# Evaluating Reliability and Load Balance in Grid Distributed Systems

S. Navaezadeh
Sama Technical and Vocational Training College,
Islamic Azad University, Mahshahr, Branch Mahshahr
Iran

**Abstract**: Grid computing systems are new and important fields, and they differ from common distributed systems due to resource subscription in large scales, high efficiency and creative applied program. Grid computing system is a wide distributed system. The main and special problem of grid concept is coordinating resource subscription and solving the problem dynamically and multi-distribution programs. Grid computing system is a distributed computing system, and in this system, communication and processors have considerable effects on reliability of grid computing. In this paper, the features of grid computing systems have been described, and some algorithms have been presented to analyze reliability of system and grid program. In this study, an intelligence algorithm has been used to create load balance in grid. Since grid environment is dynamic, firefly algorithm has been used to create load balance in sites. The results of algorithm simulation have been presented by Matlab software, and it is compared with learner particles swarm algorithm, simulated annealing and random algorithm.

**Keywords**: learner machine, grid, pso_ sa, pso_la, swarm intelligence algorithm

## 1. INTRODUCTION

In Grid computing systems, failure probability increases in these systems due to dynamic and distributed nature of grids. Therefore, error tolerance is a crucial field in computing grids, and it affects reliability. Grid is a kind of parallel distributed system providing selection and gathering non-heterogeneous distributed resources in execution time, and it is on the basis of some factors involving accessibility, capacity, efficiency and output, costs, users needs of service quality. There are many programs in grid environment, and they must be scheduled on these systems for parallel execution. Since scheduling is a NP-hard problem, we can use evaluation algorithms to solve it such as random or GA algorithm.

In computing systems, grid has been emerged as a new and important field, and it is different from distributed system due to resources subscription and creative application programs with high efficiency and performance.

Grid computing system is a kind of distributed system with wide range. The main problem of grid is to coordinate resource subscription, solving the problem dynamically and multi-distribution program. Subscription that Grid computing involved with was not only file primary exchange, but also direct accessibility to computers, software data and other sources which needed a broad spectra of problem solving strategies and resource subscription. Therefore, Grid computing reliability influenced by the relation between computing programs and important sources. [1]

Since scheduling problem is NP-hard issue, and resource management and scheduling have effective role in load balance, it influences grid efficiency [2]. The advantages of computing grid is as follows: using the resources, parallel computing, creating resources and virtual organization, management of accessing to additional resources [3].

Since artificial intelligence algorithms are efficient in optimization problems, they are considered as a good choice to solve some problems like load balance in distributed system. The purpose of load balance is to reduce the difference of load rate between the heaviest node and the lightest node [4].

## 2. LITERATURE REVIEW

Load Balance algorithm divided into different groups. On one hand, these algorithm divided into two groups static and dynamic. Load balance difference in dynamic and static mode is that in static mode the decisions related to load balance made in compiling time but in dynamic mode decisions related to load balance made in performing time.

It means that in static mode, indeed, these decisions made in the time of request for the source. And in dynamic mode, behavior of balancer varied according to parameters and policies changing. [5]

Load balancing divided into three groups; concentrated, non – concentrated and hierarchy in concentrated method, all functions scheduled by a scheduler and scheduling operation performed by the applicant source and in hierarchy method the scheduler organized in the hierarchy form. [6]

In load balance concentrated method, many studies have been conducted. In [5] in order to have concentrated load balance, genetic algorithm used and its Simulation result compared with Min – Max and Max – Min algorithm. In [2] a new genetic algorithm presented using resource fault occurrence history (rfoh) for certain scheduling in computing Grid.

This strategy keeps source fault occurrence history in Gird information server. One of these algorithms is particles swarm intelligence (pso) that is a solution for optimization problems in search apace with modeling of social behavior. A new model is proposed on the basis of PSO called PSO-SA.

In fact, simulated annealing algorithm is one of the most well-known algorithms in artificial intelligence algorithm. This algorithm was proposed by Scott Kirkaptrica in 1983. One of its applications is to solve optimization problems. In this algorithm, an arbitrary point is firstly selected from search

space, and penalty function is computed. Then, a point is selected as next point for substitution. The function is defined as follows. If heat is high, then the probability of moving toward energy increasing will increase.

Simulation results showed that proposed strategy decreased total time of programs performance. One of glowworm algorithm advantages relating to other swarm intelligence algorithms is its constriction simplicity.

Firefly Algorithm is a type of algorithm obtained from nature and collective smart algorithm which presented by yang (2008), this algorithm is a modern technique based on collective behaviors which inspired from firefly collective intelligence is a type of artificial intelligence method social behaviors in the nature based on collective behaviors in neutralized and self-organized foundations. Fireflies generate rhythmic and short beams. Optical patterns of each firefly differentiated form others. Fireflies used these beams for two reasons. 1- pairs attraction process. 2- for attracting hunt. Moreover, these beams used as a protective Mechanism for fireflies. Rhythmic beams and rate of radiation and interval rate between light signals caused two genders attract each other. Any particle of a firefly in multidimensional quest space updated by absorbing dynamically based on a knowledge of firefly and its neighbors.

Firefly optimization algorithm could be stated as follows: [7]

* all fireflies are single- gender and the factor of pairs attractiveness considered not relating to their gender.

* firefly x attracts all fireflies and attracted for all fireflies.

* attractiveness is related to their glow, so for any pair of firefly, a worm with less light is attracted toward a worm with more light. Attractiveness power related to their beam and the light intensity decreased by increasing the distance between two fireflies. If a firefly is not brighter than the others, their movement will be performed randomly.

* brighter firefly moves randomly (all fireflies could not attract them).

* firefly brightness is determined by objective function value. In maximization problem, light intensity can be easily determined by target function.

* firefly particles ae randomly distributed in quest space. According to above principles, there are two main parts in firefly algorithm, attracting firefly and movement toward attracted firefly.

## 2.1 General Form Of Firefly Algorithm
General form of firefly algorithm has been shown in the figure (1)As we can see in the figure, at first, primary coordination and light intensity rate and the distance between firefly particles are determined in quest area. Quest procedure in firefly algorithm is that any firefly isindividually compared with others. If a firefly has less light than the compared one, it will move toward a firefly with more light (the problem of finding maximum point), and in this process, particles are centralized around a particle with more light, and if in the next generation of algorithm, there is a particle with more light, particles will move toward particle with more light again. Quest stages must be generated relating to maximum number of generation. In this study, in order to optimize the problem of load Balance by using glowworm group intelligence

algorithm, a solution is presented. In this method, each node in the network is considered as a glowworm. Each glowworm tries to optimist existing load rate in itself, and this work is performed by exchanging the load among other nodes.

## 3. PROPOSED ALGORITHM
Gglowworm algorithm has been explained in the previous sections. In algorithm presented based on swarm intelligence, all nodes in Grid system are considered as a solution for finding the most optimized mode.

The location of each worm is shown by the existing rate of light. In order to determine attraction parameter for any node, its node light rate difference is measured with neighbor nodes. Any node always moves toward the best neighbor. This work is performed by attracting toward the neighbor or emitting more light from neighbor nodes during the same rate of a node light with the same rate of a node light with the neighbor nodes, its movement performed randomly. since grid is a dynamic environment, solutions always change so that, in this method, there is no need to keep information and previous history like classic glow worm, and also we have no massive particles. In this study, parameter α , β considered. 1 and 2 between 0 to 1

Particle Algorithm()
 sourceLoad
while running
Do if job Queue.size>0
Then
Lighteload    %% choose best Neighbor(entekhab avalin behtarin)
SecondLightestLoad   %% choose SecondLightest Neighbor(en entekhab dovomin behtarin)
TC (kamtarin hazine tebghe kamtarin faseleh)
threshold(tebghe avalin va dovomin)
while tc>threshold
do
Submit jobs     %%(TC)
sourceLoad        %%(currentNodeload)
velocity             %%(sourceload- lightestLoad)

**Figure 1: Running Algorithm by using glow worm.**

## 3.1 simulation Results

After simulating the proposed algorithm in MATLAB environment, results are compared with massive particles and genetic. The way of constructing the network topology for the simulation is as follows:

At first, a minimum covering tree is structured; then, considered topology is obtained by adding edges. In figure (2) time of the first job sending to the network until reaching the last job has been compared. The group showed that glowworm algorithm application relating to other algorithm will be more optimized.

In the figure (3) the difference between lightest and heaviest nodes has been demonstrated in terms of its load. (the load in glow worm shows itself by light, and due to much similarity with massive particles, we showed job comparison with this algorithm. This time consisted of the first job sending time until running the last job in the network.
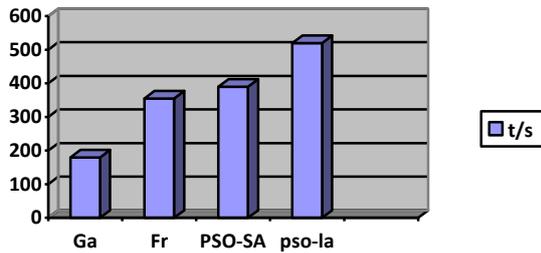
**Figure 2: executing algorithm by using glowworm algorithm**
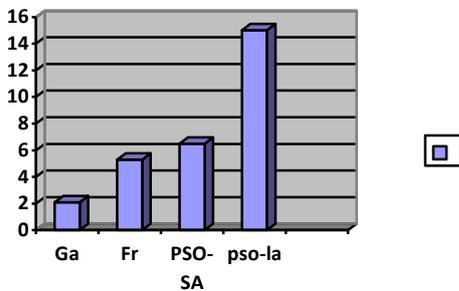


**Figure 3: the average of executing time in different algorithm**

## 4. CONCLUSION

In this study, we used glow worm algorithm for load balanced in Grid. Results showed that this algorithm has higher efficiency than other algorithms in terms of job scheduling when there are different sources. Since grid computing system is considered as a computing system with a large scale, communication and processing time cannot be ignored. The features of grid computing system have been described in this paper, and reliability of grid computing system and program has been analyzed. In addition, algorithm has been presented to evaluate reliability of grid system and program. Glow worm algorithm showed better results for the time of running and also considered as the most appropriate component of methods for optimization problems.

## 5. REFERENCES

[1]  Dai .Y.S, . Xie. M, Poh .K.L. ( 2010 ),.," Reliability Analysis of Grid Computing Systems", Department of Industrial and System Engineering, National University of Singapore.

[2]  Mohammad Khanli. Leyli,  Etminan Far . Maryam, Ghaffari . Ali. 2010," Reliable Job Scheduler using RFOH in Grid Computing", Journal of Emerging Trends in Computing and Information Sciences CIS Journal. All rights reserved.

[3]  Cao. J, Spooner. D. P, Jarvis. S. A, Nudd.v.,( 2005)._Grid load balancing using intelligent agents. Future_Gener. Comput. System.

[4]  Grosan. C, Abraham. A, and Helvik. B. ( 2007).Multiobjective evolutionary algorithms for scheduling jobs on computational grids.ADIS InternationalConference, Applied Computing, Salamanca,Spain, Nuno Guimares and Pedro Isaias (Eds).

[5]  Subrata. R., Zomaya. A. Y, and. Landfeldt. B( 2007). Arti_cial_ life techniques for load balancing in computational_grids. J. Comput. Syst. Sci.,

[6]  J. K. J and. Eberhart. R. C,( 2001). Swarm Intelligence._Morgan Kaufmann Publishers

[7]  J. K. J and R. C. Eberhart. Swarm Intelligence._Morgan Kaufmann Publishers, 2001.

# Advantages and Disadvantages of Mikrotik Nv2 Protocol on Wireless Networks

Nader chahardah chericki
ghorbani
Sama Technical and
Vocational Training College,
Islamic Azad University,
Branch of Omidiyeh,
khoozestan, Iran

**Abstract**: In today's world of wireless communications, has a special importance in communicating.and Some of the things that makes this enormous growth is more affordable than other methods, portability, installation and rapid implementation, quality and no need for physical but this method also has its own problems, is still, If at this stage we ignore some security problems, lack of communication bandwidth is a major problem facing the problem of distance communication also seen higher, In this regard, the researchers also found several solutions have been proposed, some of which follow the general protocol and some specifically for Wi-Fi connection is created between a manufacturer of products, This article describes how the workings of mikrotik Nv2 Protocol.

**Keywords**: *NV2, 802.11 , CSMA/CA, mac polling, mikrotik, polling*

## 1. Introduction

Up to Today, solutions and several methods to access media including using CSMA / CA or using multiple access with collision avoidance [1], RTS/CTS [2] or techniques, such as PCF, HCF, CSMA/CRAP, MACA[3], All techniques to send information on physical media, One of the ways which researchers and manufacturers of wireless equipment introduced and It will be capable of data transfer rates up considerably compared to previous methods used is polling. [4], Inspection method can be divided into two groups, each to be associated with the nodes in the wireless topology and the two types of the :
1. Roll Call Polling
2. Hub Polling

## 2. POLLING METHODS

This section describes the Protocol nv2 and the advantages and disadvantages that defines.

## 2.1 ROLL CALL POLLING

In this method there is a central controller and A channel out of band communication that central controller visit all nodes by send a token to them though out-band channel and Each station has data to send, send data through out-band channels and when transmit finished, send a finish message to central controller, If a station that is deemed not to send the information, the controller will inspect the other nodes. [5] In this method the control center with a series of packets containing information such as schedule and Header data to individual nodes within the network to visit them and ask to send data if has data to transfer and promulgate data transfer end.
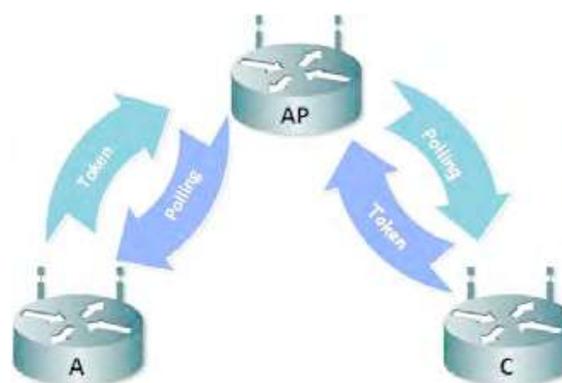


Figure 1: Roll Call Polling

## 2.2 HUB POLLING

In this method The network consists of a central station and a secondary station is zero or more. The main

station responsible is divided into several sections, which is as follow:

1- initial value and identification network

2- Invites new stations to the network and addressing them

3- Maintain and inspect the performance of the virtual ring network.

The overall effect of this method is that the base station network has to handle and store information, Each secondary station connected to the network, such as a passive entity that expects to receive poll frame from central station, Each node is added to the network must know the address of the next stop of its own, Start the process begins when the central station give access to the media by assign poll frame to the first station of its own and If the station has data to send, start data transfer Otherwise it will send the poll frame to the next station, This too will continue to poll frames to reach the central station and the cycle starts again [6].
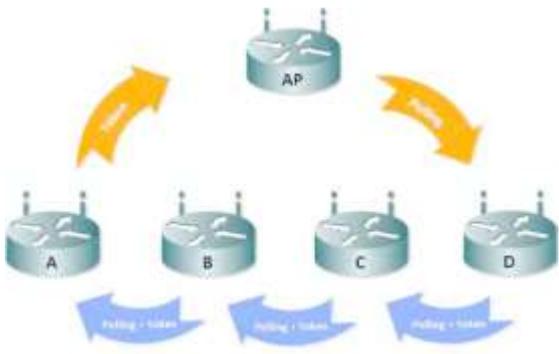


.Figure 2: Hub polling method

## 2.3 POLLING BASE METHODS DISADVANTAGES

In polling methods One of the disadvantages that reduces the time loss between the nodes of the media comes, is hands down token between stations. So that in this method, token first to be delivered to the highest station, Selected station after data transfer send token back to the central controller or an next station This problem increases the time delay between the two stations with the media which sought to reduce the amount of channels usage and Waste transfer rate yielded, However, this method can be a better alternative to using CSMA/CA, So that would be different in terms of data transfer rate of up to 05% increase, so that can increase data transfer rate up to 50% in different situation But also increases the time delay [7].

## 3. Nv2 PROTOCOL

Nv2 is one of the unique mikrotik protocol in Order to increase the rate of data transmission in wireless platform generated and Usually can be work with pci cards that has Atheros chips, this protocol use of TDMA instead of the POLLING or CSMA/CD, this method can solve hidden nodes problem in wireless communications and increase data transfer rates in Point To Point and Point To Multi Point links, Media access in Nv2 network is controlled by Nv2 Access Point. Nv2 AP divides time in fixed size "periods" which are dynamically divided in downlink (data sent from AP to clients) and uplink (data sent from clients to AP) portions, based on queue state on AP and clients. Uplink time is further divided between connected clients based on their requirements for bandwidth. At the beginning of each period AP broadcasts schedule that tells clients when they should transmit and the amount of time they can use. In order to allow new clients to connect, Nv2 AP periodically assigns uplink time for "unspecified" client - this time interval is then used by fresh client to initiate registration to AP. Then AP estimates propagation delay between AP and client and starts periodically scheduling uplink time for this client in order to complete registration and receive data from client. Nv2 implements dynamic rate selection on per-client basis and ARQ for data transmissions. This enables reliable communications across Nv2 links [8].
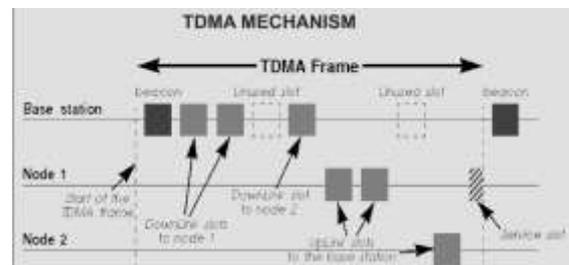


Figure 3: Nv2 structure [9].

## 3.1 ADVANTAGES

A. More Throughput
B. Lower Latency
C. Good for Point to Multipoint Networks
D. Solves Hidden Node Problem

## 3.2 DISADVANTAGES

A. Only RouterOS devices will be able to participate in Nv2 network and only RouterOS devices will see Nv2 AP when scanning.
B. Nv2 network will disturb other networks in the same channel and Nv2 network may be affected by any (Nv2 or not) other networks in the same channel.

C. Nv2 enabled device will not connect to any other TDMA based network.

D. Does not Support 802.11x standard WPA/WPA 2 Security Protocols

## 4. CONCLUSION

According to tests done on a link with a distance of about 2 km and Almost 150 meters height difference between send and receive points and In a warm region and dry air, The results show that due to the noise in the area and about 30 °C air temperature Nv2 protocol is better than 802.11 protocol on mikrotik products And increases the data transfer bandwidth, In addition, this protocol is secure because only mikrotik devices able to detect this protocol, Below pictures show results of the test in the winter and 30% of air humidity.
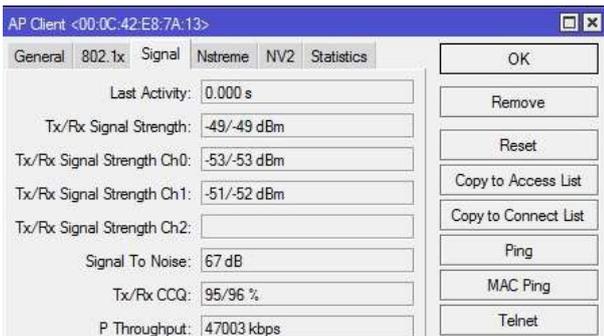


Figure. 4.  802.11 Protocol signal quality
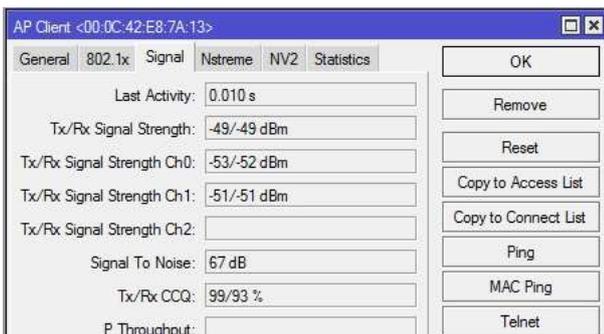


Figure. 5.  Nv2  Protocol signal quality
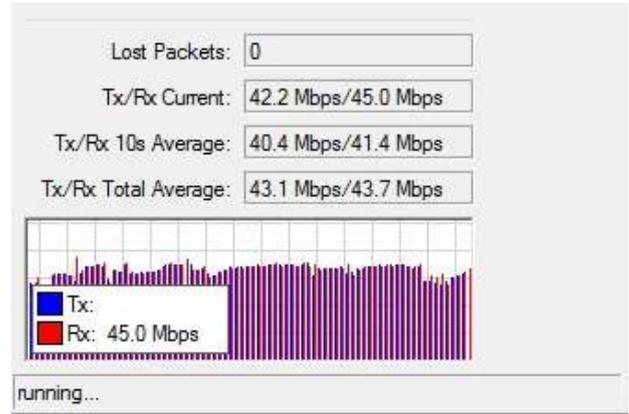


Figure. 6.  Nv2Protocol TCP bandwidth test
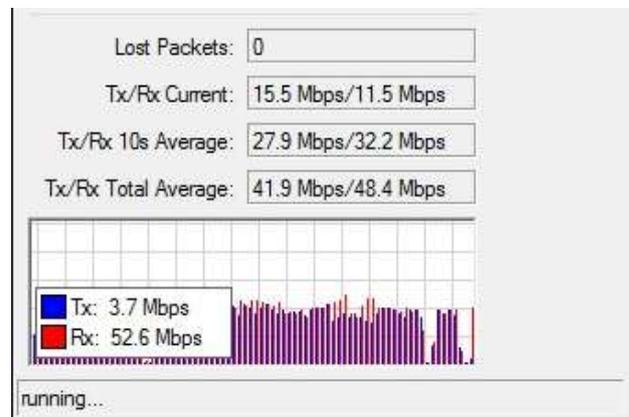


Figure.7.  802.11 Protocol TCP bandwidth test

## 5. REFERENCES

[1]  andrezuguete,improved csma/ca protocol for ieee 802.11, next generation internet network 2008.Ngi 2008 IEEE 28-30 april 2008 p 247-252,krakew

[2]  revisit of ris/cts exchange in high-speed IEEE 802.11 networks, World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a, 13-16 June 2005 ,p 240-248

[3]  http://en.wikipedia.org/wiki/media_access _control

[4]  802.11 offers a polling channel access mechanism (point coordination function)in addition to the csma/ca one. http://www.hpl_hp_com/personal/jean_tourril hes/linux/linux.wireless.mac.html

[5]  Vuong, S.T. , A low-cost and portable local area network for interconnecting PC's using electric power lines , Selected Areas in Communications, IEEE Journal on  (Volume:7 ,  Issue: 2 ), Feb. 1989 ,p 192-201.

[6] Young-Jae Kim , "Adaptive polling MAC schemes for IEEE 802.11 wireless LANs" 57th IEEE Semiannual , 2528 - 2532 vol.4, 2003Young-Jae Kim , "Adaptive polling MAC schemes for IEEE 802.11 wireless LANs" 57th IEEE Semiannual , 2528 - 2532 vol.4 , 2003

[7] Normandy Pour Reza, Moghadam Reza, NV2 protocol to increase the data transfer rate to replace the mutiny, sixteen Iranian Student Conference on Electrical Engineering, Islamic Azad University Kazeroon, 41-41 September 4311

[8] Nv2 protocol is proprietary wireless protocol developed by MikroTik http://wiki.mikrotik.com/wiki/Manual:Nv2

[9] soumilguptab haya2, Mum india 2012, prop Ritarymirdess protocols

# Mobile Personalized Notes Using Memory Package

Cheng-Chieh Chiang
Department of Information Technology
Takming University of Science and Technology
No. 56, Sec. 1, Huan-Shan Rd., Taipei 114, Taiwan

**Abstract**: Smart phones and their mobile application have been deeply affected people life in the modern society. We can find many kinds of mobile applications, either in Google Play or Apple Store, designed according to novel motivations to assist us in improving people life. This paper aims to develop a personalized note tool on smart phones to integrate different types of multimedia information to help the users record activities. In this paper we propose a structure called memory package for integrating multimedia information on smart phones. Users can create a memory package in a time period to record their activities by using multimedia contents such as text, image, video, audio, and GPS records. Our system designs a unified interface for the memory package that can help users easily generate and review their packages on phones. Users can preserve clips of user activities of life anytime and anywhere using the proposed system of personalized notes as well as the memory package.

**Keywords**: memory package; personalized notes; smart phone; mobile application; multimedia information

## 1. INTRODUCTION

Smart phones have been more popular in the modern city. Many applications on smart phones, either in Android or iOS systems, are developed with different goals to deeply change our life. Let us take a simple case for illustration. When people take a smart phone, they can capture photos of interest in life, track the GPS trajectory of their family tour, or record an important meeting in work. In past, some applications have been designed to help the users efficiently utilize sensors of mobile devices, and hence kinds of multimedia information have been generated for different goals. How to well integrate mounted sensors as well as multimedia information is a critical issue to improve the utilities of mobile devices for users.



Figure. 1 The overall concept of our proposed ideas on smart phones.

In this paper we aim to design a mobile application that can serve people as a personalized secretary tool on their smart phones to integrate different types of multimedia information to help the users record activities in life. Figure 1 presents the overall concept of our proposed ideas how to integrate multimedia information to record human activities in smart phones. Here multimedia information including image, video, and audio can be captured by hand-held smart phones to assistant the users in keeping their activities anytime and anywhere. In our design, these multimedia information generated in our life are collected and called "memory package". Users can recall their memories in the package pool in this mobile application.

Let us take the following illustrations for our proposed application. Assume John attends a two-hour meeting of the supplier presentation. He launches our proposed application in smart phone and select "meeting" scene to start on a new memory package. First, John sets the sound recording on in order to record all discussions. While this meeting is running, John can have notes for his ideas and capture photos of products. This memory package involves two-hour contents of this meeting, including audios of the presentation and discussion, images of highlighting products, and textual notes of relate ideas for this meeting. In a happy weekend, John enjoys a vacation in beach with family. Our application starts again and set as "tour" scene to create another new memory package before leaving. The GPS trajectory during this tour can be preserved in this memory package. During this two-day vacation, John can record videos for his family, take pretty images for beautiful scenery, and write some notes for funny jokes. Then, the tour memory package stores his lovable memory in two days.

