Deviation (MAD): in this method first absolute deviations from the median of utilization history data of VMs are calculated then median of these deviations is taken. This statistical method is used for determining upper utilization based on deviation strength. The reason behind choosing MAD over standard deviation is that MAD is not heavily influenced by the outliers, so the magnitude of the distances of outliers is irrelevant.
- Interquartile Range (IQR): In this method dataset is divided into quartiles and it is equal to the difference of third and first quartile ( $IQR = Q_3 - Q_1$ ). This statistical method is used to decide the upper utilization threshold.
- Local Regression (LR): It is based on Loess method. This method is used for model a relation between predictor variable and response variable. This method builds a curve to approximate the upper utilization threshold.
- Robust Local Regression (LRR): Since distribution in LR is heavily tailed due to which it is vulnerable at boundaries. To make a robust solution modification has been done by adding the robust estimation method called bisquare which transformed LR into an iterative method.

This paper proposes Median Absolute Deviation of Interquartile Range (MIQR) a technique to adjust upper utilization threshold based on statistical analysis of previous utilization of VMs. In this technique we first find the Interquartile range of utilization history data instead of the total range. Since it breaks the data into quartiles it is a robust statistics hence preferred over total range. After finding the Interquartile range we find the Median Absolute

Deviation of utilization history data of this range as this range has more influence. If the deviation is low then upper threshold value will be high and if the deviation is high then upper threshold value will be low.

Table 3.2: Proposed overload detection method

---

#### MIQR method

---

1. Sort the utilization history data of VMs ( $x_1, x_2, x_3 \dots x_n$ ) in increasing order.
  2. Find two positions Q1 and Q3 as following  
 $Q1 = 0.25 * N$  and  $Q3 = 0.75 * N$   
Where N is the total number of history data
  3. Set median = median of data from Q1 to Q3.
  4. Find deviation of data from median.  
for (i = Q1 to i = Q3)  
deviation[i] = median - data[i]
  5. Find median of deviations from Q1 to Q3.  
 $MIQR = \text{median}(\text{deviation})$   
 $MIQR$  is the median of the absolute deviations from the data's median from Q1 to Q3.
  6.  $T_u = 1 - (t * MIQR)$   
Where  $T_u$  represents upper threshold and t is a tuning parameter.
- 

### 3.2. VM selection:

Next step is selection of VMs for migration from the hosts which are detected as overloaded. Three VM selection policies are proposed by Beloglazov and Buyya [17]. When a VM is selected to migrate by using these policies, overload detection algorithm is again applied on the host. If it is still detected as overloaded then to select another VM to migrate from the host, VM selection policy is applied again. This is repeated until the host is considered as being not overloaded. VM selection policies are mentioned below.

- Minimum Migration Time Policy (MMT): In This policy a VM is selected for migration which requires minimum amount of time to complete its migration, in comparison to other VMs allocated to that host.
- Random Choice Policy (RC): This policy says which VM is selected for migration by using uniformly distributed discrete random variable  $X_d$ . Here  $X_d$  is defined by the formula  $X_d = U(0, |V_i|)$ , where  $V_i$  is a set of VMs allocated to host i.
- Maximum Correlation policy (MC): According to the proposal of Verma et al. [10], the higher the correlation between the resource usage by applications running on an over subscript server, the higher the probability of the server being overloaded. This idea is used in selecting VMs which required for migration so that selection of first VM is based on the highest correlation of CPU utilization [17]. To estimate correlation, multiple correlation coefficients were applied.

### 3.3. VM placement

After selecting VMs to be migrated new physical nodes are required to place them. These hosts should have sufficient resources for VMs and minimum increase of energy consumption. It is similar to bin packing problem with variable bin sizes and prices. The physical nodes act as the bin, VMs that is to be placed act as the items, available CPU capacities act as bin size and the power consumption by the nodes act as cost. One of the popular algorithms to solve bin packing problem is Best Fit Decreasing (BFD) algorithm. This algorithm uses bins less than  $11/9.OPT + 1$  (where OPT is the number of bins provided by the optimal solution). Beloglazov and Buyya [17] proposed modified BFD known as PABFD (power aware best fit decreasing) algorithm. The algorithm is described in Table 3.3 as Algorithm 2. It first sorts the VMs according to their CPU utilization in decreasing order in order to get proper host where the power consumption raised minimum by examining all the hosts. Finally it allocates the VM to that host.

Table 3.3: Power Aware Best Fit Decreasing (PABFD) Algorithm

Algorithm 2.

---

Input: HostList, VmList	Output: allocation of VMs
-------------------------	---------------------------

1. Sort VmList in decreasing order of utilization.
2. For each VM in VmList do
  3. Set minPower = maxUtilization
  4. Set AllocatedHost = NULL
  5. For each Host in HostList do
    6. If Host has enough resources for VM then
      7. Estimate power of Host after placing the VM
      8. If estimated power is less than minPower then
        9. Set AllocatedHost = Host
        10. Set minPower as estimated power
    11. If AllocatedHost is not NULL
    12. allocation.add(VM, AllocatedHost)
  13. return allocation

## 4. PERFORMANCE EVALUATION

### 4.1. Experimental Setup

For performing experiments physical machine has a 2.0GHz Intel(R) core(TM)2 Duo processor, 3GB RAM, 350 GB hard disk, and 32-bit Operating System Windows 10 Professional. Since, the proposed method is based on IaaS cloud service, it is necessary to estimate the proposed resource allocation method on large data centers of virtualized infrastructures by creating a view of enormous number of computing resources to the users. However, it is very difficult to perform repeatable large-scale experiments on a real infrastructure, which is required to evaluate and compare the proposed strategies. Simulation is one of the ways to evaluate the performance of the proposed method and ensure the repeatability of experiments.

CloudSim toolkit [18] latest version CloudSim 3.0.3 has been taken as a simulation platform, as it is widely used modern simulation framework focused at Cloud computing environments. Using this toolkit a data center has been simulated which has 800 heterogeneous physical nodes. Half of the nodes are HP ProLiant ML110 G4 servers, and the other half are HP ProLiant ML110 G5 servers.

CloudSim is developed by clouds lab in the University of Melbourne. It provides a Java based basic libraries (classes) for specifying datacenter, VM, applications, users, computational resources and policies for scheduling and provisioning. And also provides user friendly GUI and it is better than other simulators like gridsim or gangsim.

In CloudSim tool the cloud environment can be established with datacenter and host machines by using the java classes for each purpose. There are following basic Java classes describe the functionality of CloudSim such as: CloudSim, Datacenter, Datacenter Broker, DatacenterCharacteristics, Host, VM, VMAssignmentPolicySimple, VMAssignmentPolicy, VmScheduler, VmSchedulerTimeShared, VmSchedulerSpaceShared, Cloudlet, Cloudlet Scheduler, Cloudlet SchedulerTimeShared and Cloudlet SchedulerSpaceShared. First of all setting up the cloud environment, datacenter are created using the physical host server having its characteristics and its resources description. We have implemented our proposed work by adding Java classes MyMadiqr, MyVMAssignmentPolicy, in org.cloudbus.cloudsim.power folder and adding the proposed method MIQR in MathUtil class.

### 4.2. Simulation Steps in CloudSim

- Set the number of user.
- Initialization of common variable.
- CIS will be created by using init method.
- Datacenter will be created by using create Datacenter method.  
Ex: Power datacenter datacenter1 = create Datacenter ("Datacenter\_0").  
In this for each datacenter, we create a host with its characteristics.
- Datacenter broker instance will be created.
- Create Instance of virtual machine with PE, RAM and Bandwidth requirement.
- Now this virtual machine is submitted to broker.  
Infrastructure has been developed at this point.

- Cloudlet is created with Bandwidth and MIPS requirement.
- Now this Cloudlet will get submitted to Broker.
- Start Simulation process.
- Stop Simulation process.
- Print the status of the Simulation.

#### 4.3. Performance Metrics

To compare the efficiency of the proposed method two metrics have been used to evaluate their performance. One of the metrics is the total energy consumption by the physical nodes of a datacenter caused by the application workloads and the other metrics is percentage of service level agreement violation (SLAV).

#### 4.4. Workload Data:

For simulations data provided by PlanetLab [34] as a part of the CoMon project which is a monitoring infrastructure of 03.March.2011 and 06 March 2011 has been used. The number of hosts and VM 03.March.2011 are 800 and 1052 respectively. The number of Hosts and VM 06.March.2011 are 800 and 898 respectively. The simulation ran for 24 hours with scheduling interval of 5 minutes.

#### 4.5. Result and Analysis:

To evaluate the performance of the proposed approach, simulation of Non Power Aware (NPA), Dynamic Voltage Frequency Scaling (DVFS) and Median Absolute Deviation with MMT selection policy (MAD\_MMT), Interquartile Range with MMT selection policy (IQR\_MMT) and the proposed method Median Absolute Deviation of Interquartile Range (MIQR) with selection policy MMT has been done. In Table 4.1 simulation result of different methods for 03 March 2011 workload of Planet Lab has been given. It shows that MIQR\_MMT\_1.9 reduces energy consumption by 9.25% and 7.30% in comparison to IQR\_MMT\_1.5 and MAD\_MMT\_2.5 respectively.

Table 4.1: Simulation Results of 03 March 2011 workload

<b>Method</b>	<b>Energy Consumption</b>	<b>SLA Violation (%)</b>
NPA	2410.80 kWh	0
DVFS	803.91 kWh	0
IQR_MMT_1.5	188.86 kWh	9.98
MAD_MMT_2.5	184.88 kWh	10.18
MIQR_MMT_1.9	171.38 kWh	10.79

Comparison of three energy aware methods IQR\_MMT\_1.5, MAD\_MMT\_2.5 and MIQR\_MMT\_1.9 of 03 March 2011 workload of Planet Lab has been shown in the Figure 4.1. It shows the reduction in energy consumption by using MIQR\_MMT\_1.9 with respect to IQR\_MMT\_1.5 and MAD\_MMT\_2.5.

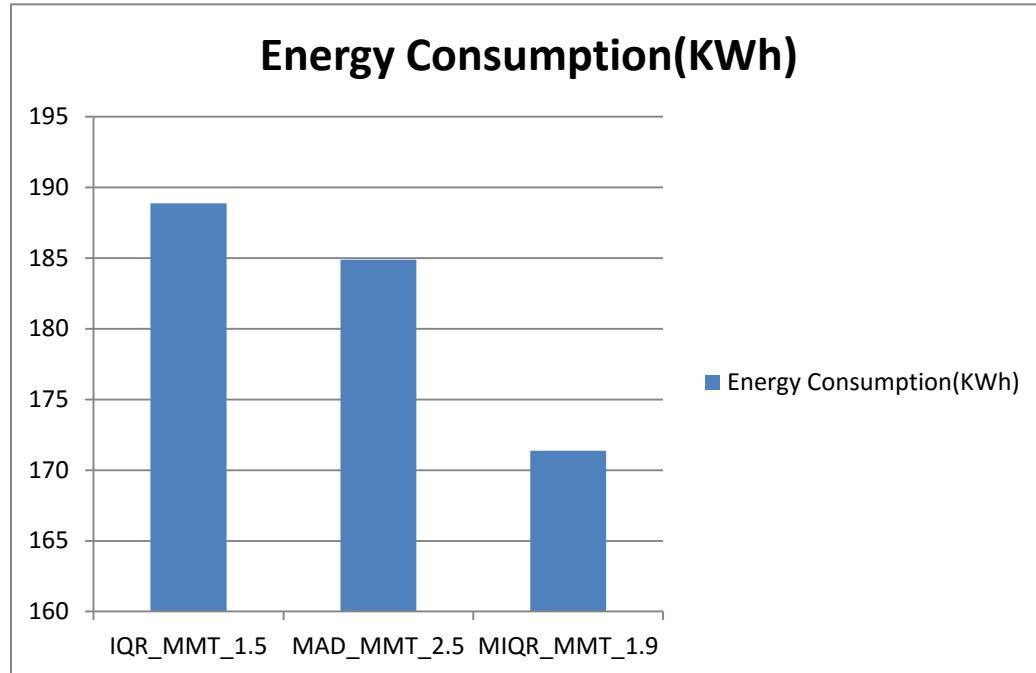


Figure 4.1 comparative analysis of 03 March 2011 workload

In Table 4.2 simulation result of different methods for 06 March 2011 workload of Planet Lab has been given. It shows that MIQR\_MMT\_1.9 reduces energy consumption by 8.96% and 6.85% in comparison to IQR\_MMT\_1.5 and MAD\_MMT\_2.5 respectively.

Table 4.2 Simulation results of 06 March 2011 workload.

Method	Energy Consumption	SLA Violation (%)
NPA	2410.80 kWh	0
DVFS	623.77 kWh	0
IQR_MMT_1.5	144.05 kWh	10.17
MAD_MMT_2.5	141.28 kWh	10.05
MIQR_MMT_1.9	131.60 kWh	10.93

Comparison of three energy aware methods IQR\_MMT\_1.5, MAD\_MMT\_2.5 and MIQR\_MMT\_1.9 of 06 March 2011 workload of Planet Lab has been shown in the Figure 4.2. It shows the reduction in energy consumption by using MIQR\_MMT\_1.9 with respect to IQR\_MMT\_1.5 and MAD\_MMT\_2.5.

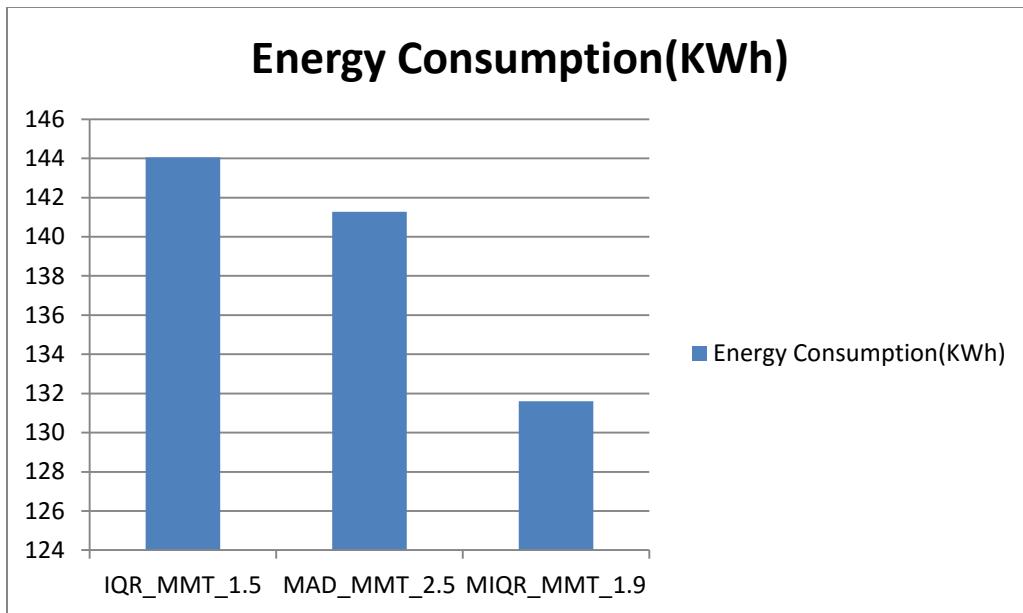


Figure 4.2 comparative analysis of 06 March 2011 workload

## CONCLUSION AND FUTURE SCOPE

Optimizing energy efficiency while satisfying Service Level Agreements (SLA) is one of the major challenges faced by cloud data centers. The study of some statistical techniques which comes under Energy-Efficient VM Consolidation Policies has been done in this work to identify overload and underloaded hosts. The main focus of this work is to propose the adaptive load detection technique to detect the overloaded hosts and adjust the upper threshold so that VMs can be placed on minimum number of hosts with maintaining SLA. The overload detection technique proposed in this work is an adaptive and based on statistical analysis of utilization history data. When MIQR load detection is applied along with MMT VM selection approaches for consolidating VMs, better optimization in energy efficiency is obtained while maintaining service level agreement.

As a future scope, to get more efficient resource provisioning which leads to higher energy efficiency it is important to investigate Cloud application workloads in order to identify common behaviors, patterns, and explore load forecasting approaches. Furthermore, to select which applications will share physical resources it is necessary to develop VM consolidation algorithms that will use the information about the historical workload patterns and application behavior.

## References

1. “Growth in data center electricity use 2005 to 2010,” Analytics Press, Tech. Report 2011.
2. Gartner, Inc., “Gartner estimates ICT industry accounts for 2 percent of global CO2 emissions,” 2007,. Available at: <http://www.gartner.com/it/page.jsp?id=503867>
3. “Scaling up energy efficiency across the data centers” August 2014 IP:14-08-A
4. Fan X, Weber WD, Barroso LA (2007) Power provisioning for a ware house sized computer. Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA 2007), ACM New York, NY, USA, pp 13–23
5. Xen and the art of virtualization. Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP 2003), Bolton Landing, NY, USA.
6. Performance Influence of Live Migration on Multi-Tier Workloads in Virtualization Environments Cloud Computing 2012: The Third International Conference on Cloud Computing, GRIDs, and Virtualization
7. R. Nathuji, K. Schwan, Virtual power: coordinated power management in virtualized enterprise systems, ACM SIGOPS Operating Systems Review 41 (6) (2007) 265–278.

8. D. Kusic, J.O. Kephart, J.E. Hanson, N. Kandasamy, G. Jiang, Power and performance management of virtualized computing environments via lookahead control, *Cluster Computing* 12 (1) (2009) 1–15.
9. Srikanthiah S, Kansal A, Zhao F. Energy aware consolidation for cloud computing. *Cluster Computing* 2009; 12:1– International Conference on Middleware, Springer, 2008, pp. 243–264.
10. A. Verma, P. Ahuja, A. Neogi, pMapper: power and migration cost aware application placement in virtualized systems, in: Proceedings of the 9th ACM/IFIP/USENIX.
11. R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, X. Zhu, No “power” struggles: coordinated multi-level power management for the data center, *SIGARCH Computer Architecture News* 36 (1) (2008) 48–59
12. M. Cardosa, M. Korupolu, A. Singh, Shares and utilities based power consolidation in virtualized server environments, in: Proceedings of the 11th IFIP/IEEE Integrated Network Management, IM 2009, Long Island, NY, USA, 2009
13. Kumar S, Talwar V, Kumar V, Ranganathan P, Schwan K. v Manage: loosely coupled platform and virtualization
14. Management in data centers. Proceedings of the 6th international conference on Autonomic computing (ICAC 2009), Barcelona, Spain, 2009; 127–136.
15. Barford P, Crovella M. Generating representative web workloads for network and server performance evaluation. *ACM SIGMETRICS Performance Evaluation Review* 1998; 26(1):151–160.
16. A. Gandhi, M. Harchol-Balter, R. Das, C. Lefurgy, Optimal power allocation in server farms, in: Proceedings of the 11th International Joint Conference on Measurement and Modeling of Computer Systems, ACM, New York, NY, USA, 2009, pp. 157–168.
17. Anton Beloglazov, Raj kumar Buyya, Energy-aware resource allocation heuristics for efficient management of datacenters for Cloud computing, *Future Generation Computer Systems* 28 (2012) 755–768
18. Calheiros RN, Ranjan R, Beloglazov A, Rose CAFD, Buyya R. CloudSim: a toolkit for modeling and simulation of Cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and Experience* 2011; 41(1):23–50.
19. Park KS, Pai VS. CoMon: a mostly-scalable monitoring system for PlanetLab. *ACM SIGOPS Operating Systems Review* 2006; 40(1):74.

# A survey on association rule hiding methods

Mahsa Roghanian<sup>1</sup>, Mohammad Naderi Dehkordi <sup>1\*</sup>

<sup>1</sup> Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran

\*Corresponding Author

## Abstract

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

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

## 1. Introduction

Data mining is finding useful algorithms from huge amounts of data in databases. In fact, data mining is a process that reveals hidden relationships and dependencies among various items. Mining association rules is one of the methods for discovering relationships among data items in data mining, which was first proposed by Agrawal et al. [1]. Mining association rules in different fields such as market basket analysis, medical, marketing, commerce, and etc. is used for retrieving valuable information. Disclosure of sensitive information in a database is among important problems of this technique which compromises security and confidentiality of this information and causes many problems. Therefore, one of the important issues in this regard is preserving privacy in data mining. In order to preserve privacy policy, sensitive information should become hidden before database is released and the sanitized database that does not have any sensitive association rules should be shared. By making some changes in the original database, hiding algorithms of association rules can prevent the mining of sensitive association rules after the process of data mining. Generally, there are three methods for hiding association

Rules: Heuristic approach, border-based approach and the exact approach. The objective of these algorithms is to hide sensitive association rules by making the least possible changes in the original database such that all sensitive association rules are hidden, no non-sensitive association rule is hidden and no artificial association rule is generated.

In this paper, different methods for knowledge protection are introduced.

The rest of this paper is organized as follows: Hiding association rules are described in section 2. In section 3, some algorithms of this field are studied. A short comparison between the studied algorithms is performed in section 4 and the article is concluded in Section 5.

## 2. Association rule hiding

Let  $I = \{I_1, I_2, \dots, I_n\}$  be a set of items. Let  $D = \{T_1, T_2, \dots, T_n\}$  be a set of transactions, where, each transaction  $T \subseteq D$  consists of a set of items. An association rule is the implication of the form  $X \rightarrow Y$ , where,  $X, Y \subseteq I$  and  $X \cap Y = \emptyset$ .  $X$  and  $Y$  are two desired

itemsets from the database in which  $X$  (left hand side itemsets of the rule) is called antecedent and  $Y$  (right hand side itemsets of the rule) is called consequent. Each association rule has two specific factors. The first factor is the support of the rule. The support of rule  $X \rightarrow Y$  is specified as support ( $X \rightarrow Y$ ) which means at least  $s\%$  of transactions includes  $X \cup Y$  itemset. The second factor is the confidence of the rule and is specified as confidence ( $X \rightarrow Y$ ) and indicates that  $c\%$  of transactions that include  $X$  itemsets, also include  $Y$  itemsets. The support of rule  $X \rightarrow Y$  can be computed by the following equation:

$$\text{Support } (X \rightarrow Y) = \frac{|X \cup Y|}{|D|} \quad (1)$$

Where,  $|X \rightarrow Y|$  is the number of transactions in the database that contains the itemset  $XY$ , and  $|D|$  denotes the number of the transactions in the database  $D$ . The confidence of rule is calculated by following equation:

$$\text{Confidence } (X \rightarrow Y) = \frac{|X \cup Y|}{|X|} \quad (2)$$

Where,  $|X|$  is the number of transactions in database  $D$  that contains itemset  $X$ .

For each database, a minimum support threshold (MST) and a minimum confidence threshold (MCT) are defined by the database owner. The only mined rules were the ones in which  $\text{Support}(X, Y) \geq \text{MST}$  and  $\text{Confidence}(X, Y) \geq \text{MCT}$ . Thus, the issue of exploring association rules could be divided into two phases. In the first phase, frequent itemsets in which their support is higher or equal to the minimum support threshold were mined. In the second phase, based on the minimum support threshold, the strong association rules were mined. A set of association rules mined from  $D$  are introduced as sensitive association rules by the database owner and the owner does not want to disclose them. The objective of hiding algorithms for sensitive association rules is to change the original database of  $D$  to the sanitized database of  $D'$  such that all sensitive association rules will be hidden and the number of missing association rules and artificial association rules will be minimized. The input of these algorithms is:  $D$  original database, minimum support threshold (MST), minimum confidence threshold (MCT), and the set of sensitive association rules that should be hidden. The output included the  $D'$  sanitized database that does not have sensitive association rules and at least one of the following objectives should be realized:

1. The number of missing association rules should be minimized.
2. The number of artificial association rules should be minimized.
3. Degree of modification, which indicates difference between the original database and the sanitized database, should be minimized.

For evaluating performance of algorithms, many evaluation metrics such as execution time, scalability, number of failures in hiding, number of missing association rules, degree of modification, and privacy level should be considered.

### 3. Categorization of different methods for hiding association rules

In order to hide sensitive association rules, either the support of the sensitive items should be reduced below the minimum support threshold or the confidence of the sensitive association rules should be reduced below minimum confidence threshold. Generally, there are three methods for hiding association rules: Heuristic approach, border-based approach and the exact approach. However, other methods such as encryption-based approach, database reconstruction approach and hybrid approach are also proposed [2].