The design of the memory package can achieve several properties. First, one package can collect memory clips that are related to the same event, one to the meeting and the other to the family tour in our illustrations mentioned above. Moreover, these memory clips are constructed by different kinds of multimedia information and then integrated in the packages. Beside raw records such as recording audio and storing GPS sequences, the users can also highlight contents in the packages by writing notes and taking images according their preferences. Note that the selection of the scene mode provides a prior concept too. Then, the memory packages can be delivered to the server site and the advanced analysis can be performed based on the integration of multimedia memory.
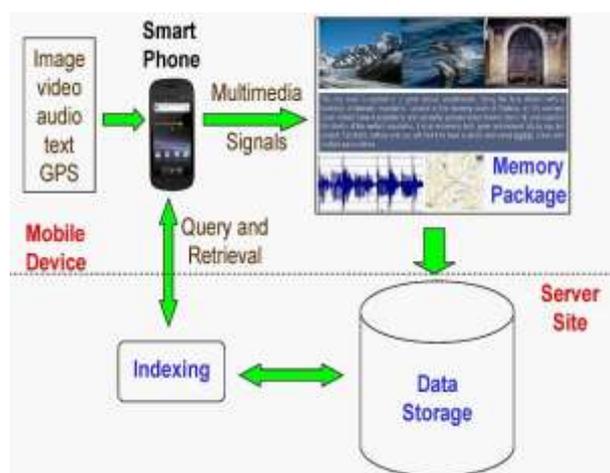
The rest of this paper is organized as the follows. Section 2 provides previous works related to this work. Section 3 presents our design of personalized memory package in this paper. Section 4 demonstrates the implementation details of the proposed system. Then finally, Section 5 makes our conclusions about this work and draws future works to improve this system.

## 2. RELATED WORKS

Personalized multimedia system has become more public in research and industrial areas in recent due to the rapid increasing of mobile devices. J. Martin and C. Trummer [7] proposed a personalized multimedia system for museums and exhibitions in a hand-held system. This system can adjust individual visitors' interests to refine the presentation of digital contents that are associated with knowledge based exhibitions and exhibition components. S. Potter et al. [8] designed a portable system called MediaPod that allows users maintaining the persistent and personalized multimedia environment on any available computer. W.-T. Chu et al. [2] proposed a tiling slideshow system to analyze video and audio information. The tiling slideshow system categorized video contents and then designed a playing scheme integrated with the audio tempo. K.-Y. Cheng et al. [1] proposed a video playing system called SmartPlayer to provide layers of events for users by analyzing motion and semantics in video. These two systems were designed for desktop or laptop, not for mobile system.

N. Sebe and Q. Tian [9] published a literature survey related to personalized multimedia retrieval in 2007. In their categorization, the possible future directions of personalized multimedia retrieval included the multimedia content description, user preferences and context, automatic content annotation, automated presentation authoring, and distributed retrieval and filtering of content descriptions and user preferences. J. Huber et al. [4] designed seven navigation approaches for video browsing on mobile device and deeply discussed the design space and the characteristics of interaction concepts for mobile video browsing. T. Karrer et al. [5] presented a mobile video system called PocketDRAGON that allows direct manipulation video navigation on mobile touch-screen devices.

In industry, many personalized image or video systems have been published in the past. For example, iPhoto [12] and iMovie [11] are two famous tools published by Apple. iPhoto can provide variant utilities for personal photos such as special-effect editing, sliding show, streaming, and event- or location-based photo management. iMovie focuses on video editing, including personal video generation, editing, template, special effect, etc. Users of Both of iPhoto and iMovie can publish their own photos and videos in Internet by use of Facebook and YouTube. Moreover, Apple also integrates these two systems with iCloud [10] such that all of images and videos can be handled over the clouding environment.

## 3. MEMORY PACKAGE

Smart phone is convenient to serve that people can record anything in anytime and anywhere. Multiple devices, such as camera and voice recording, are mounted in a smart phone to provide multimedia information for users. Thus, we always have much information stored in our smart phones such as images and videos etc., and then the information could be

considered clips of our memory that can keep parts of life. Hence we design a unified structure called memory package to collect different types of multimedia information associated with our memory in life.

One of the main goals for memory package is to deal with the integration between different types of multimedia in a smart phone. Figure 2 presents the description layer of the memory package by providing extra metadata for multimedia contents. Metadata in the description layer includes the scene mode selected by the user, the temporal information, and the original metadata of multimedia. The temporal information contains the time order of multimedia, and the original metadata contains the corresponding information of image, video, and audio such as resolution and size etc. That is to say, the description layer aims to summarize multimedia contents included in the memory package. Moreover, our proposed system designs a unified interface to control multimedia devices (camera, voice recording, GPS) mounted in a smart phone to generate contents of the memory package. The users can employ our system to easily switch to these devices to record activities in life.
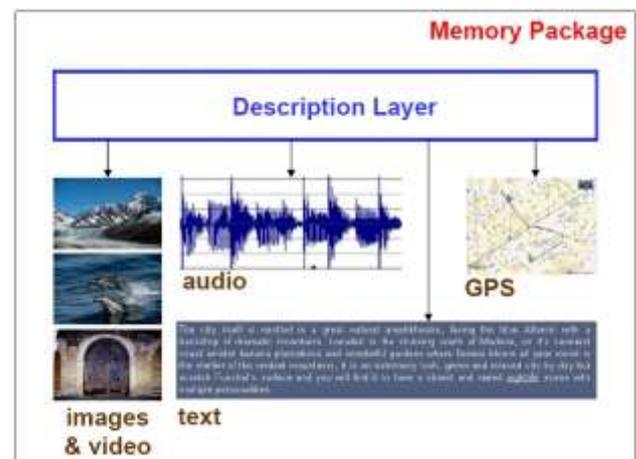


Figure. 2 The structure of the memory package. A description layer is appended to provide extra metadata of the corresponding multimedia contents.

When a new memory package is created, semantic relationships have been implicit in all multimedia contents. At least, all contents are associated with the same memory event. Indeed, there are also additional metadata in the description level that can help us construct semantic links among multimedia in the memory package. Furthermore, content-based approach [3][6] can be applied to discover relationship of multimedia information across memory packages. Hence, our design of the memory package can provide a flexible structure that has potential to construct semantic connections both inter- and outer-packages.

## 4. IMPLEMENTATION

We have implemented a beginning version of the personalized application for memory package in smart phone. Our system was implemented in Android smart phones, and we hope to extend it to Apple iOS system. This section briefly introduces our implementation of the proposed personalized memory package in Android system. When the application is launched, the users have two choices in the home screen shown in

Figure 3: new package and view package; the former to create a new memory package and the latter to browse and view existed packages.



Figure. 3  Users can create and browse memory packages in the home page of our system.
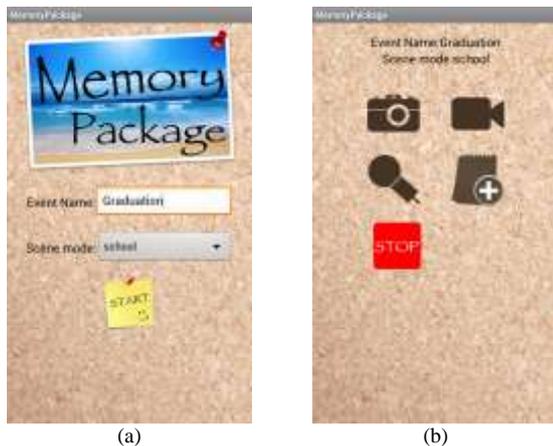


(a)                              (b)



(c)

Figure. 4  The creation of a new memory package. (a) Set the package name and its corresponding scene mode. (b) Several multimedia devices can be used for recording human activities. (c) Make a copy to the storage on cloud when the new package has been created.

Figure 4 shows some snapshots of functions for creating new memory packages. When a new package is generated, a package name and its corresponding scene mode are first

requested shown in Figure 4(a). Then, the users can select and switch multimedia devices to record anything during the package period, shown in Figure 4(b). Moreover, each of packages can be duplicated to make a copy to the storage on cloud shown in Figure 4(c) when this package creating is stopped.

The other main function, browsing packages shown in Figure 5, aims to provide an easy way for the users to find packages of interest and then look at the detailed contents of multimedia. Figure 5(a) presents several browsing functions of memory packages, containing filtered by scene modes, sorted by name/time/scene, and displayed by icon/list. When the users select a specific package, its multimedia contents are listed shown as Figure 5(b). All of multimedia contents in a package are ordered by creation time, so the users can realize the temporal relationships among events of this memory package. The users can click each of elements in Figure 5(b) to view individual media files directly. Note that the first text file is the description of this package, and GPS coordinates can be appended into multimedia contents
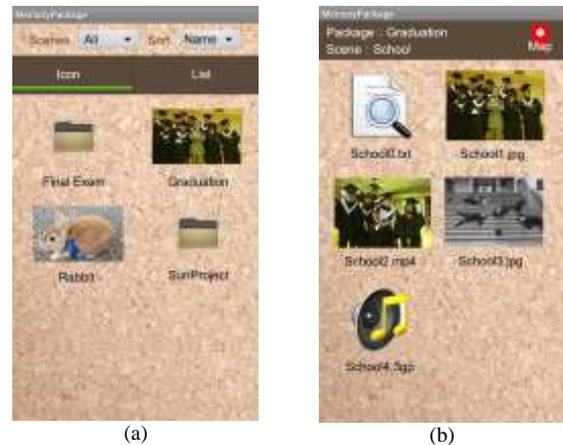


(a)                              (b)

Figure. 5 The browsing of memory packages. (a) Several functions for browsing memory packages. (b) The multimedia contents of a memory package.

# 5.  CONCLUSION

This paper presents our design of the personalized multimedia system to assist the users in recording life activities anytime and anywhere in smart phone. We propose an integrated structure called memory package to collect all of multimedia information captured by devices mounted in smart phone, such as image, video, audio, text, and GPS. This paper also describes the details of our proposed application that has been implemented in Android system. Future work mainly extends functions for our proposed system that only has the basic functions of our design at the current stage. Also, an advanced integration scheme should be considered to better visualize and represent multimedia information in memory packages.

# 6.  REFERENCES

[1]  K.-Y. Cheng, S.-J. Luo, B.-Y. Chen, and H.-H. Chu, "SmartPlayer: User-Centric Video Fast-Forwarding," in Proceedings of 27th ACM Conference on Human Factors in Computing Systems (CHI), 2009.

[2] W.-T. Chu, J.-C. Chen, and J.-L. Wu, "Tiling Slideshow: An Audiovisual Presentation Method for Consumer Photos," IEEE Multimedia, 14(3), 2007.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, Vol. 40, No. 2, pp. 1-60, 2008.

[4] J. Huber, J. Steimle, and M. Mühlhäuser, "Toward More Efficient User Interfaces for Mobile Video Browsing: An In-Depth Exploration of the Design Space," in Proceedings of ACM Multimedia (ACM MM), 2010.

[5] T. Karrer, M. Wittenhagen, and J. Borchers, "PocketDRAGON: A Direct Manipulation Video Navigation Interface for Mobile Devices," in Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (mobileHCI), 2009.

[6] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based Multimedia Information Retrieval: State of the Art and Challenges," ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 2, No. 1, pp. 1-19, 2006.

[7] J. Martin and C. Trummer, "SCALEX — A Personalized Multimedia Information System for Museums and Exhibitions," in Proceedings of 1st International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS), pp.299-302, 2005.

[8] S. Potter, R. Baratto, O. Laadan, L. Kim, and J. Nieh, "MediaPod: A Personalized Multimedia Desktop in Your Pocket," in Proceedings of 11th IEEE International Symposium on Multimedia, 2009.

[9] N. Sebe and Q. Tian, "Personalized Multimedia Retrieval: The New Trend?" in Proceedings of the 9th ACM SIGMM International Workshop on Multimedia Information Retrieval (ACM MIR), 2007.

[10] iCloud. https://www.icloud.com/.

[11] iMovie. http://www.apple.com/tw/ilife/imovie/.

[12] iPhoto. http://www.apple.com/tw/ilife/iphoto/.

# Hubness in Unsupervised Outlier Detection Techniques for High Dimensional Data –A Survey

R.Lakshmi Devi
Mother Teresa Women's University
Kodaikanal, India

R.Amalraj
Dept. of Computer Science
Sri Vasavi Arts College
Erode, India

**Abstract: --** Outlier detection in high dimensional data becomes an emerging technique in today's research in the area of data mining. It attempts to find objects that are considerably unrelated, unique and inconsistent with respect to the majority of data in an input database. It also poses various challenges resulting from the increase of dimensionality. Due to the "increase of dimensionality," distance becomes meaningless. Hubness is an aspect for the increase of dimensionality pertaining to nearest neighbors which has come to an attention. This survey article, discusses some important aspects of the hubness in detail and presents a comprehensive review on the state-of-the-art specialized algorithms for unsupervised outlier detection for high dimensional data and role of hubness.

**Keywords:** Hubness, High dimensional data, Outliers, Outlier detection, Unsupervised.

## 1. INTRODUCTION

An outlier is an observation which appears to be inconsistent with the remainder of that set of data. In data mining, detection of outliers is more interesting and an important research area. Many applications which apply outlier detection are Fraud detection for credit card, Loan application processing, Intrusion detection cyber-security, Network performance, insurance or health and so on. Most of these applications are high dimensional which means that data can contain hundreds of dimensions. Moreover in high dimensional space, the data is sparse. The sparsity of high dimensional data denotes that every point is an almost equally good outlier [1]. Many algorithms have been proposed in recent years for handling such a problem. This paper focuses a broad overview of extensive research on outlier detection techniques for high dimensional data and a role of hubness.

The various sections to be discussed are categorized in the following order: an introduction to the outlier detection is explained in Section 2, various outlier detection techniques for high dimensional data in Section 3, and Section 4 provides a detailed introduction to the phenomenon of hubness and also examines how hubness is used in various techniques. Finally the chapter is concluded in Section 5.

## 2. OUTLIER DETECTION

Outlier (anomaly) detection refers to the process of finding patterns that do not conform to standard behavior. These non-conforming patterns are often referred to as anomalies, outliers, exceptions in different application domains.

There are three categories of outliers:

*1. Point Outliers:* If specific data instance is considered as inconsistent with respect to the rest of data, then the instance is stated as a point outlier.

*2. Context Outlier:* If a data instance is inconsistent in a specific context, then it is stated as a contextual outlier.

*3. Collective Outlier:* If a collection of related data instances is inconsistent with respect to the entire data set, it is stated as a collective outlier [2].

Outliers can also be categorized into vector, sequence, trajectory and graph outliers etc., depending on the type of data from where outliers can be detected.

*Vector outlier:* Vector outliers are detected from vector like representation of Data such as the relational database.

*Sequence outlier:* In many applications, data are existing as a sequence. A good example is the computer system Call log.

*Trajectory outlier:* Recent developments in satellites and tracking facilities utilize a huge amount of trajectory data of moving objects. For eg. Animal movement data.

*Graph outlier:* Graph outliers are graph entities that are strange from their Peers. Examples are nodes, edges and sub graphs [3].
    The labels of a data instance represent if that instance is *normal* or *Anomalous.* Depending on the labels connected the data instance, anomaly detection techniques can be in one of the following three modes:

*Supervised anomaly detection.* Techniques trained in supervised mode assume the availability of a training data set which has labeled instances for normal as well as anomaly class.

*Semi-Supervised anomaly detection.* Techniques that operate in a semi supervised mode, assume that the training data has labeled instances for only the normal class. Since they do not require labels for the anomaly class, they are more widely applicable than supervised techniques.

*Unsupervised anomaly detection.* Techniques that operate in unsupervised mode do not require training data, and thus are most widely applicable [2].

Among these categories, unsupervised methods are more widely applied [4], because the other categories require accurate and representative labels that are often expensive to obtain. [4] Shows that unsupervised methods can detect outliers which are *more pronounced* in high dimensions, under the assumption that all (or most) data attributes are meaningful, i.e. not noisy.

# 3. OUTLIER DETECTION FOR HIGH DIMENSIONAL DATA

There are many applications in high dimensional domains in which data contain dozens or even hundreds of dimensions. In high dimensional space, the data are sparse and the idea of proximity fail to achieve their effectiveness. This is due to the increase of dimensionality that concentrates the high dimensional data tend to be equi-distant to each other as dimensionality increases.

To face the challenge associated with high data dimensionality, two different categories of work is to be conducted. The first category of methods project the high dimensional data to lower dimensional data. Dimensionality detection techniques such as Principle Component Analysis (PCA), Independent component Analysis (ICA), singular value Decomposition (SVD) etc. can be applied to the high dimensional data before outlier detection is performed.. Second category is to redesign the mechanism to accurately capture the proximity relationship between data points in the high dimensional space [3].

The earlier research methods for unsupervised outlier detection in high dimensional data are discussed below.

Since the distance based method for outlier detection is not effective due to various noise in high dimensional data [5] utilizes an approach named ABOD (Angle Based Outlier Detection) which uses the variance of angles between the points as an outlier degree with the assumption that angles are more stable than distances Also compares ABOD with the distance-based method LOF for a real world data set to show that ABOD to perform well on high-dimensional data.

The paper [6] proposes method for finding distance-based *outliers* based upon the k nearest neighbor points. To calculate the number of data points falling, two algorithms such as nested loop join and index join algorithms are used. Also partition-based algorithm is used. This algorithm divides the data set into different subsets and then cuts entire partitions rapidly as it is determined that they cannot contain outliers.

Distance based method to deal with the problem of finding outliers for k dimensional data sets where k >=5 is focused in paper [7]. Applying three algorithms such as index based, nested loop based, and cell based, authors come to the conclusion that cell based is for k<=4 and nested loop is the choice for k>=5 and also finds that there is no limit on the size of the dimensions.

The approach used in [8] is based on the relationship between k- Nearest Neighbor and Reverse Nearest Neighbor which involves two phases. First phase is dealing with the problem of finding a query point using kNN.The second step introduces Boolean Range Query which checks the existence of a point in a given region can be used for RNN queries problems.

Density based local outlier is identified in [9] with the help of LOF for many KDD (Knowledge discovery in databases) applications, such as detecting criminal activities in E-commerce, finding the rare instances or the outliers. Author introduces LOF (Local outlier factor) which is local meaning that it considers only the restricted neighborhood. Also proves that LOF is useful for finding the meaningful outliers in real world dataset.

A new method (LOCI—Local Correlation Integral method) for finding outliers in large, multidimensional data sets is recognized in [10] which Introduces the multi-granularity deviation factor (MDEF), to detect both remote outliers as well as outlying clusters and proposes a method which is associated with MDEF to check the existence of an outlier point by comparing MDEF value with local average. It is automatic and can be computed quickly.

LDOF (Local Distance-based Outlier Factor) approach [11] to handle KDD applications, measures the outlier-ness of an object. LDOF uses the relative location of an object to its neighbors to determine the degree to which the object deviates from its neighborhood. According to its violation degree the outlier is found. The neighborhood size is chosen only by analyzing the properties of LDOF. Top n technique is also established here to simplify the parameter setting. Also proves that it remains stable even for large range of neighborhood sizes compared to top-n KNN and top-n LOF.

The approach used in this paper [12] is to standardize the computation of an outlier score for each database object and introduces a LoOP (Local Outlier Probability) outlier detection model which is the combination of the idea of local, density-based outlier scoring with a probabilistic, statistically-oriented approach. This method is formulated to provide an outlier score which is in range of [0, 1].This is ease with which outlier score can be computed and interpreted for the comparison of datasets.

Improved K-means technique for outlier detection in high dimensional dataset is explored in [13]. This paper solves the finding of outlier detection by applying the existing Clique method for high dimensional dataset to generate subspace and then the improved k-means algorithm is applied on generated subspaces for identifying outliers.

A hybrid approach for outlier detection in high dimensional data by combining both density based and distance based approach is identified in [14] so that it can take the benefits of both density and distance based clustering methods. DBSCAN (Density based Spatial Clustering Application with noise) a density based technique and k-means are combined in this hybrid approach.

An efficient outlier detection methods has been proposed in [15] which is based on fuzzy c means clustering using Artificial Bee colony algorithm. Fuzzy clustering is used to choose the cluster heads, ABC to select the members of the clusters. When the ABC-FCM algorithm is first performed, it is found that the small clusters are the outlier clusters. Other outliers are then determined based on computing differences between objective functions values when points are temporarily removed from the data set. If a noticeable change occurred on the objective function value, the points are considered.

Reverse nearest neighbors count is recognized in unsupervised distance-based outlier detection [4]. The concept of hubness is introduced here and explores the interplay of hubness and data sparsity. Outier detection methods are implemented based on the properties of antihubs. The relationship between dimensionality, neighborhood size, and reverse neighbors are taken into account for the effectiveness of the mod

When the above outlier detection methods are analyzed, recent research proves that the concept of hubness is extensively tend to be needed for handling the problem of the increase of dimensionality in outlier detection for high dimensional data. The following section discusses characteristics of hubness and hubness based techniques for outlier detection in high dimensional data.

## 4. HUBNESS BASED OUTLIER DETECTION

### 4.1 Hubness

High dimensionality of data space resulting from the increase of dimensionality occur in many domains and face challenges for traditional data mining techniques, both in terms of effectiveness and efficiency. Main aspect of the increase of dimensionality is distance concentration, which denotes the tendency of distances between all pairs of points in high dimensional data to become almost equal [16]. Hubness is an aspect of the increase of dimensionality related to nearest neighbors.

Let $D \subset R^d$ be a finite set of n points. For point $x \in D$ and a given distance or similarity measure, the number of k-occurrences, denoted $N^k(x)$, is the number of times x occurs among the k nearest neighbors of all other points in D. For $q \in (0, 1)$, *hubs* are the $\lceil nq \rceil$ points $x \in D$ with the highest values of $N^k(x)$ [4].

It has been shown that hubness, as a phenomenon, appears in high-dimensional data as an inherent property of high dimensionality [4].

### 4.2 Role of Hubness in machine learning

Recently several papers considered the consequence of hubness in high dimensional data on different data mining and machine learning tasks [16] [18-20]. If we survey the origin of hubness phenomenon, it mainly proves that it is an essential property of high dimensional vector space, and it explore its effect on applications based on measuring distances in vector spaces, notably classification, clustering and information retrieval[18].

[16] Explores the aspect of the increase of dimensionality that is demonstrated through the usage of the concept known as hubness and thoroughly examines the emergence of it and shows that it is an intrinsic property of high dimensional data. Subsequently [16] utilizes the effect of hubness on lot of machine learning tasks belonging to supervised, semi supervised and unsupervised learning families and also discusses the interaction of hubness with dimensionality reduction.

Clustering is the process where similar elements are grouped together. This process becomes hard when the sparse of

data is high especially in high dimensional data and does not distinguish the distance between the data points properly. In [20] the authors successfully explore the concept of hubness in clustering high dimensional data by using point hubness scores to guide the search, but choose a centroid-based cluster at the end. With this concept they propose an algorithm called Global hubness-proportional K-means (GHPKM) also compares the proposed algorithm to kernel K-means and one standard density-based method, GDBScan. Finally proves that the proposed algorithm is more robust than the others.

The existence of hubness phenomenon and its applications are comprehensively experimented also in other application fields like music retrieval[21] and classification[16][19][22-24][28-30], image feature representation[25], data reduction[16][26], collaborative filtering[27] and Text retrieval[17].

Paper [22] presents various fuzzy measures for k nearest neighbor classification especially designed for high dimensional data with the usage of the concept known as hubness which express fuzziness of elements appearing in k neighborhoods of other points.

### 4.3 Role of Hubness in outlier detection

In Recent research, various papers discussed the influence of so called hubness in high-dimensional data on different data mining outlier detection tasks. Papers which concentrates on hubness is given below.

The concept of hubness is observed in [14], which affects reverse nearest-neighbor counts, i.e. k-occurrences. Hubness is demonstrated with the increase of the dimensionality of data, causing the distribution of k-occurrences to become skewed. So hubs very frequently become members of k-NN lists and, at the same time, antihubs become infrequent neighbors.

[31] explores a new important feature of the curse concerning to the distribution of k-occurrences (the number of times a point appears among the k nearest neighbors of other points in a data set) and shows that, as dimensionality increases, this distribution becomes considerably skewed and hub points emerge (points with very high k-occurrences) And also observes the importance of hubness and finds that it is an essential property of high dimensional data.

Identifying unsupervised outliers especially in high dimensional data becomes a tedious procedures. The paper [32] proposes a new approach for unsupervised outlier detection in high dimensional data. Antihub phenomenon is introduced to tackle high dimensional data and an algorithm Recursive Antihub$^2$ extending Antihub$^2$ is constructed to reevaluate the outlier score of a point by considering $N_k$ scores of the neighbors of x in addition to $N_k(x)$ itself. Also proves that this new algorithm improves the computational complexity with reduced number of iterations. In paper [4] also, hubness take place major role in finding unsupervised outliers in high dimensional data.

## 5. CONCLUSION

In this paper, the survey is discussed with different ways in which problem of unsupervised outlier detection for high

dimensional data has been formulated in literature and have attempted to provide an overview of huge literature on various techniques. Implementation of high dimensional data in most of the applications become an issue nowadays due to increase of dimensionality. Hubness is the recently known concept for handling the problems related with the increase of dimensionality and it is understood that it is an intrinsic property of the data where dimension is high. So the role of hubness has been examined in this paper. Also reviewed some of the recent advancements in unsupervised outlier detection for dealing with more complex high dimensional data with the usage of hubness. Outlier detection for high dimensional data is a fast growing emerging technique of today's research and more new methods regarding this technique will emerge in the future.

# 6. ACKNOWLEDGMENTS

# 7. REFERENCES

[1] C. C. Aggarwal and P. S. Yu, "Outlier detection for high dimensional data" ACM SIGMOD RECORD February 2002.

[2] V. Chandola, A. Banerjee, and V. Kumar, Anomaly detection: A survey ACM Comput Surv, vol. 41, no. 3, p. 15, 2009.

[3] Zhang, Ji. "Advancements of outlier detection: A survey." ICST Transactions on Scalable Information Systems 13.1 (2013): 1-26.

[4] Miloˇs Radovanovi´c, Alexandros Nanopoulos, and Mirjana Ivanovi´ "Reverse Nearest Neighbors in Unsupervised Distance-Based Outlier Detection" IEEE Transactions On Knowledge And Data Engineering, October 2014.

[5] H.-P. Kriegel, M. Schubert, and A. Zimek, "Angle-based outlier detection in high-dimensional data," in Proc 14th ACM SIGKDD Int Conf on Knowledge and Data Mining (KDD), 2008, pp. 444–452.

[6] Ramaswamy, Sridhar, Rajeev Rastogi, and Kyuseok Shim. "Efficient algorithms for mining outliers from large data sets." ACM SIGMOD Record. Vol. 29. No. 2. ACM, 2000.

[7] Knorr, Edwin M., Raymond T. Ng, and Vladimir Tucakov. "Distance-based outliers: algorithms and applications." The VLDB Journal—the International Journal on Very Large Data Bases 8.3-4 (2000): 237-253.

[8] Singh, Amit, Hakan Ferhatosmanoglu, and Ali Şaman Tosun. "High dimensional reverse nearest neighbor queries." Proceedings of the twelfth international conference on Information and knowledge management. ACM, 2003.

[9] Breunig, Markus M., et al. "LOF: identifying density-based local outliers." ACM SIGMOD record. Vol. 29. No. 2. ACM, 2000.

[10] Papadimitriou, Spiros, et al. "Loci: Fast outlier detection using the local correlation integral." Data Engineering, 2003. Proceedings. 19th International Conference on. IEEE, 2003.

[11] Zhang, Ke, Marcus Hutter, and Huidong Jin. "A new local distance-based outlier detection approach for scattered real-world data." Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, 2009. 813-822.

[12] Kriegel, Hans-Peter, et al. "LoOP: local outlier probabilities." Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009.

[13] Bhatt, Mr Pushpendra, and Tidake Bharat. "A Review Paper on Improved K-Means Technique for Outlier Detection in High Dimensional Dataset." International Journal of Engineering Sciences & Research Technology, 4(1) January, 2015.

[14] Randive, Neha, et al. "Hybrid Approach for Outlier Detection in High Dimensional Data."Int. Journal of Engineering Research and Applications, Vol. 4, Issue 4(Version 9), April 2014, pp.31-35.

[15] Chalotra, Poonam, and Maitreyee Dutta. "An Outlier Detection Method Based On Artificial Bee Colony Fuzzy Clustering." International Journal of Advanced Research in Computer Science 3.1 (2012).

[16] M. Radovanovi´c, A. Nanopoulos, and M. Ivanovi´c, Hubs in space: Popular nearest neighbors in high-dimensional data, J Mach Learn Res 11 (2010), 2487–2531.

[17] M. Radovanovi´c, A. Nanopoulos, and M. Ivanovi´c, "On the existence of obstinate results in vector space models," in Proc. 33rd Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, 2010, pp. 186–193.

[18] M. Radovanovi´c, A. Nanopoulos, and M. Ivanovi´c, Nearest neighbors in high-dimensional data: the emergence and influence of hubs, In Proceedings of the 26th International Conference on Machine Learning (ICML), Montreal, Canada, 2009. 865–872.

[19] M. Radovanovi´c, A. Nanopoulos, and M. Ivanovi´c, Timeseries classification in many intrinsic dimensions, In Proceedings of the 10th SIAM International Conference on Data Mining (SDM), Columbus, OH, 2010, 677–688.

[20] N. Tomaˇsev, M. Radovanovi´c, D. Mladeni´c, and M. Ivanovi´c, "The role of hubness in clustering high-dimensional data,"IEEE T Knowl Data En, vol. 26, no. 3, pp. 739–751, 2014.

[21] J. J. Aucouturier, "Ten experiments on the modelling of polyphonic timbre," Ph.D. dissertation, University of Paris 6, Paris, France, 2006.

[22] N. Tomaˇsev, M. Radovanovi´c, D. Mladeni´c, and M. Ivanovi´c, "Hubness-based fuzzy measures for high-dimensional k nearest neighbor classification," in Proc. 7th Int. Conf. on Machine Learning and Data Mining (MLDM), 2011, pp. 16–30.

[23] Tomašev, Nenad, et al., "A probabilistic approach to nearest-neighbor classification: Naive hubness bayesian kNN," in Proc. 20th ACM Int. Conf. on Information and Knowledge Management (CIKM), 2011, pp. 2173–2176.

[24] N.Tomaˇsev and D. Mladeni´c, "Nearest neighbor voting in High dimensional data: Learning from past occurrences," Computer Science and Information Systems, vol. 9, no. 2, pp. 691–712, 2012.

[25] N.Tomǎsev, R. Brehar, D. Mladeni´c, and S. Nedevschi, "The influence of hubness on nearest-neighbor methods in object recognition," in Proc. 7th IEEE Int. Conf. on Intelligent Computer Communication and Processing (ICCP), 2011, pp. 367–374.

[26] K.Buza, A. Nanopoulos, and L. Schmidt-Thieme, "INSIGHT: Efficient and effective instance selection for time-series classification," in Proc. 15th Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD), Part II, 2011, pp. 149–160.

[27] A. Nanopoulos, M. Radovanovi´c, and M. Ivanovi´c, "How does high dimensionality affect collaborative filtering?" in Proc. 3$^{rd}$ ACM Conf. on Recommender Systems (RecSys), 2009, pp. 293–296.

[28] Tomǎsev N, Mladeni´c D (2013a) Hub co-occurrence modeling for robust high-dimensional knn classification. In: Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science, vol 8189, Springer Berlin Heidelberg, pp. 643–659

[29] Tomašev, Nenad, and Dunja Mladenić. "Hubness-aware shared neighbor distances for high-dimensional k-nearest neighbor classification." Knowledge and information systems 39.1 (2014): 89-122.

[30] Tomašev, Nenad, et al. "Hubness-aware classification, instance selection and feature construction: Survey and extensions to time-series." Feature Selection for Data and Pattern Recognition. Springer Berlin Heidelberg, 2015. 231-262.

[31]Radovanović, Miloš, Alexandros Nanopoulos, and Mirjana Ivanović. "Nearest neighbors in high-dimensional data: The emergence and influence of hubs." Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009

[32]Sylvia, J.MichaelAntony, and T.C. Rajakumar "Recursive Antihub$^2$ Outlier Detection in High Dimensional Data.", Global Journal of Advanced Research, Vol-2, Issue-8 PP. 1269-1274

# e-Governance Implementation In Ebonyi State Nigeria: Challenges and Prospects

Fergus U. Onu
Department of Computer Science,
Ebonyi State University, Abakaliki – Nigeria.

**Abstract**

The deployment of Information and Communication Technology (ICT) in different facets of the world's economy has yielded very reasonable results. ICT has blured the barriers of hinderances in tourism, trade, healthcare, education and training. In governance and administration, ICT applications have enhanced the delivery of public services to citizens and clients not only by improving the process and management of government, but also by redefining the traditional concepts of citizenship. This paper examined the challenges facing the implementation of e-Governance in Ebonyi State – Nigeria and highlights the prospects. The research used a primary source of data by distributing, collecting and analysing a total of 500 questionnaires administered to respondents in the research area (Ebonyi State). The research found that the most difficult challenges facing the deployment of e-governance in Ebonyi State were lack of steady power supply, poor communication infrastructure and high cost of computer and internet equipments. The survey also revealed that the use of e-governance would bring improved efficiency in government operations without necessarily increasing the cost of state governance. e-Governance would strengthen democratic principles and ideologies which inturn brings good governance to the people.

**Keywords:** ICT, e-governance, good governance, internet, service delivery

## INTRODUCTION

The purpose of governance in any society or organisation is to ensure optimal benefits of resources management to the stateholders. In Nigeria, the constitution (the 1999 constitution) identified three tiers of government namely; federal, state and local governments. Each of these tiers of has defined functions they render to the citizenry. Since the emergence of civil democracy in 1999, the states in Nigeria had better autonomy because the state governors were elected by the people as against the usual appointment by the president. This brought into the scene a set of politicians with a strong private sector background, who were determined to make their marks by results rather than patronage. If this pursuit will succeed, there must be a platform to enable the citizenry to participate in governmental decision making process; a platform where everyone will be given equal opportunities to express their views about government policies and programmes without the fear of being victimized.

The use of Information and Communication Technology (ICT) in different facets of the economy has yielded unquatifiable results. ICT has blured the barriers of hinderances in tourism, trade, education and training, etc. In the realm of government, ICT applications are promising to enhance the delivery of public goods and services to citizens not only by improving the process and management of government, but also by redefining the traditional concepts of citizenship and democracy.

The use of ICTs in governance is actually a means to improve the process of governance. The framework of e-governance, enables the partial automation of the major state functions and sectors to facilitate the optimal performance and better interaction of institutions attached to these sectors of the state. In fact ICTs allows better interaction between the government, its institutions and the people.

To show that ICT is a necessity in governance, one must first consider whether the provision of the basic services are basic to the survival of the community. Basic services are "those activities that lead to the satisfaction of 'basic needs'. According to [1], basic needs range from the "minimum physiological needs" to a set or package that cover the following needs:

(a) basic necessities of life , such as food, water, clothing, fuel;

(b) needs which enhance the general welfare of the people;

(c) improve access to the means of production and economic opportunities, such as natural resources, capital (including technology), employment opportunities and income; and

(d) sense of security and freedom for decision making, such as human rights, political participation, social security, social defense and rule of law.

Based on this characterization of basic needs, ICTs can be viewed as instruments that help enable people to satisfy their basic needs. The idea that societies have survived over time even without ICT and thus need not change seems to forget the fact that societies, like all organizations, exist in an environment that is constantly changing. While ICT may not be in the category of basic needs directly, the role of ICT in development cannot be overemphasized. ICT bridges the information divide which determines the value societies place on these basic

needs. [2] said that an information society requires information age governance and citizens. ICT brings greater effectiveness and transparency in government, while creating market opportunities for goods and services.

Access to ICT can expand opportunities. This idea is based on the fact that people and organizations deal with information in all of their daily activities. All organizations handle, deliver and process information in making decisions, developing plans, giving orders, and informing other people. Oral and written communication flows between people and organizations. It could be transmitted through the phone, fax machine, computers, or through the "grapevine". There will be greater efficiency, better decisions, and more effective use of resources if more accurate and reliable information can be sent more rapidly to the people who need them. This is exactly what ICT offers. When people have access to ICT, they have expanded choices to make instead of allowing others to make those choices for them.

Consequently, we can correctly say that ICT is a conduit channel for delivering basic services like healthcare and education to the people. ICT can help maintain and build social capital by keeping people in contact and by creating networks of individuals with the same interests. It serves as a means for transferring different resources, both financial and knowledge based. It is also dramatically less expensive, less time consuming and less dangerous than the physical transportation of people and documents/information. Groups who anticipate these benefits want countries to adopt appropriate policies to take advantage of these opportunities. They envision that rapid advances in ICTs could be the answer to bridging the gap between the rich and poor and hence deliver good governance to the people.

Solutions to development issues often require changes to government processes, e.g. by decentralization of government functions. Objectives are generally to improve efficiency and effectiveness and to save costs. The driving force can also be public demand for online services and information that increase democratic participation, accountability, transparency, and the quality and speed of services. The implementation and use of ICT solutions can support governance reforms. E-governance has become more and more present around the world in this decade. Internationally, most countries are in the early stages of e-governance implementation. A good progress has been made in Europe, USA and in other Westernized countries such as Australia and Singapore. In the recent years, developing countries and their citizens have joined others to benefit from e-governance.

In the words of [3], a situation in which all interaction with the government can be done through one counter 24 hours a day, 7 days a week, without waiting in lines at government offices is possible even now if governments are willing to decentralize responsibilities and processes through the use of electronic means such as the Internet. Each citizen can then make contact with the government through a portal where all forms, legislation, news and other information will be available. Of course, at first the front office will retain several communication channels, such as physical counters, telephone, e-mail and Internet to serve everyone properly. Commercial banks in Europe, the USA, Africa and Asia have adopted this concept. Most banking transactions are now done at an ATM, by mail or through the Internet. Only in a few very special situations one has to go to a physical counter to carryout a transaction. This has saved banks an enormous amount of costs. In other words, they do more work, with less people, in less time and with less and smaller offices. Government, as a collector and source of information, can also follow this trend, in order to serve its customers (citizens, businesses, and other interest groups) better and to save costs by making internal operations more efficient.

## GOVERNANCE AND E-GOVERNANCE DEFINED

*e-governance* (Electronic governance) is a combination of the words *electronic* and *governance*. In this era of information revolution, every sphere of human activity has been revolutionized by electronic appliances in different machines especially the computer. *Electronics* in this context includes all devices like radio sets, television, telephones, computers, Global System of Mobile Communication (GSM) phones, the Internet, etc.

"Governance" has been defined by various authorities in though different, but related ways. [4] viewed governance as the exercise of economic, political and administrative authority to manage a country's affairs at all levels. It is about the process by which government, the private sector, citizens and groups articulate their interests, mediate their differences, and exercise their legal rights and obligations. [5] defined governance as the exercise of political, economic and administrative authority in the management of a country's affairs, including citizens' articulation of their interests and exercise of their legal rights and obligations. The World Bank in 1997 identified governance as having three distinct aspects thus: (1). the form of political regime; (2). the process by which authority is exercised in the management of a country's economic and social resources for development; (3). the capacity of governments to design, formulate and implement policies and discharge functions.

[6] wrote that the concept of governance denotes the use of political authority and exercise of control in a society in relation to the management of its resources for social and economic development. In the opinion of [7], governance comprises the institutions, processes and conventions in a society which determine how power is exercised, how important decisions affecting society are made and how various interests are accorded a place in such decisions. [8] defines it as the sum of the many ways individuals and institutions, public and private, manage their common affairs. It is a continuing process through which conflicting or diverse interests may be accommodated and co-operative action

may be taken. It includes formal institutions and regimes empowered to enforce compliance, as well as informal arrangements that people and institutions either have agreed or perceive to be in their interest.

From the preceding study of the different definitions by leading institutions, governance is a term by which power is exercised. It is now clear that the concept of governance denotes a "participatory process".

[9], [10], [11] in their different opinions said that e-governance refers to the processes for ensuring e-government. According to [5] e-governance is "the public sector's use of ICTs with the aim of:

1.  improving information and service delivery,
2.  encouraging citizen participation in the decision making process and
3.  making government more accountable, transparent and effective".

The African training and research centre in administration for development (CAFRAD) defines e-governance as "the use of ICTs, and especially the Internet, to adopt a new conception and attitude of governance and management where participation and efficiency are required of all the stakeholders linked in a network"[12]. The definition presents e-governance as a new way of co-coordinating, planning, formulating and implementing decisions and operations related to governance problems. It can be used by governments "to re-invent themselves, get closer to the citizenry and forge closer alliances and partnerships with diverse communities of interest, practice, expertise, conviction and interdependence within the context of national and international development agenda" [12].

In this paper, we adopt e-governance as the application of information and communication technologies (ICTs) in

1.  the delivery of government products and services, exchange of information, communication, transactions and system integration between government and citizens and government and businesses,
2.  in the internal government operations to simplify and improve democratic, government and business aspects of Governance.

In this sense, the "e" in e-governance means/points to all of the following:

*   *efficiency* in government transaction costs by using ICT to streamline their procedures, making their operations more efficient, better organised and more economical.
*   *effectiveness:* ICT can help governments to achieve better results, meet development goals through increased participation, improve the process of resource allocation, respond timely to citizens' needs and increase coverage and quality of government services.
*   *empowerment*: ICT can support increased interaction between citizens and their

governments, for citizens both to participate in the decision making process and to become more aware of their personal and community development.

*   *economic and social development:* Beyond the economic benefits that accrue to government due to efficiency and effectiveness gains, the use of ICTs in government and in its interaction
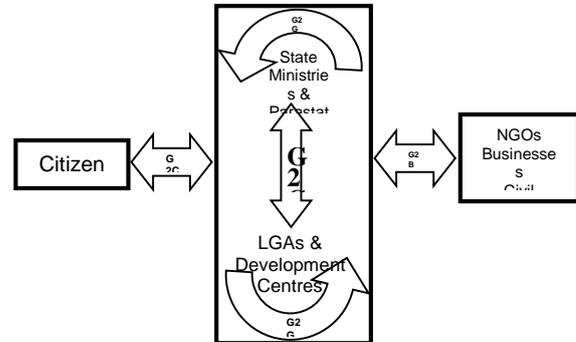


Fig. 1: Main groups interacting in e-governance

with the business community and citizens can create new businesses, attract investments and generate employment.

## E-GOVERNANCE AND GOOD GOVERNANCE

Good governance rests on the pillars of knowledge and recognition of this set of knowledge by the decision makers and the people alike. Good governance is the desire of all citizens. Digitization of this entire set of knowledge within a network which is open to all individuals opens up possibilities for all to access and use the knowledge. The Introduction of e-Governance ensures that citizens can participate in, and influence decision-making processes which affect them closely. Citizens no longer remain passive recipients of governance services provided to them, but can pro-actively decide the types and standards of governance services they want and the governance structures which can best deliver them. ICTs can influence governance processes technically, supportively and innovatively.

The strategic objective of e-governance is to support and simplify governance for all the stakeholders which include: government, citizens and businesses. The use of ICT can connect all three parties and support processes and activities. In other words, e-governance uses electronic means to support and stimulate good governance. Therefore the objectives of e-governance are similar to the objectives of good governance. Good governance can be seen as an exercise of economic, political, and administrative authorities to better manage the affairs of a country at all levels, national and local. While governance is a neutral concept, good governance addresses the allocation and management of resources to respond to collective needs. Though a universally agreed position on what constitutes good governance may be hard to come by, conceptually, it is characterised by the

principles of participation, transparency, accountability, rule of law, effectiveness, equity and strategic vision [4].

Good governance systems are participatory because they are based on democratic principles where the stakeholders in governance have a voice in the decision making process. Also the procedures and methods of decision making reflect transparency in order to ensure effective participation. The good governance system aims at "improving economic efficiency, protection and restoration of the environment and enhancing the wellbeing of the people" (adapted from International Institute of Sustainable Development, 1995). For [12], the idea of adopting ICTs in governance is to move government beyond passive information-giving to active citizen involvement in the decision making process and this is what good governance is all about. Therefore e-governance has the potentials to bring about good governance.

## The Concept of e-Governance in State Administration

The concept of e-governance is closely related to that of e-commerce. E-governance emerged in an attempt to translating the gains made in business through e-commerce into state government administration. From the definitions given above, the three main distinct groups interacting in e-governance concepts are government, citizens and businesses/interest groups. Abbreviations used in e-commerce such as B2B (business to business) and B2C (business to consumer) are used in e-governance to shortly describe which of the main groups are interacting. The most common group interactions in e-governance are G2C (Government to Citizen), G2B (Government to Business) and G2G (Government to Government). The interaction between the following groups in state governance is presented schematically in figure 1.

### Methodology

In this paper, we used questionnaires as the instrument to collect data from primary source. We distributed a total of 500 questionnaires. 479 (i.e. 95.80%) were returned with responses while 7 (1.4%) were either not returned or returned with invalid responses. Those were therefore discarded. The questionnaire was designed to assess the readiness of the stakeholders for the adoption of e-governance strategies for state administration in Ebonyi State. Table 1 shows the representation of the questionnaire administration. We addressed various aspects of e-governance in the questionnaire.

## RESULTS PRESENTATION

**Survey coverage:** The questionnaire was distributed to both indegenes and none indegenes resident in the three senatorial zones of the state at the time of the survey. The coverage of the state in the distribution of the questionnaire is shown in figure 2.
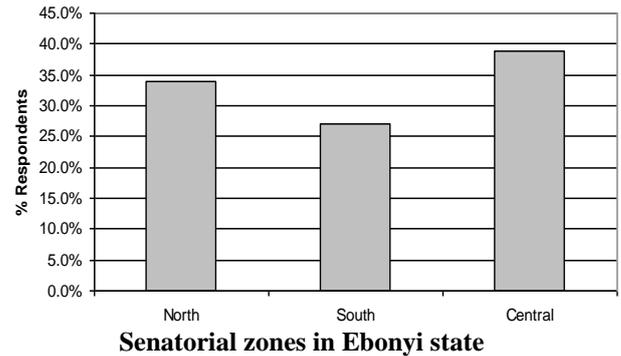


Fig. 2: Area covered in the questionnaires distribution

**Age Distribution of Respondents:** The age spread of the respondents that participated in the survey is shown in the pie chart of figure 3. An interesting observation was from the survey showed that greater number 68% (i.e. for ages between 20 – 55 as shown in pie chart) of the respondents earn some kind of salary, reside in other towns different from their home towns where there is computer business centres and as well have minimum qualifications of senior school certificate. Few respondents (15%) are below the age of 20 years and are either secondary school levers or students.
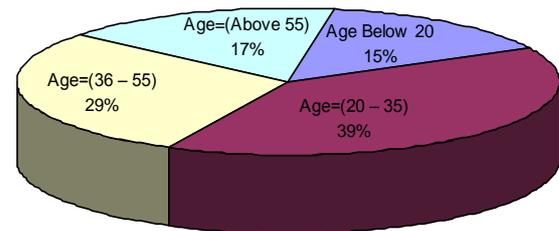


Fig. 3: Age Distribution of respondents

Possession of computer/IT skills is an important factor in e-governance implementation. When asked whether they know how to operate computer, figure 4 shows that 56.4% of the respondents said "yes", 32.2% said "just a little" while only 11.5% answered "no". Of the number that has the skill, they acquired the knowledge through private training personally funded by them or their guardians.
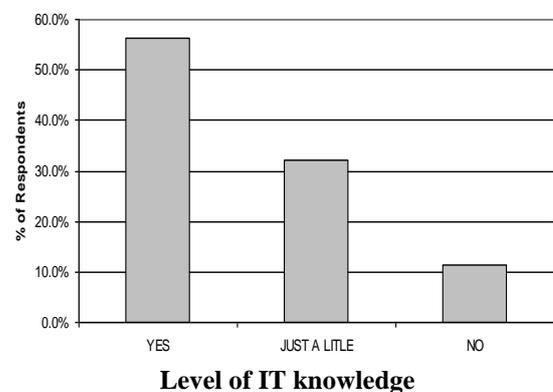


Fig. 4: Possession of computer/IT skills among respondents

Considering the frequency at which respondents use Internet services, figure 5 shows that greater number of the respondents 58.2% use Internet/Online services very often, 25.1% use it rarely (not often) while 16.7% do not use it at all.
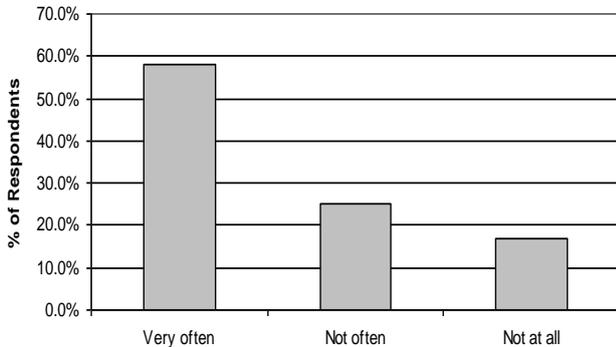


**Frequency of use of IT**
Fig. 5: Frequency of usage of internet/online services

The respondents that do not use it at all were mainly respondents from the villages that do not have computer business centers with internet services. Also their educational qualifications were below school certificate attempted.
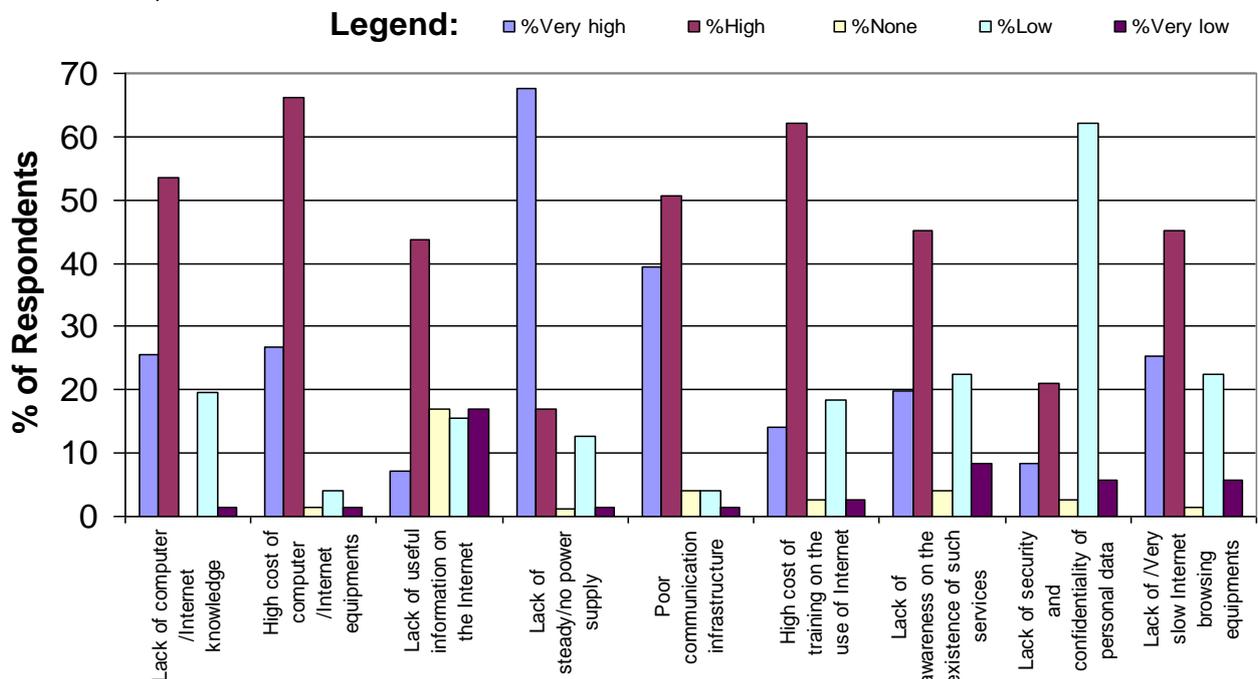


**Factors that hinder the use of e-governance services**
Fig. 6: Factors affecting the use of e-governance services in Ebonyi State.

Figure 6 is a chart of the opinions of respondents regarding the factors that hinder the use of e-governance services in Ebonyi State. These responses were studied and measured on a 5-point *likert-type* rating scale. The arithmetic mean of the responses were calculated to determine their levels and presented in table 1. The result showed that the most pronounced challenges to the implementation of e-governance in Ebonyi State were:

1. lack of steady or no power supply with a calculated mean of $\chi=4.37$,
2. poor communication infrastructure with a calculated mean of $\chi=4.23$,
3. high cost of computer/internet equipments with a calculated mean of $\chi=4.13$.

4. Lack of computer/internet knowledge with a calculated mean of $\chi=3.82$,
5. high cost of training on the use of internet with a calculated mean of $\chi=3.67$ ,
6. lack of /very slow internet browsing equipments with a calculated mean of $\chi=3.62$ and
7. lack of awareness on the existence of such services ($\chi=3.46$) were next strong hindrances.

The other factors considered namely: Lack of useful information on the internet with arithmetic mean of $\chi=3.09$ and lack of security and confidentiality of personal data ($\chi=2.64$) were low in the *likert-type* scale which shows that the respondents were of the opinion that the two factors were not strong hindrances to the use of e-governance services in the sampled communities.