Heuristic approach is one of the most widely used approaches for hiding association rules. Heuristic approach includes greedy and fast algorithms that usually find a good solution close to the best solution in a faster response time. Although heuristic methods do not ensure of providing the optimal solution, they usually provide a solution near one of the best available solutions that have fast response time [3]. Heuristic methods operate on the basis of the following two techniques [4, 5]:

- Data distortion: Data distortion means replacing an attribute value with a new value. This means in binary databases, an item is removed (replaced 1 to 0) or an item is inserted (replaced 0 to 1), so that the support or confidence for association rules are increased or decreased. These algorithms may encounter unpleasant side effects such as missing or artificial association rules.
- Data blocking: In some cases, removing or inserting an item in the database could be dangerous. For example, the technique of data distortion could not be used in medical databases. Therefore, data blocking technique is used in these cases. Blocking means replacing a value with? In an attempt to hide sensitive rules using unknown values.

Based on the heuristic approach the related algorithms can be divided into rule hiding and itemset hiding algorithms [6].

Instead of considering all non-sensitive itemsets during the sanitization process, border-based approach focuses on preserving the border of the frequent non sensitive itemsets. In this approach, sensitive association rules become hidden by changing borders in the frequent and non-frequent items in the original datasets. Itemsets that separate frequent itemsets from the ones that are not frequent are called border items. Algorithms existing in this approach are different in terms of applied methodologies for defining new borders [7, 8].

Exact approach performs the hiding process through formulation of the problem and tries to provide an optimal method for hiding. Algorithms introduced based on this algorithm, examine all available solutions using integer programming in order to obtain the best possible solution. This approach is usually slower in terms of execution time compared to heuristic algorithms. However, algorithms in this category seek to detect solutions for optimal hiding with making the least amount of modification in the database and also to detect solutions with the least amount of side effects on the original database. But this approach may not be applicable for many actual databases [9, 10].

## 4. Literature review

Clifton et al. revealed for the first time that data mining can compromise privacy [11, 12]. Since then, different algorithms have been proposed in this field, which are briefly introduced in the following.

In 1999, Attallah et al. proposed the first heuristic algorithm for hiding sensitive association rules. They could hide sensitive association rules through a lattice-like graph and hiding frequent itemsets by reducing their support [13].

Dasseni et al. provided three heuristic algorithms for single hiding of sensitive association rules, which was based on reducing only the support or the confidence each time and not both [6].

Saygin et al. were among the first researchers to propose utilization of unknown values instead of adding or deleting items in selected transactions for hiding sensitive association rules. Therefore, instead of these values, unknown values such as question mark (?) are inserted in a way that side effects on non-sensitive rules are minimized. A safety margin is defined for realizing this objective [14, 15].

Oliveira et al. were the first researchers to propose the simultaneous approaches for multiple hiding of association rules. These algorithms, called MaxFIA and MinFIA, require twice scanning of the database regardless of the number of sensitive association rules. In the first scan, sensitive transactions are detected and an index is created for them. In the second scan, the original database is modified for hiding sensitive rules by removing the victim item from transactions [16].

Oliveira et al. introduced an algorithm named SWA that requires a single scan of the database regardless of database's size and the number of sensitive rules that should be hidden [17].

Verykios et al. expanded Dasseni's work. They tested algorithms according to different database sizes and different sets of sensitive association rules and proposed two heuristic algorithms of 2.b and 2.c [18].

Pontitakis et al. presented two heuristic algorithms based on data distortion for hiding sensitive association rules. In both of these algorithms, the concept of prioritization is used for selecting transactions for sanitization. These algorithms have a high execution time due to the volume of calculations for obtaining transactions' weights [19].

Lee et al. proposed an innovative technique for hiding sensitive patterns. In their approach, a sanitization matrix is defined. By multiplying the original database that represented as a binary matrix, and the sanitization matrix, a new database, which is sanitized for privacy concern, is gotten. In this article three algorithms are presented to generate of sanitization matrix [20]. Wang et al. (proposed two data mining algorithms of ISL and DSR for hiding sensitive predictive association rules, which none of them need for data mining and selecting of the sensitive rules. In the former, a rule is hidden through reducing its confidence by an increase in the support of itemsets on the left hand side of the rule; while, in the latter, the rule is hidden through reducing the confidence by a decrease in the support of itemsets on the right hand side [21, 22].

Also Wang et al. suggested two algorithms of DCIS and DCDS. The first algorithm increases the support of itemsets on the left hand side of the rule and the second one decreases the support of itemsets on the right hand side of the rule to reducing the confidence of the rule [23].

Amiri introduced three heuristic algorithms for hiding sensitive association rules. In the 'Aggregate' algorithm, sensitive rule support is reduced by deleting some of the transactions. In the 'Disaggregate' algorithm, support is reduced by deleting some of the sensitive items. Finally the third algorithm, referred to as Hybrid, is named Hybrid, uses the 'Aggregate' algorithm to identify transactions that should be sanitized and, then, uses the 'Disaggregate' algorithm to select the items that should be removed [24].

Verykios et al. provided two WSDA and BA algorithms. In the WSDA algorithm, distortion method and in the BA algorithm blocking method was used for hiding sensitive association rules. In the WSDA algorithm, a weight is assigned to each rule

by considering the distance of each rule's confidence from the minimum support threshold. This weight is used for calculating the priority of transactions. Afterwards, transactions are sorted in ascending order based on their priority. In this algorithm, the process of hiding is performed according to reduction of the sensitive rule's support threshold. The BA algorithm sorts transactions that partially or fully support the sensitive rules in ascending order and performs the operation of making transactions unknown. Results showed that WSDA algorithm's execution time is increased by increasing the number of items and that the efficiency of this algorithm is lower in large databases [25].

Weng et al. presented the FHSAR algorithm for completely hiding sensitive association rules with limited side effects. In this algorithm, hiding sensitive rules is performed with only a single scan of the database and thus execution time is reduced. Each transaction is assigned a weight in this algorithm and transactions are sorted in descending order according to their weights. The transaction with the most weight is selected as the victim transaction [26].

Duraiswamy et al. proposed an algorithm SRH (Sensitive Rule Hiding), to hide the sensitive rules through decrease the support of the right hand side of the rule. This algorithm clustering the rules with the right hand side has the common item. It should be added that this algorithm only can hide sensitive rules with single antecedent and consequent with the sensitive item in the consequent [27].

Dehkordi et al. proposed a new method based on genetic algorithm which ensured that no missing or artificial association rules are generated. In this algorithm, before performing the hiding operation, an initial preprocessing is carried out for selecting sensitive transactions and applying modifications on them. In this method, 4 different strategies are used for calculating the fitness function and finally hiding sensitive association rules and sensitive frequent itemsets with the least side effects [28].

Modi et al. introduced a heuristic algorithm named DSRRRC (Decrease Support of R.H.S. item of Rule Clusters) which hides many sensitive association rules at a time while maintaining database quality. In this algorithm, sensitive rules are clustered based on common item on the right hand side of the rules. Then, the rules in each cluster are hidden altogether simultaneously. In this algorithm, the database should be sorted after removing each item. The DSRRRC algorithm is sensitive to the order of transaction in the database. In addition, this algorithm cannot hide rules that have multiple items on its right hand side [29]. Vijayarani et al. proposed Artificial Bee Colony optimization algorithm for hiding the sensitive association rules, which tries to model natural behavior of real honey bees in food foraging. In this algorithm first, initialized the sensitive items and number of modification required for the sensitive items. Then, the nectar amount, fitness function and probability are calculated for finding best food source [30].

Komal Shah et al. suggested two algorithms of ADSRRC and RRLR that both of them are developed to overcome limitations of the DSRRRC. In ADSRRC, rules are clustered based on common right hand side item of the rules; while the sorting of transactions is made only once depending on the sensitivity and length of each transaction. The RRLR algorithm overcomes limitation of hiding rules that have multiple right hand side items. In this algorithm, in order to hide a sensitive rule both the support and confidence are decreased. All steps are similar to ADSRRC algorithm except that in RRLR we are not creating clusters. [4].

Jain et al. proposed an algorithm that uses the data distortion technique where the position of the sensitive items is altered but their supports and the size of the database remains is never changed. It uses the idea of representative rules to prune the rules first and then hides the sensitive rules. This was achieved using the concept of representative association rule, which was introduced by Kryszkiewicz in 1998. Representative association rule is a small subset of association rules using which all sensitive rules could be deduced without accessing the original database. Advantage of this approach is that it hides maximum number of rules [31].

Domadiya et al. presented the MDSRRC algorithm for overcoming the limitations of the DSRRRC algorithm, which can also hide rules that have several items in their left or right hand side. This algorithm calculates the sensitivity of each item and each transaction at first. Afterwards, the number of occurrences for each item in the right side of the sensitive rules is obtained and the item with the most number of repetitions is removed from the transaction with the most sensitivity [32].

Sun et al. provided the first algorithm using the border-based theory, which is known by the name of BBA. The provided approach forms positive and negative item borders at first and tries to preserve the quality of borders because their quality directly influences the quality of the sanitized database [8, 33].

George et al. presented a new algorithm for hiding sensitive knowledge in the context of mining of association rules. The new approach relies on the MaxMin criterion which is a method in decision theory for maximizing the minimum gain and builds upon the border theory of frequent itemsets. In this algorithm ,for each one of the items in the sensitive itemset, the minimum support border itemsets are selected, from which the maximum support border itemset among the minimum ones (the so-called MaxMin itemset) is chosen. The MaxMin itemset indicates the victim item through which the hiding of the sensitive itemset will take place [34].

Menon et al. proposed two algorithms that formulating frequent itemset hiding as an integer program which is an exact approach. The objective here is to increase accuracy.

Divanis et al. suggested an exact method for minimized the distance between the original and sanitized database. In this algorithm by extended new transactions to the database, effects of the non-sensitive itemsets are the least [9].

Hong et al. proposed the SIF-IDF (sensitive items frequency-inverse database frequency) algorithm for hiding sensitive frequent itemsets. In this algorithm the SIF-IDF is calculated for each transaction. Then, an item with most occurrence frequencies within the sensitive itemsets will be chosen as the victim item and be deleted from the transaction with highest SIF-IDF [35].

Lin et al. presented the HMAU algorithm for hiding sensitive itemsets through transaction deletion. The HMAU is calculated for each transaction and the transaction with the lowest HMAU is deleted from the database [36].

Wei Lin et al. introduced the cpGA5DT algorithm for hiding the sensitive itemset through transaction deletion. In this algorithm, a fitness function with three adjustable weights is designed to find the appropriate transaction to be deleted such that Side effects are minimal. This proposed algorithm deletes the transactions with high probabilities in the probability vector. This approach can reduce the population size at each evaluation and also computations of database rescan [37].

Fouladfar et al. proposed FHA algorithm in for hide the association rules that hides sensitive rules with only a single scan of the database. In this algorithm for hiding sensitive rules, the position of the sensitive items is altered but its support is never changed; hence, reduced the number of missing rules [38].

Kiani Abari et al. proposed a novel algorithm for hiding sensitive association rules, named DCMHAR (Double Clustering Method in Hiding Association Rules) that the sensitive items are clustered. This algorithm can hide the association rules that have multiple items on its right or left hand side and either the support reduction or the confidence reduction or both are used in this method. In order to reduce the modifications of the database, sensitive items are clustered according to their presence in RHS or LHS, and then they are hidden as a cluster. If the items in RHS group are selected, the hiding process is done by reducing the support and confidence, and if those in LHS are selected, the hiding process is done by reducing confidence threshold [39].

## 5. Conclusions

In this paper, basic concepts in the field of preserving privacy in data mining were discussed and three fundamental heuristic, border-based and exact approaches were introduced. Furthermore, prevalent algorithms in each of these three approaches were studied. All of the presented methods made an attempt to hide sensitive association rules by minimizing side effects. The most important factor among side effects is to not have failure in hiding. The studied algorithms could be compared to each other based on various point of views. In Table 1, algorithms are categorized in terms of hiding method operation.

Table 1: Comparison of hiding sensitive association rules algorithms

Algorithm	Hiding Failure	Missing rule	Artificial rule	Scan
[13]		✓	✓	
1.a	✓	✓	✓	
1.b	✓	✓		
2.a	✓	✓		>2
2.b		✓	✓	
2.c		✓	✓	
MinFIA	✓	✓		2
MaxFIA	✓	✓		2
SWA	✓	✓	✓	1
Hidden-First		✓		
Non-Hidden-First	✓	✓		
HPCME	✓			
ISL	✓	✓	✓	
DSR		✓		
DCIS			✓	
DCDS		✓	✓	
Aggregate		✓	✓	
Disaggregate		✓	✓	
Hybrid		✓		

WSDA		✓	✓	
BA		✓	✓	
FHSAR		✓	✓	1
SRH		✓		
DSRRC		✓		
ADSRRC		✓		
RRLR	✓	✓		
[31]	✓	✓	✓	>2
MDSRRC		✓		
Artificial Bee Colony		✓	✓	
MaxMin		✓		
[38]		✓	✓	1

## References

- [1] A. Rakesh, I. Tomasz, and S. Arun, "Mining association rules between sets of items in large databases", in ACM SIGMOD Record, 1993, Vol. 22, pp. 207-216.
- [2] V. Vassilios S, "Association rule hiding methods, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 3, No 1, 2013, pp. 28-36.
- [3] L. Hai Quoc, A. I. Somjit, N. Huy Xuan, and A. I. Ngamnij, "Association rule hiding in risk management for retail supply chain collaboration", Computers in Industry, Vol. 64, No. 7, 2013, pp. 776-784.
- [4] S. Komal, T. Amit, and G. Amit, "Association Rule Hiding by Heuristic Approach to Reduce Side Effects & Hide Multiple R.H.S. Items", International Journal of Computer Applications, Vol. 45, No. 1, 2012, pp.1-7.
- [5] S. Komal, T. Amit, and G. Amit, "A study on association rule hiding approaches", International Journal of Engineering and Advanced Technology (IJEAT), Vol. 1, 3, 1, 2012, pp.151-163.
- [6] D. Elena, V. Vassilios S, E. Ahmed K, and B. Elisa, "Hiding association rules by using confidence and support", IHW '01 Proceedings of the 4th International Workshop on Information Hiding, 2001, pp. 369-383.
- [7] M. Heikki, and T. Hannu, "Levelwise search and borders of theories in knowledge discovery", Data mining and knowledge discovery, Vol. 1, No. 3, 1997, pp. 241-258.
- [8] S. Xingzhi, and Y. Philip S, "Hiding sensitive frequent itemsets by a border-based approach", Journal of Computing Science and Engineering, Vol. 1, No. 1, 2007, pp. 74-94.
- [9] M. Syam, S. Sumit, and M. Shbnath, "Maximizing Accuracy of Shared Databases when Concealing Sensitive Patterns", Information System Research, Vol. 16, No. 3, 2005, pp. 256– 570.
- [10] G.D. Aris, and V. Vassilios S, "An integer programming approach for frequent itemset hiding", in Proceedings of the 15th ACM International Conference on Information and knowledge management, 2006, pp. 748-757.
- [11] C. Chris, and M. Don, "Security and privacy implications of data mining", in ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1996, pp. 15-19.
- [12] C. Chris, K. Murat,, and V. Jaideep, "Defining privacy for data mining", in National Science Foundation Workshop on Next Generation Data Mining , 2002, Vol. 1, pp. 191 –207.
- [13] A. Mike, B. Elisa, E. Ahmed, I. Mohamed, and V. Vassilios S, "Disclosure Limitation of Sensitive Rules", Knowledge and Data Engineering Exchange, 1999.(KDEX'99) Proceedings. 1999 Workshop on, 1999, pp. 45-52.
- [14] S. Yücel, V. Vassilios S, and C. Chris, "Using Unknowns to Prevent Discovery of Association Rules", Acm Sigmod Record, Vol. 30, No. 4, 2001, pp. 45–54.
- [15] Y. Saygin, V. S. Verykios, and A. K. Elmagarmid, "Privacy preserving association rule mining", in Proceedings of the 2th International Workshop on Research Issues in Data Engineering: Engineering e-Commerce/ e-Business Systems, 2002, pp. 151-158.
- [16] O. Stanley RM, and Z. Osmar R, "Privacy Preserving Frequent Itemset Mining", in Proceedings of the IEEE international conference on Privacy, security and data mining, 2002, Vol.14, pp. 43–54.
- [17] O. Stanley RM, and Z. Osmar R, "An efficient one-scan sanitization for improving the balance between privacy and knowledge discovery", Department of Computing Science, University of Alberta, Canada, Tech. Rep, 2003, pp. 1-20.
- [18] V. Vassilios S, E. Ahmed K, B. Elisa, S. Yücel, and D. Elena, "Association rule hiding", IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 4, 2004, pp. 434–347.
- [19] P. Emmanuel D, T. Achilleas A, and V. Vassilios S, "An experimental study of distortion-based techniques for association rule hiding", Research Directions in Data and Applications Security XVIII, 2004.
- [20] L. Guanling, C. Chien-Yu, and C. Arbee LP, "Hiding sensitive patterns in association rules mining", in Computer Software and Applications Conference, 2004, Vol. 1, pp. 424-429.
- [21] W. Shyue-Liang, and J. Ayat, "Hiding Sensitive Predictive Association Rules", in IEEE International Conference on Systems, Man and Cybernetics, 2005, Vol. 1, pp. 164 – 169.
- [22] W. Shyue-Liang, P. Bhavesh, and J. Ayat, "Hiding informative association rule sets", Expert Systems with pp ications, Vol . 33, No. 2, 2007, pp. 316–323.

- [23] W. Shyue-Liang, P. Dipen, J. Ayat, and H. Tzung-Pei, "Hiding Collaborative Recommendation Association Rules", *Applied Intelligence*, Vol. 27, No. 1, 2007, pp. 67-77.
- [24] A. Ali, "Dare to share: Protecting Sensitive Knowledge with ata Sanitization", *Decision Support Systems*, Vol. 43, No. 1, 2007, pp. 181–191.
- [25] V. Vassilios S, P. Emmanuel D, T. Yannis, and C. Liwu, "Efficient algorithms for distortion and blocking techniques in association rule hiding", *Distributed and Parallel Databases*, Vol. 22, No. 1, 2007, pp. 85–104.
- [26] W. Chih-Chia, C. Shan-Tai, and L. Hung-Che, "A Novel Algorithm for Completely Hiding Sensitive Association Rules", in Eighth International Conference on Intelligent Systems Design and Applications, 2008, Vol. 3, pp. 202-208.
- [27] D. Kalyani, M. Datta, and N. Maheswari, "A New Approach to Sensitive Rule Hiding", *Computer and Information Science*, Vol. 1, No. 3, 2008, pp. 107–111.
- [28] N. D. Mohammad, B. Kambiz, and K. Ahmad, "A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms", *Journal Of Software*, Vol. 4, No. 6, 2009, pp. 555–562.
- [29] M. Chirag N, R. Udai Pratap, and P. Dhiren R, "Maintaining privacy and data quality in privacy preserving association rule mining", in Computing Communication and Networking Technologies (ICCCNT), 2010, pp. 1-6.
- [30] S. Vijayarani, and P. M Sathiya, "Association Rule Hiding using Artificial Bee Colony Algorithm", *International Journal of Computer Applications*, Vol. 33, No. 2, 2011, pp. 41–47.
- [31] J. Dhyanendra, S. Amit, G. Neetesh, N. Priusha, S. Deepika, and P. Amit, "Hiding Sensitive Association Rules Without Altering the Support of Sensitive Item(S)", *International Journal of Artificial Intelligence & Applications (IJAIA)*, Vol. 3, No. 2, 2012, pp. 75–84.
- [32] D. Nikunj H, and R. Udai Pratap, "Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database", in Advance Computing Conference (IACC), 2012, pp.1306-1310.
- [33] S. Xingzhi, and Y. Philip S, "A border-based approach for hiding sensitive frequent itemsets", in Fifth IEEE International Conference on Data Mining (ICDM'05), 2005, pp. 1-8.
- [34] M. George V, and V. Vassilios S, "A border-based approach for hiding sensitive frequent itemsets", *Data & Knowledge Engineering*, Vol. 65, No. 1, 2008, pp. 75-89.
- [35] H. Tzung-Pei, L. Chun-Wei, Y. Kuo-Tung, and W. Shyue-Liang, "Using TFIDF to hide sensitive itemsets", *Applied Intelligence*, Vol. 38, No. 4, 2013, pp. 502-510.
- [36] L. Chun-Wei, H. Tzung-Pei, and H. Hung-Chuan, "Reducing Side Effects of Hiding Sensitive Itemsets in Privacy Preserving Data Mining", *The Scientific World Journal*, Vol. 2014, 2014, pp. 1-14.
- [37] L. Chun-Wei, Z. Binbin, Y. Kuo-Tung, and H. Tzung-Pei, "Efficiently Hiding Sensitive Itemsets with Transaction Deletion Based on Genetic Algorithms", *The Scientific World Journal*, Vol. 2014, 2014.
- [38] F. Maryam, and N. D. Mohammad, "A heuristic algorithm for quick hiding of association rules, " *Advances in Computer Science: an International Journal (ACSIJ)*, 2015.
- [39] K. A. Zahra, and N. D. Mohammad, "Double Clustering Method in Hiding Association Rules", *Journal of Advances in Computer Research*, Vol. 7, No. 1, 2016, pp. 67-87.

# A LIGHT WEIGHT SECURE PROTOCOL FOR DATA TRANSMISSION IN VANET

Shawar Gul, Noor Ul Amin, Faisal Bahadur, Insafullah, Muazzam Ali Khan

*Department of Information Technology, Hazara University, Mansehra , Pakistan, <sup>e</sup> NUST Pakistan*

## Abstract

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

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

## 1. Introduction

Vehicular ad hoc network (VANET) is a form of mobile ad hoc network (MANET) [1], which tries to deploy the basic concept of unbroken varying vehicular motion. In VANET each and every vehicle acts as nodes. In the current era VANET is an open active research area designed to improve safe driving, traffic optimization and some services through Vehicle to vehicle communication (V2V) or vehicle to infrastructure communication (V2I) [2]. In this network, each vehicle communicates with one another or in other words sends and receives messages by GPS, On Board Unit (OBU) and equipped with Event Data Recorder (EDR), trusted component, etc, while the responsibility of Road Side Unit (RSU) is broadcasting safety messages periodically. With the fast growth in the development of Wireless communication protocols [3], and reducing hardware cost needed, in conjunction with the automobile industry try to enhance road safety level and get competitive edge in the market, so modern vehicles are manufactured with the latest communication devices like GPS, OBU, SENSOR etc. On the other hand, wireless communication is

itself at risk for different assaults and the security of vehicular ad hoc network should not be damaged. Some dishonest nodes try to inject false information into the network so that to get unfair advantage on road or to cause serious accidents, so receiver should be authenticated the sender vehicle before taking any action based on the received safety message. Now RSUs are fixed infrastructure deployed on the road, whose work to periodically broadcast safety related information, usually the deployment distance among each RSUs is approximately 300m to 1 km and they broadcast at the interval of every 300ms. Hence the deployment of RSUs along with a high way to provide universal infrastructure is not possible economically for the time being. So vehicle should have the property to authenticate other vehicles with limited help from a Trusted Authority (TA). Also, symmetric keys are used for identification of vehicle; hence an eavesdropper may track the sending vehicle, so we want to secure communication in VANET and to reduce computational cost. Hence, in proposed research work, we propose A Light Weight Secure Protocol for Data Transmission in VANET.

## 2. PRELIMINARIES

### 2.1. Discrete Logarithm Problem (DLP)

Discrete Logarithm Problem is the hot topic in cryptography due to its hardness. Discrete Logarithm Problem uses in many cryptographic schemes such as ElGamal [4] are based on the hardness of DLP. Discrete Logarithm Problem state that: Suppose  $x_1, y_1 \in Z$  implies  $y_1 \neq 0$  and  $x$  is the generator of  $Z_p$ , finding a unique integer  $k$ ,  $0 \leq k \leq n - p$  implies that  $y = x^{1^k}$ .

### 2.2. Elliptic Curve Discrete Logarithm Problem (ECDLP)

Suppose  $p$  is prime number which is greater than 3, EC over finite field  $F_p$  can be defined by the Eq. 1:

$$B^2 = A^3 + x_1 A + y_1 \quad (1)$$

Where  $x, y$  belong to  $F_p$  &  $4x_1^3 + 27y_1^2 \not\equiv 0 \pmod{p}$  &  $E(F_p)$  is a set which containing all the points  $(a, b)$  belong to  $F_p$  which satisfy Eq. 1, with the point of infinity  $\mathcal{O}$ .