Table 1: Analysis of responses on the factors that hinder the use of e-governance services in Ebonyi State

| Factors | | × | f | f× | χ | % |
|---|---|---|---|---|---|---|
| a | Lack of computer /Internet knowledge | | | | | |
| | Very high | 5 | 122 | 610 | 3.82 | 25 |
| | High | 4 | 256 | 1024 | | 53 |
| | None | 3 | 0 | 0 | | 0 |
| | Low | 2 | 94 | 188 | | 20 |
| | Very low | 1 | 7 | 7 | | 1 |
| b | High cost of computer /Internet equipments | | | | | |
| | Very high | 5 | 128 | 640 | 4.13 | 27 |
| | High | 4 | 317 | 1268 | | 66 |
| | None | 3 | 7 | 21 | | 1 |
| | Low | 2 | 20 | 40 | | 4 |
| | Very low | 1 | 7 | 7 | | 1 |
| c | Lack of useful information on the Internet | | | | | |
| | Very high | 5 | 34 | 170 | 3.09 | 7 |
| | High | 4 | 209 | 836 | | 44 |
| | None | 3 | 81 | 243 | | 17 |
| | Low | 2 | 74 | 148 | | 15 |
| | Very low | 1 | 81 | 81 | | 17 |
| d | Lack of steady/no power supply | | | | | |
| | Very high | 5 | 324 | 1620 | 4.37 | 68 |
| | High | 4 | 81 | 324 | | 17 |
| | None | 3 | 6 | 18 | | 1 |
| | Low | 2 | 61 | 122 | | 13 |
| | Very low | 1 | 7 | 7 | | 1 |
| e | Poor communication infrastructure | | | | | |
| | Very high | 5 | 189 | 945 | | 39 |
| | High | 4 | 243 | 972 | 4.23 | 51 |
| | None | 3 | 20 | 60 | | 4 |
| | Low | 2 | 20 | 40 | | 4 |
| | Very low | 1 | 7 | 7 | | 1 |
| f | High cost of training on the use of Internet | | | | | |
| | Very high | 5 | 68 | 340 | 3.67 | 14 |
| | High | 4 | 297 | 1188 | | 62 |
| | None | 3 | 13 | 39 | | 3 |
| | Low | 2 | 88 | 176 | | 18 |
| | Very low | 1 | 13 | 13 | | 3 |
| g | Lack of awareness on the existence of such services | | | | | |
| | Very high | 5 | 95 | 475 | 3.46 | 20 |
| | High | 4 | 216 | 864 | | 45 |
| | None | 3 | 20 | 60 | | 4 |
| | Low | 2 | 108 | 216 | | 23 |
| | Very low | 1 | 40 | 40 | | 8 |
| h | Lack of security and confidentiality of personal data | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Very high | | 5 | 40 | 200 | 2.64 | 8 |
| High | | 4 | 101 | 404 | | 21 |
| None | | 3 | 13 | 39 | | 3 |
| Low | | 2 | 298 | 596 | | 62 |
| Very low | | 1 | 27 | 27 | | 6 |
| i | Lack of /Very slow Internet browsing equipments | | | | | |
| Very high | | 5 | 121 | 605 | 3.62 | 25 |
| High | | 4 | 216 | 864 | | 45 |
| None | | 3 | 7 | 21 | | 1 |
| Low | | 2 | 108 | 216 | | 23 |
| Very low | | 1 | 27 | 27 | | 6 |

## DISCUSSION

The number of questionnaires distributed and returned with valid information were 479 i.e. 95.80% of the total number of questionnaires distributed. This number is a statistically acceptable size of the population for analysis. The distribution of the questionnaires was spread over the area considered for the survey in an acceptable proportion thus Ebonyi North (34%), Ebonyi South (27%) and Ebonyi Central (39%). We observed and would want to report that 100% of the respondents owned at least a GSM phone and also have a phone call centre in their communities no matter how rural their community was. Other interesting findings include that 98% of the respondents were of the opinion that government services can be delivered to government clients namely: citizens, organizations and other government departments through GSM phones which incidentally forms the most common means of information communication now among the respondents. 98.8% do not know whether Ebonyi State government has a functional website or not and 98.7% like to know about the activities of government functionaries.

## CONCLUSION

The survey has shown that the most difficult challenges facing the deployment of e-governance in Ebonyi State were: lack of steady power supply, poor communication infrastructure and high cost of computer and internet equipments. The survey has also revealed that the use of e-governance would bring improved efficiency in government operations without necessarily increasing the cost of state administration. E-governance will induce accountability and prudence in the management of public fund. There will be easy access to information and government services, greater access to the people's representatives, greater participation of the people in the governance processes and hence e-governance strengthens democratic principles and ideologies. These are the real features of good governance.

So, the implementation of e-governance in Ebonyi State-Nigeria has very high prospects. There are truly some challenges, but the challenges are sur-mountable. Therefore, e-governance is a sure way of bringing good governance to the people.

## REFERENCES

[1] Alampay E., Richard Heeks and Peter Paul A. Soliva (2003), Bridging the Information Divide: A Philippine Guidebook on ICTs for Development University of the Philippines, University of Manchester, Supported by AMIC, IDRC Canada and UNDP-APDIP *online at http://www.apdip.net/elibrary (retrieved on September 26, 2008).*

[2] Steven Clift (2002) *E-Democracy: The Promise of the Future is a Reality Today* online at http://www.publicus.net (retrieved on September 26, 2008).

[3] Michiel Backus (2001), E-Governance and Developing Countries *Introduction and examples* online @ *www.iicd.org/about/publications/egovernance-and-develoing-countries-introduction-andexamples/report3.pdf* accessed september 25, 2009).

[4] UNDP APDIP (2007) Identifying 'Killer' e-Governance Applications for Telecentres Lessons from the Philippines online at *http://www.apdip.net/apdipenote/21.pdf* (accessed on September 24, 2008).

[5] UNESCO (2003); World Public Sector Report 2003: E-Government at the Crossroads New York

[6] Organisation for Economic Co-operation and Development (OECD) (1995), participatory development and good governance online @ *www.oecd.org/dac/governance-development/31857685.pdf* retrieved September 24, 2012.

[7]     Institute On Governance (2003) Principles for Good Governance in the 21st Century Policy Brief No. 15 online @ *http://unpan1.un.org/intradoc/groups/public/documents/ UNPAN/UNPAN011842.pdf* accessed on July 18, 2012.

[8]     Commission on Global Governance (1995) Our Global Neighborhood report of the Commission on Global Governance ISBN 0-19-827998-1; Published by Oxford University Press.

[9]     Okot-Uma R. W. (2000), ELECTRONIC GOVERNANCE: Re-inventing Good Governance London: Commonwealth Secretariat

[10]    Holmes, D. (2001) eGov: eBusiness Strategies for Government. London, U.K.: Nicholas Brealey publishing.

[11]    Tapscott, D. & Caston, A. (1993). Paradigm Shift: The New Promise of Information Technology New York: McGraw-Hill

[12]    Misuraca Ginluca (2007), e-governance in Africa from theory to action: a handbook on ICTs for local governance, *IDRC-New Jersey*, Africa World Press

# PricePoints: A System for Aiding Consumers with Grocery Market Shopping

Hmood Al-Dossari, Maha Al-Yahya, Khaled Al-Quraini, Yahya Al-Qahtani, Sultan Al-Barrag

College of Computer and Information Sciences

King Saud University, Riyadh, Saudi Arabia

**Abstract**: In grocery markets, price variations are often observed for the same item offered by several groceries. This has led to many consumers paying high prices because of the search costs associated with obtaining price information. In this paper, we present PricePoints, a system to help consumers determine the best available offer and thus save their money. To do so, PricePoints compares several items in a number of grocery stores using multiple criteria in order to guide the consumers to select the best product based on their preferences. The system uses a novel algorithm that is based on multi-criteria including item price, its availability, grocery distance, and reputation. The paper presents the design, implementation, and evaluation of the PricePoints system. The obtained results indicate that the proposed system provides justified and reasonable recommendations for consumers in selecting their preferred groceries.

**Keywords**: E-Commerce; Recommender System, Decision Making.

## 1. INTRODUCTION

People need to buy goods and services constantly, and the search costs to find the best offer are high. Moreover, consumers vary in the way they interpret what constitutes a "best offer" for them. For some, it might be the lowest price; for others, it might be the availability of an item, store location, or store reputation. In a typical grocery item purchase scenario, the consumer usually makes decisions on two levels. The first level consists of determining the preferred item from among many alternatives, and the second level involves finding the best place to purchase it. Usually, this decision is based on price as well as other factors. The process of finding the best offer is highly costly, and services that help save time, effort, and money for the consumer are highly desirable. Moreover, the large number of grocery products available and the different brands have made the decision-making process more complex.

In this paper, we present PricePoints, a system that supports consumers who must decide which grocery items to purchase depending on their preferences (i.e., item price and availability as well as market location and reputation). The system uses a novel algorithm to compare the grocery items, taking into consideration a number of factors (i.e. criteria) that are preferred by that specific consumer.

The rest of the paper is organized as follows: Section 2 presents relevant work. Section 3 presents the research methodology. Sections 4, and 5 present the system design and components. Sections 6, 7, and 8 present system implementation, evaluation, and discussion, respectively. Section 9 presents the conclusions and directions for further work.

## 2. RELATED WORK

A considerable amount of literature has been published in the area of consumer assistance in e-commerce. One example is a Comparison Shopping Agent (CSA) or shopbot. CSAs are described as systems that "collect product and service information, especially price-related information, from multiple online vendors, aggregate them, and then provide value-added service to online shoppers to assist their online shopping" [1]. One of the earliest shopbots was BargainFinder [2] that enables users to compare music CD prices from several online stores. It mainly operated by comparing prices from several vendors without considering other product information or attributes. ShopBot [3] is an agent that can query vendor websites and extract relevant product information. Unlike BargainFinder, ShopBot initiates the query from the user's browser and not the agent's server. Early CSAs used search engines and screen scrapers to extract relevant product information. In addition, vendor sites were entered by humans as static links.

Comparison shopping sites are usually portals that rely on vendors to provide product information. There might be special relations or partnerships between the CSA and vendors to provide product information. Therefore, although they provide price comparisons, they do not necessarily provide unbiased recommendations (i.e., free of advertisements and commissions). Although the product information is stored in a structured machine-processable form in the vendor databases, shopbots can only access this information by retrieving and parsing web pages, using various methods to extract it. Semantic Web shopbots address the shortcomings of traditional shopbots that are based on screen-scraping. Traditional shopbots extract product information by parsing pages retrieved using screen scrapers [4]. Parsing limits the type, quality, and accuracy of the information being retrieved.

Semantic Web technologies can extract the semantics of web pages so that extracted product information is more accurate and meaningful. A proposed approach for Semantic Web shopbots was presented in [4]. The approach is based on Web services, Web service registries, and ontologies. The approach was envisioned to address current shopbot problems, mainly the "inability to provide comparisons on attributes other than price and unbiased services."

CSAs have usually focused on comparing individual item prices to find the best price for consumers. However, the comparison agent proposed in [5] considers the total value discounts that could be gained by buying more products from the same retailer. The authors view this problem as an optimization problem of choosing the retailers and items to buy in order to minimize the total buying cost, taking into account the total value discounts offered. They propose a formulation for this problem in a model that could be integrated with CSAs/search engines for multiple item purchases.

In [6], an architecture named ConsiderD was presented that maintains rules in comparison shopping. The rules are extracted from online retailers by humans, automatically expressed in eXtensible Rule Markup Language (XRML), and used to include not just the item price for comparison, but also other information such as delivery cost, delivery options, and shipping rates. The architecture is implemented for the online bookstore domain and was tested on five online bookstores. Results of the evaluation indicate that rule-based computation significantly outperforms the average data-retrieved delivery cost.

Another area relevant to our research is the quality and reputation of services in e-Commerce systems. A reputation-based approach is described in [7] that uses a reputation-based approach to facilitate the protection of consumers. To address the inherent subjectivity of consumer ratings, consumers were asked to provide their expectations alongside their ratings. This approach incorporates expectations into the reputation calculation in order to distinguish between subjective and unfair ratings. The authors observed ratings over time and then divided them into different windows to capture the dynamic behavior of service providers. They then computed the quality of a service for each window. The study presented in [8] evaluates the quality of open web services that are provided by Internet companies in the telecommunications sector to help users choose the best provider depending on the following criteria: execution price, execution time, and reputation. The researchers presented two methods based on those criteria and believe that successful methods depend on the willingness of all users to provide feedback that is correct and accurate about the level of services provided by the suppliers. The researchers recommend gathering more information to better calculate the quality of open web services and predict the best service.

In the area of web services, the study in [9] focuses on methods for computing quality of service (QoS) to enable consumers to make informed choices about the best service. It uses a comparative approach to verify the authenticity, accuracy, and extent of compliance of the published information with the consumers' evaluation of the QoS attributes. The researchers presented methods for service matching, ranking, and selection based on several factors to compute the quality of web services. The approach is also capable of predicting the value of unavailable attributes.

Similarly, the approach presented in [10] computes a personalized QoS ranking based on a prediction framework for cloud services that takes advantage of past consumer experiences. This QoS ranking prediction method does not require additional service invocations to rank. The authors presented two algorithms: cloudRank1 and cloudRank2 based on employed service, full service, preference function, and confidence value. These algorithms calculate the reference values for these factors and display the arrangement of cloud services to help users to choose the best services.

With regards to consumer behavior when buying online, [11] described factors that affect it. The study compared price and confidence to determine which is the most important for consumers. The researchers studied consumers who were interested in purchasing books by monitoring the opinions of buyers and those who attempted to buy items on book-selling websites. An important finding of this study is that its results might not generalized to other products because different products have different factors that affect it.

While sharing the same aim as the previous studies (i.e., supporting consumer decision-making), our proposed system attempts to address their limitations. It is clear that the success of the majority of the previous proposals mainly depends on the willingness of end-users to give their feedback on the quality of the services that they consume. Additionally, these approaches assume identical items (i.e., all required items are available in all potential markets) when comparing prices, which is unlikely to happen. In doing so, the recommendation and guidance of these approaches could be misleading and inaccurate. In our study, we aim to compare available offers based on our proposed algorithm rather than consumer evaluations. In proposing our solution, we address the problem of having a non-identical list of items, hence providing more useful information to support the consumer decision-making process.

## 3. METHODOLOGY

We elicited and identified requirements by studying the factors that might influence consumers' decisions when purchasing grocery items. An online survey was designed and distributed to over 770 participants (62% male, 38% female) to identify the most influential factors that affected their purchase behavior. The survey consisted of eleven multiple choice questions about their behavior when purchasing grocery items.

The results of the survey indicated that more than 70% of consumers would like an application that helps them purchase grocery items and provides detailed information in terms of price, availability, and market distance.

## 4. SYSTEM DESIGN

The system was designed as an online web-based system. In the following subsections, we describe the design considerations and decisions.

## 4.1 Ranking based on Multi-criteria

One of the major challenges that we faced in the system design was how to rank the markets based on multiple criteria, specifically, an item's price and availability as well as the market's distance and reputation.

Another important issue is the link between an item's availability and price. That is, we cannot separate one from the other because the price variation of items between different markets depends not only on the total price of the products, but also on the availability of these products in the market. Hence, in order to equitably compare prices, there must be an equivalent availability of the products in all markets. Moreover, a comparison that depends only on the total value of a market is incorrect because there could be a difference between the item's price in these markets (e.g., the price of a packet of rice could be Saudi Riyal (SR) 150 in one market and SR 230 in another) and if we calculate the price of rice with other item prices, we will not notice the large difference between its prices in many markets. To conclude, a traditional comparison of many items based on multiple criteria without considering the relations among them (e.g., the price-availability relation) may produce a misleading result. Hence, a new mechanism is required to provide a reasonable and realistic comparison.

For the purpose of comparing more than one factor at the same time and giving more accurate and realistic outcome, the design and implementation of our system is based on the proposed PricePoints algorithm as described in the next section.

## 4.2 PricePoints Algorithm

In this paper, we propose a PricePoints, an algorithm to compare the prices of items between different markets. The main idea of the PricePoints algorithm is to compare the items of the candidate markets individually rather than collectively. That is, the algorithm starts by taking the first item, compares its price for all markets, and then gives a point to the cheapest market. It then does the same for the next item and continues until it reaches the end of the shopping list.

After comparing all items, the algorithm collects the total points of each market in order to determine the winner (the market with the highest points). The algorithm then removes the market with the highest score from the list and repeats the same process with the remaining markets. Finally, the PricePoints arranges the candidate markets in descending order by score. Figure 1 lists the pseudo-code of the PricePoints algorithm.



**Figure 1. PricePoints Algorithm**

By comparing the shopping list item-by-item, we are able to not only consider the relation between availability and price, but also capture any variation in item prices.

## 5. SYSTEM COMPONENTS

Our proposed system comprises three major components: Data Collection, Product Selection and Ranking Calculation. We will describe the three components in details on the following sections.

## 5.1 Data Collection

We assume that there is direct access to market databases. These databases contain the price and number of existing products as well as all relevant information such as category, provider, expiration date, and specification. The Data Collection component extracts product information from several databases and stores them in one central database.

## 5.2 Product Selection

In our system, the consumer can choose a list of grocery products that they wish to purchase and then the system displays all relevant information. Product information includes the four major influential factors: price, availability, distance, and rating.

### Price

Price is the amount money that the consumer must pay the supplier for grocery products such as milk, corn, or rice. The price is the most important factor that influences the consumer when deciding where to buy the required products [12].

### Availability

Availability indicates whether a required product is available in a specific grocery or not.

### Distance

Distance is computed as the distance from the location of the consumer to the candidate grocery. There are many approaches that can be used to calculate the distance such as Global Positioning System (GPS) or the Google Maps API.

### Reputation

Reputation represents the total points earned by a grocery during the system's usage. At every invocation, the system gives one point to the grocery that gets the highest score when the prices of products in all groceries are compared (i.e., the cheapest grocery). As price is the most important factor in buying decisions, the system computes the grocery reputation based on price. This is a new and unique approach that differs from other reputation approaches that are based on user ratings [13], [14], [15]. These ratings, unfortunately, suffer from subjectivity as well as unfairness. Our proposed approach addresses these limitations by generating the ratings automatically without humans' involvement which may provide fake evaluations [16].

## 5.3 Ranking Calculation

Comparing many items based on multiple criteria is not straightforward. This is because different customers may have different preferences when selecting the best grocery. For example, some may be interested in the cheapest price, while others may be looking for the nearest grocery. The ranking calculation is the final process in our system. After all computations are made, the system displays the best possible offer for the consumer based on their preferences. The following example demonstrates this process.

To demonstrate the process of the PricePoints algorithm, suppose that we have four groceries where each has five products that a consumer SC1 looking for, as shown in Table 1.

**Table 1. Example of groceries with five products.**

|  | SP1 | SP2 | SP3 | SP4 |
|---|---|---|---|---|
| **Milk** | 7 | 8 | 9 | - |
| **Rice** | 1 | 1 | - | 1.5 |
| **Tissues** | 85 | 70 | 90 | - |
| **Detergent** | 15 | - | 15 | 14.25 |
| **Soda** | 5 | 6 | 5 | - |
| *Points* | *3* | *2* | *1* | *1* |
| **Total (SR)** | 113 | 85 | 119 | 15.75 |

It is clear that SP1 is the best grocery in terms of price because it has three product prices that are the lowest (milk, rice, and soda). Hence, SP1 is given three points. In addition, SP2, SP3, and SP4 are given two, one, and one points, respectively. Note that instead of using the absolute total points for each grocery, PricePoints calculates the cheapest as a percentage of the total points as follows:

*Grocery percentage = (Grocery Points/Total Points) × 100*     *(1)*

For example, the percentage of SP1 can be calculated as: SP1 percentage = $100 \times (3/5) = 60\%$. The system gives SP1 60% of the total points and the percentages of other groceries are

calculated with respect to the remaining percentage (40%). Table 2 shows the other groceries after SP1 was removed from the list.

**Table 2. Remaining candidate groceries.**

|  | SP2 | SP3 | SP4 |
|---|---|---|---|
| **Milk** | 8 | 9 | - |
| **Rice** | 1 | - | 1.5 |
| **Tissues** | 70 | 90 | - |
| **Detergent** | - | 15 | 14.25 |
| **Soda** | 6 | 5 | - |
| *Points* | *3* | *1* | *1* |
| **Total (SR)** | 84 | 119 | 15.75 |

Again, from Table 2, it is obvious that SP2 is the second best grocery, as it gets the highest number of points after comparing the prices of all three groceries. To calculate the percentage of SP2, we modified equation (1) as follows:

*Grocery percentage = (Grocery Points/Total Points) × 100 × remaining percentage*     *(2)*

Using equation (2), the percentage of SP2 is: $(3/5) \times 100 \times 0.40 = 24\%$. The PricePoints repeats the same process to determine the percentage of SP3 and SP4, and the results are shown in Figure 2.



**Figure 2. Ranking of the four groceries based on price**

After calculating the percentage of each candidate grocery, the PricePoints algorithm arranges the four candidate groceries in descending order and returns the best (cheapest) one to the customer.

### Availability

Ranking the candidate groceries based on availability is a straightforward process. The availability of items for each grocery equals the number of available items divided by the total number of items.

*Availability = Available Items/Total Items*     *(3)*

For example, the availability of items in SP4 is computed by dividing two (the number of available items: rice and detergent) by five (number of all items) to equal 40%. Table 3 shows the calculation of availability for all groceries. It is obvious that if SC1 concerns on having all items on a grocery, then the system will nominate SP1 as the best candidate.

**Table 3. Availability of the candidate groceries.**

|  | SP1 | SP2 | SP3 | SP4 |
|---|---|---|---|---|
| **Availability** | 100% | 80% | 80% | 40% |

### *Distance*

There are many applications that can calculate the distance such as GPS and the Google Maps API. Table 4 shows an example for the distance between the current customer and the four groceries. If the most important factor for SC1 is how far she is from the candidate groceries, the proposed system will nominate SP3 as the best choice.

**Table 4. Distance to the candidate groceries.**

|  | SP1 | SP2 | SP3 | SP4 |
|---|---|---|---|---|
| **Distance (KM)** | 1.7 | 2.4 | 0.3 | 4.0 |

### *Reputation*

Each time a grocery wins, the system increases its rating factor by 1. On the other words, the proposed system gives a point to each market that obtains the highest percentage of savings. For example, suppose that we have historical ratings for four markets, as listed in Table 5. It is clear that after 73 transactions, SP1 won 40 times, SP2 won 18 times, SP3 won only once, and SP4 won 14 times.

**Table 5. Reputation of the candidate groceries.**

|  | SP1 | SP2 | SP3 | SP4 |
|---|---|---|---|---|
| **Rating** | 40 | 18 | 1 | 14 |

In our running example, after consumer SC1 finishes their comparison, the system adds one point to the SP1 grocery, if it is the winner, and stores this information in the database. Hence, the rating of SP1 will become 41.

## 6. IMPLEMENTATION

The system was implemented as a web-based system and a detailed test was carried out to ensure that the system functionality is aligned with the proposed design and requirements. The following subsections discuss the ranking of markets based on user preferences: price, availability, distance, and rating.

## 6.1 Price

Assume that the prices of a sample grocery list in four markets are as shown in Table 6.

**Table 1. Consumer Requirements**

|  | SP1 | SP2 | SP3 | SP4 |
|---|---|---|---|---|
| **Milk** | - | 4 | 7 | 1 |
| **Carrots** | - | 7 | 15 | 20 |
| **Onion** | 9 | 9 | 9 | - |
| **Video** | - | 60 | 140 | 280 |
| **Yogurt** | 6 | 7 | 8 | 12 |
| **Total (SR)** | 15 | 87 | 179 | 313 |

We note that the total prices of the items are dissimilar. SP1 has the lowest cumulative total price (SR 15), however, there are three unavailable items (milk, carrots, and a video). This means that SP1 is not necessarily the best choice. For SP2, the cumulative total price (SR 87) is higher than SP1 but it has all the required items. SP4 is the market with the highest cumulative total, furthermore, one of the items (onion) is not available. In contrast SP2 and SP3 have lower cumulative totals and all the required items are available. The results shown in Figure 3 were based on the data in Table 6.
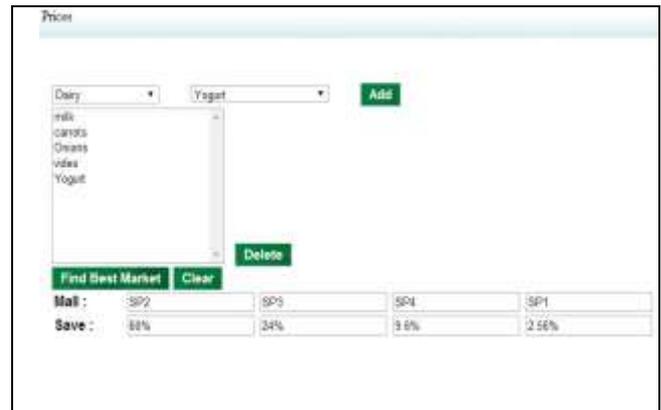


**Figure 3. Groceries' Ranking based on Price**

The results of the system show that SP2 has the highest percentage of savings (60%). Next highest are SP3 (24%), SP4 (9.6%), and SP1 (2.56%). These results are more realistic and can better guide the consumer to the most preferred grocery. In our walking example SP2 is the best choice with regards to the price factor.

## 6.2 Availability

Based on our survey, more than 55% of participants prefer item availability in the market where they are purchasing grocery items. Our implementation computes availability as the percentage of items that exist in a specific grocery with respect to the total items selected by the consumer. The results from our system are shown in Figure 4.
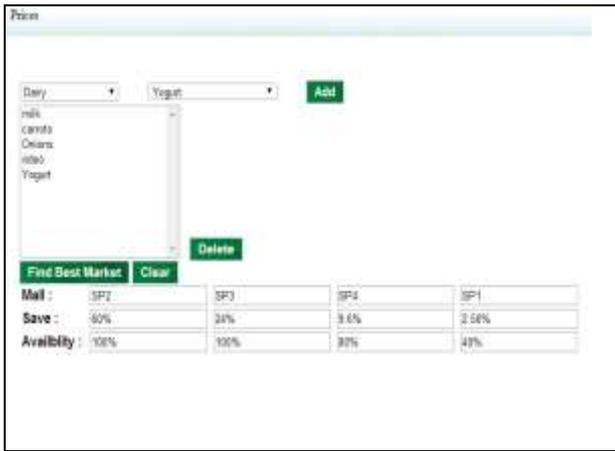
**Figure 4. Groceries' Ranking based on Availability**

As can be seen in this figure, SP2 and SP3 have 100% availability because they stock all the grocery items, while SP4 gets 80% because it does not have any onions and SP4 gets only 40% because it does not have milk, carrots, or a video.