ECDLP states that: suppose  $p$  &  $Q$  be the two points on  $E$  &  $n$  is the order  $P$ . Now finding the integer  $d$ , where  $0 \leq d \leq n - 1$  Such that  $P = dQ$ .

### 2.3. Hyper elliptic Curve

HEC is generalized form of EC and state that HC  $HEC(F_n)$  on finite field  $F_n$  can be define the equation (1)

$$HEC: Y_1^2 + H(x_1)Y_1 = f(x) \pmod{n} \quad (1)$$

Where  $H(x_1)$  belong to  $F[x_1]$  is the polynomial. The degree of this polynomial is  $H(x_1)$  is less or equal to  $g$ . where  $f(x_1)$  is the monic polynomial which belong to  $F[x_1]$ . the degree of  $f(x_1)$  is less than or equal to  $2gs + 1$ .

### 2.4 Divisor

The divisor are the sum of points formally on HEC .

$$D1 = (A(x1), B(x1)) = \left( \sum_{a=0}^{gs} A_a x1^a, \sum_{a=0}^{gs-1} B_a x1^a \right)$$

## 2.5. Jacobian Group

The Divisors of the HEC make the Abelian group. Also called Jacobian of  $HEC(F_{n1})$ . the following equation define the  $o(J_c(F_{n1}))$

$$\text{Order of Jacobian } |(\sqrt{n} - 1)^{2gs}| \leq o(J_{HEC}(F_{n1})) \leq |(\sqrt{n} + 1)^{2gs}|$$

## 2.6. Hyper Elliptic Curve Discrete Logarithm Problem

Suppose  $D2$  is called the divisor of Hyper Elliptic Curve & the order of the divisor  $D2$  isn't in the group of the Jacobian  $J_{HEC}(F_{n1})$ , finding the random integer  $ds \in F_{n1}$ , such that:  $D_1 = ds \cdot D2$

## 3. Related Work

Lei Zhang et al [5] proposed a basic security - preserving authentication protocol, which proficiently abuses the development of vehicle, street constraint and circulated RSUs in VANETs. However the execution rate diminished in this scheme and additionally the activity load expanded.

Grover et al [6], give a thought to propose a straightforward scheme in which parameter angle is utilized for the discovery of Sybil nodes, where the angle esteem, stay same at any interim of time for every node, and the recreation result demonstrates that the roughly 0.5% mistake rate and 99% precision. However the handling time and capacity limit are not characterized in the current plan so that to increase 100% discovery precision

The author in [7] has proposed the usage of the digital signature and PKI yet it's not giving the mechanism required to the endorsement renouncement, despite the fact that it is the compulsory piece of any public key infrastructure based arrangement.

Y. Peksen and T. Acarman in [8] depicted the organization of vehicle to vehicle communication to telecast the time - basic security messages out and about. Multi jumping is basically used to maximize dissemination range. In the proposed scheme the vehicular node is at the most extreme separation as for another sender and must be available to the media.

C. C. Lee and Y. M. Lai In this paper demonstrate that Lee and Lai's IBV [9] scheme contains a few security dangers in VANET. The primary anybody can follow the genuine identity of the sender who simply has the freely known framework parameters. So this scheme is neglecting to accomplish the prerequisite of securely saving. Second a deceptive vehicle disperses the message for lawful vehicle furthermore utilizes a fake personality to abstain from being followed, so for aforementioned restriction Lee and Lai's IBV plan is not secure and suitable for VANET.

Q. Zhao et al [10] proposed another plan scheme (3GDD) 3G-helped information conveyance so it first allot 3G solving so as to spend plan to every last time, space the ILP detailing of the first enhancement issue, after this it picks those bundle which generally conveyed through VANET for 3G transmission. However the unified whole

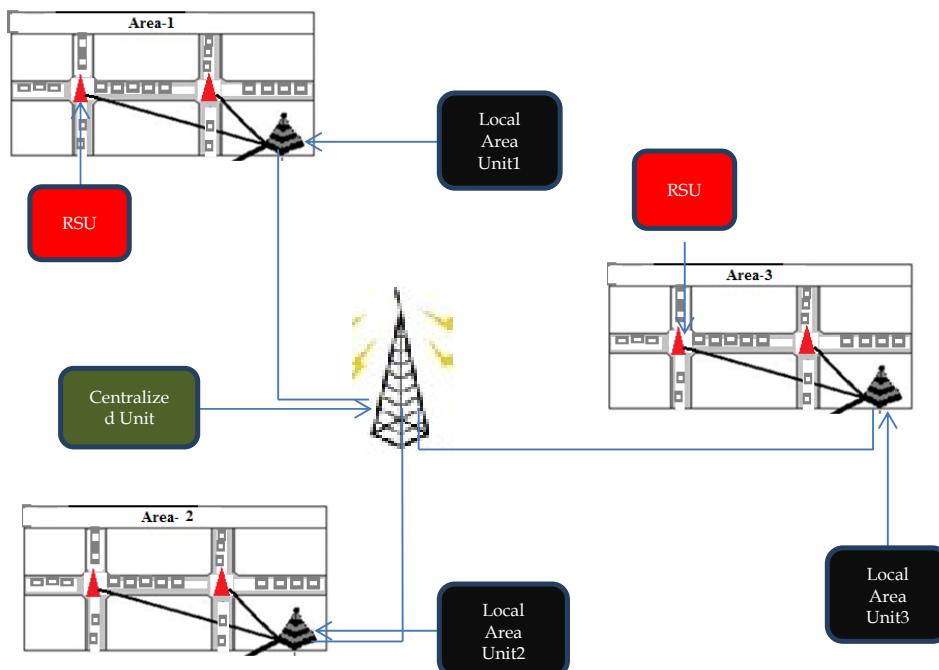
number straight programming-based arrangement does not have capacity to scale for the huge \_scale system with 3G access point.

The author in this paper [11] , proposed TDMA based MAC protocol that presents another multichannel TDMA MAC protocol which is exceptionally proposed for a VANET situation, that backings effectively multi jump show and one-bounce administrations in the control channel by eliminating so as to utilize affirmations furthermore the shrouded terminal issue. VeMAC are attempting to decrease the transmission crashes created because of the node versatility on the control channel by doling out the arrangement of time spaces to vehicles which are preceding onward both headings. In any case, the impediment of this paper In VeMAC protocol every node is capable to transmit a bundle amid its time space under such circumstance if the node has no information to incorporate into the high-need short applications field.

The ADDHV [12] diminishes pointless, repetitive transmissions and in this manner merges the usefulness of sweet spots acquired the GEDDAI. Additionally, it makes utilization of a strategy of autonomic processing to choose when the node or not spreading data in light of the issue of network partition. This choice depends on the topographical area of the vehicle and an idea of individual proficiency of proliferation. Besides, distinctive least separations between a source node and the start of development of the sweet spots are utilized. The spreading done that path with ADDHV in an urban situation is then contrasted with the simulation results with the AID and the DBRS.

The author in [13] has proposed another scheme which expects to forward information between the vehicles in 2015. The strategy utilized as a part of the scheme is ECC. In this article, first it has been contended that ECC is the most reasonable mechanism for securing VANET, yet later a few impediments of ECC have likewise been distinguished by them. The ECC has a high computational cost and communication overhead and can't give any security prerequisites, for example, are an authentication, distributed and fair revocation.

#### 4. PROPOSED DIAGRAM



## MAIN COMPONENTS OF OUR PROPOSED DIAGRAM

- **User/ Vehicle**
- **Road Side Unit(RSU)**
- **Local Sector Unit(LSU)**
- **Centralized Unit(CU)**

### 4.1. User/ Vehicle

It is the principle component of the VANET which is used in the registration with Road Side Unit. This vehicle should support the software installation.

### 4.2. Road Side Unit

Road Side Unit is in charge of communicating specifically with vehicle. The fundamental obligations of Road Side Unit incorporate to get identity information of vehicle and send to Local Sector Units for validation. Additionally, Road Side Unit is mindful to convey time basic data with respect to movement circumstance ahead, warning, climate information and entertainment facilities to automobiles. These RSU are fixed at each intersection or circuitous where they are associated with separate Local Sector Units by means of a fast wired association.

### 4.3. Local Sector Unit

Local Sector Unit comprises of huge and productive database for the whole entire city vehicles. Furthermore one local sector unit is installed in each city, LSU is operated to authenticate vehicles utilizing in that zone. In the proposed model, Local Sector Units are required to keep up predictable data about vehicles to enhance performance delay in data sharing aiming to evacuate the conceivable odds of clog which may happen in the entire city centralized database. Local Sector Unit accomplishes the higher performance by partitioning endeavors needed to authenticate vehicles. Further , significant role of Local Sector Unit is to telecast city- wide information to automobiles by means of Road Side Unit, such as movement information in the specific part furthermore inform centralized Unit to return traffic to different route if there should be an occurrence of substantial movement in that area. Local Sector Unit is needed to be advanced with up-to-date information of vehicle and entirely Local Sector Units ought to have reliable information.

### 4.4. Centralized Unit

Centralized Unit is one unit or more relying on density furthermore, territory of city. The principle obligations of Centralized Unit are to inter link all Local Sector Unit for the transmission city-wide information like traffic data, climate news and so forth. Centralized Unit additionally keeps up huge information which is utilized by security analyst keeping in mind the end goal to track different attacker into the system for examination purposes.

## 5. Proposed Algorithm

1. User A chooses the random number  $k_{A1}$  from [1, n-1],

Compute the public key  $RA1 = kA1D$ . And then, send  $ID_{A1}, R_{A1}$  to user B.

2. User B will perform the following actions: After receiving message 1

2.1. Generate public key  $RBi = kBvD$ .

2.2. Compute  $sB = (kBv + RBi' dB) \bmod n$  and  $Z = hsB$

$(RA1 + RA1' QA)$ , where  $R_{Bi'}$  and  $R_{A1}$  shows the x coordinate of  $R_{Bi}$  and  $R_{A1}h$ , is HEC domain parameter.

2.3.  $(k1_v, K2_v) \leftarrow KDF(x_z)$  where  $x_z$  is the x-coordinate of Z and the key derivation function is KDF.

2.4. Compute  $tB = MAC k1v(IDB2, IDA1, RBi, RA1, 2)$ .

2.5. Send  $ID_{B2}, R_{Bi}, t_B$  to user A.

3. User B does the following: After receiving message 2,

3.1. Compute  $sA = (kA1 + RA1' dA) \bmod n$  and  $Z = hsA (RBi + RBi' QB)$ .

3.2.  $(k1_v, k2_v) \leftarrow KDF(x_z)$

3.3. Compute  $t = MAC k1v (IDB2, IDA1, RBi, RA1, 2)$  to verify  $t = tB$ .

3.4. Compute  $tA = MAC k1v (IDA1, IDB2, RA1, RBi, 3)$ .

3.5. Sends  $t_A$  to B.

4. User B computes the following: After receiving message 3

$t = MAC k1v (IDA1, IDB2, RA1, RBi, 3)$  and verifies that  $t = tA$ .

5. Z will be the shared secret key which is used for encrypting messages further, If the verification is successful.

## 6. Security Analysis

### 6.1. Authenticity

Authenticity means to ensure that the information is gotten from that node which it claims. The proposed scheme provides the security property of authentication; because before sending the message, the sender of the message first generate digital signature of a message such as  $(k_Bv + R_Bi \cdot D_B, (k_A + R_A1 \cdot d_A))$ . Hence uncovering an individual distinguishing identification number  $(k_Bv + R_Bi \cdot d_B)$  is a hard undertaking for the attacker.

## 6.2. Unforgeability

The proposed scheme provides the security property of Unforgeability; when the attacker wants to generate a forge signature, then in such cases, the attacker needs the sender private key ( $k_{1v}$ ), ( $k_{2v}$ ). Now to generate key ( $k_{1v}$ ), ( $k_{2v}$ ), to do this, the attacker must solve the elliptic curve discrete logarithm problem which is in feasible for attacker.

## 6.3. Confidentiality

It intends to keep secure the important data out of range from unapproved user. The proposed scheme ensure security requirement such as confidentiality, for revealing the contents of message, the attacker needs key  $k$  ( $k_{1v}$ ,  $k_{2v}$ ). On the other hand, the secret key is only known to sender and receiver, so the attacker cannot reveal the original message content.

## 6.4. Integrity

The proposed light weight secure protocol meet the security property of integrity. When the sender of the message send the message then it first calculate the one way hash function of a message  $h(R_{A1} + R_{A1}, Q_A)$ , then it send for integrity. At receiving side, when the receiver receive the message and  $h(R_{B1} + R_{B1}, Q_B)$  hash, then it decrypt the message and calculate the hash of a receiving message  $h(R_{A1} + R_{A1}, Q_A)$ , after compare this  $h(R_{A1} + R_{A1}, Q_A) = h(R_{B1} + R_{B1}, Q_B)$ . If it is equal then the message contents cannot change otherwise the message is not in original form.

## 6.5. Non repudiation

Non repudiation confines sender from denying the encrypted text he/she sent. In the event that the sender denies, recipient send encrypted text to judge/trusted authority. By utilizing verification strategy judge can choose that the message is sent by him/her by.

## 7. Cost Analysis

In the cost analysis phase we show key comparisons and compare computation cost of the proposed encryption scheme with existing scheme.

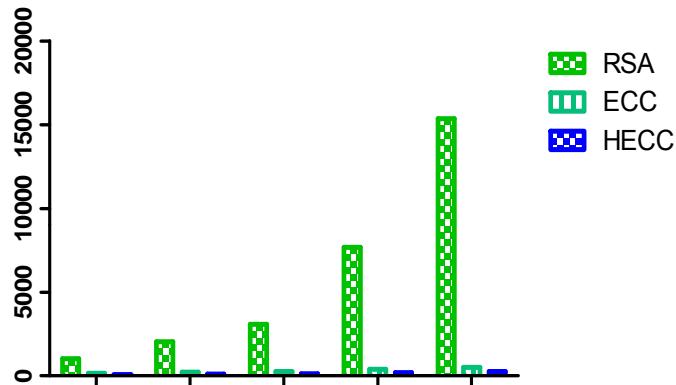
### 7.1. Key comparison

So that to achieves the same security level, NIST has recommended key length of 80 bits for HECC, 160 bits for ECC and 1024 bits for RSA.

**Table 1 NIST Recommended Key Size**

RSA	Elliptic Curve	Hyper Elliptic Curve
1024	160	80
2048	224	112
3072	256	128
7680	384	192
15360	512	256

The below figure 5 shows the comparison among the RSA, ECC and HEC. The below graph shows increasing the key length strongly affects the decoding time. So ECC is not sufficient for encrypting long messages. Y represents key size and x shows the time.



**Figure 1 key Comparison**

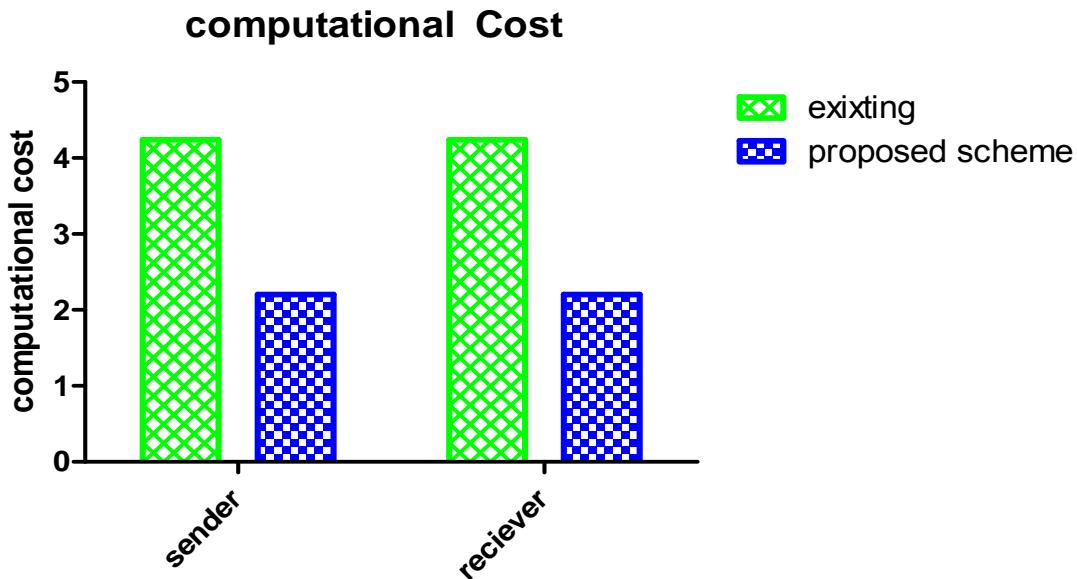
## 8. Computational Cost Analysis

### 8.1. Computational cost:

The computational cost implies the measure of computational efforts to be contributed both sender and receiver of a message. By and large, computational cost is assessed by checking the quantity of overwhelming operations included. Commonly these operations incorporate private key encryption and decryption, hashing, particular multiplication, division, addition, and exponentiation. The Elliptic curve point multiplication ECPM and HECDM is the two major operations. The proposed scheme HECDM is compared with existing scheme ECPM, as result HECPM has less computational cost as compare to the existing scheme. The single scalar multiplication consumes 4.24 ms for ECPM and 2.2 ms HECDM on a PC running jdk 1.6 having two centers of Intel CPU with preparing speed of 2.00 GHz and essential memory limit of 4 GB working with Microsoft Windows vista [14].

**Table.2 Comparison of Computational cost**

Schemes	Participants	Major operation		Cost Reduction in %
		ECC	HECC	$\frac{8.48 - 4.4}{8.48} * 100$ =48.11%
Proposed scheme	Sender A	-	2.2	
	Receiver B	-	2.2	
Kuldeep SinghEt al [13]	Sender A	4.24	-	
	Receiver B	4.24	-	



## 9. Conclusion

Our proposed scheme satisfies the security properties like confidentiality, authenticity, integrity, unforgeability, verifiability and non-repudiation with minimal cost. The limitation of the existing elliptic curve cryptography scheme [13] has high computation and communication cost because the existing scheme is based on elliptic curve cryptography (ECC). Our proposed scheme is based on hyper elliptic curve cryptography (HECC) which is computationally more difficult for an attacker further more uses low parameters than elliptic curve cryptography (ECC). The computation and communication cost of the proposed scheme is comparatively lower than the existing scheme based on elliptic curve cryptography (ECC). Because of saving computational and communications cost, our scheme is suited for restricted computational devices. The protocol also reduces the computational cost 48.11% as compared to the existing scheme.

## Reference

- [1] Y. Toor, P. Muhlethaler, and A. Laouiti, “Vehicle Ad Hoc networks: applications and related technical issues,” *Communications Surveys Tutorials, IEEE*, vol. 10, no. 3, pp. 74 –88, quarter 2008.
- [2] F. Li and Y. Wang, “Routing in vehicular ad hoc networks: A survey,” *Vehicular Technology Magazine, IEEE*, vol. 2, no. 2, pp. 12 –22, Jun. 2007.
- [3] J.Bernsen, “A\_Reliability-Based Routing Protocol for Vehicular AdHoc Networks,” *Master’s Theses*, Jan. 2011.
- [4] G. Gamage et al (1998) an efficient scheme for secure message transmission using proxy signcryption, Technical report 9801, Monash University
- [5] Grover, J., Gaur, M. S., & Laxmi, V. (2010, Nov.): A novel defense mechanism against Sybil attacks in VANET. In 3rd Security of information and networks conference, SIN (pp. 249–255).

[6]Security Analysis of Vehicular Ad Hoc Networks (VANET) 2010 Second International Conference on Network Applications, Protocols and Services

[7]Y. Peksen, T. Acarman, "Multihop safety message broadcasting in VANET: A distributed medium access mechanism with a relaying metric," in Proc of 2012 Int. Symp: On Wireless Communication Systems (ISWCS), Paris, France, August 28-31, 2012, pp. 346-350.

[8] C. C. Lee and Y. M. Lai, "Toward a secure batch verification with group testing for VANET," Wireless Networks, vol. 19, no. 6, pp. 1441-1449, 2013

[9] Q. Zhao, Y. Zhu, C. Chen, H. Zhu, and B. Li, "When 3G meets VANET: 3G-assisted data delivery in VANETs," IEEE Sensors J., vol. 13, no. 10, pp. 3575–3584, Oct. 2013

[10] H. Omar, W. Zhuang, and L. Li, "VeMAC: A TDMA-based MAC protocol for reliable broadcast in VANETs," IEEE Trans. Mobile Comput, Sept. 2013.

[11]Jae-Han Lim et al, "Interplay between TVWS and DSRC: Optimal Strategy for Safety Message Dissemination in VANET", IEEE Journal on Selected Area in Communications, Vol.32, No.11, November. 2014.

[12] Mechanism in VANETs"International Journal of Advanced Trends in Computer Science and Engineering, Vol.3, No.4, Pages: 55-60 (2014) **Special Issue of ICCEIT 2014 - Held on September 01, 2014 in The Solitaire Hotel, Bangalore, India**

[13] K. Singh, P. Saini, AK. Singh, "Authentication and Privacy Preserving Message Transfer Scheme for Vehicular Ad Hoc Networks"Proceedings of the 12th ACM International Conference on Computing Frontiers Article No. 58, 2015

[14]M. Blaze et al (1998) Divertible protocols and atomic proxy cryptography, in Eurocrypt'98, LNCS 1403, Springer-Verlag.

# Dynamic Edge Detection in a Digital Video Stream Using Kirsch Filters

S.Aparna

Department of Computer Science Asst. Professor  
GITAM University Research Scholar JNTUH

Dr.M. Ekambaram Naidu

Principal, ARJUN College of Engineering and  
Technology, Hyderabad

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

**Key words:** Edgedetection, Kirsch compass kernel, Video image processing

## 1. Introduction

Digital video is essentially an ordered sequence of still images which are temporally correlated. One can employ algorithms used for still image processing for processing digital video frames also. The only disturbing point to be noted here is that the algorithms meant for video processing should be computationally very efficient, reliable, robust and less time consuming. In this context, this paper proposes Kirsch edge detection algorithm for converting each and every frame in a video sequence without any time loss. The paper is organized in the following manner. Basic details of the Kirsch compass kernel algorithms are given in section 2. Section 3 shows the results of applying eight masks of Kirsch algorithm on a sample low resolution digital image and the final result of applying ‘pairwise dyadic operation of maximum between two’ on all eight edge maps is presented. Then Kirsch algorithm masks are applied on a high resolution video frame and the final result of applying ‘pairwise dyadic operation of maximum between two’ on all eight edge maps is presented in section 3. Section 4 presents the results of applying eight masks of Kirsch algorithm on a stream of digital video frames and section 5 outlines the results of applying mask #8 of Kirsch algorithm on another stream of digital video frames. Concluding remarks on the research presented in the paper are given in section 6.