## 6.3 Distance

The results of our survey showed that 19% of the participations are concerned by market distance. As we explained in Section 5.2 the system only shows the distance of each grocery. Hence, the consumer determines the nearest grocery. The results of the distance computation in our implementation are shown in Figure 5. As can be seen, SP4 is the nearest at 9 km, followed by SP2 (12 km), SP1 (15 km), and SP3 (16 km).
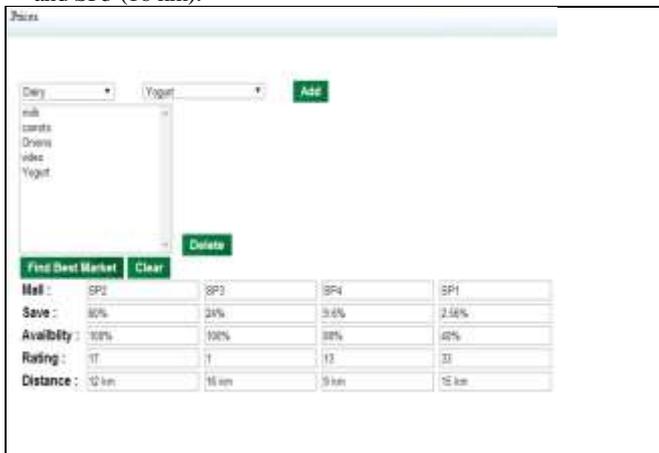


**Figure 5. Groceries' Ranking based on Distance**

## 6.4 Rating

As explained in Section 5.2, the system gives a point to each grocery that obtains the highest percentage of savings. The results of our survey showed that more than 47% of participants do not trust the opinions of others, therefore our approach computes these ratings automatically without human intervention.

To illustrate how the system produces the ratings automatically, assume that the current situation for four groceries is shown in Figure 6. It is obvious that SP1 is the best grocery, as it has 35 points. Assume also that we have two consumers SC1 and SC2 each has a different shopping list. The system guides them to the best grocery based on their requirements and consequently adds one point to the best grocery's rating. We set a fixed availability of 100% for all markets and intentionally assign SP1 the cheapest prices for most of the items.
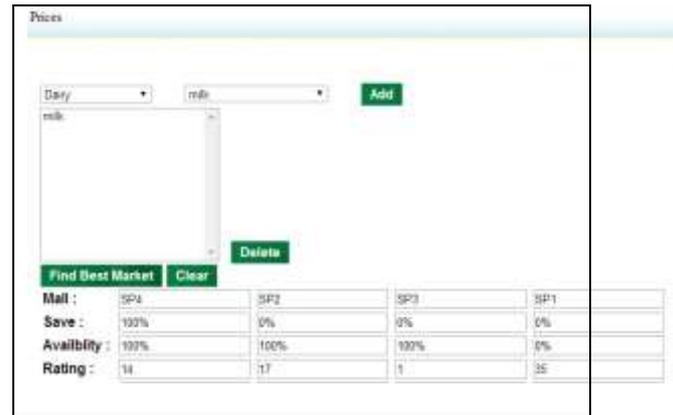


**Figure 6. The current rating for the four groceries**

Now assume that SC1 has the following shopping list: yogurt, cheese, and soup. When the consumer submits his/her request to the system, the ratings are shown as in Figure 7. It is clear that SP1 obtains 100% for the savings, indicating that it is the cheapest price for all selected items. Consequently, the system presents it to SC1 as the best choice and increases its rating by one point to 36, as shown in Figure 7.
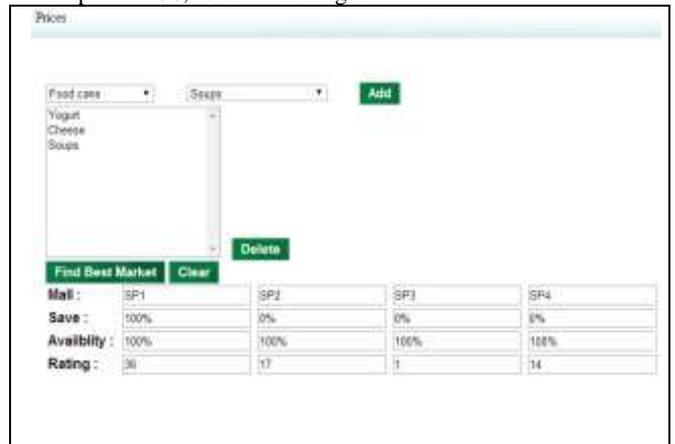


**Figure 7. Groceries' rating update after SC1's transaction**

The third consumer SC2 submits a request that contains: beans, canned fruit, a TV, and a microwave. The system compares the prices and chooses SP1 as the best grocery in terms of price, as shown in Figure 8. Again, the saving rate for SP1 is the highest (75%) among the four markets, so it is the winner. The system again increments SP1's rating by one point to 37.
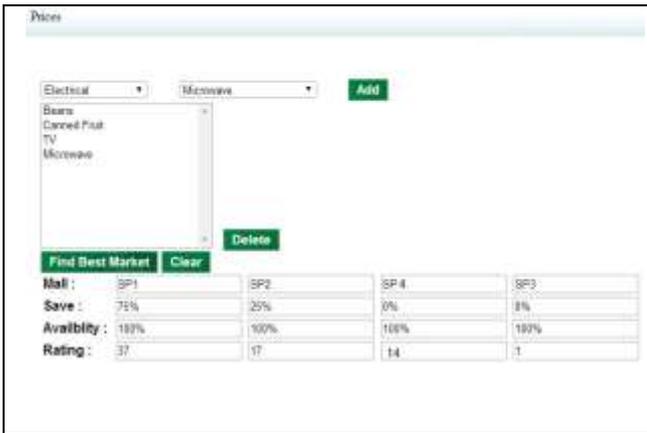
**Figure 8. Groceries' rating update after SC2's transaction**

| 13 | The system addresses the deficiencies and poor performance of regulatory organizations for consumer protection. |
| 14 | I would recommend the system to friends. |
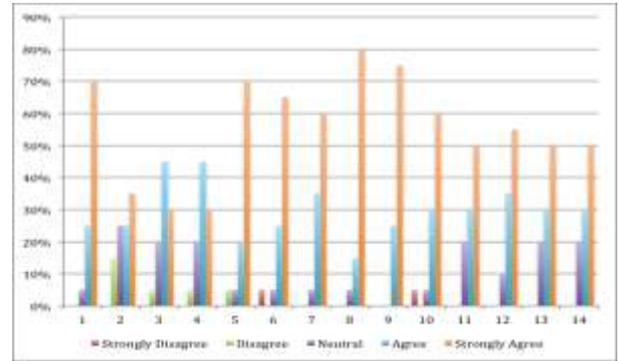


**Figure 9. Respondents' perceptions on PricePoints**

## 7. Evaluation and Results

After system development and testing, we evaluated the system to obtain feedback from the consumers on its effectiveness in helping them find the best offer available in a grocery. We designed a survey with 14 statements to elicit participants' views and feedback with respect to the system. A total of 20 users participated in the survey. The statements are shown in Table 7, and the results are shown in the chart in Figure 9.

**Table 2. Perception Statements in the survey**

| Item | Statement |
|---|---|
| 1 | The system idea is important for the grocery consumer. |
| 2 | The system is easy to use. |
| 3 | The system covered the most important factors for grocery consumers. |
| 4 | The system raises the level of awareness among consumers. |
| 5 | The system uses traditional methods in the search for the cheapest market. |
| 6 | The system contributes to saving money. |
| 7 | This system search is better than traditional methods for finding the availability of food. |
| 8 | The system contributes to saving time and effort. |
| 9 | The system helps identify the nearest market. |
| 10 | The ratings method used in the system provides a more credible approach than ratings based on consumers' opinions and reviews. |
| 11 | Rating affects the decision to choose a market. |
| 12 | The system helps the consumer get a fair deal from the market owner. |

## 8. DISCUSSION

The results obtained from our evaluation indicate that the system as designed and implemented provides justified and reasonable recommendations for consumers when purchasing grocery items. The results of the survey show that the majority of respondents agreed that there is a need for such a system to help them with their purchases. When asked whether the system was able to consider all relevant factors for consumers when making purchases, the replies varied, but the majority agreed that major factors were considered. When asked about the major issues that concern consumers, that is money savings, time, and effort, and whether the system was able to help them find the best offer, the majority of the respondents agreed.

With regards to the issue of the cheapest market and the availability of items, the majority of respondents agreed that the system excelled over traditional methods. With regards to guidance to the closest available market, survey results show that the majorities indicate that system was able to provide this service. When asked to compare the rating model provided in the system with other traditional rating models based on consumer opinions and reviews, the majority of respondents indicated that this model is better suited to the problem at hand, is more credible, and does have an effect on the decision to choose a specific market.

With regards to the interface design and ease of system use, answers varied. In general, the system was not very complicated, but improvements were needed to the interface to simplify its use and access for non-technical users. When asked if the system provides a solution to address the deficiencies of the regulatory organizations for consumer protection, the majority of respondents agreed. In general, results of the evaluation were positive, as the majority of the respondents (80%) indicated that they would recommend such an application to friends.

## 9. CONCLUSION AND FUTURE WORK

In this paper, we presented a system that helps consumers select the best grocery market based on their preferences (i.e. price, availability, and distance).

The system design is based on incorporating several relevant factors for consumers when making decisions, and combines them in a novel algorithm that we call PricePoints. The system was designed, implemented, and tested with real consumers. An evaluation was carried out to obtain consumer views on the system. Results of the evaluation indicate that the system was useful and provided relevant information that helped them make the best choices for grocery items purchases.

During the design and implementation of the system, we were faced with several challenges. First, it is difficult to get the actual price of grocery items from groceries. A new method of real-time data acquisition and extraction should be included in future designs of our system. Second, we were faced with the challenge of multiplicity, as companies can produce the same product with different specifications and prices. Finally, a major challenge we were faced with is that not all groceries have their data available online, and it was not easy to persuade them to supply us with data about their products and prices.

Results of our evaluation as well as the challenges we faced provide directions we can explore for future. In the current system implementation, we dealt with a small number of items and a small number of factors without information about the quantity required by each item. The PricePoints algorithm can be extended to consider quantity. Another vital improvement to obtain real time data would be to employ "crowdsourcing" or "collaborative participation" methods for data collection. In this approach, the data is not obtained from the vendors, instead ordinary people who actually visit the grocery participate using their mobile devices to add product and cost information to the central database and get, in response, immediate comparison information.

Although our implementation is focused on grocery item shopping, the design of the system is generic and can easily be adapted to other domains such as books, clothing, and hardware.

## REFERENCES

[1]   Y. Wan, Comparison-Shopping Agents and Online Small Business, Khosrow-Pour, 2006.

[2]   B. Krulwich, The BargainFinder agent: Comparison price shopping on the Internet, Bots and Other Internet Beasties, SAMS NET, 1996.

[3]   R.B. Doorenbos, O. Etzioni and D. S. Weld, A scalable comparison-shopping agent for the World-Wide Web, Proc. 1st International Conference on Autonomous Agents, 1997, pp.39–48.

[4]   M. Fasli, The next generation of shopbots: Semantic interoperability and personalization, Comp.-Shopp. Serv. Agent Des. Idea Publ. Group, 2009, pp.19–36.

[5]   S. Kameshwaran and L. Benyoucef, Optimal buying from online retailers offering total value discounts, Proc. 10th International Conference on Electronic Commerce, 2008, p.21.

[6]   G.G. Lim, J.Y. Kang, J.K. Lee and D.C. Lee, Architecture and performance of the rule based comparison shopping: Delivery cost experience, Proc. 10th International Conference on Electronic Commerce, 2008, pp.32.

[7]   H. Al-Dossari, A reputation-based approach for consumer protection in Saudi Arabia, Proc. 4th International Conference on Computational Aspects of Social Networks (CASoN), 2012, pp.321–326.

[8]   Y. Liu, A.H. Ngu and L.Z. Zeng, QoS Computation and Policing in Dynamic Web Service Selection, Proc. 13th International World Wide Web Conference, New York, 2004, pp.66–73.

[9]   Z. Xu, P. Martin, W. Powley and F. Zulkernine, Reputation-Enhanced QoS-based Web Services Discovery, Proc. IEEE International Conference on Web Services, 2007, pp.249–256.

[10]  Z. Zheng, X. Wu, Y. Zhang, M.R. Lyu and J. Wang, QoS Ranking Prediction for Cloud Services, IEEE Trans. Parallel Distrib. Syst., Vol.24, No.6, 2013, pp.1213–1222.

[11]  H.-W. Kim, Y. Xu and S. Gupta, Which is more important in Internet shopping, perceived price or trust?, Electron. Commer. Res. Appl., Vol.11, No.3, 2012, pp.241–252.

[12]  S.A.G. Katherine Isselmann Disantis, What 'Price' Means When Buying Food: Insights From a Multisite Qualitative Study With Black Americans., Am. J. Public Health, 2013.

[13]  T.D. Huynh, N.R. Jennings and N.R. Shadbolt, An integrated trust and reputation model for open multi-agent systems, Auton. Agents Multi-Agent Syst., Vol.13, No.2, 2006, pp.119–154.

[14]  S. Reece, A. Rogers, S. Roberts and N.R. Jennings, A Multi-Dimensional Trust Model for Heterogeneous Contract Observations, Proc. 22nd AAAI Conference on Artificial Intelligence, 2007, pp.128–135.

[15]  G. Zacharia, Trust management through reputation mechanisms, Appl. Artif. Intell., Vol.14, 2000, pp.881–907.

[16]  A. Jøsang, R. Ismail and C. Boyd, A Survey of Trust and Reputation Systems for Online Service Provision, Decis. Support. Syst., Vol.43, No.2, 2007, pp.618–644.

# Pairwise Ordered $\zeta$ – Extremely Disconnected Spaces

Sharmila.S
Department of Mathematics
Nirmala College for Women
Coimbatore, India

I.Arockiarani
Department of Mathematics
Nirmala College for Women
Coimbatore, India

**Abstract**: In this paper, a new class of intuitionistic fuzzy topological spaces called pairwise ordered intuitionistic fuzzy $\zeta$ – extremally disconnected space is introduced. We also apply these notions of $\zeta$ – extremally disconnectedness to discuss Tietze's extension theorem and several other properties.

**Keywords:** intuitionistic fuzzy $\zeta$ – extremally disconnected space, intuitionistic fuzzy $\zeta$ – space, intuitionistic fuzzy real line, intuitionistic fuzzy unit interval, intuitionistic fuzzy continuous function.

## 1. INTRODUCTION

After the introduction of the concept of fuzzy sets by Zadeh [12], several researches were conducted on the generalisations of the notion of fuzzy set. The concept of "Intuitionistic fuzzy sets" was first published by Atanassov and many works by the same author and his colleagues[1,2 ] appeared in the literature. An introduction to intuitionistic fuzzy topological space was introduced by Dogan Coker[5]. In this paper a new class of intuitionistic fuzzy topological spaces namely, pairwise ordered intuitionistic fuzzy $\zeta$ – extremally disconnected spaces is introduced by using the concepts of ordered fuzzy topology and fuzzy bitopology.

## 2. PRELIMINARIES

**Definition 2.1**.[1]. An intuitionistic fuzzy set (IFS, in short) A in X is an object having the form $A = \left\{ x, \mu_A(x), \upsilon_A(x) / x \in X \right\}$ where the functions $\mu_A : X \to I$ and $\upsilon_A : X \to I$ denote the degree of membership (namely $\mu_A(x)$) and the degree of non-membership (namely $\upsilon_A(x)$) of each element $x \in X$ to the set A on a nonempty set X and $0 \le \mu_A(x) + \upsilon_A(x) \le 1$ for each $x \in X$. Obviously every fuzzy set A on a nonempty set X is an IFS's A and B be in the form $A = \left\{ x, \mu_A(x), 1 - \mu_A(x) / x \in X \right\}$

**Definition 2.2.[1].** Let X be a nonempty set and the IFS's A and B be in the form $A = \left\{ x, \mu_A(x), \upsilon_A(x) / x \in X \right\}$, $B = \left\{ x, \mu_B(x), \upsilon_B(x) / x \in X \right\}$ and let $A = \{ A_j : j \in J \}$ be an arbitrary family of IFS's in X. Then we define

(i) $A \subseteq B$ if and only if $\mu_A(x) \le \mu_B(x)$ and $\upsilon_A(x) \ge \upsilon_B(x)$ for all $x \in X$.

(ii) A=B if and only if $A \subseteq B$ and $B \subseteq A$.

(iii) $\overline{A} = \left\{ x, \upsilon_A(x), \mu_A(x) / x \in X \right\}$.

(iv) $A \cap B = \left\{ x, \mu_A(x) \cap \mu_B(x), \upsilon_A(x) \cup \upsilon_B(x) / x \in X \right\}$.

(v) $A \cup B = \left\{ x, \mu_A(x) \cup \mu_B(x), \upsilon_A(x) \cap \upsilon_B(x) / x \in X \right\}$

(vi) $1_{\sim} = \{ \langle x, 1, 0 \rangle x \in X \}$ and

$0_{\sim} = \{ \langle x, 0, 1 \rangle x \in X \}$.

**Definition 2.3.[5].** An intuitionistic fuzzy topology (IFT, in short) on a nonempty set X is a family $\tau$ of an intuitionistic fuzzy set (IFS, in short) in X satisfying the following axioms:

(i) $0_{\sim}, 1_{\sim} \in \tau$.

(ii) $A_1 \cap A_2 \in \tau$ for any $A_1$, $A_2 \in \tau$.

(iii) $\bigcup A_j \in \tau$ for any $A_j : j \in J \subseteq \tau$.

In this paper we denote intuitionistic fuzzy topological space (IFTS, in short) by $(X, \tau), (Y, \kappa)$ or X,Y. Each IFS which belongs to $\tau$ is called *an intuitionistic fuzzy open set* (*IFOS*, in short) in X. The complement $\overline{A}$ of an IFOS A in X is called an *intuitionistic fuzzy closed set*(*IFCS*, in short). An IFS X is called *intuitionistic fuzzy clopen* (*IF clopen*) iff it is both intuitionistic fuzzy open and intuitionistic fuzzy closed.

**Definition 2.4.[5].** Let $(X, \tau)$ be an IFTS and $A = \left\{ x, \mu_A(x), \upsilon_A(x) \right\}$ be an IFS in X. Then the fuzzy interior and closure of A are denoted by

(i) $cl(A) = \bigcap$ {K: K is an IFCS in X and $A \subseteq K$ }.

(ii) $\text{int}(A) = \bigcup$ {G: G is an IFOS in X and $G \subseteq A$ }.

Note that, for any IFS A in $(X, \tau)$, we have $cl(\overline{A}) = \overline{\text{int}(A)}$ and $\text{int}(\overline{A}) = \overline{cl(A)}$.

**Definition 2.10.[10].** Let A be an IFTS $(X, \tau)$. Then A is called an *intuitionistic fuzzy $\zeta$ open* set( $IF\zeta OS$, in short) in X if $A \subseteq bcl(\text{int}(A))$.

**Definition 2.11.[10].** Let A be an IFTS $(X, \tau)$. Then A is called an *intuitionistic fuzzy $\zeta$ closed* set ( $IF\zeta CS$, in short) in X if $b\,\text{int}(cl(A)) \subseteq A$.

**Definition 2.12.[10].** Let $f : X \rightarrow Y$ from an IFTS X into an IFTS Y. Then $f$ is said to be an Intuitionistic fuzzy $\zeta$ continuous (IF$\zeta$ cont, in short)[11] if $f^{-1}(B) \in IF\zeta OS(X)$ for every $B \in \kappa$.

**Definition 2.12.[8].** An ordered set on which there is given a fuzzy topology is called an ordered fuzzy topological space.

**Definition 2.12.[7]**. A fuzzy bitopological space is a triple $(X, \tau_1, \tau_2 \leq)$ where X is a set and $\tau_1, \tau_2$ are any two fuzzy topologies on X.

# 3. PAIRWISE ORDERED INTUIIONISTIC FUZZY $\zeta$ – EXTREMALLY DISCONNECTED SPACES

**Definition 3.1**. Let $(X, \tau)$ be an IFTS. Let A be any intuitionistic fuzzy $\zeta$ open set (in short, $IF\zeta OS$ ) in $(X, \tau)$. If $IF\zeta cl(A)$ is $IF\zeta$ open, then $(X, \tau)$ is said to be intuitionistic fuzzy $\zeta$ − extremally disconnected (in short, $IF\zeta$ − extremally disconnected).

**Proposition 3.2.** Let $(X, \zeta)$ is an intuitionistic fuzzy $\zeta$ space. Then the following statements are equivalent.

(i) $(X, \tau)$ is an intuitionistic fuzzy $\zeta$ extremally disconnected space.

(ii) For each $IF\zeta CS$ set A, we have $IF\zeta \text{int}(A)$ is intuitionistic fuzzy $\zeta$ closed.

(iii) For each $IF\zeta OS$ set A , we have $IF\zeta cl(IF\zeta \text{int}(\overline{A})) = \overline{IF\zeta cl(A)}$.

(iv) For each pair of $IF\zeta OS$ A and B in $(X, \tau)$, we have
$$\overline{IF\zeta cl(A)} = B, IF\zeta cl(B) = I\overline{F\zeta cl(A)}$$

**Proposition 3.3.** Let $(X, \tau)$ be an IFTS. Then $(X, \tau)$ is intuitionistic fuzzy $\zeta$ extremally disconnected space if and only if for any $IF\zeta OS$ A and $IF\zeta CS$ B such that $A \subseteq B$, $IF\zeta cl(A) \subseteq IF\zeta \text{int}(B)$.

**Notation 3.4.** An IFS which is both $IF\zeta OS$ and $IF\zeta CS$ is called intuitionistic fuzzy $\zeta$ clopen set.

**Remark 3.5.** Let $(X, \tau)$ is intuitionistic fuzzy $\zeta$ basically disconnected space. Let $\{A_i, \overline{B_i} / i \in N\}$ be collection such that $A_i$'s are $IF\zeta OS$ and $B_i$'s are $IF\zeta CS$ sets. If $A_i \subseteq A \subseteq B_j$ and $A_i \subseteq B \subseteq B_j$ for all $i, j \in N$, then there exists an $IF\zeta COGF$ set C such that $IF\zeta cl(A_i) \subseteq C \subseteq IF\zeta \text{int}(B_j)$ for all $i, j \in N$.

**Theorem 3.6.** Let $(X, \tau)$ is intuitionistic fuzzy $\zeta$ basically disconnected space. Let $\{A_q\}_{q \in Q}$ and $\{B_q\}_{q \in Q}$ be monotone increasing collections of an $IF\zeta OS$ sets and $IF\zeta CS$ of $(X, \tau)$. Suppose that $A_{q1} \subseteq B_{q2}$ whenever $q_1 < q_2$ (Q is the set of all rational numbers). Then there exists a monotone increasing collection $\{C_q\}_{q \in Q}$ of an $IF\zeta COGF$ sets of $(X, \tau)$ such that $IF\zeta cl(A_{q1}) \subseteq C_{q2}$ and $C_{q1} \subseteq IF\zeta \text{int}(B_{q2})$ whenever $q_1 < q_2$.

**Notation 3.7**. $I^0(A)$ denotes increasing intuitionistic fuzzy interior of A , $I(A)$ denotes increasing intuitionistic fuzzy closure of A .

**Definition 3.8.** Let $(X, \tau, \leq)$ be an ordered IFTS and let A be any IFS in $(X, \tau, \leq)$, A is called increasing $IF\zeta$ open if $A \subseteq I (I^0(A))$. The complement of an increasing $IF\zeta OS$ is called decreasing $IF\zeta$ closed.

**Definition 3.9.** Let $(X, \tau)$ be an IFTS. For any IFS A in $(X, \tau, \leq)$,

$I^{IF\zeta}(A)$ = increasing intuitionistic fuzzy $\zeta$ closure of A

$= \bigcap \{B / B$ is an increasing intuitionistic fuzzy $\zeta$ closed set and $B \supseteq A \}$,

$D^{IF\zeta}(A)$ = decreasing intuitionistic fuzzy $\zeta$ closure of A

$= \bigcap \{B / B$ is an decreasing intuitionistic fuzzy $\zeta$ closed set and $B \supseteq A \}$,

$I^{0IF\zeta}(A)$ = increasing intuitionistic fuzzy $\zeta$ interior of A

$= \bigcup \{B / B$ is an increasing intuitionistic fuzzy $\zeta$ open set and $B \subseteq A \}$,

$D^{0IF\zeta}(A)$ = decreasing intuitionistic fuzzy $\zeta$ interior of A

$= \bigcup \{B / B$ is an increasing intuitionistic fuzzy $\zeta$ open set and $B \subseteq A \}$,

Clearly, $I^{IF\zeta}(A)$ (resp. $D^{IF\zeta}(A)$ ) is the smallest increasing (resp. decreasing) intuitionistic fuzzy $\zeta$ closed set containing A and $I^{0IF\zeta}(A)$ (resp. $D^{0IF\zeta}(A)$ ) is the largest increasing (resp. decreasing) intuitionistic fuzzy $\zeta$ open set contained in A.

**Proposition 3.10.** For any IFS A of an ordered intuitionistic fuzzy topological space $(X, \tau, \leq)$, the following statements hold :

   (i)   $\overline{I^{IF\zeta}(A)} = D^{0IF\zeta}(\overline{A})$

   (ii)   $\overline{D^{IF\zeta}(A)} = I^{0IF\zeta}(\overline{A})$

   (iii)   $\overline{I^{0IF\zeta}(A)} = D^{IF\zeta}(\overline{A})$

   (iv)   $\overline{D^{0IF\zeta}(A)} = I^{IF\zeta}(\overline{A})$

**Definition 3.11.** Let $(X, \tau_1, \tau_2 \leq)$ be an ordered IF bitopological space. Let A be any $\tau_1 -$increasing (resp. decreasing) $IF\zeta OS$ in $(X, \tau_1, \tau_2 \leq)$. If $I_{\tau_2}^{IF\zeta}(A)$ (resp. $D_{\tau_2}^{IF\zeta}(A)$ is $\tau_2 -$increasing (resp. decreasing) $IF\zeta OS$ in c, then $(X, \tau_1, \tau_2 \leq)$ is said to be $\tau_1 -$upper ($\tau_1 -$lower) $IF\zeta -$extremally disconnected. Similarly we can define $\tau_2 -$upper ($\tau_2 -$lower) $IF\zeta -$extremally disconnected. An IFTS $(X, \tau_1, \tau_2 \leq)$ is said to be *pairwise upper $IF\zeta -$extremally disconnected* if it is both $\tau_1 -$upper $IF\zeta -$extremally disconnected and $\tau_2 -$upper $IF\zeta -$extremally disconnected. Similarly we can define *pairwise lower $IF\zeta -$extremally disconnected.* An IF bitopological space $(X, \tau_1, \tau_2 \leq)$ is said to be *pairwise ordered $IF\zeta -$extremally disconnected* if it is both pairwise upper $IF\zeta -$extremally disconnected and pairwise lower $IF\zeta -$extremally disconnected.

**Proposition 3.12.** For an ordered IF bitopological space $(X, \tau_1, \tau_2 \leq)$ the following statements are equivalent:

(i)      $(X, \tau_1, \tau_2 \leq)$ is pairwise upper $IF\zeta -$*extremally disconnected*

(ii)     For each $\tau_1 -$decreasing $IF\zeta CS$ A, $D_{\tau_2}^{0IF\zeta}(A)$ is $\tau_2 -$decreasing $IF\zeta CS$ A, $D_{\tau_1}^{0IF\zeta}(A)$ is $\tau_1 -$decreasing $IF\zeta$ closed.

(iii)    For each $\tau_1 -$increasing $IF\zeta OS$ A, $D_{\tau_2}^{IF\zeta}\overline{(I_{\tau_2}^{IF\zeta}(A))} = \overline{(I_{\tau_2}^{IF\zeta}(A)}$. Similarly, for each $D_{\tau_1}^{IF\zeta}\overline{(I_{\tau_1}^{IF\zeta}(A))} = \overline{(I_{\tau_1}^{IF\zeta}(A)}$.

(iv)    For each pair of a $\tau_1 -$increasing $\tau_2 -$increasing $IF\zeta OS$ A and $\tau_1 -$decreasing $IF\zeta OS$ B in $(X, \tau_1, \tau_2 \leq)$ with $\overline{I_{\tau_2}^{IF\zeta}(A)} = B$, $D_{\tau_2}^{IF\zeta}(B) = \overline{I_{\tau_2}^{IF\zeta}(A)}$. Similarly, for each pair of a $\tau_2 -$increasing $IF\zeta OS$ A and

$\tau_2 -$decreasing $IF\zeta OS$ B in $(X, \tau_1, \tau_2 \leq)$ with $\overline{I_{\tau_1}^{IF\zeta}(A)} = B$, $D_{\tau_1}^{IF\zeta}(B) = \overline{I_{\tau_1}^{IF\zeta}(A)}$.

**Proposition 3.13.** Let $(X, \tau_1, \tau_2 \leq)$ be an ordered IF bitopological space. Then $(X, \tau_1, \tau_2 \leq)$ is pairwise ordered $IF\zeta -$*extremally disconnected* space if and only if for a $\tau_1 -$decreasing $IF\zeta OS$ A and $\tau_2 -$decreasing $IF\zeta CS$ B such that $A \subseteq B$, we have $D_{\tau_1}^{IF\zeta}(A) \subseteq D_{\tau_1}^{0IF\zeta}(B)$.

**Notation 3.14.** An ordered IFS which is both decreasing (resp. increasing) $IF\zeta OS$ and $IF\zeta CS$ is called a decreasing (resp. increasing) $IF\zeta COS$.

**Remark 3.15.** Let $(X, \tau_1, \tau_2 \leq)$ be a pairwise upper $IF\zeta -$*extremally disconnected* space. Let $\{A_i, \overline{B_i} / i \in N\}$ be a collection such that $A_i$'s are $\tau_1 -$decreasing $IF\zeta OS$, $B_i$'s are $\tau_2 -$decreasing $IF\zeta CS$ and let $A, \overline{B}$ be $\tau_1 -$decreasing $IF\zeta OS$ and T2-increasing $IF\zeta OS$ respectively. If $A_i \subseteq A \subseteq B_j$ and $A_i \subseteq B \subseteq B_j$ for all $i, j \in N$, then there exists a $\tau_1$ and $\tau_2 -$decreasing $IF\zeta COS$ C such that $D_{\tau_1}^{IF\zeta}(A_i) \subseteq C \subseteq D_{\tau_1}^{0IF\zeta}(B_j)$ for all $i, j \in N$.

**Proposition 3.16.** Let $(X, \tau_1, \tau_2 \leq)$ be a pairwise ordered $IF\zeta -$*extremally disconnected* space. Let $(A_q)_{q \in Q}$ and $(B_q)_{q \in Q}$ be the monotone increasing collections of $\tau_1 -$decreasing $IF\zeta OS$ and $\tau_2 -$decreasing $IF\zeta CS$ of $(X, \tau_1, \tau_2 \leq)$ respectively and suppose that $A_{q_1} \subseteq B_{q_2}$ whenever $q_1 < q_2$ (Q is the set of rational numbers). Then there exists a monotone increasing collection $(C_q)_{q \in Q}$ of $\tau_1$ and $\tau_2 -$decreasing $IF\zeta COS$ of $(X, \tau_1, \tau_2 \leq)$ such that $D_{\tau_1}^{IF\zeta}(A_{q_1}) \subseteq C_{q_2}$ and $C_{q_1} \subseteq D_{\tau_1}^{0IF\zeta}(B_{q_2})$ whenever $q_1 < q_2$.

**Definition 3.17.** An intuitionistic fuzzy real line is the set of all monotone decreasing IFS $A \in \zeta_R$ satisfying $\bigcup\{A(t) : t \in \Re\} = 1_{\sim}$ and $\bigcap\{A(t) : t \in \Re\} = 0_{\sim}$ after the identification of an IFSs $A, B \in \Re(I)$ if and only if A(t-)=B(t-) and A(t+)=B(t+) for all $t \in \Re$ where $A(t-) = \bigcap\{A(s) : s < t\}$ and $A(t+) = \bigcup\{A(s) : s > t\}$.

The intuitionistic fuzzy unit interval $I(I)$ is a subset of $\Re(I)$ such that $[A] \in I(I)$ if the membership and

nonmembership of an IFS line $\Re(I)$ $A \in \zeta_R$ are defined by

$$\mu_A(t) = \begin{cases} 1, t < 0 \\ 0, t > 1 \end{cases} \quad \text{and} \quad \upsilon_A(t) = \begin{cases} 0, t < 0 \\ 1, t > 1 \end{cases}$$

respectively.

The natural intuitionistic fuzzy topology on $\Re(I)$ is generated from the subbasis $\{L_t, R_t : s < t\}$ where

$\text{Lt}, \text{Rt} : \Re(I) \rightarrow I(I)$ are given by $L_t[A] = \overline{A(t-)}$ and $R_t[A] = A(t+)$ respectively.

**Definition 3.18.** Let $(X, \tau_1, \tau_2 \le)$ be an ordered IF bitopological space. A function $f : X \rightarrow \Re(I)$ is said to be lower (resp. upper) intuitionistic fuzzy $\zeta$ continuous function if $f^{-1}(\Re_t)(f^{-1}(L_t))$ is an $IF\zeta OS$ set, for each

**Notation 3.19.** Let X be any nonempty set and $A \in \zeta^X$. Then for $x \in X$, $<\mu_A(x), \upsilon_A(x)>$ is denoted by A˜.

**Proposition 3.20.** Let $(X, \tau_1, \tau_2 \le)$ be an ordered IF bitopological space, $A \in \zeta^X$ be an $\tau_1 -$IFS, and let $f : X \rightarrow \Re(I)$ be such that

$$f(x)(t) = \begin{cases} 1˜, t < 0 \\ A˜, 0 \le t \le 1 \\ 0˜, t > 1 \end{cases} \text{ for all } x \in X \text{ and } t \in \Re.$$

Then $f$ is $\tau_1 -$ lower(resp.upper) intuitionistic fuzzy $\zeta$ continuous function if and only if A is an $\tau_1 -$increasing or $\tau_1 -$decreasing $IF\zeta OS$.

**Definition 3.21.** Let $(X, \tau_1, \tau_2 \le)$ be an ordered IF bitopological space. The characteristic function of IFS A in X is the function $\chi_A : X \rightarrow I(I)$ defined by

$\chi_A(x) = A˜, x \in X$.

**Proposition 3.22.** Let $(X, \tau_1, \tau_2 \le)$ be an IFTS and let $A \in \zeta^X$ be an $\tau_1 -$IFS. Then $\chi_A$ is $\tau_1 -$lower (resp. $\tau_1 -$upper) continuous if and only if A is an $\tau_1 -$increasing or $\tau_1 -$decreasing $IF\zeta OS$.

**Proof.** The proof follows from Proposition 4.1.

**Definition 3.23.** Let $(X, \tau_1, \tau_2 \le)$ and $(Y, \kappa_1, \kappa_2 \le)$ be ordered IF bitopological spaces. A function $f : (X, \tau_1, \tau_2) \rightarrow (Y, \kappa_1, \kappa_2)$ is called $\tau_1 -$increasing (resp. $\tau_1 -$decreasing) intuitionistic fuzzy strongly $\zeta$ continuous (in short, $\tau_1 -$increasing(resp. $\tau_1 -$decreasing)IFstrongly strongly $\zeta$ continuous ) if $f^{-1}(A)$ is $\tau_1 -$increasing (resp. $\tau_1 -$decreasing)

IF $\zeta$ clopen in $(X, \tau_1, \tau_2 \le)$ for every S₁ and S₂ $IF\zeta OS$ in $(Y, \kappa_1, \kappa_2 \le)$. If $f$ is both $\tau_1 -$increasing and $\tau_1 -$decreasing IFstrongly $\zeta$ continuous, then it is called ordered $\tau_1 -$IF strongly $\zeta$ continuous.

**Proposition 3.24.** Let $(X, \tau_1, \tau_2 \le)$ be an ordered IF bitopological space. Then the following statements are equivalent.

(i) $(X, \tau_1, \tau_2 \le)$ is pairwise ordered
$IF\zeta - extremally\ disconnected$.

(ii) If $g, h : X \rightarrow R(I)$, $g$ is $\tau_1 -$lower
IF $\zeta$ continuous function, $h$ is $\tau_2 -$upper
IF $\zeta$ continuous function and $g \subseteq h$, then there
exists an $\tau_1$ and $\tau_2$ increasing       IF strongly
$\zeta$ continuous function, $f : (X, \tau_1, \tau_2, \le) \rightarrow R(I)$
such that $g \subseteq f \subseteq h$.

(iii) If $\overline{A}$ is $\tau_2 -$ increasing $IF\zeta OS$ and B is
$\tau_1 -$decreasing $IF\zeta OS$ such that $B \subseteq A$,
then there exists an $\tau_1$ and $\tau_2 -$ increasing IF
strongly $\zeta$ continuous function
$f : (X, \tau_1, \tau_2, \le) \rightarrow I(I)$ such that
$B \subseteq f^{-1}(\overline{L_1}) \subseteq f^{-1}(R_0) \subseteq A$.

# 4. TIETZE EXTENSION THEOREM FOR PAIRWISE ORDERED INTUITIONISTIC FUZZY $\zeta -$EXTREMALLY DISCONNECTED SPACE

**Notation 4.1.** Let $(X, \tau_1, \tau_2 \le)$ be an ordered IF bitopological space and $A \subset X$. Then an IFS $\psi_A^*$ is of the form $<x, \psi_A(x), 1 - \psi_A(x)>$.

**Proposition 4.2.** Let $(X, \tau_1, \tau_2 \le)$ be a pairwise ordered intuitionistic fuzzy $\zeta$ extremally disconnected space and let $A \subset X$ such that $\psi_A^*$ is $\tau_1$ and $\tau_2 -$increasing $IF\zeta OS$ in $(X, \tau_1, \tau_2 \le)$. Let $f : (A, \tau_1 / A, \tau_2 / A) \rightarrow I(I)$ be $\tau_1$ and $\tau_2 -$increasing IF strongly $\zeta$ continuous function. Then $f$ has an $\tau_1$ and $\tau_2 -$ increasing IF strongly $\zeta$ continuous extension over $(X, \zeta)$.

# 5. ACKNOWLEDGMENTS

# 6. REFERENCES
[1] K. Atanassov, Intuitionistic fuzzy sets, Fuzzy Sets and Systems.Vol 20, No.1, 1986, pp.87-96.

[2]     K. Atanassov and S.Stoeva, Intuitionistic fuzzy sets,
        in: Polish symp. On internal and Fuzzy
        Mathematics, Poznam 1983, pp.23-26.

[3]     G. Balasubramanian, Fuzzy disconnectedness and
        its stronger forms, Indian J.Pure Appl. Math. Vol.
        24,1993,pp. 27-30.

[4]     C.L. Chang, Fuzzy topological spaces, J. Math.
        Anal. Appl. Vol.24, 1968, pp.182-190.

[5]     D. Coker, An introduction to Intuitionistic fuzzy
        topological spaces, Fuzzy Sets and Systems. Vol.
        88,No.1, 1997, pp.81-89.

[6]     D. Coker and M. Demirci, On intuitionistic fuzzy
        points, NIFS. Vol.1, No.2, (1995, pp.79-84.

[7]     A. Kandil,Biproximities &fuzzy bitopological
        spaces, Simen Stevin. Vol.63, 1989, pp.45-66.

[8]     Katsaras A.K, Ordered fuzzy topological spaces,
        J.Math.Anal.Appl. Vol.84,1981,pp. 44-58.

[9]     T. Kubiak, L.fuzzy normal spaces and Tietze
        extension theorem, J. Math. Annal. Appl. Vol.25,
        1987, pp.141-153.

[10]    Sharmila.S and I. Arockiarani., On intuitionistic
        Fuzzy $\zeta -$ open sets, Jour.Global Research in
        Mathematical Archieve. Vol.2, No.10, 2015.

[11]    G. Thangaraj and G. Balasubramanian, On fuzzy
        basically disconnected spaces, J. Fuzzy Math. Vol.
        9, 2001,pp.103-110.

[12]    L.A. Zadeh, Fuzzy Sets, Information and Control.
        Vol. 9, 1965,pp. 338-353.

# An Empirical Analysis of Security on Nigerian's Internet Banking Platform: an end user's Perspective

Simon Enoch Yusuf
Federal University, Kashere
Gombe State
Nigeria

**Abstract**: The speedy advancement of Internet business has stirred-up the banking and financial sectors towards encouraging customers to embark on banking on the internet. On the other hand, confidentiality, security, and privacy of online banking activities and users basic information are the main concerns for both the online banking customers and to the banking industry. In this study, we investigated security on Nigerian's internet banking platform of 10 selected financial institutions in Nigeria. The research finding uses a modified  security checklist to analyse the security features and reliability of the Nigerian's selected banking industry. The results shows sufficient strength in all the basic security features available on the various internet banking platform, however, few of the platform are not sufficiently fault tolerant built.

**Keywords**: Electronic banking; Phishing; Online; Privacy, Reliability; Authentication

## 1.  INTRODUCTION

The ubiquitous nature of the internet has caused Nigerian banks to conduct daily banking activities from anywhere easily and conveniently over the internet. The development in the information technology sector and also the development in the banking industry has drastically increased communication and transactions between banks and customers [1].

The speedy advancement of Internet business has stirred-up the banking and financial sectors towards encouraging customers to embark on banking on the internet. This latest banking environment; Internet banking is increasingly managed as an operational activity and an important component of a multi-channel strategy [2].  Internet banking is defined as the use of the Internet to deliver banking activities such as funds transfer, paying bills, viewing current and savings account balance, paying mortgages and purchasing financial instruments and certificates of deposits [3] [4].

Technology based banking is now bringing more choices than ever before in the financial sector. It is now becoming possible to access our bank account in multiple ways and take care of our financial affairs quickly and efficiently [5].

Confidentiality, privacy and security of internet banking transactions and personal information are the major concerns for both the banking industry and internet banking customers [6] [7] [8][9]. Security threats such as keyloggers, phishing, pharming,  adware, malware,  Trojans, viruses and spyware are currently the most common online banking security threats and risks [6] [7].

Majority of published research on internet banking majorly focused on security on the network side and on also on the server's side  by creating a safe and secured communication channels for the user computer and the bank server's side computer [10] and most of the attacks on Internet Banking used today are based on fraudulent internet users stealing users login data.

In Nigeria, identity fraud is the most common type of attack on the internet banking platform [11].  This identity fraud is called phishing and pharming. Others include; Cross-site scripting and Keylogger/Trojan horses. In this type of attack, a software is use to manipulate legitimate user to transact with fake banking platform [11].

One of the most recent types of internet attack is the man-in-the-browser (MITB) attack [12], in this attack, malwares lives inside customers internet browsers, it then changes what customers see on their browser, with this, the attackers now has the ability to steal and modify the destination account number, amount and  any information that the customer entered and display whatever they choose.

Generally, Phishing, Adware, pharming, keylogger, malware, spyware, Trojans and viruses are the most frequent online banking security threats and risks [13] available on the internet banking platform.

## 2.  RELATED WORK

As the electronic banking market matures, Internet Banking as well as Mobile Banking has shifted from being merely a distribution channel towards becoming a central part of the bank's growth strategy. Banks have increasingly started to view and leverage their Internet Banking and/or Mobile Banking platform as their new virtual branch, through which (almost) the same services can be delivered to the client as via a physical branch [14].

Hamid   et al. [15] Compared and analysed the internet banking system in Malaysia and Thailand. In their analysis, they used comparison as the theoretical base and secondary data to determine the differences between both countries with respect to Internet banking. Their analysis indicates that both nations are dissimilar in providing basic services offered by their commercial banks.

Subsorn and Limwiriyakul  [16] Examined Internet banking security systems in Australian banks by creating a proposed Internet banking security. The following were uncovered by their research work; lack of Internet banking security in all the 16 selected Australian bank, Better Internet banking security

information, two-factor authentication and stronger encryption in use by those banks reviewed.

In another paper, Subsorn and Limwiriyakul [7] examined 12 (twelve) Thai commercial banks, they revealed that there was a distinct lack of internet banking security information provided on all the selected Thai banks' websites as compared to the selected Australian banks which provided better internet banking security information.

Mueni and Muchiri [17] used multi case study design to define a framework that can be used for assessing Internet banking system reliability. They collected data from a key informant using a questionnaire and document analysis guide and analysed using both descriptive and inferential statistics. In their study, they found out that documentation and size internal properties had significant positive effect on reliability.

An Analysis of Internet Banking Security of Foreign Subsidiary Banks in Australia was also conducted by [8]. They examined Internet banking security in nine (9) foreign subsidiary banks in Australia, they compared their results and findings with a previous research [16] [7] in order to produce a more practical and comprehensive guideline, as well as to include a weight rating of security related website information for the banking industry in Australia. The findings of both the previous research paper [16] [7] [18] revealed that there was lack of related Internet banking security information in all the selected Australian owned and foreign owned banks' websites which have the potential to impact on the confidentiality of the banks and its customers as well as future potential customers.

## 3. METHODOLOGY

This study applied a qualitative research method by employing analytical methods. The Analysis was conducted by investigating the availability and reliability of internet banking in Nigeria. The basic features deployed by each bank were also considered.

Currently, there are 1662 financial institutions in Nigeria, which comprises of 726 Bureau De Change, 21 Commercial banks, 1 non interest bank, 5 Development finance, 3 Discount House, 64 Finance company, 2 Merchant Banks, 792 Microfinance Bank, 1 Microfinance institution, 40 Primary mortgage institution, 7 unclassified financial institutions [19].

The research was carried out in Nigeria – North East. It involved a sample size of seven (7) commercial banks and three (3) Microfinance banks. Three (3) branch from each selected bank was randomly selected and ten (10) IT experts from those banks who were interviewed to determine their perspective of the level of availability and reliability provided.

Sources of data included both primary and secondary data which assisted the researcher to make a thorough analysis of the study problem at hand.

Primary data was collected through personal interviews and use of questionnaires to gather accurate information. While, Secondary data was obtained from available sources such as text books, journals, on-line published articles, information from the local newspapers and internet search engines among others.

In addition, a security checklist by [20] was adopted and modified for the purposes of evaluating the security features of the selected banks. A list of the checklist is presented below:

**Table 1. Modified Subsorn & Limwiriyakul Security Checklist.**

| | Category | Descriptions |
|---|---|---|
| 1 | General online security, privacy and reliability | Provide internet banking security information<br>Privacy in transaction<br>Bank security mechanism system |
| 2 | Fault Tolerant | Operate despite failure<br>Roll back in case of system failure<br>Use of redundant system<br>Reliability in terms of accepting wrong input<br>Internet banking system is recoverable from errors and failure<br>Meeting required software standards |
| 3 | Security features on the internet banking platform | Required standard for inactivity system timeout<br>Not meeting the required standard for inactivity system time out.<br>Inactivity system timeout not in use<br>Limited daily online transfer amount to third party<br>Limit can be increase by approval of the bank<br>No limit for internet transfer.<br>Login information alert<br>i. last login<br>ii. activity log<br>iii. SMS or email alert login<br>not available<br>Use session management<br>Do not use session management |
| 4 | Authentication Technology | User Site<br>Password login restriction<br>Enforce good password practice (combining letters, numbers, and special characters)<br>Enforce login pin length<br>Uses two factor authentication for transaction e.g. addition of Token, SMS etc.<br>Two factor authentication not in use<br>Logon requirement: using user ID, email address, password, CAPTCHA etc<br>3 times logon failure limitations<br>Logon failure limitation not use<br>On screen logon user input<br>Keypad logon user input<br>Bank site<br>Use of encryption and digital technology |
| 5 | IT Assistance, Support and Monitoring | Provide 24/7 customer care line<br>24/7 customer care line not available<br>Uses secured email<br>Provide frequent ask question support form<br>Monitor internet banking transaction |
| 6 | System Requirement | Internet banking platform compatible with popular web browsers.<br>Security software tools available to customers |

## 3.1  Analysis

**Table 2.  Availability of internet banking features**

| S/N | Security feature category | Commercial Banks | | | | | | | Microfinance Banks | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | First Bank Plc | Zenith Bank | GT Bank | UBA | Diamond Bank | EcoBank | Sky Bank | Gombe Microfinance Bank Limited | Jewel Coop Microfinance Bank Limited | Adamawa Savings & Loans Limited |
| 1 | General online security, privacy and reliability | | | | | | | | NA | NA | NA |
| | a. Provide internet banking security information | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | b. Privacy in transaction | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | c. Bank security mechanism system | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| 2 | Fault Tolerant | | | | | | | | | | |
| | a. Operate despite failure | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | b. Roll back in case of system failure | | | | | | | | | | |
| | c. Use of redundant system | | | | | | | | | | |
| | d. Reliability in terms of accepting wrong input | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | e. Internet banking system is recoverable from errors and failure | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| 3 | Security features on the internet banking platform | | | | | | | | NA | NA | NA |
| | a. Required standard for inactivity system timeout | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | b. Not meeting the required standard for inactivity system time out. | | | | | | | | ✓ | ✓ | ✓ |
| | c. Inactivity system timeout not in use | | | | | | | | ✓ | ✓ | ✓ |
| | d. Limited daily online transfer amount to third party | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | e. Limit can be increase by approval of the bank | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | f. No limit for internet transfer. | | | | | | | | | | |
| | g. Login information alert | | | | | | | | | | |
| | i. last login | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | ii. activity log | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | iii. SMS or email alert login | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | iv. not available | | | | | | | | ✓ | ✓ | ✓ |
| | h. Use session management | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | i. Do not use session management | | | | | | | | ✓ | ✓ | ✓ |
| 4 | Authentication Technology | | | | | | | | NA | NA | NA |
| | User Site | | | | | | | | | | |
| | a. Password login restriction | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | b. Enforce good password practice (combining letters, numbers, and special characters) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | c. Enforce login pin length | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | d. Uses two factor authentication for transaction e.g. addition of Token, SMS etc. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | e. Two factor authentication not in use | | | | | | | | | | |
| | f. Logon requirement: using user ID, email address, password, CAPTCHA etc | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | g. 3 times logon failure limitations | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | h. Logon failure limitation not use | | | | | | | | | | |
| | i. On screen logon user input | | | ✓ | ✓ | | | ✓ | | | |
| | Keypad logon user input | ✓ | ✓ | | | ✓ | ✓ | | | | |
| | Bank site | | | | | | | | | | |
| | a. Use of encryption and digital technology | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| 5 | IT Assistance, Support and Monitoring | | | | | | | | | | |
| | a. Provide 24/7 customer care line | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | b. 24/7 customer care line not available | | | | | | | | | | |
| | c. Uses secured email | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | d. Provide frequent ask question support form | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | e. Monitor internet banking transaction | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| 6 | System Requirement | | | | | | | | | | |
| | a. Internet banking platform compatible with popular web browsers. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | NA | NA | NA |
| | b. Security software tools available to customers | | | | | | | | | | |
| | c. Meeting required software | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |

| | standards | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

Note:        NA – Not Available        ✔ -Yes

## 3.2. General online security, privacy and reliability

All the selected commercial banks provide basic internet security tips through email, SMS and their bank web pages. In addition, other security mechanism that provides extra layer of security on the internet platform was communicated to the end-user. However, all selected Microfinance banks do not provide such services and therefore, not available.

## 3.3. Fault Tolerant

Seven (7) selected commercial institution were examined for fault tolerant capability in terms of failure and reliability. None of the banks provide roll back capability in terms of failure. However, they all have supporting or redundant system in case of system failure and meets the requirement in terms of continuity despite inputting wrong input.

## 3.4. Security features on the internet banking platform

All selected commercial banks provide inactivity system timeout capability which happen atmost 3 minutes of inactivity. As part of their security features, a limit transfer amount of N1,000,000.00 to third part account in another bank per transaction is enforced on all platform. All banks equally provide last login information, activity login over a period of time, SMS or email alert login information and also session time out.

## 3.5. Authentication Technology

The following authentication technology is enforce by all the internet banking platform, password login restriction, two factor authentication,  3 times logon failure limitation, use of encryption and digital technology for transaction. However, only 3 of the bank uses on screen logon user password input.

## 3.6. IT Assistance, Support and Monitoring

Customers care and monitoring banking service is been provided by Nigeria commercial banks to her customers all through the week 24/7. They also monitor all internet transactions on their platform. In addition, FAQ is been used online to support and help bank customers in getting answers to some of their questions.

## 3.7. System Requirements

All the banks declared that they do not provide security software for their customer uses. They also declared that their internet platform is compatible to all popular browsers, however, one (1) stated that it bank only allow Firefox for it internet banking, they disabled other browsers for security purposes.

## 4. Conclusions

Implementing appropriate sets of control, polices, processes, procedures, legal framework, software and hardware functions,  and organizational structures will assist in achieving greater level of security on internet banking platform. These holistic approach need to be monitored, reviewed, established, implemented, and improved, where necessary, to ensure that the specific security and business objectives of the bank are met.

All selected commercial banks are currently actively using internet banking for transaction with sufficient level of security on their various platforms. However, for additional security on the use of keyboard for user input on the internet platform, the use of on screen logon should be adopted. This is because, hackers use a tool to capture keyboard key strokes and then later use it for illegitimate transaction.

Unfortunately, all the existing selected Microfinance bank do not use internet banking from the user end and therefore analyzing those banks was not possible. Finally, financial institutions are encourage to always meet up with all the novel security challenges in an online banking transaction by continuous monitoring of latest threat on the banking platform and also by increasing the various security features available.