## 2. Basics of Kirsch Compass Kernel Algorithm

The KirschCompass Kernel Algorithm is a reliable edge operator, which achieves a tradeoff between preserving edge details and suppressing noise components. Kirsch edge operator employs eight convolution kernels. Each point in the image is convolved with all eight masks. Each mask when convolved with image yields edges oriented in a particular direction. The maximum value over all eight orientations is the output value for the edge magnitude image. Basic details of Kirsch Compass Kernel Algorithm edge operator are explained below. A digital image  $I_0$  is convolved independently with all eight masks and eight individual edge maps obtained. The convolution is carried out in an image pixel wise. Let  $F_m(n,k)$  be the value of convolution carried out on a pixel with image coordinates  $(n,k)$  by one of the eight masks, say the  $m^{\text{th}}$  mask. Then,  $F_m(n,k)$  is given by

$$F_m(n, k) = \sum_{j=-1}^{+1} \sum_{i=-1}^{+1} G_0(n + j, k + i) M_m(j, i)$$

where  $G(n+j, k+i)$  denotes 0, the gray or color value at point  $(n+j, k+i)$  in the image  $I_0$ ,  $i = -1, 0, 1; j = -1, 0, 1$ ;  $n = 0, \dots, H-1$  and  $k = 0, \dots, W-1$ .  $H$  and  $W$  respectively denote the height and width of the image  $I_0$ .  $M_m(j, i)$  is the coefficient at point  $(j, i)$  in the  $m^{\text{th}}$  mask. Now, the maximum convolution,  $F_{\max}(n, k)$  is defined as

$$F_{\max}(n, k) = \max \left\{ |F_1(n, k)|, |F_2(n, k)|, \dots, |F_8(n, k)| \right\}$$

where  $F(n, k)$  represents filter value. So, the filtered image  $FI$  consisting of  $F(n, k)$  is written as

$$FI = \begin{bmatrix} F_{\max}(0,0) & \dots & F_{\max}(0,W-1) \\ \vdots & \ddots & \vdots \\ F_{\max}(H-1,0) & \dots & F_{\max}(H-1,W-1) \end{bmatrix}$$

All three neighborhood points are examined line by line and if the gray or color of the middle point is the local maximum value, then, it is the extremum point. In this way, the extreme image  $MI$  in the horizontal direction can be obtained. Thus  $MI$  is given by

$$MI_w = \begin{bmatrix} MI_w(0,0) & \dots & MI_w(0,W-1) \\ \vdots & \ddots & \vdots \\ MI_w(H-1,0) & \dots & MI_w(H-1,W-1) \end{bmatrix}$$

where  $MI(n, k)$  is the gray or color value at point,  $(n, k)$ , in the Image  $MI$ . Now,  $MI(n, k)$  is given by

$$MI_w(n, k) = \begin{cases} F_{\max}(n, k) & F_1 < F > F_2 \\ 0 & others \end{cases}$$

where  $F = F(n, k-1)$ ,  $F = F(n, k)$ ,  $F = F(n, k+1)$ . Similarly, the extremum image  $MI$  in the vertical direction is defined as

$$MI_h = \begin{bmatrix} MI_h(0,0) & \dots & MI_h(0,W-1) \\ \vdots & \ddots & \vdots \\ MI_h(H-1,0) & \dots & MI_h(H-1,W-1) \end{bmatrix}$$

where  $MI(n, k)$  is the gray or color value at point,  $(n, k)$ , in the Image  $MI$ .  $MI(n, k)$  is given by

$$MI_h(n, k) = \begin{cases} F_{\max}(n, k) & F_3 < F > F_4 \\ 0 & others \end{cases}$$

$F_3 = F_{\max}(n-1, k)$ . Similarly, the final extremum image  $MI$  is written as

$$MI = \begin{bmatrix} MI(0,0) & \dots & MI(0,W-1) \\ \vdots & \ddots & \vdots \\ MI(H-1,0) & \dots & MI(H-1,W-1) \end{bmatrix}$$

where  $MI(n, k)$  denotes the gray value at point,  $(n, k)$ , in the Image  $MI$  and defined as

$$MI(n, k) = \max \{MI_w(n, k), MI_h(n, k)\}$$

We examine each point in the image  $MI$ , if the gray value is larger than the preset threshold  $T$ , then, the corresponding point is labeled as edge point, in this way, the initial edge image  $EI$  is written as

$$EI = \begin{bmatrix} EI(0,0) & \dots & EI(0,W-1) \\ \vdots & \ddots & \vdots \\ EI(H-1,0) & \dots & EI(H-1,W-1) \end{bmatrix}$$

where  $EI(n, k)$  is the gray value at point,  $(n, k)$ , in the image  $EI$  and defined as

$$EI(n, k) = \begin{cases} 255 & MI(n, k) > T \\ 0 & MI(n, k) \leq T \end{cases}$$

Kirsch filter is essentially a direction based filter. With the help of Kirsch Compass Masks we can find edges in the following eight directions (i) North, (ii) North West, (iii) West, (iv) South West, (v) South, (vi) South East, (vii) East and (viii) North East. Fig. 1 shows all the eight masks

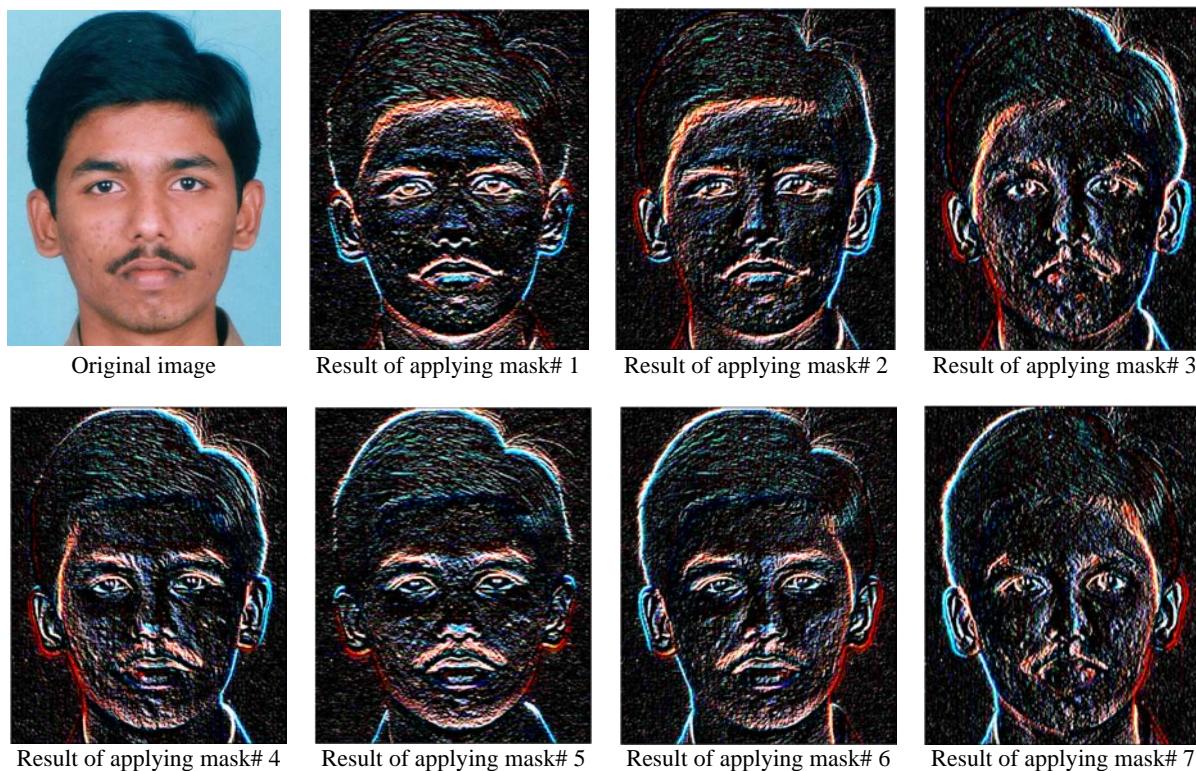
$\begin{array}{ccc} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{array}$	$\begin{array}{ccc} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{array}$	$\begin{array}{ccc} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{array}$	$\begin{array}{ccc} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{array}$	$\begin{array}{ccc} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{array}$	$\begin{array}{ccc} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{array}$	$\begin{array}{ccc} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{array}$	$\begin{array}{ccc} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{array}$
North Direction Mask #1	North West Direction Mask #2	West Direction Mask #3	South West Direction Mask #4	South Direction Mask #5	South East Direction Mask #6	East Direction Mask #7	North East Direction Mask #8

Fig.1: Eight masks of Kirsch compass kernel algorithm

With reference to Fig.1, one would observe that each mask will give edges of its own direction. In order to highlight the significance of these masks, different use cases have been studied and results presented the following sections.

### 3.Use Case 1 –Processing of low resolution digital images using Kirsch algorithm

A low resolution digital image is considered here for the case study. All the eight masks have been tried on this image and results shown in Fig.2. With reference to Fig.2, one can observe that the edge maps as well as the composite edge map do not exhibit the qualitative content of the original image. The question that arises here is whether Kirsch algorithm would yield quality content once it is applied to a stored video frame. Fig. 3 demonstrates the unlikelihood of such a result.

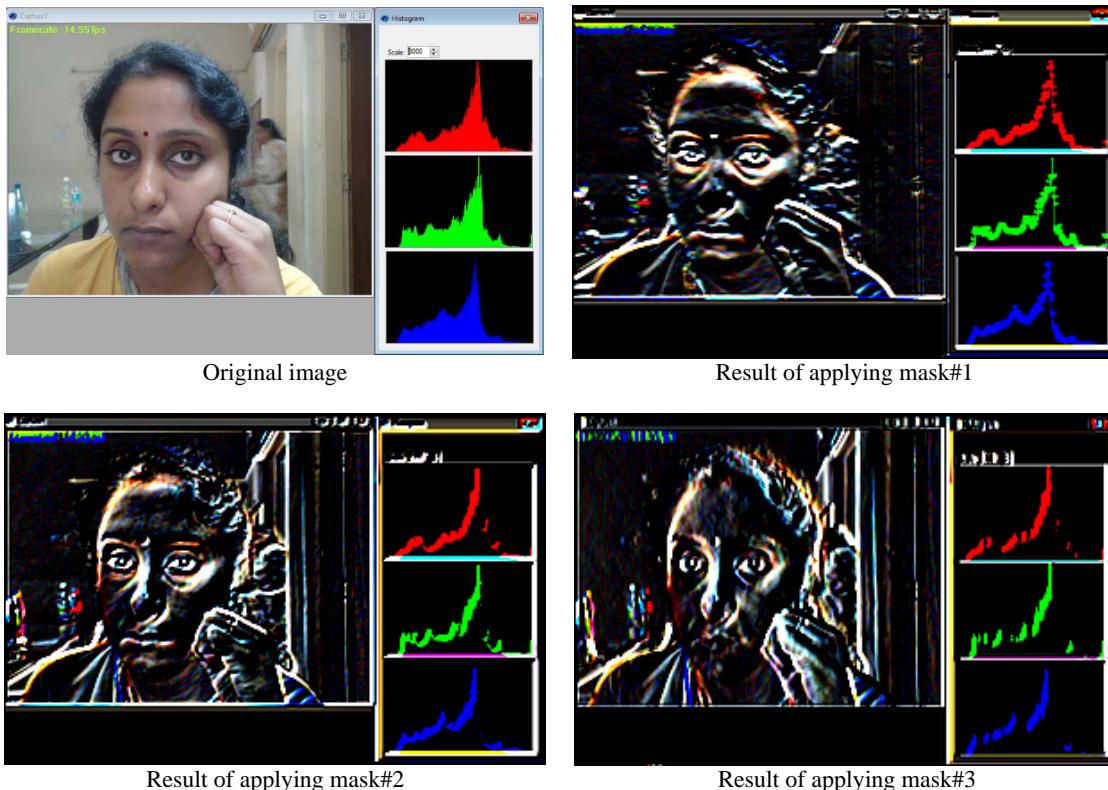




Result of applying mask# 8      Maximum of all edge maps  
Fig. 2: Results of applying masks on a low resolution digital image

#### 4. Use Case 2—Processing of a stored video frame using Kirsch algorithm

The experiment was repeated for a stored video frame and results obtained shown in Fig. 3. It was observed that the edge maps do not exhibit quality content. The term ‘quality content’ refers to presence of various features of the original image such that an observer can recognize at least partially the original image in edge maps.



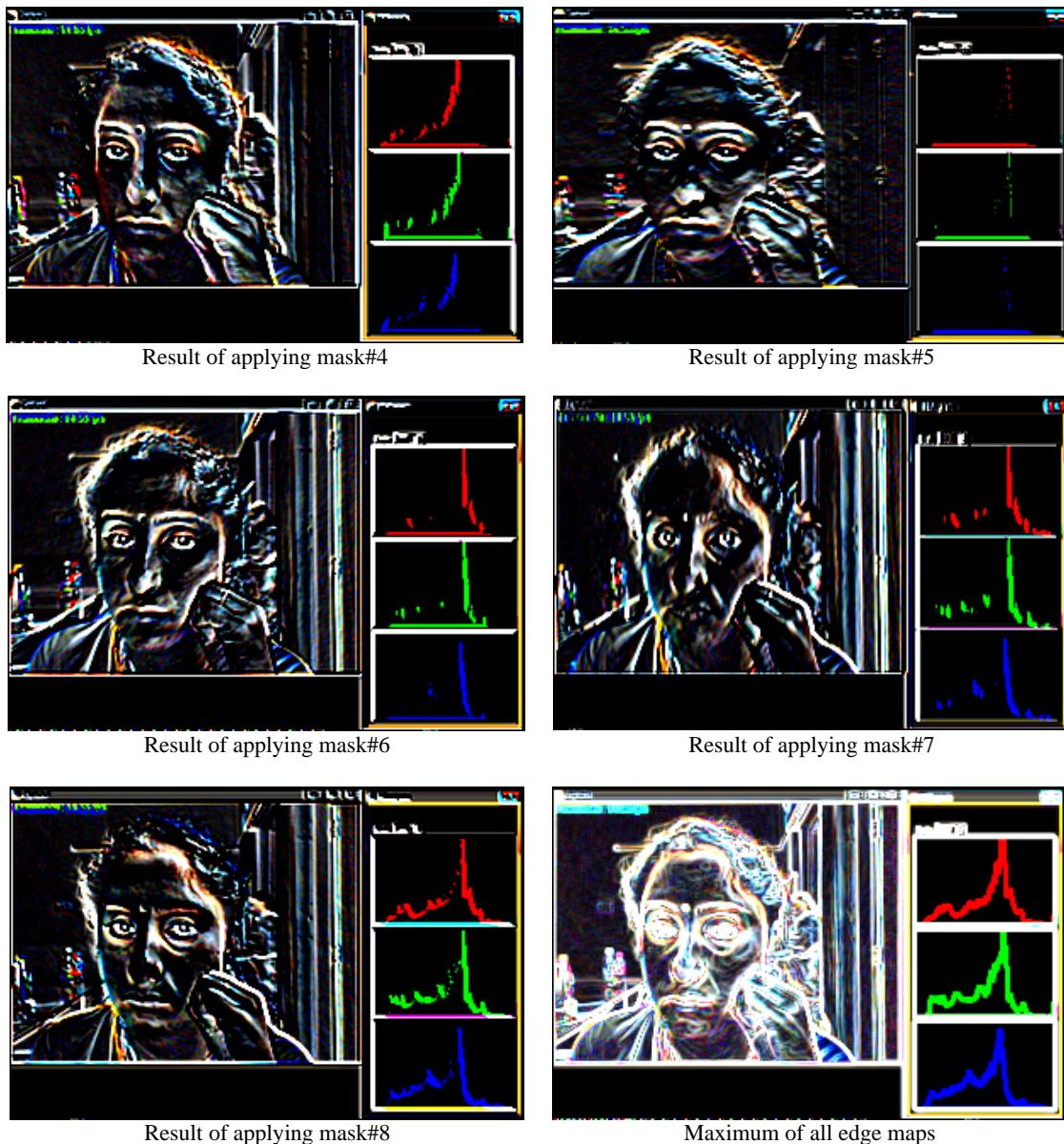
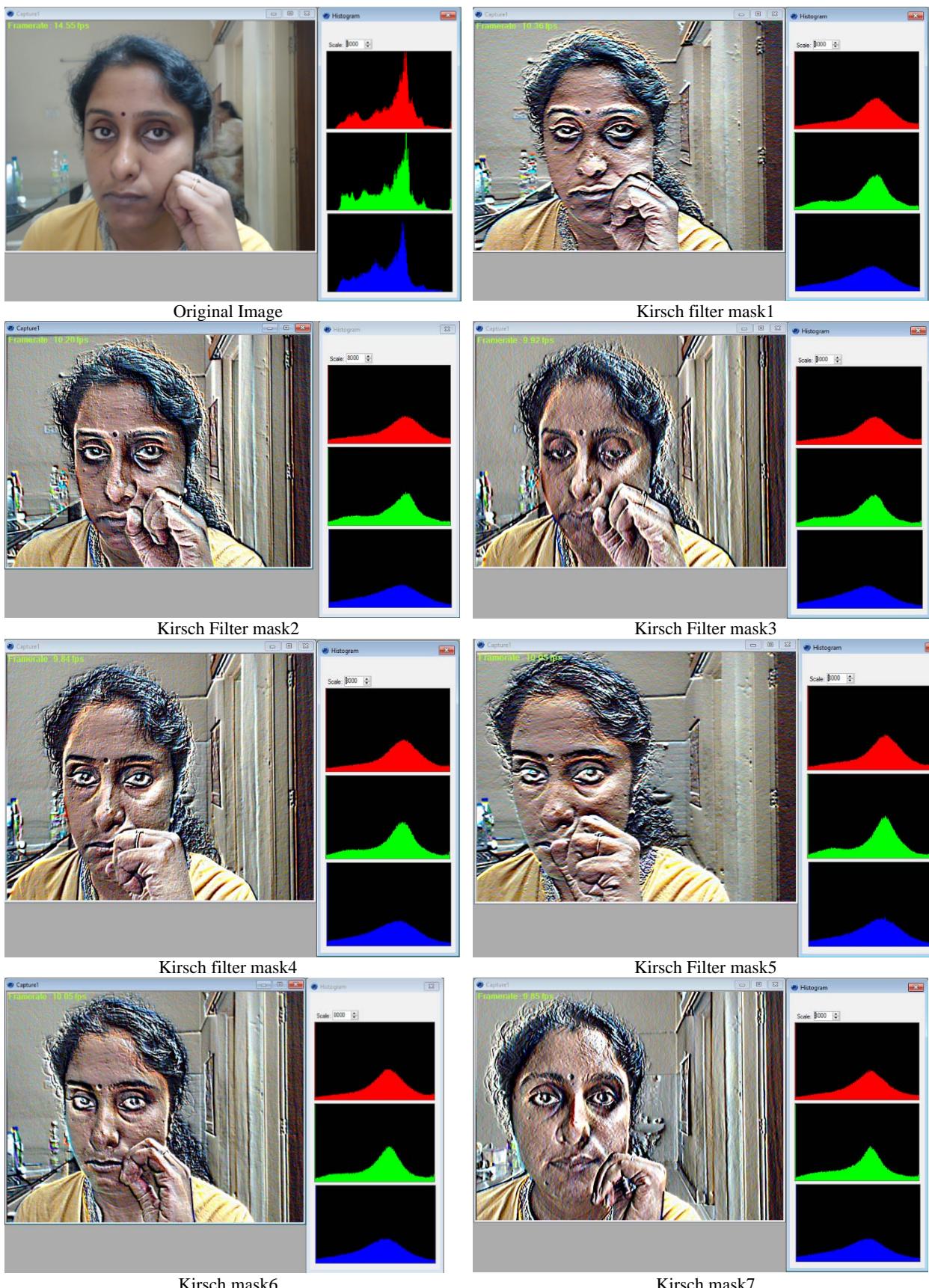


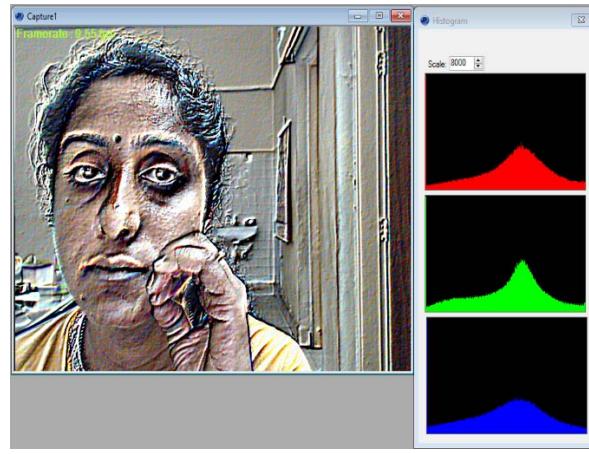
Fig. 3: Results obtained by using Kirsch masks on a stored video frame

Based on the above results shown in Fig. 3, one would come to an understanding that Kirsch filter will not retain quality content when a stored image is processed using relevant Kirsch masks. Consequently, an effort was made to study the usefulness of this algorithm on a stream of digital video. Results obtained are shown in Fig. 4.

### Use Case 3—Processing of stream of video frames using Kirsch algorithm

Using a web cam, a streaming video is observed and one such frame is captured and shown in Fig. 4. Kirsch masks are applied directly to the video and the edge maps captured and displayed in Fig. 4. It is to be noted that the Kirsch masks are applied to streaming video frames and not on stored video frames. Resulting video frames are captured for the purpose of analysis. Each edge map shown in Fig. 4 could be seen to preserve quality content. Another interesting observation made during the experiment was that one could apply a single Kirsch mask on a streaming video and the resulting edge maps do not change with respect to changes in the time domain. A video stream is basically a spatio-temporal activity.





Kirsch mask8

Fig. 4: Results obtained by using Kirsch masks on a streaming video

#### Use Case 4—Processing of stream of video frames using a single mask of Kirsch algorithm

As a special case, a single Kirsch mask (Mask #8) was applied to a streaming video and the resulting edge map stream recorded. Some of the edge map frames are shown in Fig. 5.

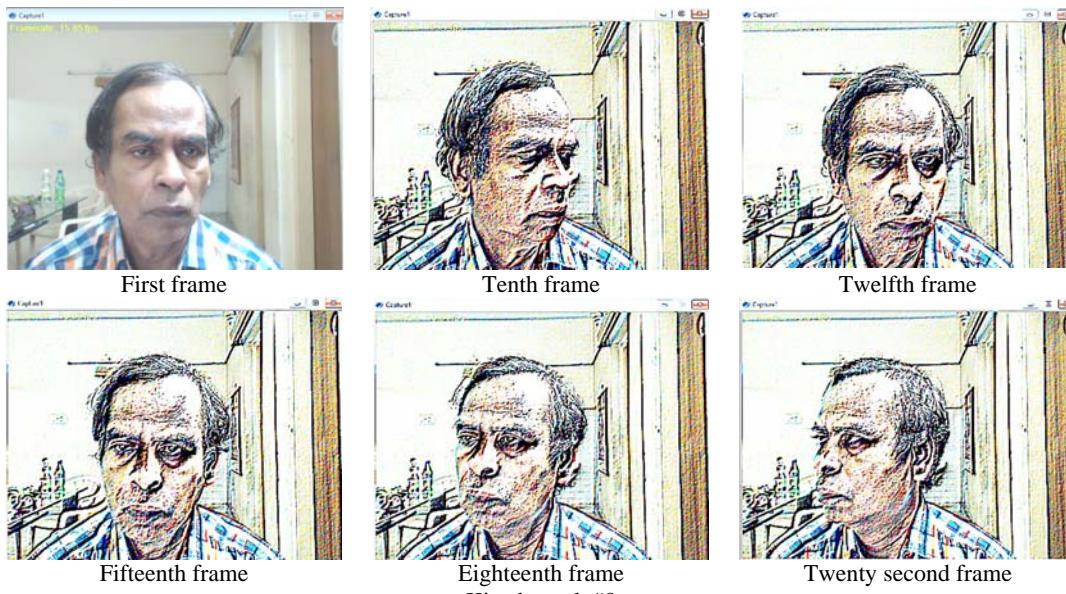


Fig. 5: Results obtained by using Kirsch mask #8 on a streaming video

With reference to Fig. 5, one can observe that the quality content of the streaming video is retained by applying a single Kirsch mask. One can choose any of the eight masks for obtaining stream of edge maps in a desired direction.

#### 6. Conclusion

This paper advocates the use of Kirsch compass algorithm to process streaming video rather than stored images or frames. The reason for this is that a running video contains fixed or insignificantly changing background and significantly changing foreground.

## References

1. McLoughlin, K.J., P.J. Bones and N. Karssemeijer 2004. Noise equalization for detection of microcalcification clusters in direct digital mammogram images, IEEE Transactions on Medical Imaging, 23(3): 313–320.
2. Nakayama, R., Y. Uchiyama, K. Yamamoto, R. Watanabe and K. Namba, 2006. Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms, IEEE Transactions on Biomedical Engineering 53(2) 273–283.
3. Eichel, P.H., E.J. Delp and K. Koral, 1998. A method for a fully automatic definition of coronary arterial edges from cineangiograms, IEEE Transactions on Medical Imaging, 7(4): 313–320.
4. Jing Wang, Jie Tian, 2001. Improved multiresolution sequential edgelinking and its application in medical image, Chinese J. Computers, 24(8): 825–829.
5. Yuqian Zhao, eihua Gui and Zhencheng Chen, 2005. Medical image edge detection based on mathematical morphology, 2005 IEEE 27<sup>th</sup> Annual International Conference of the Engineering in Medicine and Biology Society, 1(1): 6492–6495.
6. Dawei Qi, Lei Yu, 2008. Multifractal spectrum theory used to medical image from CT testing, 2008 IEEE International Conference on Advanced Intelligent Mechatronics, 1(1): 68–73.
7. Chuanyu Chang, 2004. A contextual-based Hopfield neural network for medical image edge detection, 2004 IEEE International Conference on Multimedia and Expo. 2(1): 1011–1014
8. Gudmundsson, M., E.A. El-Kwae and M.R. Kabuka, 1998. Edge detection in medical images using a genetic algorithm, IEEE Transactions on IEEE Transactions on Medical Imaging, 17(3): 469–474.
9. Li, H., K.J. Ray Liu and Shih-Chung B. Lo, 1997. Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms, IEEE Transactions on Medical Imaging, 16(6): 785–798.
10. Castleman, K.R., 1996. Digital image processing, Englewood Cliffs, NJ: Prentice-Hall.
11. Zhang, G., G. Gong and W. Zhu, 2006. Automatic video object segmentation based on clone algorithm and fuzzy mathematics, Chinese Journal of Electronics, 15(3): 482–486
12. Liu, G. and R.M. Haralick, 2000. Two practical issues in canny's edge detector implementation, the 15 International Conference on Pattern Recognition, 3(1): 676–678.

# An approach for scheduling problem on single machine

Omar Selt<sup>1</sup>, Tarak Benslimane<sup>2</sup> and Thameur Abdelkrim<sup>3</sup>

<sup>1</sup>Laboratory of pure and applied mathematics, Department of Mathematics, University of M'sila, Algeria

<sup>2</sup>Department of Electrical Engineering, University of M'sila, Algeria

<sup>3</sup>Unité de Recherche Appliquée en Energies Renouvelables, URAER, Centre de Développement des Energies

**Abstract—** This paper considers the elaboration of tabu search approach to solve a scheduling problem of  $n$  tasks on single machine. This problem is strongly NP-difficult, which makes finding an optimal solution looks impossible. To improve the performance of this approach, we used, on one hand, different diversification strategies ( $T_1$  and  $T_2$ ) with the aim of exploring unvisited regions of the solution space, and on the other hand, we proposed three types of neighborhoods (neighborhood by swapping, neighborhood by insertion and neighborhood by blocs). It must be noted that tasks movement can be within one period or between different periods. Besides that, all data in this problem are supposed to be integer and deterministic. The weighted sum of the end dates of tasks constitutes the optimization performance criterion for the problem treated in this paper.

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

## I. INTRODUCTION

It is well known that the majority of scheduling problems are of NP-difficult type, which is practically impossible to find a polynomial algorithm to solve this sort of problems [1-4]. Therefore, many research works were carried out, in this frame, for tasks scheduling problem for one machine or different machines ([5-13]).

Zribi and *al* have studied the problem  $1//N-C//\sum_{j=1}^n w_j C_j$  and

have compared two exact methods, the Branch and Bound method and the integer programming one. They have concluded that Branch and Bound method has better performance and it allows resolving instances of more than 1000 tasks. Chang et al. proposed a genetic algorithm (GA) enhanced by dominance properties for single machine scheduling problems to minimize the sum of the job's setups and the cost of tardy or early jobs related to the common due date. Selt and zitouni have studied the following problem

$(P_m // N-C // \sum_{j=1}^n w_j C_j)$ , carrying out a comparative study of

heuristic and metaheuristic for three identical parallel machines.

In this paper, the tasks scheduling problem for single machine under availability constraint is investigated, by adopting a polynomial approach with optimization of the solution in terms of execution time and cost.

## II. PROBLEM STATEMENTS

The objective, to be attained, is to determine the input sequence of the machine's tasks, such as the weighted sum of tasks' end-dates  $\sum w_j C_j$  to be minimal [1].

### A. formal statement 1

It must be noted that there is  $(n!)$  possibility to assign  $n$  tasks to the machine I.

### B. Neighborhood structure

Neighborhood determination constitutes the most important stage in metaheuristic methods elaboration. In the following part, we use three neighborhoods (neighborhood by swapping, neighborhood by insertion and neighborhood insertion by blocs).

## III. TABU SEARCH

Tabu search is the metaheuristic that keeps track of the regions of the solution space that have already been searched in order to avoid repeating the search near these areas [14]. It starts from a random initial solution and successively moves to one of the neighbors of the current solution. The difference between tabu search and other Metaheuristic approaches is based on the notion of tabu list, which is a special short term memory, storing of previously visited solutions including prohibited moves. In fact, short term memory stores only some of the attributes of solutions instead of whole solution. So, it gives no permission to revisit solutions, and then, avoids cycling and being stuck in local optima.

During the local search, only those moves that are not tabu will be examined, if the tabu move does not satisfy the predefined aspiration criteria. These aspiration criteria are used, because the attributes in the tabu list may also be shared by unvisited good quality solutions. A common aspiration criterion is better fitness, i.e. the tabu status of a move in the tabu list is overridden, if the move produces a better solution [14-16].

The process of Tabu Search (TS) can be represented as follows:

### Algorithm 1

**Step 1** Generate initial solution  $x$ .

**Step 2** Initialize the Tabu List.

**Step 3** While set of candidate solutions  $X''$  is not complete.

**Step 3.1** Generate candidate solution  $x''$  from current solution  $x$ .

**Step 3.2** Add  $x''$  to  $X''$  only if  $x''$  is not tabu or if at least one Aspiration Criterion is satisfied.

**Step 4** Select the best candidate solution  $x^*$  in  $X''$ .

**Step 5** If  $\text{fitness}(x^*) > \text{fitness}(x)$ , then  $x = x^*$ .

**Step 6** Update Tabu List and Aspiration Criteria

**Step 7** If termination condition met, then finish; otherwise, go to Step 3.

#### IV. FORMAL STATEMENT 2

It must be noted that tasks movement can be within one period or between different periods.

**Notations:** sort

We denote by :

$J = \{1, 2, \dots, n\}$  : The set of tasks.

$p_h$  : Execution time of the task  $h$ .

$I$  : Single machine

$k$  : Number of availability zones.

$Z = \{1, 2, \dots, k\}$  : Availability zones.

$E_z$  : Period of unavailability zones.

$\sigma$  : Sequence assigned to machine  $I$ .

$w_h$  : Weight of the task  $h$

$C_h$  : Execution time of the task  $h$  by the machine  $I$ .

$C_z$  ( $z \in Z$ ) : Execution time of the task  $j \in J_z$ , allocated to the zone  $z$ .

$f(\sigma)$  : Objective function cost.

$f_{swapp}$  : Swapping algorithm cost.

$f_{insert}$  : Insertion algorithm cost.

$f_{ins\_bloc}$  : Insertion par bloc algorithm cost

$f_{best}$  : Minimal cost.

T : Tabu List.

L : Tabu Liste Size

#### V. HEURISTIC

To get an initial solution, we used the following heuristic:

To assign, to the first free position, the task with the smallest ratio

$\frac{p_j}{w_j}$  among non-scheduled tasks that verify the availability constraint [17-19].

**Remark:**

The complexity of this algorithm is  $O(n \log n)$  [20].

#### A. Algorithm 2

**Initialization**

$j = \{1, 2, \dots, n\}$ ;  $E_1 = 0$ ;  $\sigma = \emptyset$ ;  $C_z = 0$ ;  $f(\emptyset) = 0$ ;

$p_j = \text{random}(1..99)$

$w_j = \text{random}(1..10)$ ;  $z = 1$

**Begin**

Sort tasks  $h \in J$  in increasing order according to the criterion

$\frac{p_j}{w_j}$  in a list  $U$

**While** ( $U \neq \emptyset$  and  $z_k \geq p_h$ ) **do**

**Begin**

Set  $p_h = p_h / w_h$  from the top list of  $U$

**Endif**

Assigned the task  $h$  to the machine I;

Delete the task  $h$  from the list  $U$ ;

Compute  $C_z = \sum p_j + E_z$ ;

Determine  $\sigma = \sigma \cup \{h\}$  and  $f_\sigma = f_\sigma + w_h C_z$ ;

**End**

**Else**

**Begin**

Set  $z = z + 1$ ;

**End**

**End if**

**End**

#### B. Algorithm 3

**Step 1** Get an initial solution  $\sigma$  and  $T[1]=0$ ;

**Step 2** Do permutation by swapping

**Step 3** Do permutation by insertion

**Step 4** Do permutation by insertion by bloc

**Step 5** Compute:  $f_1 = f_{swapp}$ ;  $f_2 = f_{insert}$ ;  $f_3 = f_{ins\_bloc}$

**Step 6** Consider  $L = \sqrt{N}$  (Tabu list size)

**Step 7** **for**  $k=1$  to  $3$  **Do**

If  $f_{init} < f_k$  Do

$T[1] = f_{init}$ ;

**else**  $T[1] = f_k$ ;

**End if**

$T_k = T[1]$ ;

**End**

**Step 7.1**  $f_{best} = \min(T_1, T_2, T_3)$

**End if**

**Step 7.2** Display  $\sigma(f_{best})$

#### Example

Consider the problem P1 with the following data:

TABLE I. 6 TASKS SCHEDULING RESULTS

$j$	1	2	3	4	5	6
$P_j$	11	36	88	10	91	31
$W_j$	3	6	8	7	4	1
$P_j/W_j$	3.66	6	11	1.42	22.75	31

Results of initial algorithm are :  $f= 2666$  ; execution time = 0,156 s

Results of tabou (swapping) are :  $f= 2145$  ; execution time = 0,991 s

Results of tabou (insertion) are :  $f= 2431$  ; execution time = 1,024 s

Results of tabou (inser. by bloc) are :  $f= 2567$  ;

execution time = 0,306 s

The best realuts are obtained by using by tabou (swapping) for  $f=2145$ .

## VI. COMPUTATIONAL ANALYSIS

### Data generation

The proposed approaches were tested on problems generated with 1000 tasks similar to those used in previous studies (Ho and Chang,1995; M'Hallah and Bulfin, 2005). For each task  $j$ , an integer processing time  $p_j$  was randomly generated in the interval (1,99) with a weight  $w_j$  randomly chosen in interval (1,10).

The table 2 below presents :

- 1-The initial mean values of objective function corresponding to initial sequence.
- 2-The initial mean values of objective function
- 3-The average times corresponding to the three neighborhoods.
- 5-The best costs.

## VII. RESULTS

The results listed in table (02) show clearly that the tabu method based on neighborhood by insertion and neighborhood by swapping presents the best (lowest) costs compared with tabu method based on neighborhood by blocks. This is due to the fact that the first neighborhoods ensures a faster tasks movement; besides that, the search space is richer with optimal partial sequences in each availability zones. This can also be explained by the nature of adopted neighborhoods. The results show that execution time obtained by the proposed neighborhoods is acceptable.

On the other hand, the heuristic amelioration rate between the three neighborhoods is remarkable (Fig.1, Fig.2).

In table 2 below, following abbreviations are used:

AC: Average costs

AT : Average time

## VIII. CONCLUSION

In this paper,  $n$  tasks scheduling problem for one machine under availability constraint is discussed, with the aim of minimizing both of execution time and cost. The approach based on tabu search allowed solving this problem, with the enhancement of initial solution obtained by a heuristic of complexity  $O(n \log n)$ . By considering three types of neighborhoods, tabu list and diversification strategy, the results of tabu search method were encouraging, and they will be more encouraging if good neighborhood based on problem's data is defined.

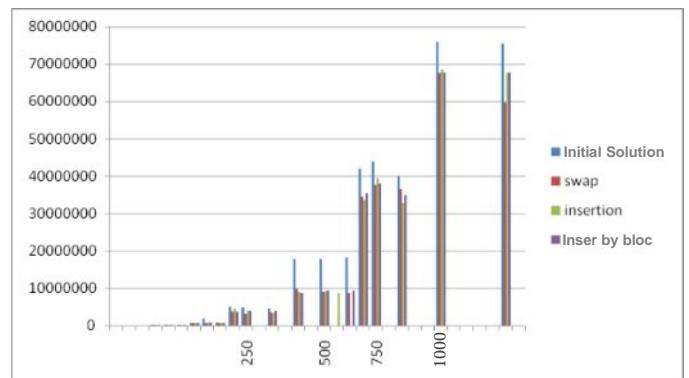


Figure 1. Histogram of heuristic cost amelioration based on tabu search for different N values.

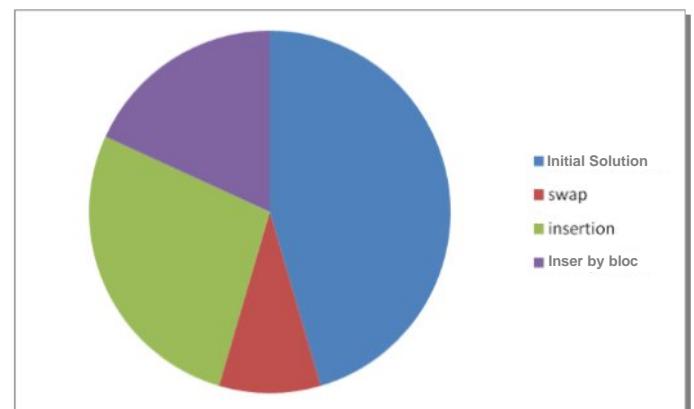


Figure 2. Circle graph of heuristic cost amelioration based on tabu search.

## REFERENCES

- [1] M. Sakarovitch , “Optimisation combinatoire : Programmation discrète”, Hermann, France, 1984.
- [2] P. Hansen, “The steepest ascent mildest descent heuristic for combinatorial programming”, Proceedings of the Congress on Numerical Methods, Capri, Italy, (2006).
- [3] M. Haouari and T. Ladhari, “Branch and bound-based local search method for the flow shop problème”, J. Oper. Res. Soc. Vol. 54. Pp. 1076-1084, 2003.
- [4] J. Ho, and YL. Chang, “Minimizing the number of tardy jobs for m paralell machines”, Eur. J. Oper. Res. Vol. 84, pp. 334-355, 1995.

- [5] C. Sadfi, "Problèmes d'ordonnancement avec minimisation des encours" Thèse PhD., Institut National Polytechnique de Grenoble, France, 2002.
- [6] N. Zribi, I. Kacem, A. El-Kamel and P. Borne. "Minimisation de la somme des retards dans un jobshop flexible", Revue e-STA (SEE). Vol. 6(6).pp. 21-25, 2005.
- [7] P. C. Chang, Shih-Hsin, Lie, Ting, Liu, Julie Yu-Chih, "A genetic algorithm enhanced by dominance properties for single machine scheduling problems with setup costs", International Journal of Innovative Computing Information and Control.Vol. 7(5A), pp. 2323–2344, 2011.
- [8] O. Selt and R. Zitouni, "A comparative study of heuristic and metaheuristic for three identical parallel machines", Cjpas. Vol. 8. pp. 3147-3153, 2014.
- [9] R. Zitouni and O .Selt, "Metaheuristics to solve tasks scheduling problem in parallel identical machines with unavailabilite periodes", RAIRO. Oper. Res. Vol. 50(1), pp. 83-90, 2016.
- [10] C. Y. Lee, "Machine scheduling with an availability constraints", J. Global Optim., Vol. 9, pp. 395-416, 1996.
- [11] P. King-Wah, "A genetic algorithm based heuristic for two machine nowait flowshop scheduling problems with class setup times that minimizes maximum lateness", International Journal of Production Economics,Vol. 141, pp. 127–136, 2013.
- [12] C. Y. Lee, "Minimising the makespan in two machines flow shop scheduling problem with availability constraints", Oper. Res. Lett. Vol. 20, pp. 129-139, 1997.
- [13] L. Yun-Chia, H.Yu-Ming, and Chia-Yun, "T Metaheuristics for drilling operation scheduling in Taiwan PCB industries", Int. J. Prod. Econ., Vol. 141(1), pp. 189-198, 2013.
- [14] F. Glover, "Future paths for integer programming and links to artificial intelligence", Comput. Oper. Res.Vol. 13, pp. 533-549, 1986.
- [15] F. Glover, and S. Hanafi, "Tabu Search and Finite Convergence", Special Issue on Foundations of heuristics in Combinatorial Optimization. Discrete Appl. Math.Vol 119, pp. 3-36, 2002.
- [16] G. Abdenacer, "A comparative study of metaheuristic algorithms for solving quadratic assignment problem", IIACSA. Vol. 5, N. 1, 2014.
- [17] C. Y. Lee, "Two machines flow shop scheduling problem with availability constraints", European J. Oper. Res. Vol 114, pp. 420-429, 1999.
- [18] R. M'Hallah, and RL.Bulfin, "Minimizing the weighted number of tardy jobs of paralell processors", Eur. J. Oper. Res. Vol. 160, pp. 471-4847, 2005.
- [19] G. Schmidt, "Scheduling with limited machine availability", European J. Oper. Res. Vol. 121, pp. 1-15, 2000.
- [20] W. E. Smith, "Various optimizes for single-stage production", Naval Res. Logist. Vol. 3, pp. 59-66, 1956.

TABLE II. RESULTS OBTAINED BY HEURISTIC AND TABU SEARCH

N	Initial Solution (average of 3 instances)		Tabu search by swap		Tabu search by insertion		Tabu search by blocs		Best costs
	AC	AT (second)	AC	AT (second)	AC	AT (second)	AC	AT (second)	
N=20	29190	0,185	41647	0,654	31779	0,307	35466	0,13	29190
	45768	0,201	40772	0,578	29096	0,224	32259	0,166	29096
	30731	0,194	29720	0,576	37763	0,447	33526	0,161	29720
N=50	205130	0,583	189551	0,634	223921	1,524	176484	0,328	176484
	207358	0,429	218017	0,603	219091	1,019	229906	0,437	218017
	214071	0,919	209126	0,645	194182	0,973	208584	0,437	194182
N=100	734976	2,212	682309	3,682	593040	5,179	707554	1,617	593040
	1994099	2,191	702905	3,385	786843	5,356	761648	1,472	702905
	839684	1,997	707977	3,297	685365	5,808	703850	1,700	685365
N=250	5090272	5,90	3845619	7,73	4659201	9,18	3697427	6,56	3697427
	4897215	6,14	3201037	7,61	4061839	9,63	3982569	6,49	3201037
	4658617	6,142	3597109	6,95	3312510	8,89	4045917	6,92	3012010
N=500	17866722	8,120	9730402	9,789	8964538	10,876	8763401	7,871	8763401
	17986067	7,128	9056321	9,562	9254175	10,501	9342104	7,549	9056321
	18410307	7,298	8657831	9,861	8765109	10,612	9297364	7,724	8657831
N=750	41931982	12,20	34537327	15,87	33762437	16,87	35318023	14,73	33762437
	43858415	13,01	37612943	16,08	39576182	16,49	38003183	13,98	37612943
	39895222	12,44	36578139	15,83	32789193	16,81	34987710	14,01	32789193
N=1000	75763079	16,80	67451280	20,563	68673189	25,675	67645329	17,977	67451280
	75377034	17,12	59672815	20,986	67563821	25,560	67832961	18,286	59672815
	74190765	17,44	66347819	20,926	58976897	25,241	68936103	18,672	58976897

# RICA: Reform based Imperialist Competitive Algorithm for Mapping Applications to Network on Chip based, Many-core architectures

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

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

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

**Keywords-** Imperialist competitive algorithm, many-core processor, task mapping

## I. INTRODUCTION:

The increase in the number of IP cores in SoC and also the high optimality and efficiency required by modern applications have caused the old bus- or point to point-based methods to be replaced by a new communicational method. The method which is called NoC, was proposed for a concurrent execution of the components of the system and to achieve more efficiency, less energy consumption and optimization increase[3].

The architecture of NoC consists of a combination of the network with some nodes, each of which contains a source and a switch or router. The switch navigates the traffic in the network and the source can be everything similar to memory (for storing data), or a processor which processes the data. The most common network topology is the mesh which consists of m rows and n columns.

## II. Related works

The mapping of an application on NoC architecture is a NP-hard problem. If the number of cores in architecture equals the number of applied cores, there are  $n!$  Different solutions. So far, different solutions with different purposes have been proposed for this

problem[1,4,5-17,19-21]. Some of these methods have dealt with reducing the performance time [9,11,14,15] and used some of the other optimization techniques like Integer Linear Programming (ILP) [20], innovative methods [7, 16, 21], simulated annealing [10], boundaries and branches [5] and genetic algorithm [1, 4, 13, 19] to decrease the communication cost or the energy consumed by the system. Some of these methods [6, 12, 20] have too much complexity and calculation time and some of them [1, 4, 10, 13, 19] don't guarantee the best result and it is possible to not reach a desirable answer.

CastNet [21], ILP [20], MOCA [16], NMAP [12] and GAMR [4] are the selected solutions that will be compared with our approach.

## III. Energy model

In [22], the criterion offered for the energy consumed by the router is the energy that is used for passing a bit of datum through router. Also, in [18], a model for bit energy is presented which is appropriate for the architecture of 2D mesh based NOC. In this model, the consumed energy for sending a bit of datum from i vertex to j vertex is shown by the Equation-1

$$E_{\text{bit}}^{i,j} = (h_{i,j} + 1) \times E_{\text{Rbit}} + h_{i,j} \times E_{\text{Lbit}} \quad (1)$$

In this formula, ERbit and ELbit are the consumed energy of the router and the conjunction, respectively. The variable  $h_{i,j}$  shows the distance between the two routers I and j that, depending on the number of route finders, is in the shortest route. Since the 1 formula is a linear function of variable of  $h_{i,j}$  and the constants of ERbit and ELbit, we can claim that minimizing the average of distances can result in minimizing the consumed energy. In a mesh based NoC,  $h_{i,j}$  equals the Manhattan distance between the vertex  $i(x_i, y_i)$  and the vertex  $j(x_j, y_j)$  that is shown in Equation-2

$$h_{i,j} = |x_i - x_j| + |y_i - y_j| \quad (2)$$

#### IV. Definitions

Now for a better analysis of the problem, some definitions and hypotheses are presented:

Graph of the core communication is a directed graph that is shown by CCG (C, A). Every vertex of  $c_i \in C$  indicates a core of IP and every edge shows a connection between IP  $c_i$  and IP  $c_j$ . The weight of each edge that is shown by  $comm_{i,j}$ , is the required width for the connection between two vertexes of  $c_i$  and  $c_j$

Architectural graph of NOC is an directed graph which is indicated by NAG(R, P). Each vertex of  $r_i \in R$  indicates a source vertex in architecture and each edge of  $p_i, j \in P$  shows a connection route between  $r_i$  and  $r_j$ . The weight of a edge that is shown by  $e_{i,j}$ , indicates the average consumed energy for sending a bit of datum from  $r_i$  to  $r_j$

**Core mapping:** The function of mapping of the connection core graph CCG(C, A) to the architectural graph of NOC, NAG(R,P) is one by one and is shown by map (C ) in Equation-3 :

$$\begin{aligned} \text{map} :& \rightarrow R \Rightarrow r_i = \text{map}(C_i) \forall_{c_i} \\ & \in C \exists r_i \in R, \forall c_i \in C, r_i \neq r_j \end{aligned} \quad (3)$$

The mapping is defined when  $|C| \leq |R|$ . The goal of mapping function is to find the amount of Equation-4:

$$\text{mix} \left\{ \sum \forall_{a_i, j} comm_{i,j} \times e_{\text{map}}(c_i), \text{map}(C_j) \right\} \quad (4)$$

In addition to the different goals that exist in the designing network on chip, different algorithms and methods have been presented to solve these problems. Since the mapping problem is one of NP\_hard problems and obtaining the best answer maybe very expensive, most of the presented algorithms are heuristic. Some of these methods are presented on the basis of genetic algorithm and have dealt with solving the mapping problem by making some changes in it [4, 8, 9, 13, 14, 17, 19].

For solving most of optimizing problems, we often use evolutionary methods such as genetic algorithm, optimization of ant's colony, assimilated re-cooking. Genetic algorithms are the most known evolutionary algorithms and nowadays have been applied for many applications and systems. What is obvious is that the mental and cultural evolution of mankind is so quicker than his genetic and physical evolution. Therefore, the cultural and attitudinal evolution of humankind, also, is not ignored and a group of algorithms, called cultural algorithms have been introduced. In fact, Cultural algorithms are not a totally new class of algorithms, but the basic idea is that these algorithms expectedly increase their convergence rate by adding the capability of cultural evolution (by enhancing the possibility of information exchange among members of the population) to the existing algorithms.

#### V. ICA

A new algorithm in the field of evolutionary computations which has been founded based on social and political evolution of human being, is the imperialist competitive algorithm and its steps are depicted in Fig. 1 Imperialist competitive algorithm is suitable for non discrete functions. In other word, parameters of a country can have floating-point or duplicate values.