## 5. ACKNOWLEDGMENTS

## 6. REFERENCES

[1] Giannakoudi. S. 1999 "Internet banking: the digital voyage of banking and money in cyberspace" Information and Communication Technology Law, Vol. 8 No. 3, pp. 205-43.

[2] Black, N.J.,  Lockett, A., Winklhofer, H and McKechnie, S. 2002 "Modelling consumer choice of distribution channels: An illustration from financial services. The International Journal of Bank Marketing, 20(4): 161-173.

[3] Singhal, D. and Padhmanabhan, V. 2008 "A Study on Customer Perception Towards internet Banking: Identifying major contributing factors. The Journal of Nepalese Business Studies: V (1), 101 – 111.

[4] Haque, A., Ahmad, H. I., and Daraz, A. H. 2009. "Issues of E – Banking transaction: An Empirical Investigation on Malaysian Customers perception" Journal of Applied Sciences, 9(10), 1870 – 1879.

[5] Aychiluhim, D. and Tibebe, B. 2014 "Internet Banking Security Framework: The case of Ethiopian Banking Industry" HiLCoE Journal of Computer Science and Technology, Vol. 2, No. 2. Pp. 7 – 13.

[6] Subsorn, P., and  Limwiriyakul, S. 2011 "A comparative analysis of the security of Internet banking in Australia: A customer perspective"  Presented in 2nd International Cyber Resilience Conference (ICR2011). Perth, Australia: Edith Cowan University, 2011a.

[7] Subsorn, P. and Limwiriyakul, S. 2011. "A comparative analysis of internet banking security in Thailand: A customer perspective" Presenting in 3rd International Social Science, Engineering and Energy Conference 2011b (ISEEC2011). Nakhon Pathom, Thailand.

[8] Hutchinson, D. and Warren, M. 2003 "Security for Internet banking: A framework" Logistics Information Management, 16(1), 64 -73.

[9] Hutchinson,D. and Warren, M. 2001 "A framework of security authentication for internet banking" Paper presented at the International We-B Conference (2nd), Perth.

[10] Peotta,L. and Holtz, M. D. 2011 "A formal Classification of Internet Banking Attacks and Vulnerabilities",

International Journal of Computer Science and Information Technology (IJCSIT), Vol. 3, No 1, Feb .

[11] Yusuf, S. E. , Adebayo, K. J. and Adetula, E. O. 2013 "Mitigating Cyber Identity Fraud using Advanced Multi Anti-Phishing Technique" International Journal of Advanced Computer Science and Applications (IJACSA) Vol. 4, No. 3. pp 156 – 164. Available online athttp://www.thesai.org/Downloads/Volume4No3/Paper_25-Mitigating_Cyber_Identity_Fraud_using_Advanced_Multi_Anti-Phishing_Technique.pdf

[12] Cain, C. "Analyzing Man-in-the-Browser (MITB) Attacks" SANS Institute InfoSec Reading Room, 2014. Available online at https://www.sans.org/reading-room/whitepapers/forensics/analyzing-man-in-the-browser-mitb-attacks-35687

[13] BankMuscat. "Internet banking security threats" Retrieved October, 2015, from http://www.bankmuscat.com/enus/ConsumerBanking/bankingchannels/internetbanking/Pages/InternetBankingSecurityThreats.aspx

[14] Dijkstra, M. and Esajas, M. O. "Security for Internet Banking and in Mobile Banking is the key to Customer Loyalty", retrieved July, 2015, Available online at ww.ibis-management.com.

[15] Hamid, M. R. A., Amin, H., Lada, S., Ahmad, N. 2007 "A Comparative Analysis of Internet Banking in Malaysia and Thailand. Journal of Internet business" Issue 4

[16] Subsorn, P. and Limwiriyakul, S. A. 2012 "Comparative analysis of the security of Internet banking in Australia: A customer perspective". Procedia Engineering 32. Elsevier 260 – 272. www.sciencedirect.com

[17] Mueni, M. F. and Muchiri, M. G. 2014. "An assessment framework for Internet banking system reliability" International Journal of Technology in Computer Science & Engineering, Volume 1(3) , September 2014 , pp 88-100.

[18] Subsorn, P.and Limwiriyakul, S. 2012 "An Analysis of Internet Banking Security of Foreign Subsidiary Banks in Australia: A Customer Perspective" ( IJCSI) International Journal of Computer Science Issues, Vol. 9, Issue 2, No 2.

[19] Central Bank of Nigeria, "Financial Institution" Retrieved August, 2015, from http://www.cenbank.org/Supervision/Inst-DM.asp

[20] Stawowski, M. "Client side Vulnerability Assessment", retrieved from http://www.clico.pl/services/Clientside_Vulnerability_Assessment.pdf, Last accessed on 10/09/2015.

# Effects of Object-Oriented Programming on Modern Software Development

Onu Fergus U.
Ebonyi State
University, Abakaliki
Nigeria

Okorie Kingsley
10 Jideofor Street
Thinkers Corner
Enugu – Nigeria

Ugwa Chioma U
43/45 Unity Street off
Nike Road Trans-
Ekulu,
Enugu – Nigeria

**Abstract:** Object oriented Programming (OOP) entered the world of programming in the 1980s, but it actually became very popular from 1990s till date. What changes has it actually introduced into the programming world when placed in a balance with other programming techniques? This paper answers this question by taking a detailed and analytical look into the history and evolution of programming Languages. The paper unraveled the comparative advantages of OOP over a few earlier programming techniques. We collected data from both primary and secondary sources, analyzed the data and found out that stakeholders in software development industry have felt the tremendous impact of OOP on modern software development process. We have also noted that despite all the good and desirable features offered by Object Oriented programming, it is obvious that stakeholders in software development still expect easier and more flexible features than those the Object Oriented Programming currently presents.

## 1. INTRODUCTION

Herbsleb and Moitra (2001), declared that software has become a vital component of almost every business in recent times. Success in all businesses increasingly depends on using software as a competitive weapon. This fact has made software development a construction of strategy for business success. Software development is a serious engineering concept that needs team work. The team members may be collocated or dispersed depending on where the needed expertise is found. The team-based approach to software development has made it a global concern. This is what Herbsleb and Moitra termed Global Software Development (GSD). In GSD distance is a major issue leading to coordination, communication, and management issues. To survive the problems posed by the globalisation of software development, collaborative tools among other technologies will play a key role concludes Herbsleb and Moitra (2001). Object Oriented Programming presents this collaborative feature and hence poses the tendency to provide great success in modern software development.

## 2 REVIEW OF RELATED LITERATURE

## 2.1Brief History of Computer Programming

Computer programming has undergone both tumultuous and turbulent moments which subjected programmers to rough learning and application terrain. It used to be so bad that one programmer may not and (most times) cannot comprehend what another programmer presented with their codes. To make the situation worse, the programmer who authored a set of codes would end up forgetting completely what the codes were to do few years later. The consequence of was a software or application which is very difficult (if not impossible) to maintain. Going down the memory lane, let us recall that programming the computer started with the novel works of the following: John Von Neumann, Charles Babbage, etc. The Assembly language, other procedural-based programming of 1950s, Structured programming techniques of late 1960s, Modular programming concepts 1970s and Object-Oriented programming of late 1980s have taken their turns to influence the computer programming and software development landscape.

**Assembly Language:** Assembly language as stated by Wikipedia, is a programming Language that can be used to directly tell the computer what to do. An assembly language is almost exactly like the machine language that a computer can understand, except that it uses words called *mnemonics* in place of numbers. A computer cannot really understand an assembly program directly. However, an assembler can easily change the program into machine code, replacing the mnemonic with the binary patterns they stand for. It is a programming language that is a close approximation of the binary machine code (Mifflin, 2009).

**Procedure-Based programming:** Many of today's programmers may not be used to some words such as routines, subroutines in programming as they are with functions and objects. These are what procedural programming paradigm, derived from structured programming which is built on the concept of the procedure calls are based on. In procedural programming, program codes are organized into small "procedures" that use and change our data. In ColdFusion for example, we write our procedures as either custom tags or functions. These functions typically take some input, carry out a process, and then produce some output. Ideally your functions would behave as "black boxes" where input data goes in and output data comes out. The key idea here is that our functions and procedures have no intrinsic relationship with the data they operate on. As long as you provide the correct number and type of arguments, the function/procedure will do its work and faithfully return the expected result. So in a procedural system our functions/procedures use data "given" to them (as parameters) but also directly access any shared data they need. (KevanStannard, 2011). Procedural programming presents a list or set of instructions telling a computer what to do step by step and how to move from the first line of code to the second line of code. Some examples include BASIC, Fortran, Pascal, C and Go.

**Modular programming:** Modular programming is a software design technique that emphasizes the separation of the functionality of a program into independent, interchangeable segments called *modules*. Theoretically, modules represent a separation of concerns, and improve maintainability by enforcing logical boundaries between components. Modules are typically incorporated into the program through interfaces. In the modular programming methodology, one of the most important principles is encapsulation, whereby the data contained within a module is accessible to the rest of the program only via behaviours of the module. This makes it much easier to control what happens when data is modified. Modular programming is a method for designing software by way of breaking up components of

a large software program into manageable pieces. Those pieces, or "modules," can then be independently developed, tested, and refined. It's a process that generally helps shorten development time and avoid replicating code (WiseGEEK)

**Structured Programming:** Structured Programming is a technique for organizing computer program codes in a hierarchy of modules/segments showing how they are used. Each segment has a single entry and a single exit point. At the entry point, control is passed downward through the structure to higher levels of the structure without unconditional branches. Three types of control flow are used namely: sequential, test/selection, and iteration. Structured programming which includes modular programming, is a subset of procedure based programming enforces a logical structure on the program being written to make it more efficient and easier to understand and modify (Rouse, 2014)

**Object-Oriented programming:** Object Oriented Programming arrangements are in Classes and in Objects. An instance is classes of Human beings. These classes can have sub-classes; the sub-class can be Ebonyi girls. Classes inherit properties and behaviours so, Ebonyi girls like money because Ibo girls like money; Ibo girls like money because human beings like money. Object-oriented methodologies are an extension of modular ones, the additional ingredient being inheritance. With this ingredient, the code can refer directly to kinds of interactive things. The Object Oriented Programming ideals are. In object-oriented programming, concepts are directly molded in code employing the ideas of classes and inheritance. A distinct member of a class is called an object.

## 2.2 Features of Object-Oriented Programming

The concept of Object-Oriented Programming is very interesting and creates a lot of ease in the programming process. It includes the concept of:

- Objects
- Classes
- Data Abstraction and Encapsulation
- Inheritance
- Polymorphism

*Objects:* Objects are the basic run-time entities in an object-oriented system. Programming problem is analyzed in terms of objects and nature of communication between them. When a program is executed, objects interact with each other by sending messages. Different objects can also interact with each other without knowing the details of their data or code.

*Classes:* A class is a collection of objects of similar type. Once a class is defined, any number of objects can be created which belong to that class.

*Data Abstraction and Encapsulation:* Abstraction refers to the act of representing essential features without including the background details or explanations. Classes use the concept of abstraction and are defined as a list of abstract attributes. Storing data and functions in a single unit (class) is encapsulation. Data cannot be accessible to the outside world and only those functions which are stored in the class can access it.

*Encapsulation*: Once an Object is created, knowledge of its implementation is not necessary for its use. In older programs, coders needed understand the details of a piece of code before using it. Objects have the ability to hide certain parts of themselves from programmers. This prevents programmers from tampering with values they shouldn't. Additionally, the object controls how one interacts with it, preventing other kinds of errors. For example, a programmer (or another program) cannot set the width of a window to -400. (Popyack, 2012)

*Inheritance:* Inheritance is the process by which objects can acquire the properties of objects of other class. In object oriented

programming, inheritance provides reusability, like, creating additional features to an existing class without modifying the class. This is achieved by deriving a new class from the existing one. The new class will have combined features of both the classes.

*Polymorphism:* Polymorphism means the ability to take more than one form. An operation may exhibit different behaviors in different instances. The behavior depends on the data types used in the operation. Polymorphism is extensively used in implementing Inheritance.

## 2.3 Inherent Pros of Object-Oriented Programming

*Code Reuse and Recycling*: Objects created for Object Oriented Programs can easily be reused in other programs.

*Design Benefits*: Large programs are very difficult to write. Object Oriented Programs force designers to go through an extensive planning phase, which makes for better designs with fewer flaws. In addition, once a program reaches a certain size, Object Oriented Programs are actually easier to program than non-Object Oriented ones. ( Jeffrey L. Popyack )

*Software Maintenance*: Programs are not disposable. Legacy code must be dealt with on a daily basis, either to be improved upon (for a new version of an exist piece of software) or made to work with newer computers and software. An Object Oriented Program is much easier to modify and maintain than a non-Object Oriented Program. So although a lot of work is spent before the program is written, less work is needed to maintain it over time. .(Jeffrey L. Popyack)

## 2.4 Procedure-Based Versus Object-Based Programming

Procedural Programming lays emphasis on identification and specification of a set of steps to solve a given task and the precise way to execute it to reach the desired outcome. For example, if you want to calculate the month-end closing balance of the departmental imprest, the steps would follow thus:

- Assign the initial monthly-allocation

- Sum up all the expenses within the month

- then subtract the sum of the expenses from the initial monthly allocation

- the subtraction will give you the month-end closing balance

A procedure which can be a subroutine or a function contains an ordered list of instructions to be carried out. A procedure can be called at any time during the execution the program by any other procedure or by itself. Let us note at this point that Procedure-based and Object-based programming are two ways of showing problems to be solved and how to go about solving them.

The key difference between Object Oriented Programming and Procedural Programming is that the focus of Procedural Programming is to break down the programming task into a collection of variables and subroutines while, the focus of Object Oriented Programming is to break down the programming task into objects, which encapsulate data and methods. Most notable difference could be that while Procedural Programming uses procedures to directly operate on data structures, Object Oriented Programming will bundle the data and methods together so that an object will operate on its own data (KevanStannard, 2011).

## 3. METHODOLOGY

We studied the effect of OOP on modern software development with data from two main sources as discussed below.

**Primary source:** We conducted oral interviews with a total of 125 interviewees out of which 10 were Lecturers, 15 were freelance indigenous software developers in Enugu Nigeria and 100 were final year students of computer Science. These respondents were selected from three different Universities

namely: Ebonyi State University Abakiliki - Nigeria, Nnamdi Azikiwe University, Awka - Nigeria and Enugu State University of Science and Technology, Enugu - Nigeria. The questions sought the views of the above named groups of persons' on the effects of Object-Oriented Programming on modern software development.

**Secondary source:** we consulted existing documents like computer science journals, text books, laboratory manuals and manuscripts, etc. The Internet was a major source of the secondary data source. Most of our journal articles were sourced through the internet. We studied various reviews and comments from people of all walks of life.

## 4.0 RESULTS PRESENTATION & SUMMARY OF FINDINGS

Table 1 shows the occupational distribution of the interviewee. A total of 125 respondents' opinions were sampled and responses collected and analyzed on a 5-point *linker type* scale as shown subsequently.

| S/n | Respondents' Occupation | No. | Percentage |
|-----|------------------------|-----|------------|
| 1. | Software developers | 15 | 8% |
| 2. | Lecturers | 10 | 12% |
| 3. | Final year students | 100 | 80% |
| | **Total** | **125** | **100%** |

**Table 1: Occupation distribution of interviewed respondents**
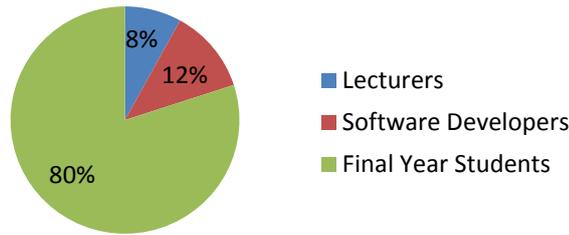
**Figure 1: Pie Chart Showing the Occupational distribution of interviewed respondents**

Table 2 presents the opinion of the respondents to the questions presented by the interviewer. These questions were presented to indigenous software developers, Lecturers and final year students of Computer Science selected from three universities mentioned earlier.

**Table 2: Questions and responses by respondents**

| S/N | Questions | x | F | fx | $\chi$ (mean) | % |
|---|---|---|---|---|---|---|
| 1 | Object oriented programming languages enable programming in modules? | | | | | |
| | Strongly Agree | 5 | 100 | 500 | 4.80 | 80.00 |
| | Agree | 4 | 25 | 100 | | 20.00 |
| | Undecided | 3 | 0 | 0 | | 0.00 |
| | Disagree | 2 | 0 | 0 | | 0.00 |
| | Strongly Disagree | 1 | 0 | 0 | | 0.00 |
| 2 | Program can be divided into units of task and tested differently in OOP? | | | | | |
| | Strongly Agree | 5 | 83 | 417 | 4.60 | 66.67 |
| | Agree | 4 | 33 | 133 | | 26.67 |
| | Undecided | 3 | 8 | 25 | | 6.67 |
| | Disagree | 2 | 0 | 0 | | 0.00 |
| | Strongly Disagree | 1 | 0 | 0 | | 0.00 |
| 3 | Object oriented programming languages have the concept of objects, class, data abstraction, inheritance, and polymorphism? | | | | | |
| | Strongly Agree | 5 | 100 | 500 | 4.73 | 80.00 |
| | Agree | 4 | 17 | 67 | | 13.33 |
| | Undecided | 3 | 8 | 25 | | 6.67 |
| | Disagree | 2 | 0 | 0 | | 0.00 |
| | Strongly Disagree | 1 | 0 | 0 | | 0.00 |
| 4 | It is easy to debug programs written with object oriented programming than procedure oriented programming? | | | | | |
| | Strongly Agree | 5 | 83 | 417 | 4.53 | 66.67 |
| | Agree | 4 | 25 | 100 | | 20.00 |
| | Undecided | 3 | 17 | 50 | | 13.33 |
| | Disagree | 2 | 0 | 0 | | 0.00 |
| | Strongly Disagree | 1 | 0 | 0 | | 0.00 |
| 5 | Tasks are broken into smaller units in object oriented programming? | | | | | |
| | Strongly Agree | 5 | 17 | 83 | 3.40 | 13.33 |
| | Agree | 4 | 17 | 67 | | 13.33 |
| | Undecided | 3 | 92 | 275 | | 73.33 |

| | | | | | |
|---|---|---|---|---|---|
| | Disagree | 2 | 0 | 0 | | 0.00 |
| | Strongly Disagree | 1 | 0 | 0 | | 0.00 |
| 6 | Object oriented programming languages have graphical interface that allow for the design of a more attractive program interface than in procedure oriented programming? | | | | | |
| | Strongly Agree | 5 | 25 | 125 | 3.33 | 20.00 |
| | Agree | 4 | 42 | 167 | | 33.33 |
| | Undecided | 3 | 17 | 50 | | 13.33 |
| | Disagree | 2 | 33 | 67 | | 26.67 |
| | Strongly Disagree | 1 | 8 | 8 | | 6.67 |
| 7 | Object oriented programming aids team work better than procedure oriented programming? | | | | | |
| | Strongly Agree | 5 | 67 | 333 | 4.53 | 53.33 |
| | Agree | 4 | 58 | 233 | | 46.67 |
| | Undecided | 3 | 0 | 0 | | 0.00 |
| | Disagree | 2 | 0 | 0 | | 0.00 |
| | Strongly Disagree | 1 | 0 | 0 | | 0.00 |
| 8 | Object oriented programming makes it easy to modify existing codes for a re-use? | | | | | |
| | Strongly Agree | 5 | 67 | 333 | 4.33 | 53.33 |
| | Agree | 4 | 42 | 167 | | 33.33 |
| | Undecided | 3 | 8 | 25 | | 6.67 |
| | Disagree | 2 | 8 | 17 | | 6.67 |
| | Strongly Disagree | 1 | 0 | 0 | | 0.00 |
| 9 | Objects created for object oriented programs can easily be reused in other programs? | | | | | |
| | Strongly Agree | 5 | 58 | 292 | 4.13 | 46.67 |
| | Agree | 4 | 42 | 167 | | 33.33 |
| | Undecided | 3 | 17 | 50 | | 13.33 |
| | Disagree | 2 | 0 | 0 | | 0.00 |
| | Strongly Disagree | 1 | 8 | 8 | | 6.67 |
| 10 | In object oriented programming, once an object is created, knowledge of its implementation is not necessary for its use? | | | | | |
| | Strongly Agree | 5 | 50 | 250 | 3.87 | 40.00 |
| | Agree | 4 | 33 | 133 | | 26.67 |
| | Undecided | 3 | 17 | 50 | | 13.33 |
| | Disagree | 2 | 25 | 50 | | 20.00 |
| | Strongly Disagree | 1 | 0 | 0 | | 0.00 |

## 4.     DISCUSSION

We can see from table 2 that all the questions/factors listed for the opinion of the respondents were turned-in with high arithmetic mean. Exceptions to these were two factors namely: *presence of graphical interface that allow for the design of a more attractive program interface than in procedure oriented programming* and *breaking of tasks into smaller units*. These factors have arithmetic mean of 3.33 and 3.40 respectivey. That shows that respondents do not see these factors as a positive impact that OOP has brought to software development. Those factors really do exist in the previous programming techniques.

All the other factors that were tested had high positive impact as they presented with high arithmetic mean. These showed that Object-Oriented Programming has made a very remarkable positive impact on modern software development. The impact has brought about resultant increase in the production of millions of software on daily basis all over the world. Object-Oriented Programming has been very successful and possibly more successful than other conventional programming approaches.

The later popularity of OOP notwithstanding, a few issues affected its popularity in its early stage. These issues were: technological shortcomings such as disc space and energy/time dissipated in program planning and design. OOP has brought a lot of changes into computer programming in particular and the world of software development in its entirety. Suffice it to say in a nut shell that Object Oriented Programming has brought the dawn of a new epoch in the software development world.

*Encapsulation, inheritance, instance, abstraction* the *reuse property,* etc were some of the key characteristics of OOP that made it stand out from the earlier programming techniques.   These properties were so unique and interesting that their usage in software development heralds a revolution in the software industry. Deploying Object Oriented Programming paradigm in software development saves a lot of code through the reusable of components, frameworks and designs. Object Oriented Languages made available generic templates and saved the time and space needed for code duplication.

## 5. CONCLUSIONS

There is so much software now than has ever been in the world of software development and in the world in general. Is it possible that this is as a result of time and chance? What do we attribute these daily churning of software to? We attribute these changes and the ease of software development to object oriented programming and all the easy manipulations of data and functions. A glimpse at our oral interview result shows in every area that Object Oriented Programming is preferred by all the three groups of respondents. In view of the above we affirm that the effects of Object Oriented Programming on modern software Development is very positive and makes software development very fast to development.

## 6. REFERENCES

1. Abadi, Martin; Cardelli, Luca (1996). A Theory of Objects. Springer-Verlag New York, Inc.
2. Ambler, Scott (1st January 1998). "A Realistic Look at Object-Oriented Reuse". www.drdobbs.com.
3. Armstrong, Joe. In Coders at Work: Reflections on the Craft of Programming. Peter Seibel, ed. Codersatwork.com.
4. Booch, Grady (1986). Software Engineering
5. Boronczyk, Timothy (11 June 2009). "What's Wrong with OBJECT ORIENTED PROGRAMMING", zaemis.blogspot.com.
6. Brooks, Fred P. (April 1987). "No Silver Bullet — Essence and Accidents of Software Engineering".
7. C. J. Date, Hugh Darwen. Foundation for Future Database Systems: The Third Manifesto (2nd Edition)
8. C. J. Date, Introduction to Database Systems, 6th-ed., Page 650
9. Cambridge: Prentise Hall International Series in Computer Science. p. 23.
10. Cardelli, Luca (1996). "Bad Engineering Properties of Object-Oriented Languages".
11. Computation Center and Research Laboratory. p. 88f. "In the local M.I.T. patois, association lists [of atomic symbols]
12. Conway, Richard (1978). A primer on disciplined programming using PL/I, PL/CS, and PL/CT. Winthrop Publishers.
13. Dr. Alan Kay on the Meaning of "Object-Oriented Programming"". 2003.
14. Graham, Paul. The Emerald Programming Language". 2011-02-26
15. 1995 Reviewers Guide to Visual FoxPro 3.0: DFpug.de
16. Hoare, C. A. (Nov 1965). "Record Handling". ALGOL Bulletin (21): 39–69. doi:10.1145/1061032.1061041.
17. Holmevik, Jan Rune (1994). "Compiling Simula: A historical study of technological genesis". IEEE Annals of the History of Computing 16.
18. Jacobsen, Ivar; Magnus Christerson; PatrikJonsson; Gunnar Overgaard (1992). Object Oriented Software Engineering.  Addison-Wesley ACM Press. pp. 43–69.
19. James, Justin (1 October 2007). "Multithreading is a verb not a noun". Techrepublic.com. Archived from the original on 2 January 2013.
20. John C. Mitchell, Concepts in programming languages, Cambridge University Press, 2003, p.278

**21.** Kay, Alan. "The Early History of Smalltalk". September 2007.

**22.** Kindler, E.; Krivy, I. (2011). Object-Oriented Simulation of systems with sophisticated control**.** *International Journal of General Systems*. pp. 313–343.

**23.** Lewis, John; Loftus, William (2008). Java Software Solutions Foundations of Programming Design 6th ed. Pearson Education**.**

**24.** M.Trofimov, OBJECT ORIENTED PROGRAMMING - The Third "O" Solution: Open OBJECT ORIENTED PROGRAMMING. First Class, OMG, 1993, Vol. 3, issue 3, p.14.

**25.** Meyer, Bertrand (1988). Object-Oriented Software Construction.. p. 105. "Object - a synonym for atomic symbol" Meyer, Second Edition, p. 230

**26.** Michael Lee Scott, Programming language pragmatics, Edition 2, Morgan Kaufmann, 2006, p. 470

**27.** Neward, Ted (26 June 2006). "The Vietnam of Computer Science". Interoperability Happens.

**28.** Pierce, Benjamin (2002). Types and Programming Languages. MIT Press. section 18.1 "What is Object-Oriented Programming?"

**29.** Poll, Erik. "Sub typing and Inheritance for Categorical Data types".

**30.** Potok, Thomas; MladenVouk; Andy Rindos (1999). "Productivity Analysis of Object-Oriented Software Developed in a Commercial Environment". Software

**31.** Rich Hickey, JVM Languages Summit 2009 keynote, Are We There Yet? November 2009.

**32.** Robert Harper (17 April 2011). "Some thoughts on teaching FP". Existential Type Blog.

**33.** Ross, Doug. "The first software engineering language". LCS/AI Lab Timeline:.

**34.** Shelly, Asaf (22 August 2008). "Flaws of Object Oriented Modeling". Intel Software Network.

**35.** Shelly, Asaf (22 August 2008). "HOW TO: Multicore Programming (Multiprocessing) Visual C++ Class Design Guidelines, Member Functions". support.microsoft.com.