Such values are not appropriate for the problem of task mapping, because each parameter of the country is an indicator of the location of an IP in NOC architecture that should be a positive integer. In addition, regarding our definition of the problem, the parameters of a country cannot have duplicate values. For matching the Imperialist competitive algorithm with the problem of task mapping, we should make some changes in the steps of production of primary countries, matching policy and revolution in order to prevent the production of inappropriate values.

#### VI. Proposed Approach

We use the ICA for mapping applications on the mesh based NoC via following steps:

**Production of primary countries (initializing):** The parameters of a country must have integer and non-recurring values. These values should be in a certain range that is dependent on the size of NoC. Number of parameters equals the number of vertexes in CCG application graph. For example, Fig. 2-4, respectively shows MPEG4 application graph, the randomly generated country and the architecture of network on the chip on which the graph is supposed to be mapped.

As you can see, the graph related to MPEG4 application has 12 vertexes (Fig. 2) and for this reason, the country shown in Fig. 3 has 12 parameters. The existing values in parameters are in a range from 1 to 16 which depends on the NoC dimensions. The existing values in parameters are indicative of a place in the architecture of NoC. For example, the existing value in the first parameter of the country is 3 which mean the first node of the MPEG4 graph should be placed in the third tile of NOC. Similarly, the second node should be placed in the fifth tile and you can see these cases in Fig. 4.

**Matching policy:** As mentioned in the introduction of imperialist competitive algorithm, in the phase of matching, countries are moved towards the colony. At this step, we apply the following changes to prevent floating-point values:

A random number (n) is selected in a range from zero to the number of parameters of the country.

A parameter from the imperialist is randomly selected and copied in the corresponding location of the colony (Fig. 5)

The duplicated value generated in the colony is modified (Fig. 6)

The 2 and 3 steps are repeated n times.

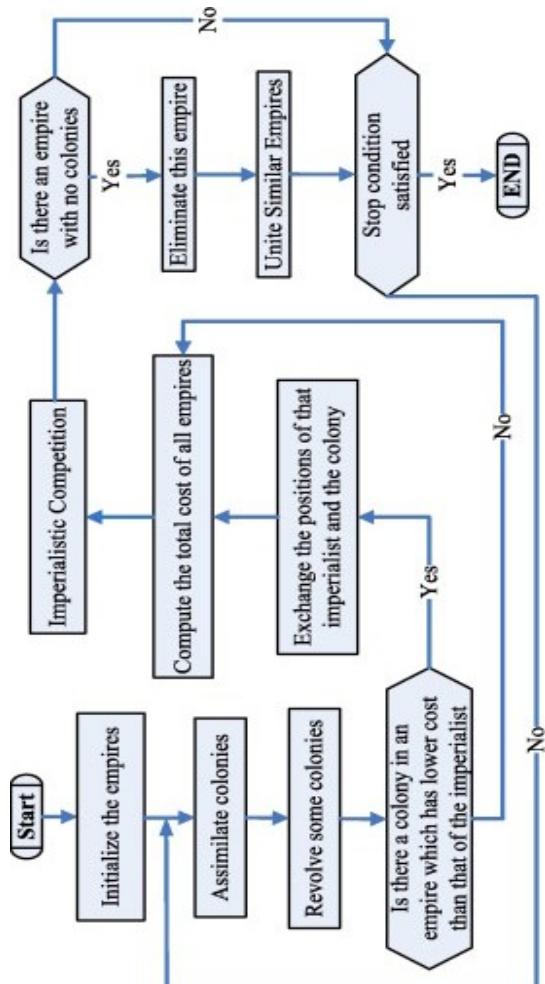


Figure.1: Flowchart of Imperialist competitive algorithm [2]

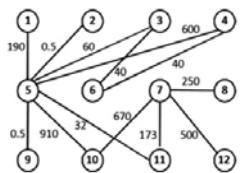


Figure. 2: Application graph of MPEG4



Figure. 3: Example of a country

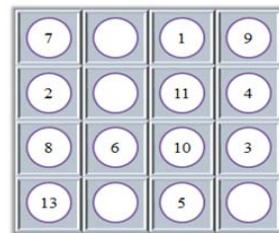


Figure. 4: Mapped country to the related NoC



Figure. 5: Coping parameter from empire to country



Figure. 6: Exchange the duplicated value and eliminated one in the country

## VII. Revolution or reform

In a revolution, sudden changes in the parameters of a country occur. This step can be implemented in different ways. By observing the developments in the countries and their results we can reach to the conclusion that creating a revolution in a country imposes large expenses to the country and compensation for these losses may take years. If these changes happen in a gradual manner, not only such costs are prevented, but the country can gradually reach a better position by reforming in various fields. Therefore, to implement this step of the Imperialist competitive algorithm, instead of suddenly changing all the parameters of parameters of a country, as you can see in Fig. 7, we randomly choose two parameters of the colony and replace their values. By examining the results obtained from the implementation of both methods (the reform and revolution), we observe that with the development of revolution and its progress in the country, the time of achievement increases, while making reforms and its progress accelerates the convergence process and a better achievement (Table 1 and 2).

To compare the results of the Imperialist competitive algorithm, we implement some commonly used applications and finally, we compare the consumed energy with other algorithms. The applications and their details are in Table 3. In the first two graphs which

were taken from [11], the weight of edge is based on MB/S and in the other three graphs; it is in terms of KB/sec [15].

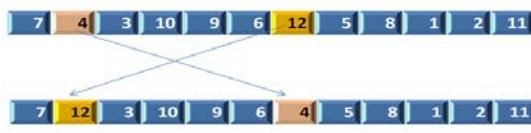


Figure 7: Reform process

Table 1: Comparison of revolution and reform for MPEG4

Execution number	Reform		Revolution	
	Consumed energy	Decade length	Consumed energy	Generation length
1	3685.7	17	3660.1	37
2	3619.3	12	3739.1	26
3	3619.3	11	3619.3	42
4	3619.3	18	3685.7	46
5	3619.3	9	3682.8	46

Table 2: Comparison of revolution and reform for VOPD

Execution number	Reform		Revolution	
	Consumed energy	Decade length	Consumed energy	Generation length
1	4092.8	17	4143.4	50
2	4092.8	25	4086.5	47
3	4086.5	36	4167.4	24
4	4072.0	10	4192.7	44
5	4082.8	13	4192.7	47

In Fig 8 you can see the two applications of VOPD and MPEG4 and the NOC resulting from their mapping with the Imperialist competitive algorithm.(Fig8-a,c shows the graph related to MPEG4 and VOPD. Fig8-b, d shows NoC from the mapping with RICA algorithm).

The initial conditions of imperialist competitive algorithm:

- Technology: 100 nm
- Consumed energy in the input port of router : 328 nj mb/sec□1
- Consumed energy in the output port of router: 65.5 nj mb/sec□1
- Consumed energy of Physical link: 79.6 nj mb/sec□1 [18]
- Topology of NoC: 4×4 mesh
- The number of decades: 50
- Rate of revolution: 90%
- The number of primary empires: 10
- Development Environment: MATLAB

Note that the number of initializing countries: depended on the kind of application for example for the MPEG4 we define 500 countries.

Table 3: Application graph characteristics

Application	Number of the vertexes	Number of the edges
VOPD	16	20
MPEG4	12	13
263Dec	14	15
263Enc	12	12
MP3Enc	13	13

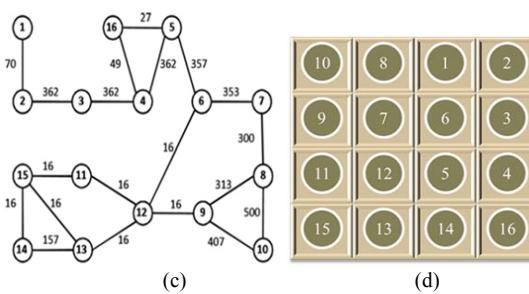
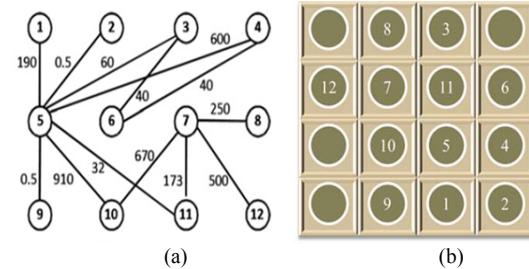


Figure 8: RICA results for mapping MPEG4 and VOPD

It is noteworthy that regarding the type of the implemented application to achieve the best result we have increased the number of countries to the extent that this rise does not have a significant influence on the final result. Of course, the response time decreases with increasing primary countries. For example, in MPEG4 application, with 500 countries we can reach to a result in 10-20 decades, with 1000 countries in 5-10 decades.

## VIII. RICA performance evaluation

At this time we compare the results of RICA with algorithms of CastNet [21], ILP [20], MOCA [16] and NMAP [12].

We implement these algorithms at the same situation which is defined in previous section and you can see in Fig. 9, the energy diagram of these two applications (MPEG4 and VOPD) in different algorithms.

In the diagram, the consumed energy of different applications caused by different algorithms is presented in terms of  $\mu_j$ .

As you can see in Table 4 and Fig. 9, in most Implemented applications, the results obtained from of the RICA algorithm is better than those from the other algorithms.

We will have a separate comparison between the algorithms RICA and GAMR (based on genetic algorithms) [4].

The reason is that by repeating the execution of these two algorithms and/or changing the parameters of algorithms, we can come to the same final conclusions, but these two methods can be compared from other aspects such as convergence time.

Before presenting the results of two algorithms, it is noteworthy that both of the algorithms are belong to optimization algorithms and changing the parameters of the algorithm and various executions of these algorithms will follow different results due to having different initial populations that are randomly generated.

Initial conditions of GAMR algorithm have been determined with mutation rate of 1% and the rate of crossover 90% like RICA algorithm.

In Table 5, you see a comparison between the results obtained from the application mapping of MPEG4 with execution of each algorithm for 5 times and in Table 6, you are provided with the results of the execution of the application VOPD. The number of the initial population in both algorithms for the first application is 500 and for the second application is 1000.

**Table 4:** Consumed energy in micro jowl ( $\mu\text{j}$ )

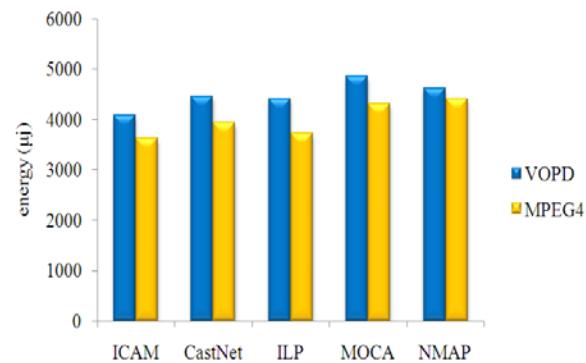
Graph	RICA	CastNet	ILP	MOCA	NMAP
VOPD	4072.60	4456.23	4413.27	4854.60	4623.43
MPEG4	3619.30	3933.67	3719.27	4311.00	4398.98
263Dec	18.33	20.43	20.43	20.43	20.43
263Enc	236.27	236.24	236.24	236.24	274.70
MP3Enc	17.26	17.75	17.75	19.17	17.75

**Table 5:** Comparison of RICA and GAMR implementation for MPEG4 application graph in micro jowl ( $\mu\text{j}$ )

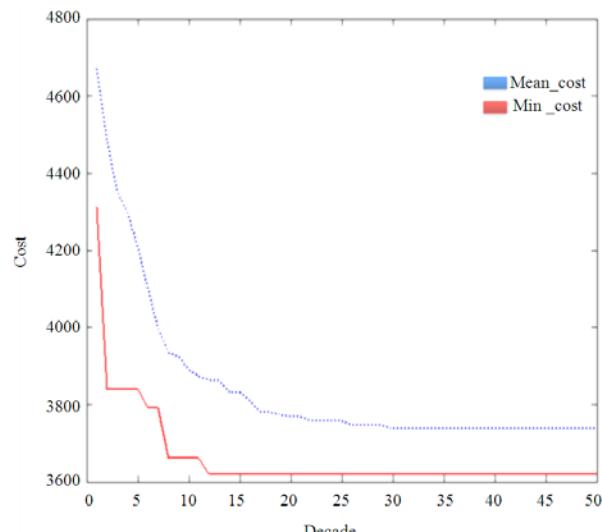
Execution number	RICA		GAMR	
	Consumed Energy	Decade length	Consumed Energy	Generation length
1	3619.6	16	3670.2	39
2	3619.3	11	3619.3	47
3	3619.3	16	3740.4	31
4	3660.1	19	3619.9	19
5	3619.3	10	3619.3	36

**Table 6:** Comparison of RICA and GAMR implementation for VOPD application graph in micro jowl ( $\mu\text{j}$ )

Execution number	RICA		GAMR	
	Consumed energy	Decade length	Consumed Energy	Generation length
1	4092.8	27	4112.7	24
2	4086.5	25	4123.2	44
3	4086.5	36	4147.2	43
4	4072.6	10	4113.1	39
5	4082.7	13	4123.2	47



**Figure 9:** The diagram of comparing the consumed energy



**Figure 10:** Average and minimum cost for MPEG4 in RICA

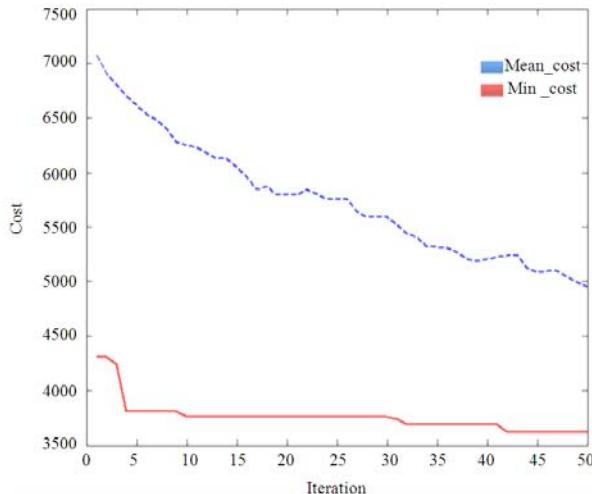


Figure.11: Average and minimum cost for MPEG4 in GAMR

By repeating the genetic algorithm or increasing the initial population and the number of generations we can come to conclusions similar to those from the execution of the RICA, but at the same conditions, RICA succeeds in fewer decades than the generations of the genetic algorithms and its results are better too.

In Fig. 10, the diagram of the average and minimum cost of RICA algorithm for MPEG4 application has been depicted and Fig. 11 shows the diagram of the average and minimum cost in GAMR algorithm for the same application. As you see the convergence time in the RICA is less than the GAMR.

## IX. Conclusion

This study presents a new method for task mapping on chip network aimed to reduce energy consumption. This method (RICA) has been formed based on imperialist competitive algorithm. Since imperialist competitive algorithm has been basically proposed for solving constant optimization problems, for its application in solving the mapping problem, we need to do some changes and modifications in different steps of the algorithm. To do so, there have been some changes in the stages of producing primary countries, absorption policy (matching) and revolution and a discrete version of this algorithm was presented.

Results of the RICA implementation prove that this approach is more efficient than other methods which are based on the similar optimization algorithms such as genetic algorithm. Also according to the experiences and obtained results it seems that imperialist competitive algorithm could be useful for other goals of NoC design such as communication scheduling, network assignment and routing path allocation.

## X. Acknowledgment

Special thanx to the research deputy of Islamic Azad University(Gorgan branch), for their support of our project.

## XI. References

- [1] Ascia, G., V. Catania and M. Palesi, 2004. Multi-objective mapping for mesh-based NoC architectures, CODES-ISS'04,
- [2] Atashpaz-Gargari, E.C., 2007. Lucas, imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic. Proceeding of the IEEE Congress Evolutionary Computation, pp: 4661-4667.
- [3] Dally, W.J. and B. Towles, 2001. Route Packets, not wires: On-chip Interconnection Networks. Proceedings of the on Design Automation Conference, Las Vegas, Nevada, USA, pp: 684-689.
- [4] Fen, G. and W. Ning, 2010. Genetic algorithm based mapping and routing approach for network on chip architectures. Chinese. J. Electron., 19: 1.
- [5] Hu, J. and R. Marculescu, 2003. Energy-aware mapping for tile-based NoC architectures under performance constraints. Proceedings of the Asia and South Pacific Design Automation Conference, Kitakyushu, Japan, pp: 233-239.
- [6] Hu, J. and R. Marculescu, 2003. Exploiting the routing flexibility for energy/performance aware mapping of regular NoC architectures. Proceedings of the Conference on Design, Automation and Test in Europe,
- [7] Janidarmian, M., A. Khademzadeh and M. Tavanpour, 2009. Onyx: A new heuristic bandwidth-constrained mapping of cores onto tile-based Network on Chip, IEICE Electron. Express pp: 1-7.
- [8] Jena, R.K. and G.K. Sharma, 2007. A multi-objective evolutionary algorithm based optimization model for network-on-chip synthesis. Proceeding of the international Conference Information Technology Las Vegas, Nevada, USA, pp: 977-982.
- [9] Lei, T. and S. Kumar, 2003. A two-step genetic algorithm for mapping task graphs to network on chip architecture. Proceeding of Romicro Symposium Digital System. Design, Belek-Antalya, Turkey, pp: 180-187
- [10] Marcon, C., A. Borin, A. Susin, L. Carro and F. Wagner, 2005. Time and energy efficient mapping of embedded applications onto NoCs, ASP-DAC'05
- [11] Mehran, A., A. khademzadeh and S. Saeidi, 2008. DSM: A heuristic dynamic spiral mapping algorithm for network on chip. IEICE Electr. Express, 5: 464-471.
- [12] Murali, S. and G. De Micheli, 2004. Bandwidth-constrained mapping of cores onto NoC architectures. Proceeding of the DATE, IEEE Computer Society, pp:
- [13] Nedjah, N., M. Silva and L. Mourelle, 2010. Customized computer-aided application mapping on NoC infrastructure using multi-objective optimization. J. Syst. Architecture,
- [14] Ruxandra, P., 2011. Mapping concurrent applications to multiprocessor systems with multithreaded processors and network on chip-based interconnections. BS Thesis, Linköping University, Sweden,
- [15] Shi, Z. and A. Burns, 2010. Schedulability analysis and task mapping for real-time on-chip communication, Springer Science + Business Media, LLC,
- [16] Srinivasan, K. and K.S. Chatha, 2005. A technique for low energy mapping and routing in network-on-chip architectures,

Proceedings of the 2005 International Symposium on Low Power Electronics and Design, San Diego, CA, USA,

- [17]Srinivasan, K. and K.S. Chatha, 2005. ISIS: A genetic algorithm based technique for custom on-chip interconnection network synthesis, Proceedings of the International Conference on VLSI Design, Washington, DC, USA, pp: 623-628
- [18]Srinivasan, K., K.S. Chatha and G. Konjevod, 2006. Linear-programming-based techniques for synthesis of network-on-chip architectures. IEEE Trans. Very Large Scale Integr. Syst., 407-420.
- [19]Tavanpour, M., A. Khademzadeh, S. Pourkiani and M. Yaghobi, 2010. GBMAP: An Evolutionary Approach to mapping cores onto a mesh-based NoC architecture. J. Commun. Comp.,
- [20]Tosun, S., O. Ozturk and M. Ozen, 2009. An ILP formulation for application mapping onto network-on-chips. Proceeding of the 3rd International Conference on Application of Information and Communication Technologies, AICT2009, Baku,
- [21]Tosun, S., 2010. New heuristic algorithms for energy aware application mapping and routing on mesh-based NoCs. J. Syst. Architecture,
- [22]Ye, T.T., L. Benini and G. De Micheli, 2002. Analysis of power consumption on switch fabrics in network routers. Proceedings of the IEEE Design and Automation Conference, New Orleans, LA, USA, pp: 524-529.



**Alireza Mahini**, received the Master degree in computer systems architecture in 2006 from Iran University of Science and Technology (IUST), Iran. Now he is a PhD candidate in the Islamic Azad University, sciences and researches branch. He is academic staff of Islamic Azad University, Gorgan branch and president of Gorgan SAMA College. His current researches focus on distributed systems, high performance IP route lookup, network processor architecture and Network on Chip.



**Dr. Hossein Pedram**, received PhD in computer Engineering (1992), Electrical and Computer Engineering Department, Washington State University, Pullman, Washington, USA. Now he is Associate Professor, of Computer Engineering & IT Department in Amirkabir University of Technology (Tehran Polytechnic), Iran. His current researches focus on Computer Architecture, Asynchronous Design, and Network on Chip, Innovations in Computer Architecture, Distributed Systems and Networking.

# An overview of Service Oriented Architecture, Cloud Computing and Azure Platform

Kamran Shaukat, Muhammad Umair Hassan, Haider Ali, Muhammad Shah Zaib, Muhammad Muhibb Ullah

Department of Information Technology,  
University of the Punjab, Jhelum Campus  
Jhelum, Pakistan

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

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

## I. INTRODUCTION TO CLOUD COMPUTING

One of the most recent approach in small and enterprise-sized IT industry is the critical requirement for a fundamental change in the IT architecture. Cloud computing gives a noteworthy movement to the IT industry. Cloud technology requires the high scalability, rapid provisioning, and virtualized environments and is driven by internet service. In this era, the main reason to adopt cloud over standardized IT deployment is its stability, flexibility, reliability, rapid provisioning, and scalability [1].

The concept of cloud computing can trace its roots back to grid computing that provides rapid provisioning of resources. It is not mandatory that grid computing should be in cloud; actually it depends on the types of users, whether they are clients or administrators. Grid computing requires software that can be divided and computed or serviced on a single or multiple systems [2].

### A. Grid Computing

A grid is made up of a number of resources and layers with different levels of implementation. A computational grid takes a somewhat different approach to high performance computing. A grid typically consists of a collection of heterogeneous and geographically distributed machines [2]. There are many types of grid that are organized as:

- Information Grids: These are aimed to provide an efficient and simple access to data without worries about platforms, location and performance [3].
- Compute Grids: These exploit the distributed collection of systems in the perspectives of processing power.
- Services Grid: These provide the scalability and reliability across different servers.

Conceptually we can imagine three layers in grid, the lower being the physical one. In the second, there are different operating systems and upper layer is the application.

Thus, the Grid Computing is comprised of those applications that are used for computational problems which lies in parallel networking environment.

### B. Grid-The Way to Cloud

As previously said, the cloud traces its roots back to grid computing but it is not mandatory that grid should be mentioned in cloud computing. The following subsections will help us to learn the benefits of grid when it is deployed for cloud computing.

- Storage/Data/Information: Provides logical views of data without having to understand where the data is located or whether it is replicated.
- System Management: Defines, controls, configures, and removes components and/or

services on a grid using automated or physical methods.

## II. WHY CLOUD COMPUTING?

The cloud computing has acquired a great option for remote services as well as for local services. There are many advantages of cloud for which we are using its services in our daily life. It is thus taking part in health sector, education sector, and government organizations. This model changing the way the infrastructure these organizations are using this model [4].

Although, the cloud is being used by many IT organizations. Some of them are Google, Amazon and Microsoft. These organizations approach to cloud which are discussed in this paper.

### A. Google and Cloud Computing

Google is known for its growing services including their exceptionally well known internet searcher, email, profitability, and mapping applications. Google has distributed numerous adaptable, and superior assortment of research papers on applications. We will discuss now the BigTable being used by google.

*BigTable:* Google's another way to deal with distributed computing is their Bigtable system of information storage [5]. In numerous regards, Bigtable externally looks like a relational database management system (RDBMS). Both store information in plain shape with named lines and sections, and they permit information to be sought utilizing the line name (and conceivably the segment name) as keys.

### B. Amazon and Cloud Computing

Amazon is known for their offering books on the web, however they are additionally effectively putting resources into administrations that permit engineers to exploit their figuring innovation. Amazon Web Services give engineers utilization of open APIs to get to Amazon's boundless foundation in a way they use services. Software designers, new businesses, and organizations needing dependable figuring force are individuals from an extensive and developing group utilizing Amazon services and that is Amazon Elastic Compute Cloud (EC2) [6]. The EC2 gives virtualization to engineers that they can load Amazon machines by favored custom applications and programming situations.

### C. Cloud in Health Sector

Social insurance associations like Healthcare organizations (HCOs) are relied upon to give as good as ever patient consideration abilities while all the while constraining medicinal services cost increments. In any case, regardless of the critical points of interest for the use of grid computing as a component of Healthcare IT (HIT), security and protection, dependability, coordination and information versatility are a percentage of the huge difficulties and obstructions to execution that are in charge of its moderate appropriation [7].

*Benefits of Cloud Technology in Health:* Cloud computing offers critical advantages to the medicinal services area: doctor's facilities, specialist's centers, and wellbeing facilities. In addition, human services information should be shared crosswise over different geologies and settings which assist the social insurance supplier and the patient bringing on remarkable deferral in treatment and loss of time [7].

## III. CLOUD AND DYNAMIC INFRASTRUCTURE

Using the cloud services, users can access standard IT resources for the deployment of new applications, and computing resources rapidly without re-engineering their entire infrastructure, thus cloud is dynamic. Cloud dynamic infrastructure is based on an architecture that have the following combination of terms initially:

- Service Management
- Security
- Virtualization and Consolidation
- Energy Efficiency
- Resilience
- Asset Management
- Information Infrastructure

We have not purposed the further explanation of the described dynamic infrastructures. As people only need to use the cloud services not to re-build the whole application or infrastructure, thus cloud is dynamic in all respects.

## IV. SERVICE ORIENTATION IN CLOUD

Cloud computing has a core model which is known as service orientation. This methodology embraces the idea of services as a system development and application. It supports the development of rapid, low cost, interoperable, evolvable, and flexible applications. The service orientation arises the two concepts which are Quality of Service (QoS) and the other one is Software as a Service [8]. These are further explained in the following context.

- Quality of service: This indicates a section of those attributes that are functional and non-functional and that we can use to assess the service in numerous point of views.
- Software as a Service: it is a delivery model for the applications and have been acquired from the universe of ASPs. This is used to provide software based solutions over large area network from a central data center and thus making them accessible on rental or subscription basis.

## V. SERVICE ORIENTED ARCHITECTURE

SOA is a way to deal with design that is expected to advance adaptability through epitome and free coupling. SOA capacities are characterized and uncovered as "administrations" and there is one and only occasion of every administration usage, either at every administration, for instance conversion scale figuring,

sent in one spot and one place just, and is remotely conjured by anything that necessities to utilize it.

There are two noteworthy parts inside SOA: the service consumer and the provider. The service provider is maintainer of the administration, and the association that makes accessible one or more administrations for others to utilize. The service consumer can find the administration metadata in the registry, and builds up the required customer segments to tie and utilize the administrations. Inside this context the contracts of service, the standards of service, and interoperability play their role [9] for enterprises and that principles are:

- Standardized Service Contract
- Discoverability
- Abstraction
- Lake Of State
- Loose Coupling
- Reusability
- Autonomy
- Composability

Together with these standards, their assets manage the utilization of SOA for enterprise application integration. Likewise demonstrating systems and philosophy, for example, the service oriented orientated modeling framework (SOMF), presented by the Organization for Advancement Structured Information Standards (OASIS) [10], give intends to viably acknowledging service oriented architecture.

## VI. COMPUTING TECHNOLOGIES AND PLATFORMS

Cloud applications are improved by utilizing stage and structures that give diverse sorts of services, and these kind of services are described below:

### A. Amazon Web Services (AWS)

AWS offers comprehensive cloud IaaS services, ranging from virtual compute, storage, and networking to complete computing stacks. AWS is mostly known for its compute and storage on demand services, namely Elastic compute cloud (EC2) and simple storage services (S3).

### B. Google AppEngine

Google AppEngine is a scalable runtime environment mostly devoted to executing Web applications. These take advantage of the large computing infrastructure of google to dynamically scale as the demand varies over time. AppEngine provides both a secure execution environment and a collection of services that simplify the development of scalable and high-performance Web applications.

### C. Microsoft Azure

Microsoft azure is a cloud operating system and a platform for developing applications in the cloud. It provides a scalable runtime environment for web applications and distributed applications in general. Applications in Azure are organized around the concept of

roles, which identify a distribution unit for applications and embody the applications logic.

### D. Hadoop

Apache Hadoop is an open source system that is suited for handling huge information sets on ware equipment. Hadoop is an execution of MapReduce, an application programming model created by google, which gives two central operations to information handling: map and reduce.

### E. Manjrasoft Aneka

Manjrasoft Aneka is a cloud application platform for rapid creation of scalable applications, and their development on various types of clouds in a seamless and elastic manner. It also supports the framework of application development and a runtime environment that we can implant on hardware equipment like networked PCs, clusters, and cloud resources.

### F. Force.com

It is a platform of cloud computing on which we can make social applications. This service is provided by the SalesForce.com and with the coordination of Software as a Service.

## VII. AZURE THE CLOUD APPLICATION PLATFORM

With the approach of the distributed computing stage, we are encountering an outlook change in the improvement range, where cloud abilities are additionally considered while outlining and fabricating new applications. The key levers impacting this movement in facilitating applications from on-reason to cloud are adaptability, versatility, decreased capital costs, 24/7 accessibility, geological get to and lower aggregate expense of proprietorship. Then again, it is clear that not all applications are prepared to be put in the cloud. Rather regular, non-mission basic applications that are not driven by strict consistence or administrative needs are the most prompt arrangement of utilizations that can be moved to the cloud. Microsoft's Windows Azure Platform gives a well-known and adaptable environment to drive and bolster particular needs and administrations of the advancement group, and clients. The Windows Azure stage gives a uniform ordeal as it empowers engineers and clients to utilize existing Microsoft advancements to create or utilize applications on-reason or in the cloud [11].

### A. An Overview of the Windows Azure Platform

Windows Azure is of cloud developments innovation, every giving a specific course of action of organizations to application architects. The figure also describes that Window Azure can be used on both running applications and on premise applications both. Following are the Windows Azure platform parts:

- Windows Azure: Provides a Windows-based environment for fleeing data on servers in Microsoft server farms.
- SQL Azure: Provides data organizations in the cloud in perspective of SQL Server.

- Windows Azure platform AppFabric: Provides cloud organizations for joining applications running in the cloud or on premises.

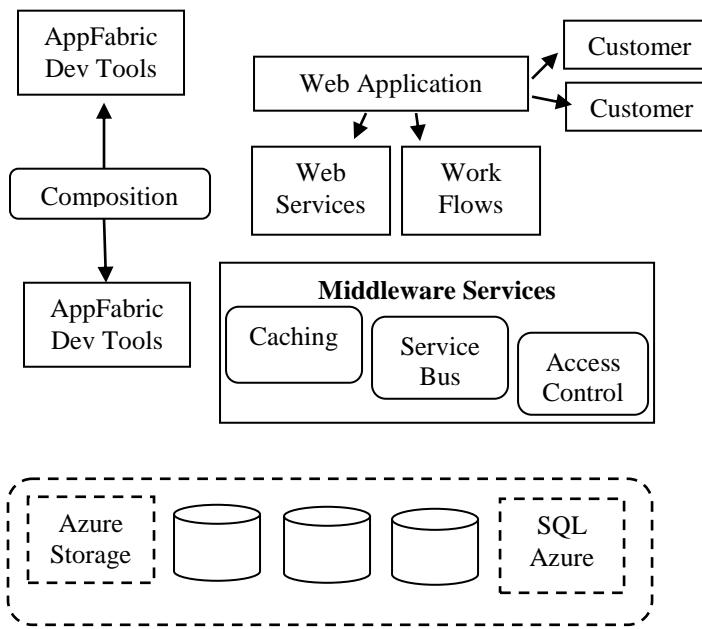


Figure 1

All aspects of the Windows Azure stage has its own part to play.

*1) Window Azure:* At an abnormal state, Windows Azure is anything but difficult to grasp: It's a phase for running Windows applications in addition, securing their data in the cloud. Figure 1 demonstrates its basic sections. Windows Azure continues running on a considerable numerous machines as shown in figure, all arranged in Microsoft data centers and accessible by method for the Internet. Windows Azure register and limit organizations depend on top of this fabric [12].

Both Windows Azure applications and on-premises applications can get to the Windows Azure capacity administration, and both do it similarly: utilizing a Restful methodology. The hidden information store is definitely not Microsoft SQL Server, in any case. Indeed, Windows Azure capacity isn't a social framework, and its question dialect isn't SQL. Since it's essentially intended to bolster applications based on Windows Azure, it gives less complex, more versatile sorts of capacity. Appropriately, it permits putting away binary large objects (blobs), gives lines to correspondence between segments of Windows Azure applications, and even offers a type of tables with a clear question dialect. Running applications and putting away their information in the cloud can have clear advantages. As opposed to purchasing, introducing, and working its own frameworks, for instance, an association can depend on a cloud supplier to do this for them. Additionally, clients pay only for the figuring and capacity they utilize, instead of keeping up a huge arrangement of

servers just for crest burdens. Also, on the off chance that they're composed accurately, applications can scale effortlessly, exploiting the gigantic server farms that cloud suppliers offer [12].

*2) SQL Azure Database as a Service (DaaS):* One of the key characteristics of any application inside of the undertaking, over web, is information. Over the long haul, information becomes tremendously credited to various sources, diverse gadgets and distinctive frameworks. To address the developing size and size of the information – we require an answer which can address the essential information challenges related to reasonability, adaptability and accessibility. SQL Azure is the cloud-based innovation answer for manage social and different sorts of information as a major aspect of Windows Azure platform. It has two important parts i. SQL Azure Data Sync and ii. SQL Azure Database [13].

### B. About Windows Azure

Windows Azure is the Microsoft's Cloud Computing platform. It includes a few segments:

- Register administration: servers to run your code
- Capacity administration: to store unstructured information
- SQL Azure: to store organized information
- Windows Azure AppFabric: for security and availability

A few samples of where to utilize it include:

- Social: Backend Facebook/social applications
- Portable: One stockpiling and benefits answer for iPhone/Android/Win Phone applications
- WebSites that will get spikes in activity
- HPC: Simulations, demonstrating, and so forth.
- A "managed home" for Access databases

### C. Building Cloud Infrastructure with Microsoft Azure

The cloud arrives to sit tight. With the business sector for cloud arrangements developing quickly, most endeavors are currently utilizing some sort of cloud business arrangement some place in their association - infrequently without the information of the IT office. Market specialists from IDC anticipate overall spending on cloud IT foundation will "develop by 21% year over year to \$32 billion in 2015, representing roughly 33% of all IT base spending, which will be up from around 28% in 2014." In May 2015 Analysts from Gartner said that worldwide investment over Infrastructure as a Service (IaaS) "is relied upon to reach just about US\$16.5 billion in 2015, an expansion of 32.8 percent from 2014, with a compound annual growth rate (CAGR) from 2014 to 2019 figure at 29.1 percent" [14].

*Azure: Five Scenarios for High Value Deployment of Cloud:* There are five scenarios to deploy a high valued cloud, by knowing these scenarios we can implement an efficient cloud system with the Microsoft Azure.

*a) Scenario One: Extend your Datacenter with Azure Storage:* Azure Storage is distributed storage that furnishes clients with anyplace and at whatever time access. It is exceedingly solid, profoundly accessible and enormously versatile. Azure Storage effectively scales from megabytes to exa-bytes, and you pay just what you utilize when you use it. Evaluating depends on the quantity of capacity exchanges, information put away, information departure and the kind of replication. This makes it appealing for new companies, little to moderate sized organizations and venture associations alike. For instance, a new business could plan an application and dispatch it without worrying about supporting development on a worldwide scale. Microsoft brings up that "Azure Storage is open from anyplace on the planet, from an application, whether it's running in the cloud, on the desktop, on an on-premises server, or on a versatile or tablet gadget" [15].

*b) Scenario Two: Extend your Datacenter with Virtual Network and Site-to-Site VPN:* Azure Virtual Network permits you to amplify an on-reason system into the cloud through site-to-site VPN. You can oversee it like an on-reason infrastructure, and control the system topology and setup of DNS and IP address ranges. A site-to-site VPN association with Azure is secured with industry standard IPSec innovation and the endpoint at your site will no doubt be the firewall you as of now have. A distinct option for uniting over people in general Internet is given by Azure ExpressRoute, which empowers you to unite your on-reason or arranged infrastructure with Azure server farms. This is basically a MPLS association. ExpressRoute associations are speedier and more solid, and have lower latencies also, higher security. There are two sorts of network choices: through a trade supplier and through a system administration supplier. ExpressRoute is advertised as a team with various Exchange Providers and Network Service Provider accomplices to make the private associations [15].

*c) Scenario Three: Extend your Datacenter with Azure Backup and Disaster Recovery:* Backup and restore choices are an essential for any business association. Azure Backup is a basic and solid information security arrangement which empowers clients to move down their on-premises information to Microsoft Azure. It is based on top of Azure's powerful worldwide framework and stores reinforcement information in geo-duplicated capacity which keeps up 6 duplicates of information crosswise over two Azure datacenters. Azure Site Recovery secures essential applications by planning the replication what's more, recovery of physical or virtual machines. You can reproduce to your own datacenter, to a facilitating administration supplier, or even to Azure to evade the cost and multifaceted nature of building and dealing with your own particular optional area. At last, Azure StorSimple gives an incorporated integrated storage that oversees capacity assignments between on-premises gadgets and Microsoft Azure distributed storage to improve catastrophe recuperation ability and effectiveness [15].

*d) Scenario Four: Virtual Machines, Using Azure for On-Demand Development and Test:* If your organization creates custom applications, you require an advancement and test environment for those applications. This incorporates introducing designer apparatuses, for example, Visual Studio and making a test situation that duplicates a genuine creation situation. To pick a current picture, go to the exhibition on the Microsoft site, where pictures are accessible from Microsoft and numerous accomplice arrangements including Oracle, SAP, IBM, Cloudera, Hortonworks and numerous others. Furthermore, the open source group offers pictures at VM Depot.

*e) Scenario Five: Single Sign-On with Azure Active Directory: Scenario and Benefits:* Azure Active Directory (AD) gives personality and access administration in the cloud. It offers "a powerful arrangement of abilities to oversee clients and assembles and secures access to on-premises and cloud applications including Microsoft online administrations like Office 365". It is the cloud partner to Active Directory, which offers on-premises personality administration through Windows Server. Utilizing Azure AD permits you to amplify single sign-on abilities to Office 365 and in addition other Microsoft and outsider arrangements. Azure AD can be utilized as a standalone administration as a part of the cloud or incorporated with on-premise Active Directory. Stretching out existing on-premises registries to. It can additionally make an incorporated and bound together experience for the administration of client and gadget characters, including improved client access to Windows, Mac, iOS and Android gadgets.

## VIII. CONCLUSION

We have defined the cloud computing in detailed in this paper, also defined the service orientation and its techniques that are being used by many platforms. The main purpose of this discussion was to build a scenario using the service oriented cloud computing to enhance the data centers and usability of cloud provided by the number of organizations. We described the scenario using the Microsoft Azure services and how it can be efficiently used within an organization. As the world is growing very fast and also the technology is being advanced, so to keep this advancement we need to implement such systems which we can use to enhance scalability, interoperability, confidentiality and advancement.

## REFERENCES

- [1] "Cloud Computing" Second Edition by Dr. Kumar Saurabh
- [2] Naidila Sadashiv, S. M. Dilip Kumar, "Cluster, grid and cloud computing: A detailed comparison", Computer Science & Education (ICCSE), 2011 6th International Conference, Publisher: IEEE, pp. 477 – 482 , 3-5 Aug. 2011.
- [3] Jin-Qiang Ma, "Study on the Application of Agent in Information Grid", Computational Intelligence and Software Engineering (CiSE), 2010 International Conference, Publisher: IEEE, pp. 1 - 4, E-ISBN: 978-1-4244-5392-4, 10-12 Dec. 2010
- [4] Pragaladan R, Suganthi P, "A Study on Challenges of Cloud Computing in Enterprise Perspective", International Journal of

- Advanced Research in Computer and Communication Engineering,  
Vol. 3, Issue 7, July 2014
- [5] Shalini Ramanathan, Savita Goel, Subramanian Alagumalai, "Comparison of Cloud database: Amazon's SimpleDB and Google's Bigtable", Recent Trends in Information Systems (ReTIS), 2011 International Conference, Publisher: IEEE, pp. 165 - 168, DOI: 10.1109/Retis.2011.6146861, 21-23 Dec. 2011
  - [6] Mehrin Gilani, Catherine Inibhunu, Qusay H. Mahmoud, "Application and network performance of Amazon elastic compute cloud instances", Cloud Networking (CloudNet), 2015 IEEE 4th International Conference, pp. 315 - 318, Publisher: IEEE, DOI: 10.1109/CloudNet.2015.7335328, 5-7 Oct. 2015.
  - [7] "Impact of Cloud Computing on Healthcare", [Online Available] at [www.cloud-council.org/cscchealthcare110512.pdf](http://www.cloud-council.org/cscchealthcare110512.pdf)
  - [8] Valentino Baraldo, Alberto Zuccato, Tullio Vardanega, "Reconciling Service Orientation with the Cloud", Service-Oriented System Engineering (SOSE), 2015 IEEE Symposium, pp. 195 – 202, Publisher: IEEE, DOI: 10.1109/SOSE.2015.26, Date of Conference: March 30 2015-April 3 2015.
  - [9] T. Erl. "Service-Oriented Architecture: Concepts, Technology, and Design", Prentice Hall PTR, 2009.
  - [10] OASIS. "Reference Architecture Foundation for Service-Oriented Architecture", Version 1.0, October, 2009. [Online Document] Available at: <http://docs.oasis-open.org/soa-rm/soa-ra/v1.0/soa-ra-cd-02.pdf>
  - [11] Dinesh Agarwal, Sushil K. Prasad, "AzureBench: Benchmarking the Storage Services of the Azure Cloud Platform", Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International, Publisher: IEEE, pp. 1048 - 1057, Date of Conference: 21-25 May 2012.
  - [12] Eduardo Roloff, Francis Birck, Matthias Diener, Alexandre Carissimi, Philippe O. A. Navaux, "Evaluating High Performance Computing on the Windows Azure Platform", Cloud Computing (CLOUD), 2012 IEEE 5th International Conference, pp. 803 - 810, Publisher: IEEE, DOI: 10.1109/CLOUD.2012.47, 29 Jun 2012.
  - [13] David Chappell. *Introducing The Windows Azure Platform*, December 2009.
  - [14] Press release: *Gartner Says Worldwide Cloud Infrastructure-as-a-Service Spending to Grow 32.8 Percent in 2015* at (accessed Jan 2016) <http://www.gartner.com/newsroom/id/3055225>
  - [15] Building Your Cloud Infrastructure with Microsoft Azure: *Five High Value IaaS Scenarios for your Business* by Brian Bourne

## IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA  
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia  
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA  
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway  
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India  
Dr. Amogh Kavimandan, The Mathworks Inc., USA  
Dr. Ramasamy Mariappan, Vinayaka Missions University, India  
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China  
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA  
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico  
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India  
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania  
Dr. Junjie Peng, Shanghai University, P. R. China  
Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia  
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India  
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain  
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India  
Dr Li Fang, Nanyang Technological University, Singapore  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia  
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India  
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand  
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India  
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia  
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India  
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India  
Dr. P. Vasant, University Technology Petronas, Malaysia  
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea  
Dr. Praveen Ranjan Srivastava, BITS PILANI, India  
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong  
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia  
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan  
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria  
Dr. Riktesh Srivastava, Skyline University, UAE  
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia  
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt  
and Department of Computer science, Taif University, Saudi Arabia  
Dr. Tirthankar Gayen, IIT Kharagpur, India  
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan  
Prof. Ning Xu, Wuhan University of Technology, China  
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen  
& Universiti Teknologi Malaysia, Malaysia.  
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India  
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan

Prof. Syed S. Rizvi, University of Bridgeport, USA  
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan  
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghata (HP), India  
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal  
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P  
Dr. Poonam Garg, Institute of Management Technology, India  
Dr. S. Mehta, Inha University, Korea  
Dr. Dilip Kumar S.M, Bangalore University, Bangalore  
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan  
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University  
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia  
Dr. Saqib Saeed, University of Siegen, Germany  
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]  
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt  
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India  
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India  
Dr. Muhammad Sohail, KUST, Pakistan  
Dr. Manjaiah D.H, Mangalore University, India  
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India  
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada  
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia  
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India  
Dr. M. Azath, Anna University, India  
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh  
Dr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia  
Dr. Suresh Jain, Devi Ahilya University, Indore (MP) India,  
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia  
Dr. Hanumanthappa. J. University of Mysore, India  
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)  
Dr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria  
Dr. Santosh K. Pandey, The Institute of Chartered Accountants of India  
Dr. P. Vasant, Power Control Optimization, Malaysia  
Dr. Petr Ivankov, Automatika - S, Russian Federation  
Dr. Utkarsh Seetha, Data Infosys Limited, India  
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal  
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore  
Assist. Prof. A. Neela madheswari, Anna university, India  
Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India  
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh  
Dr. Atul Gonsai, Saurashtra University, Gujarat, India  
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand  
Mrs. G. Nalini Priya, Anna University, Chennai  
Dr. P. Subashini, Avinashilingam University for Women, India  
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat  
Mr Jitendra Agrawal, : Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal  
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India  
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof, Kanwalvir Singh Dhindsa, B.B.S.B.Engg. College, Fatehgarh Sahib (Punjab), India  
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah  
Mr. Nitin Bhatia, DAV College, India  
Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India  
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia  
Assist. Prof. Sonal Chawla, Panjab University, India  
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India  
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia  
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia  
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India  
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France  
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India  
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology, Durban, South Africa  
Prof. Mydhili K Nair, Visweswaraiah Technological University, Bangalore, India  
M. Prabu, Adhiyamaan College of Engineering/Anna University, India  
Mr. Swakkhar Shatabda, United International University, Bangladesh  
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan  
Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India  
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India  
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India  
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran  
Mr. Zeashan Hameed Khan, Université de Grenoble, France  
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow  
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria  
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India  
Dr. Maslin Masrom, University Technology Malaysia, Malaysia  
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India  
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City  
Dr. Mary Lourde R., BITS-PILANI Dubai, UAE  
Dr. Abdul Aziz, University of Central Punjab, Pakistan  
Mr. Karan Singh, Gautam Budtha University, India  
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India  
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia  
Assistant Prof. Yasser M. Alginahi, Taibah University, Madinah Munawwarah, KSA  
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India  
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India  
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India  
Dr. M. Munir Ahmed Rabbani, B. S. Abdur Rahman University, India  
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India  
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia  
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India  
Asst Prof. Jasmine. K. S, R.V. College of Engineering, India  
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius  
Prof. Mahesh Goyani, G H Patel College of Engg. & Tech, V.V.N, Anand, Gujarat, India  
Dr. Mana Mohammed, University of Tlemcen, Algeria  
Prof. Jatinder Singh, Universal Institution of Engg. & Tech. CHD, India

Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim  
Dr. Bin Guo, Institute Telecom SudParis, France  
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia  
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia  
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius  
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore  
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India  
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India  
Dr. C. Arun, Anna University, India  
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India  
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran  
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology  
Subhabrata Barman, Haldia Institute of Technology, West Bengal  
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan  
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India  
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India  
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand  
Dr. P. Chakrabarti, Sir Padampat Singhania University, Udaipur, India  
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.  
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran  
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India  
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA  
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India  
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India  
Mr. Serguei A. Mokhov, Concordia University, Canada  
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia  
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India  
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA  
Dr. S. Karthik, SNS Collegeof Technology, India  
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain  
Mr. A.D.Potgantwar, Pune University, India  
Dr. Himanshu Aggarwal, Punjabi University, India  
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India  
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai  
Dr. Prasant Kumar Pattnaik, KIST, India.  
Dr. Ch. Aswani Kumar, VIT University, India  
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA  
Mr. Arun Kumar, Sir Padam Pat Singhania University, Udaipur, Rajasthan  
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia  
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA  
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia  
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India  
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India  
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia  
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan

Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA  
Mr. R. Jagadeesh Kannan, RMK Engineering College, India  
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India  
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh  
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India  
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia  
Mr. R. Mohammad Shafi, Madanapalle Institute of Technology & Science, India  
Dr. F.Sagayaraj Francis, Pondicherry Engineering College, India  
Dr. Ajay Goel, HIET, Kaithal, India  
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India  
Mr. Suhas J Manangi, Microsoft India  
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded, India  
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India  
Dr. Amjad Rehman, University Technology Malaysia, Malaysia  
Mr. Rachit Garg, L K College, Jalandhar, Punjab  
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India  
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan  
Dr. Thorat S.B., Institute of Technology and Management, India  
Mr. Ajay Prasad, Sir Padampat Singhania University, Udaipur, India  
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India  
Mr. Syed Raifiul Hussain, Ahsanullah University of Science and Technology, Bangladesh  
Mrs Fazeela Tunnis, Najran University, Kingdom of Saudi Arabia  
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India  
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA  
Mr. Anand Kumar, AMC Engineering College, Bangalore  
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India  
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India  
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India  
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India  
Dr. V V S S Balaram, Sreenidhi Institute of Science and Technology, India  
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India  
Prof. Niranjana Reddy. P, KITS, Warangal, India  
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India  
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India  
Dr. A. Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai  
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India  
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan  
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India  
Dr. Tossapon Boongoen, Aberystwyth University, UK  
Dr. Bilal Alatas, Firat University, Turkey  
Assist. Prof. Jyoti Praaksh Singh, Academy of Technology, India  
Dr. Ritu Soni, GNG College, India  
Dr. Mahendra Kumar, Sagar Institute of Research & Technology, Bhopal, India.  
Dr. Binod Kumar, Lakshmi Narayan College of Tech.(LNCT)Bhopal India  
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan  
Dr. T.C. Manjunath, ATRIA Institute of Tech, India  
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan

Assist. Prof. Harmunish Taneja, M. M. University, India  
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India  
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India  
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad  
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India  
Mr. G. Appasami, Dr. Pauls Engineering College, India  
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan  
Mr. Yaser Miaji, University Utara Malaysia, Malaysia  
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh  
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India  
Dr. S. Sasikumar, Roever Engineering College  
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India  
Mr. Nwaocha Vivian O, National Open University of Nigeria  
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India  
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India  
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore  
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia  
Dr. Dhuha Basheer abdullah, Mosul university, Iraq  
Mr. S. Audithan, Annamalai University, India  
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India  
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India  
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam  
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India  
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad  
Mr. Deepak Gour, Sir Padampat Singhania University, India  
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India  
Mr. Ali Balador, Islamic Azad University, Iran  
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India  
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India  
Dr. Debojyoti Mitra, Sir padampat Singhania University, India  
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia  
Mr. Zhao Zhang, City University of Hong Kong, China  
Prof. S.P. Setty, A.U. College of Engineering, India  
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India  
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India  
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India  
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India  
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India  
Dr. Hanan Elazhary, Electronics Research Institute, Egypt  
Dr. Hosam I. Faiq, USM, Malaysia  
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India  
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India  
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India  
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan  
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India  
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia  
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India

Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India  
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India  
Prof Anupam Choudhary, Bhilai School Of Engg.,Bhilai (C.G.),India  
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya  
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.  
Dr. Kasarapu Ramani, JNT University, Anantapur, India  
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India  
Dr. C G Ravichandran, R V S College of Engineering and Technology, India  
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia  
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia  
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India  
Dr. Nikolai Stoianov, Defense Institute, Bulgaria  
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode  
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India  
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh  
Mr. Hemanta Kumar Kalita , TATA Consultancy Services (TCS), India  
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria  
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela  
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India  
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia  
Dr. Nighat Mir, Effat University, Saudi Arabia  
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India  
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore  
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore  
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US  
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India  
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India  
Mr. P. Sivakumar, Anna university, Chennai, India  
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia  
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India  
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia  
Mr. Nikhil Patrick Lobo, CADES, India  
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India  
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India  
Assist. Prof. Vishal Bharti, DCE, Gurgaon  
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India  
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India  
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India  
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India  
Mr. Hamed Taherdoost, Tehran, Iran  
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran  
Mr. Shantanu Pal, University of Calcutta, India  
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom  
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria  
Mr. P. Mahalingam, Caledonian College of Engineering, Oman  
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt

Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India  
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India  
Mr. Muhammad Asad, Technical University of Munich, Germany  
Mr. AliReza Shams Shafiqh, Azad Islamic university, Iran  
Prof. S. V. Nagaraj, RMK Engineering College, India  
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India  
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia  
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India  
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India  
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco  
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India  
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India  
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India  
Mr. Sunil Taneja, Kurukshetra University, India  
Mr. Fauzi Adi Rafrastra, Dian Nuswantoro University, Indonesia  
Dr. Yaduvir Singh, Thapar University, India  
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece  
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore  
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia  
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia  
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran  
Assoc. Prof. Dhirendra Mishra, SVKM's NMIMS University, India  
Prof. Shapor Zarei, UAE Inventors Association, UAE  
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India  
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India  
Prof. Anant J Umbarkar, Walchand College of Engg., India  
Assist. Prof. B. Bharathi, Sathyabama University, India  
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia  
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India  
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India  
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore  
Prof. Walid Moudani, Lebanese University, Lebanon  
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India  
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India  
Associate Prof. Dr. Manuj Darbari, BBD University, India  
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India  
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India  
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India  
Dr. Abhay Bansal, Amity School of Engineering & Technology, India  
Ms. Sumita Mishra, Amity School of Engineering and Technology, India  
Professor S. Viswanadha Raju, JNT University Hyderabad, India  
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India  
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India  
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia  
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia  
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India  
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia

Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India  
Mr. Shervan Fekri Ershad, Shiraz International University, Iran  
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh  
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh  
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India  
Ms. Sarla More, UIT, RGTU, Bhopal, India  
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India  
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India  
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India  
Dr. M. N. Giri Prasad, JNTUCE,Pulivendula, A.P., India  
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India  
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India  
Assist. Prof. Navnish Goel, S. D. College Of Enginnering & Technology, India  
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya  
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh  
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India  
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh  
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan  
Mr. Mohammad Asadul Hoque, University of Alabama, USA  
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India  
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan  
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA  
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India  
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina  
Dr S. Rajalakshmi, Botho College, South Africa  
Dr. Mohamed Sarrab, De Montfort University, UK  
Mr. Basappa B. Kodada, Canara Engineering College, India  
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India  
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India  
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India  
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India  
Dr . G. Singaravel, K.S.R. College of Engineering, India  
Dr B. G. Geetha, K.S.R. College of Engineering, India  
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon  
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran  
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India  
Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)  
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India  
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India  
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)  
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India  
Assist. Prof. Maram Balajee, GMRIT, India  
Assist. Prof. Monika Bhatnagar, TIT, India  
Prof. Gaurang Panchal, Charotar University of Science & Technology, India  
Prof. Anand K. Tripathi, Computer Society of India  
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India  
Assist. Prof. Supriya Raheja, ITM University, India

Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.  
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India  
Prof. Mohan H.S, SJB Institute Of Technology, India  
Mr. Hossein Malekinezhad, Islamic Azad University, Iran  
Mr. Zatin Gupta, Universti Malaysia, Malaysia  
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India  
Assist. Prof. Ajal A. J., METS School Of Engineering, India  
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria  
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India  
Md. Nazrul Islam, University of Western Ontario, Canada  
Tushar Kanti, L.N.C.T, Bhopal, India  
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India  
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh  
Dr. Kashif Nisar, University Utara Malaysia, Malaysia  
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA  
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan  
Assist. Prof. Apoorvi Sood, I.T.M. University, India  
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia  
Mr. Swapnil Soner, Truba Institute College of Engineering & Technology, Indore, India  
Ms. Yogita Gigras, I.T.M. University, India  
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College  
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad  
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India  
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad  
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India  
Mr. Mehdi Golsorkhtabar Amiri, Islamic Azad University, Iran  
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India  
Assist. Prof. P.Oliver Jayaprakash, Anna University,Chennai  
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India  
Dr. Asoke Nath, St. Xavier's College, India  
Mr. Masoud Rafighi, Islamic Azad University, Iran  
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India  
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India  
Mr. Sandeep Maan, Government Post Graduate College, India  
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India  
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India  
Mr. R. Balu, Bharathiar University, Coimbatore, India  
Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India  
Prof. P. Senthilkumar, Vivekanandha Institue of Engineering and Techology for Woman, India  
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India  
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India  
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India  
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran  
Mr. Laxmi chand, SCTL, Noida, India  
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad  
Prof. Mahesh Panchal, KITRC, Gujarat  
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode

Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India  
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhania University, India  
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India  
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India  
Mr. Srikanta Kumar Mohapatra, NMIFT, Orissa, India  
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan  
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India  
Prof. Elboukhari Mohamed, University Mohammed First, Oujda, Morocco  
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia  
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.  
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India  
Mr. G. Premsankar, Ericsson, India  
Assist. Prof. T. Hemalatha, VELS University, India  
Prof. Tejaswini Apte, University of Pune, India  
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia  
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran  
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India  
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India  
Mr. Vorugunti Chandra Sekhar, DA-IICT, India  
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia  
Dr. Aderemi A. Atayero, Covenant University, Nigeria  
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan  
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India  
Mr. Hassen Mohammed Abdulla Alsaifi, International Islamic University Malaysia (IIUM) Malaysia  
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan  
Mr. R. Balu, Bharathiar University, Coimbatore, India  
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar  
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India  
Prof. K. Saravanan, Anna university Coimbatore, India  
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India  
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN  
Assoc. Prof. S. Asif Hussain, AITS, India  
Assist. Prof. C. Venkatesh, AITS, India  
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan  
Dr. B. Justus Rabi, Institute of Science & Technology, India  
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India  
Mr. Alejandro Mosquera, University of Alicante, Spain  
Assist. Prof. Arjun Singh, Sir Padampat Singhania University (SPSU), Udaipur, India  
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad  
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India  
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India  
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia  
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India  
Mr. Hassen Mohammed Abdulla Alsaifi, International Islamic University Malaysia (IIUM)  
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA  
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu  
Dr. K. Reji Kumar, , N S S College, Pandalam, India

Assoc. Prof. K. Seshadri Sastry, EIILM University, India  
Mr. Kai Pan, UNC Charlotte, USA  
Mr. Ruikar Sachin, SGGSIET, India  
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India  
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India  
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt  
Assist. Prof. Amanpreet Kaur, ITM University, India  
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore  
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia  
Dr. Abhay Bansal, Amity University, India  
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA  
Assist. Prof. Nidhi Arora, M.C.A. Institute, India  
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India  
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India  
Dr. S. Sankara Gomathi, Panimalar Engineering college, India  
Prof. Anil Kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India  
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India  
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology  
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia  
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh  
Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India  
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India  
Dr. Lamri LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept. Computer Science, UBO, Brest, France  
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India  
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India  
Mr. Ram Kumar Singh, S.V Subharti University, India  
Assistant Prof. Sunish Kumar O S, Amaljyothi College of Engineering, India  
Dr Sanjay Bhargava, Banasthali University, India  
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India  
Mr. Roohollah Etemadi, Islamic Azad University, Iran  
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria  
Mr. Sumit Goyal, National Dairy Research Institute, India  
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India  
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur  
Dr. S.K. Mahendran, Anna University, Chennai, India  
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab  
Dr. Ashu Gupta, Apeejay Institute of Management, India  
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India  
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus  
Mr. Maram Balajee, GMR Institute of Technology, India  
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan  
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria  
Mr. Jasvir Singh, University College Of Engg., India  
Mr. Vivek Tiwari, MANIT, Bhopal, India  
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India  
Mr. Somdip Dey, St. Xavier's College, Kolkata, India

Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China  
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh  
Mr. Sathyaprakash P., S.K.P Engineering College, India  
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India  
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India  
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India  
Mr. Md. Abdul Ahad, K L University, India  
Mr. Vikas Bajpai, The LNM IIT, India  
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA  
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India  
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai  
Mr. A. Siles Balasingh, St.Joseph University in Tanzania, Tanzania  
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India  
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India  
Mr. Kumar Dayanand, Cambridge Institute of Technology, India  
Dr. Syed Asif Ali, SMI University Karachi, Pakistan  
Prof. Pallvi Pandit, Himachal Pradesh University, India  
Mr. Ricardo Verschueren, University of Gloucestershire, UK  
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India  
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India  
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India  
Dr. S. Sumathi, Anna University, India  
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India  
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India  
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India  
Assist. Prof M. Anand Kumar, Karpagam University, Coimbatore, India  
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex  
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India  
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India  
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat  
Mr. Sivakumar, Codework solutions, India  
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran  
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA  
Mr. Varadala Sridhar, Varadhaman College Engineering College, Affiliated To JNTU, Hyderabad  
Assist. Prof. Manoj Dhawan, SVITS, Indore  
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India  
Dr. S. Santhi, SCSVMV University, India  
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran  
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh  
Mr. Sandeep Reddivari, Mississippi State University, USA  
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal  
Dr. Hazra Imran, Athabasca University, Canada  
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India  
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India  
Ms. Jaspreet Kaur, Distance Education LPU, India  
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman  
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India

- Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India  
Mr. Khaldi Amine, Badji Mokhtar University, Algeria  
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran  
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India  
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India  
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia  
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India  
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India  
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India  
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany  
Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India  
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India  
Dr. Nadir Bouchama, CERIST Research Center, Algeria  
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India  
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco  
Dr. S. Malathi, Panimalar Engineering College, Chennai, India  
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India  
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India  
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan  
Dr. G. Rasitha Banu, Vel's University, Chennai  
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai  
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India  
Ms. U. Sirthuja, PSG college of arts &science, India  
Dr. Ehsan Saradar Torshizi, Urmia University, Iran  
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India  
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India  
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim  
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt  
Dr. Nishant Gupta, University of Jammu, India  
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India  
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India  
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus  
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India  
Dr. Rahul Malik, Cisco Systems, USA  
Dr. S. C. Lingareddy, ALPHA College of Engineering, India  
Assistant Prof. Mohammed Shuaib, Interl University, Lucknow, India  
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India  
Dr. T. Thambidurai, Sun Univercell, Singapore  
Prof. Anandkumar Telang, BKIT, India  
Assistant Prof. R. Poorvadevi, SCSVMV University, India  
Dr Uttam Mande, Gitam University, India  
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India  
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India  
Dr. Mohammed Zuber, AISECT University, India  
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia  
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India

Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India  
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India  
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq  
Dr. Urmila Shrawankar, G H Raisoni College of Engineering, Nagpur (MS), India  
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India  
Dr. Mukesh Negi, Tech Mahindra, India  
Dr. Anuj Kumar Singh, Amity University Gurgaon, India  
Dr. Babar Shah, Gyeongsang National University, South Korea  
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India  
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India  
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India  
Assistant Prof. Ankit Garg, Amity University, Haryana, India  
Assistant Prof. Rajashe Karappa, SDMCET, Karnataka, India  
Assistant Prof. Varun Jasuja, GNIT, India  
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India  
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India  
Dr. Faouzi Hidoussi, UHL Batna, Algeria  
Dr. Naseer Ali Husieen, Wasit University, Iraq  
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai  
Dr. Ahmed Farouk Metwaly, K L University  
Mr. Mohammed Noaman Murad, Cihan University, Iraq  
Dr. Suxing Liu, Arkansas State University, USA  
Dr. M. Gomathi, Velalar College of Engineering and Technology, India  
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia  
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India  
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India  
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran  
Dr. Thiyyagu Nagaraj, University-INOU, India  
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe  
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India  
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India  
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India  
Dr. Shenshen Liang, University of California, Santa Cruz, US  
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia  
Mr. Snehasis Banerjee, Tata Consultancy Services, India  
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania  
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia  
Dr. Ying Yang, Computer Science Department, Yale University, USA  
Dr. Vinay Shukla, Institute Of Technology & Management, India  
Dr. Liviu Octavian Mafteiu-Scai, West University of Timisoara, Romania  
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq  
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India  
Dr. Timothy Powers, University of Hertfordshire, UK  
Dr. S. Prasath, Bharathiar University, Erode, India  
Dr. Ritu Shrivastava, SIRTS Bhopal, India  
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India  
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania

Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India  
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India  
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India  
Assistant Prof. Mallikarjun C Sarsamba Bheemnna Khandre Institute Technology, Bhalki, India  
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India  
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India  
Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq  
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco  
Dr. Parul Verma, Amity University, India  
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco  
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India  
Assistant Prof.. G. Selvavinayagam, SNS College of Technology, Coimbatore, India  
Assistant Prof. Madhavi Dhingra, Amity University, MP, India  
Professor Kartheesan Log, Anna University, Chennai  
Professor Vasudeva Acharya, Shri Madhwa vadira Institute of Technology, India  
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia  
Assistant Prof., Mahendra Singh Meena, Amity University Haryana  
Assistant Professor Manjeet Kaur, Amity University Haryana  
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt  
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia  
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India  
Assistant Prof. Dharmendra Choudhary, Tripura University, India  
Assistant Prof. Deepika Vodnala, SR Engineering College, India  
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA  
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India  
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan  
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India  
Assistant Prof. Chirag Modi, NIT Goa  
Dr. R. Ramkumar, Nandha Arts And Science College, India  
Dr. Priyadarshini Vydhalingam, Harathiar University, India  
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka  
Dr. Vikas Thada, AMITY University, Pachgaon  
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore  
Dr. Shaheera Rashwan, Informatics Research Institute  
Dr. S. Preetha Gunasekar, Bharathiyar University, India  
Asst Professor Sameer Dev Sharma, Uttarakhand University, Dehradun  
Dr. Zhihan Lv, Chinese Academy of Science, China  
Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, Amritsar  
Dr. Umar Ruhi, University of Ottawa, Canada  
Dr. Jasmin Cosic, University of BiHac, Bosnia and Herzegovina  
Dr. Homam Reda El-Taj, University of Tabuk, Kingdom of Saudi Arabia  
Dr. Mostafa Ghobaei Arani, Islamic Azad University, Iran  
Dr. Ayyasamy Ayyanar, Annamalai University, India  
Dr. Selvakumar Manickam, Universiti Sains Malaysia, Malaysia  
Dr. Murali Krishna Namana, GITAM University, India  
Dr. Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India  
Professor Vimalathithan Rathinasabapathy, Karpagam College Of Engineering, India

Dr. Sushil Chandra Dimri, Graphic Era University, India  
Dr. Dinh-Sinh Mai, Le Quy Don Technical University, Vietnam  
Dr. S. Rama Sree, Aditya Engg. College, India  
Dr. Ehab T. Alnfrawy, Sadat Academy, Egypt  
Dr. Patrick D. Cerna, Haramaya University, Ethiopia  
Dr. Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), India  
Associate Prof. Dr. Jiliang Zhang, North Eastern University, China  
Dr. Sharefa Murad, Middle East University, Jordan  
Dr. Ajeet Singh Poonia, Govt. College of Engineering & technology, Rajasthan, India  
Dr. Vahid Esmaeelzadeh, University of Science and Technology, Iran  
Dr. Jacek M. Czerniak, Casimir the Great University in Bydgoszcz, Institute of Technology, Poland  
Associate Prof. Anisur Rehman Nasir, Jamia Millia Islamia University  
Assistant Prof. Imran Ahmad, COMSATS Institute of Information Technology, Pakistan  
Professor Ghulam Qasim, Preston University, Islamabad, Pakistan  
Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women  
Dr. Wencan Luo, University of Pittsburgh, US  
Dr. Musa PEKER, Faculty of Technology, Mugla Sitki Kocman University, Turkey  
Dr. Gunasekaran Shanmugam, Anna University, India  
Dr. Binh P. Nguyen, National University of Singapore, Singapore  
Dr. Rajkumar Jain, Indian Institute of Technology Indore, India  
Dr. Imtiaz Ali Halepoto, QUEST Nawabshah, Pakistan  
Dr. Shaligram Prajapat, Devi Ahilya University Indore India  
Dr. Sunita Singhal, Birla Institute of Technologyand Science, Pilani, India  
Dr. Ijaz Ali Shoukat, King Saud University, Saudi Arabia  
Dr. Anuj Gupta, IKG Punjab Technical University, India  
Dr. Sonali Saini, IES-IPS Academy, India  
Dr. Krishan Kumar, MotiLal Nehru National Institute of Technology, Allahabad, India  
Dr. Z. Faizal Khan, College of Engineering, Shaqra University, Kingdom of Saudi Arabia  
Prof. M. Padmavathamma, S.V. University Tirupati, India  
Prof. A. Velayudham, Cape Institute of Technology, India  
Prof. Seifeide Kadry, American University of the Middle East  
Dr. J. Durga Prasad Rao, Pt. Ravishankar Shukla University, Raipur  
Assistant Prof. Najam Hasan, Dhofar University  
Dr. G. Suseendran, Vels University, Pallavaram, Chennai  
Prof. Ankit Faldu, Gujarat Technological Universiry- Atmiya Institute of Technology and Science  
Dr. Ali Habiboghi, Islamic Azad University  
Dr. Deepak Dembla, JECRC University, Jaipur, India  
Dr. Pankaj Rajan, Walmart Labs, USA  
Assistant Prof. Radoslava Kraleva, South-West University "Neofit Rilski", Bulgaria  
Assistant Prof. Medhavi Shriwas, Shri vaishnav institute of Technology, India  
Associate Prof. Sedat Akleylek, Ondokuz Mayis University, Turkey  
Dr. U.V. Arivazhagu, Kingston Engineering College Affiliated To Anna University, India  
Dr. Touseef Ali, University of Engineering and Technology, Taxila, Pakistan  
Assistant Prof. Naren Jeeva, SASTRA University, India  
Dr. Riccardo Colella, University of Salento, Italy  
Dr. Enache Maria Cristina, University of Galati, Romania  
Dr. Senthil P, Kurinji College of Arts & Science, India

Dr. Hasan Ashrafi-rizi, Isfahan University of Medical Sciences, Isfahan, Iran  
Dr. Mazhar Malik, Institute of Southern Punjab, Pakistan  
Dr. Yajie Miao, Carnegie Mellon University, USA  
Dr. Kamran Shaukat, University of the Punjab, Pakistan  
Dr. Sasikaladevi N., SASTRA University, India  
Dr. Ali Asghar Rahmani Hosseiniabadi, Islamic Azad University Ayatollah Amoli Branch, Amol, Iran  
Dr. Velin Kralev, South-West University "Neofit Rilski", Blagoevgrad, Bulgaria  
Dr. Marius Iulian Mihailescu, LUMINA - The University of South-East Europe  
Dr. Sriramula Nagaprasad, S.R.R.Govt.Arts & Science College, Karimnagar, India  
Prof (Dr.) Namrata Dhanda, Dr. APJ Abdul Kalam Technical University, Lucknow, India  
Dr. Javed Ahmed Maher, Shah Abdul Latif University, Khairpur Mir's, Pakistan  
Dr. B. Narendra Kumar Rao, Sree Vidyanikethan Engineering College, India  
Dr. Shahzad Anwar, University of Engineering & Technology Peshawar, Pakistan  
Dr. Basit Shahzad, King Saud University, Riyadh - Saudi Arabia  
Dr. Nilamadhab Mishra, Chang Gung University  
Dr. Sachin Kumar, Indian Institute of Technology Roorkee  
Dr. Santosh Nanda, Biju-Pattnaik University of Technology  
Dr. Sherzod Turaev, International Islamic University Malaysia  
Dr. Yilun Shang, Tongji University, Department of Mathematics, Shanghai, China  
Dr. Nuzhat Shaikh, Modern Education society's College of Engineering, Pune, India  
Dr. Parul Verma, Amity University, Lucknow campus, India  
Dr. Rachid Alaoui, Agadir Ibn Zohr University, Agadir, Morocco  
Dr. Dharmendra Patel, Charotar University of Science and Technology, India  
Dr. Dong Zhang, University of Central Florida, USA  
Dr. Kennedy Chinedu Okafor, Federal University of Technology Owerri, Nigeria  
Prof. C Ram Kumar, Dr NGP Institute of Technology, India  
Dr. Sandeep Gupta, GGS IP University, New Delhi, India  
Dr. Shahanawaj Ahamad, University of Ha'il, Ha'il City, Ministry of Higher Education, Kingdom of Saudi Arabia  
Dr. Najeed Ahmed Khan, NED University of Engineering & Technology, India  
Dr. Sajid Ullah Khan, Universiti Malaysia Sarawak, Malaysia  
Dr. Muhammad Asif, National Textile University Faisalabad, Pakistan  
Dr. Yu Bi, University of Central Florida, Orlando, FL, USA  
Dr. Brijendra Kumar Joshi, Research Center, Military College of Telecommunication Engineering, India

# **CALL FOR PAPERS**

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

**IJCSIS 2016**  
**ISSN: 1947-5500**  
**<http://sites.google.com/site/ijcsis/>**

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

### ***Track A: Security***

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity  
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.,) Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on

its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

#### **Track B: Computer Science**

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com). Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes>.

**© IJCSIS PUBLICATION 2016  
ISSN 1947 5500  
<http://sites.google.com/site/ijcsis/>**