**36.** Stepanov, Alexander. "Stlport: An Interview with A. Stepanov". Sutherland, I. E. (30 January 1963). "Sketchpad: A Man-Machine Graphical Communication System"

# Cheater identification in Visual secret sharing schemes using SHA Algorithm and Alpha channel.

Deepika M P
Cochin University of Science and Technology,
Cochin 22, India

A Sreekumar
Cochin University of Science and Technology,
Cochin 22, India

**Abstract:** Secret sharing or secret splitting refers to methods for distributing a secret among a group of participants, each of whom is allocated a share of the secret. The secret can be reconstructed only when a sufficient number of shares are combined together. Individual shares are of no use on their own. Visual secret sharing (VSS) is an alternate form of secret sharing, where the visual information to be shared is distributed /encrypted in such a way that decryption done using Human Visual System. Visual secret is efficient since secret decoding only depends on the human vision system .There are several visual cryptographic schemes to encode the secret image into shares. In these schemes, normally a participant holds one share, and when the participants stack a sufficient number of shares, the secret image /the original image can be reconstructed through the human visual system. As we know where security is enforcing, there will be some provision to spoil the system. So in this case the dishonest participant, aka cheater, can provide a fake share to cheat other participants thereby he may reconstruct the original one and can use the same for any illegal purpose. In this paper, we propose a cheater identification method by adding an extra alpha channel with each share that contains the authentication information related to that particular share. Moreover the proposed scheme can be used with any visual cryptographic scheme to identify the cheaters.

**Keywords:** Secret Sharing, Visual Cryptography, Alpha Channel, Cheater Identification, SHA

## 1. INTRODUCTION

Nowadays digitized personal data is common since technology progresses day by day. Even more of emphasis is on data security today than there has ever been. So protecting these types of data in a secure way which does not impede the access of an authorized authority is an immensely difficult and very interesting research problem even more of emphasis is on data security today than there has ever been.

Visual cryptography is very special data security method which provides a very powerful technique by which one secret can be distributed into two or more shares. When shares on transparencies are super imposes exactly together, the original secret can be discovered without any complicated algorithm or computer participation.

In this paper, many types of visual cryptographic schemes are examined, from the very first type of traditional visual cryptography, up to the latest developments including the positive and negative points of the respective schemes. Cheating is possible in visual cryptography because protection of secret sharing participants is not the main concern. Since there is no restriction on the behavior of the participants, any participant, called a cheater, can reveal a forged share on purpose. Cheating identification in visual cryptography is a main thing in such situations. So we propose a method which uses conventional cryptographic hash function to identify the cheater and prevent such participants.

## 2. VISUALCRYPTOGRAPHIC SCHEMES:

Visual Cryptography is the scheme which can decode concealed secret image without any cryptographic computations. This scheme is very secure and easy to implement. The basic model consists of a printed page of ciphertext (which can be sent by mail or faxed) and a printed transparency (which serves as a secret key).The original text is revealed by placing the transparency with the key over the page with the ciphertext, even though each one of them is indistinguishable from each other. The system is similar to a one time pad in the sense that each page of ciphertext is decrypted with a different transparency. Due to its simplicity, the system can be used by anyone without any knowledge of cryptography and without performing any cryptographic computations.

Visual cryptography (VC) is first introduced by Naor and Shamir [1] in 1994, VC is a cryptographic technique which allows visual information (picture, text, etc) to be encrypted in such a way that the decryption can be performed by the human visual system.However, it is distinguished from traditional secret sharing technique [2], in that the decryption of an image encrypted by a visual cryptography scheme requires no mathematical computations or knowledge of cryptography. Instead, the original image becomes visible to the naked eye simply by overlaying cipher transparencies – known as shares – created during the encryption process. Naor and Shamir [1] establish visual cryptography as a visual variant of the k out of n secret sharing problem. In this scheme, one wishes to randomly divide a secret amongst a group of n individuals in such a way as to allow any k < n of them (or, in certain cases, only a qualified subset of them), to recover the secret from their individual shares. However, any number of individuals $k^0 < k$ should be prevented from obtaining any information about the original secret by combining their individual shares.

Visual cryptography schemes are typically lossy and produce decrypted images that are often noisy or suffer from diminished contrast and resolution. A number of factors can affect the quality of the resulting decrypted image in a VC scheme. Typically, as the number of shares n is increased, the contrast of the resulting decrypted image worsens. Furthermore, many schemes produce shares in which each pixel of the original image is represented by multiple pixels

in each share, diminishing the resolution of the decrypted image.

It can be tempting to think of visual cryptography as a form of Steganography, but it is important to understand the distinction between the two. In Steganography, one seeks to conceal the existence of a message, perhaps by composing the message using invisible ink. By contrast, visual cryptography – like its true cryptographic counterparts – seeks only to conceal the message itself. It is, however, possible to combine Steganography and visual cryptography to produce two benign -looking images that, when superimposed, reveal a third hidden image.

## 2.1 A (2, 2) Visual Secret Sharing

The simplest VC algorithm was given by Naor and Shamir [1] in their introductory paper on visual cryptography. They presented a 2 out of 2 scheme, in which 2 shares would be generated (n = 2) for each image encrypted, while decryption would require these 2 shares (k = 2) to be super-imposed. At its most basic level, the 2 out of 2 algorithm works by representing each pixel in the original image by 2 pixels in each share. Each pixel in the original image is read and, if a white pixel is encountered, one of the first two rows in Figure 1 is selected with equal probability, and each share is assigned a 2 pixel block as shown in the third and fourth columns. Similarly, if a black pixel is encountered, one of the last two rows is selected with equal probability, and is assigned to each share.

If two white pixels overlap when two shares are superimposed, the resulting pixel will be white. By contrast, if a black pixel in one share overlaps with either a white or black pixel in the other share, the resulting pixel will be black. This implies that the superimposition of the shares represents the Boolean OR functions. The last column in Figure 1 shows the resulting subpixel when the subpixels of both shares in the third and fourth columns are superimposed.

As demonstrated in Figure 1, if a pixel in the original image was black, the subpixel in the superimposition of the two shares will be fully black. Similarly, if a pixel in the original image was white, the subpixel in the superimposition of the two shares will be black and white. However, because the pixels are small and situated very close together, the human eye averages the relative contributions of the black and white pixels, resulting in a grey pixel.

Figure 2 shows the encryption and decryption of the University of Western Ontario logo using Naor and Shamir's 2 out of 2 algorithm, in which 2 subpixels are used for each original pixel. Neither share generated reveals any information about the original image, but when the two are superimposed as shown in Figure 2(d), a representation of the original image can be seen. The aspect ratio of the original image is distorted in the decrypted version due to the fact that the use of 2 subpixels per original pixel doubles the width of the decrypted image while retaining its original height.

Figure 1: 2 out of 2 using 2 subpixels per original pixel



Figure 2: 2 out of 2 encryption/decryption using 2 subpixels per original pixel



| | |
|---|---|
| (a) Original Image | (b) Share 1 |
| (c) Share 2 | (d) Shares Super-Imposed |
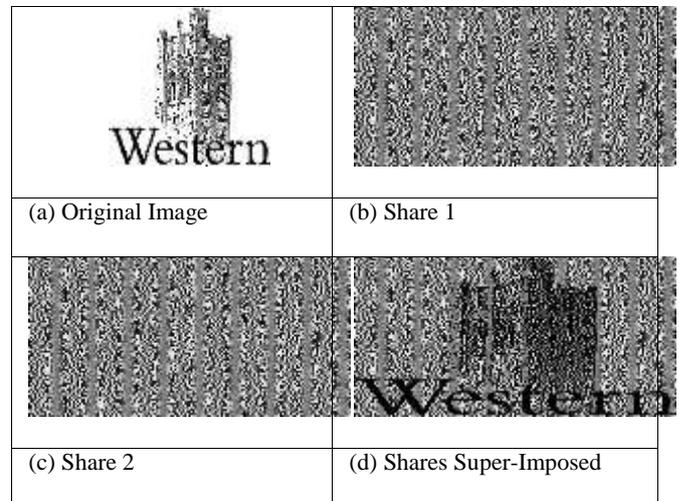
To compensate for the distortion of the aspect ratio of the original image, Naor and Shamir [1] recommend using a 2 × 2 subpixel block to represent each original pixel. This produces an image that is four times the size of the original image, but retains the aspect ratio of the original image. Figure 3 shows the subpixels used in this new variant of the 2 out of 2 algorithm.

Figure 3: 2 out of 2 using 4 subpixels per original pixel



Figure 4 shows an encryption and decryption cycle on the same image used in Figure 2, this time using the 4 subpixel variant of the 2 out of 2 algorithm. It is clear that while the image is four times as large as the original, its original aspect ratio has been preserved, producing a clearer and more natural looking result.

Figure 4: 2 out of 2 encryption/decryption using 4 subpixels per original pixel



(a) Original Image

(b) Share 1

(c) Share 2

(d) Shares Super-Imposed
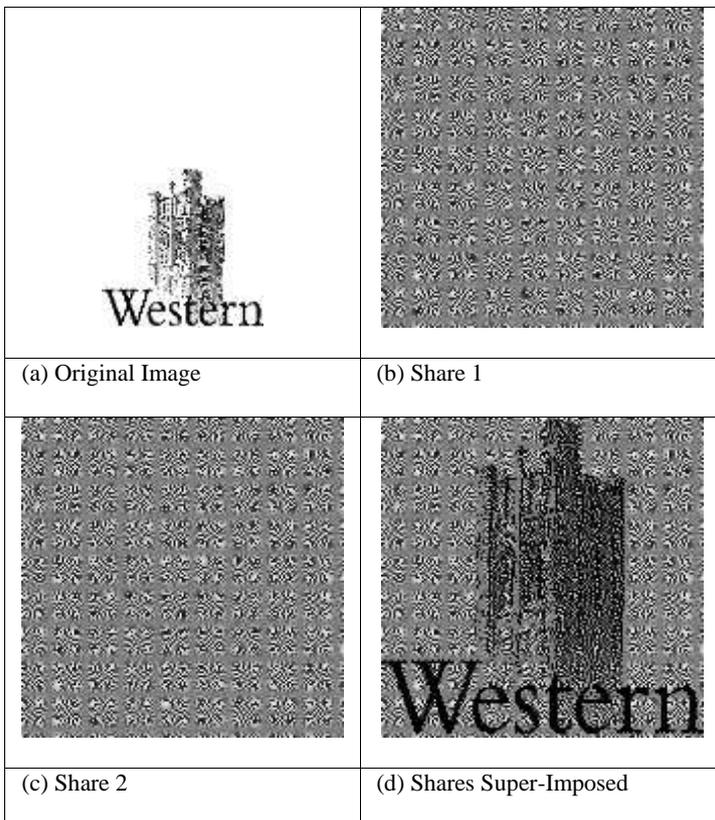
## 2.2 Extended Visual Cryptography:

Extended VC takes the idea of visual cryptography further by creating shares which are meaningful to anyone who views them. This helps to alleviate suspicion that any encryption has taken place and also presents visually pleasing shares which incorporate all the previously mentioned features of VC. Extended visual cryptography schemes allow the construction of shares in which the shares are meaningful as opposed to having random noise on the shares. After the sets of shares are superimposed, the meaningful information disappears and the secret is recovered. This is the basis for the extended form of visual cryptography.

Figure 5.shows an example of a (2, 2) EVCS. As can be seen from the figure, two meaningful shares are generated from the base images. During this share creation, the secret is encoded between each of the shares. After superimposing each share, the secret is completely recovered while the meaningful information on each share completely disappears.

In order to use this extended visual cryptography scheme, a general construction needs to be defined. Ateniese et al. [19] have devised a mechanism by which we can generate the shares for the scheme. A stronger security model for EVCS is one in which the shares associated with a forbidden subset can be inspected by the user, meaning that the secret image will still remain totally hidden even if all n shares are previously known by the user. A symmetric approach to fully address a general (k, n) problem was also proposed in [19].

For each set of access structure, let $P = \{1, 2, \dots n\}$ represents the set of elements called participants, and let $2^P$ denote the set of all subsets of P. Let $\Gamma_{Qual}/\Gamma_{Forb}$ be the collection of Qualified / Forbidden sets. The pair is called the access structure of the scheme. Any qualified set can recover the shares image by stacking its participant's transparencies, while any forbidden set has no information on the shared image. This extension generalizes the original secret sharing problem by [2]. In [19] the authors propose a new technique to realize (k, n) VCS, which is better with respect to the pixel expansion than the one proposed by Naor and Shamir.

Figure 5: The result of (2, 2) –EVCS encryption process



(a) Secret (192 × 192).

(b) Base image 1 (192 × 192).

(c) Base image 2 (192 × 192).

(d) Extended share 1 (384 × 384).

(e) Extended share 2 (384 × 384).

(f) Recovered secret after superimposing share 2 atop share 1 (384 × 384).
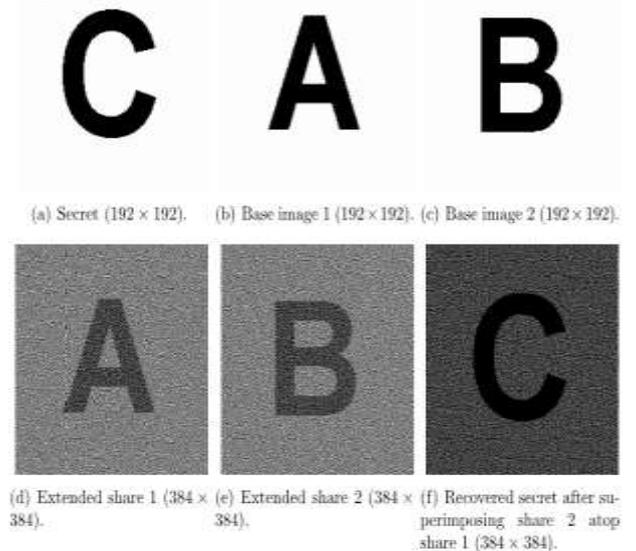
One of the most potentially useful types of visual cryptography scheme is colour visual cryptography. The reason for this is that the majority of people nowadays are

more used to colour images and interact with them more frequently. Natural colour images can be used to share secrets; this provides a very helpful cover for unsuspicious hiding the fact that any encryption has taken place at all. However, some of these schemes do not work without a computer, which does defeat the main purpose of visual cryptography. Other colour schemes do try to keep with the main ethos of instantaneous decryption without a computer.

## 2.3. Colour Visual Cryptography:

Visual cryptography schemes were applied to only black and white images till year 1997. A very primitive example of color image sharing appears in [16].Verheul and Van Tilborg proposed an important color visual cryptography scheme [17]. In this visual cryptography scheme one pixel is distributed into m sub pixels, and each sub pixel is divided into c color regions.  In each sub pixel, there is exactly one color region colored, and all the other color regions are black. In 2000, Yang and Laih [18] proposed a different construction mechanism for the colored visual cryptography scheme. They argued that their method can be easily implemented and can get much better block length than Verheul and Van Tilborg's scheme.

F.Liu, C.K.Wu, X.J. Lin [20] proposed a new approach for colored visual cryptography scheme. They proposed three different approaches for color image representation:

• In first approach, colors in the secret image can be printed on the shares directly. It works similar to basic visual cryptography model. Limitations of this approach are large pixel expansion and quality of decoded image is degraded.

•In second approach separate three color channels are used. Red, green, blue for additive model and cyan, magenta, yellow for subtractive model. Then normal visual cryptography scheme for black and white images is applied to each of the color channels. This approach reduces the pixel expansion but quality of image gets degraded due to half toning process.

•In third approach, binary representation of color of a pixel is used and secret image is encrypted at bit-level. This results in better quality of image.

A major common disadvantage of the above reviewed colored VCS schemes is that the number of colors and the number of subpixels determine the resolution of the revealed secret image. If many colors are used, the subpixels require a large matrix to represent it. Also, the contrast of the revealed secret image will go down drastically. Consequently, how to correctly stack these shared transparencies and recognize the revealed secret image are the major issues.

Recently, more and more applications of visual cryptography, such as authentication, human identification, copyright protection, watermarking, visual signature checking etc. are introduced. The print and scan application of VCS is also introduced by researchers. In this application, scan the shares into a computer system and then digitally superimpose their corresponding shares. This would make possible secure verification of e-tickets or other documents. The developments and the research works done by other researchers in the different perspectives on visual cryptography, such as access structure, generation of shares and other aspects are already reported by different authors.

## 3.  CHEATING IN VISUAL CRYPTOGRAPHY:

In VC, all participants who hold shares are assumed to be semi-honest, that is, they will not present false or fake shares during the phase of recovering the secret image. Thus, the image shown on the stacking of shares is considered as the real secret image. Nevertheless, cryptography is supposed to

guarantee security even under the attack of malicious adversaries who may deviate from the scheme in any way. For cheating, a cheater presents some fake shares such that the stacking of fake and genuine shares together reveals a fake image.

Horng [6] proposed that cheating is possible in (k, n)-V SS where k < n. The cheating activity of Horng is that the n-1 cheaters collusively use their transparencies to know the secret and infer the victim's transparencies $T_v$; thus they can generate fake transparencies FTs to make the victim to accept the cheating image by stacking FTs + $T_v$.

Consider (2, 3)-V SS scheme as an example in Figure 6. As secret image is encoded into three distinct transparencies, denoted $T_1$, $T_2$, $T_3$, Then, the three transparencies are respectively delivered to Alice, Bob, and Carol. Without lose of generality, Alice and Bob are assumed to be the collusive cheaters and Carol is the victim. In cheating, $T_1$ and $T_2$ to create forged transparency $T_2'$ such that superimposing $T_2'$ and $T_3$ will visually recover the cheating image. Precisely, by observing the following collections of $3 \times 3$ matrices which are used to generate transparencies, the cheaters can predict the actual structure of the victim's transparency so as to create $T_2'$.

$$\text{Create T2'. } C^0 = \begin{bmatrix} 100 \\ 100 \\ 100 \end{bmatrix} \text{ and } C^1 = \begin{bmatrix} 100 \\ 010 \\ 001 \end{bmatrix}$$

By observing the above matrices, two rows of above $C^0$ or $C^1$ matrix are determined by the collusive cheaters. Therefore, the structure of each block in $T_3$ is exact the remaining row. For presenting a white pixel of cheating image, the block in $T_2$ is set to be the same structure of $T_3$. For presenting a black pixel of cheating image, the block in $T_2$ is set to be the different structure of $T_3$. Figure 1 and Figure 2 Shows the whole cheating process. Shows the cheaters create to change the decoded image. If the block in $T_3$ is [010], then $T_2$ is set to be [010] for a white pixel or it is set to be [001] for a black pixel. Formally, the cheaters can construct a sub-base matrix (SBM) by $T_1$ and $T_2$ then infer $T_3$.

Figure 6: Cheating in visual cryptographic scheme.
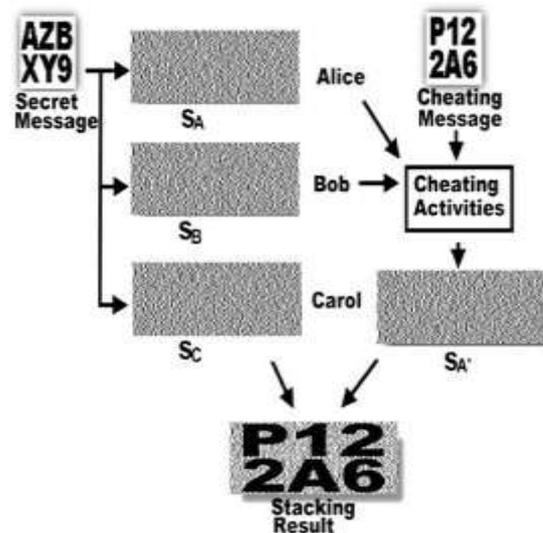
Figure 7.The basic concept of cheating process in 2 out of 3 VCS

| | Pixel in Secret message | Block in Share S₁ | Block in Share S₂ | Block in Share S_C | Pixel in Cheating message | Block in Share S'₁ | Block in Share S'₁ |
|---|---|---|---|---|---|---|---|
| Case 1 | White | [1 0 0] | [1 0 0] | [1 0 0] | White | [1 0 0] | [1 0 0] |
| Case 2 | white | [1 0 0] | [1 0 0] | [1 0 0] | Black | [0 1 0] | [0 0 1] |
| Case 3 | Black | [1 0 0] | [0 1 0] | [0 0 1] | White | [0 0 1] | [0 0 1] |
| Case 4 | Black | [1 0 0] | [0 1 0] | [0 0 1] | Black | [1 0 0] | [0 1 0] |

### 3.1 Cheating Prevention Using Authentication Based

There are several schemes that solves the cheating problem by using verification shares to ensure from other participants are authentic and hence the recovered secrete image is authentic. However each participant is burden with a verification share.

An authentication based cheating prevention scheme consists of shares Si and verification shares Vi. Shares Si are generated by any visual cryptographic scheme. Verification shares Vi, for i =1, 2, . . . ,n, generated by the verification shares generation process are used to verify the correctness of the shares Sj, for j =1, 2, . . . ,n and i !=j . Each participant Pi should provide the dealer with a distinct verification logo Li to be used for verifying the authenticity of other shares. All logos are confidential. The verification shares generation process is based on a 2-out-of-2 VC. Each verification share Vi is divided into n−1 regions, R i,j where 1≤j ≤n, j !=i so that when stacking Vi and Sj the logo Li appears in R i,j.

The main limitations in this type of scheme are; each participant, however, was burdened with an extra verification share.VC requires total number of $n^2$ subpixels in all transparencies and this scheme requires total number of $2n^2$ subpixels. Finally in this scheme there is a possibility to create a forged share without modifying any blocks within the victim's region to pass the verification process when the number of n is becoming large.

In [7] Tsai, Chen, Horng scheme, use Generic Algorithms (GA for short) to solve the cheating problem. The proposed scheme does against the cheating attack in VC. The GA based share construction method provides another direction for creating shares.

There are different cheating prevention and cheater identification methods in [8][9][10][11][12][13][14], each having its own advantages and disadvantages.

## 4. PROPOSED METHOD FOR CHEATER IDENTIFICATION:

The proposed method can be used with all kind of secret sharing scheme in visual cryptography. After share construction, using any of the visual cryptographic schemes, an authentication signal is generated for each block(here each raw) of the share using SHA algorithm (Secure Hash Algorithm) and the generated signal is embedded in the alpha channel of the shares. The authentication signal generation phase is described in the Algorithm-I. There are mainly two requirements in the proposed method:

- The size of the secret image as well as shares should be rows $X$ 1024.
- The image format of the share should be PNG image. Actually PNG image support alpha channel.

Figure 8. Authentication signal embedding while share construction phase.
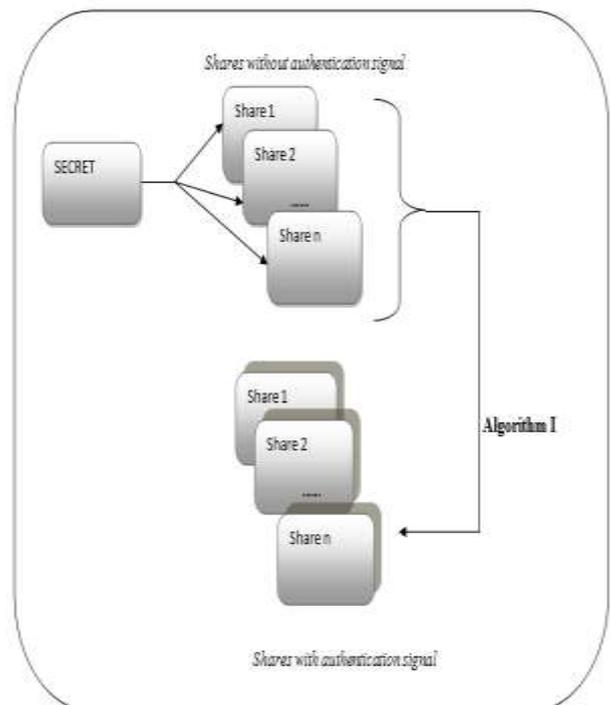


Algorithm-I can again used in the generation of authentication signal for each block of the share at reconstruction phase and the generated signal is compared with the authentication signal that is already embedded in the alpha channel of each share. If both signals are same then we can say that the shares are genuine, otherwise fake shares. By this way we can identify that the shares modified or not. While modifying the shares the alpha channel will not be affected.

*Algorithm-I*

[Each share is having a size *row* × 1024; because in preprocessing stage we are converting the secret image into some size, say *row × 1024*]

***Input:*** share

***Output***: share with authentication signal embedded in alpha channel

Step1. Take one row at a time

Step2. For (*i=0* to *row*) do the step 3 and step 4

Step3.  Apply SHA- 512 on the bits (*1024*) of the ith row.

Step4. Resulting *512* bit hash value is embedded as the ith row of alpha channel

Step5. Return the modified share.

Step6. Stop

Figure 9: Secret Sharing Phase
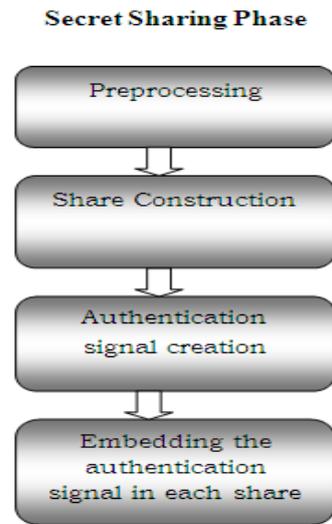


The Algorithm II describes the process of secret reconstruction. In the secret reconstruction phase, the proposed system uses the SHA -512 algorithm to find out the authentication signal of each share once again. Then the generated authentication signal is verified against the received authentication signal with each share in alpha channel.  If the share is a fake share then the generated authentication signal (using the SHA- 1 algorithm) then the verification fails and the cheating can be identified and prevented at this point.

Figure 10: Secret Reconstruction Phase



In secret sharing phase the main sub phases are; (a) preprocessing the secret image, (b) share construction (c) Authentication signal creation and (d) Embedding the authentication signal in each share. In preprocessing phase the secret image is converted into the prescribed size, n X 1024 pixels, because the proposed method requires the image to be in n X 1024 pixel size. This makes the application of the SHA -512 algorithm for generating the authentication signal simpler.  In the second phase, share construction, any size invariant secret sharing scheme can be used. One of the most important things that should be considered here is the formats of the shares. It should be in PNG image format, because the PNG image format provides the transparent channel called alpha channel with the image. In the proposed system we have used the progressive visual cryptography [4][5] as well as size invariant (2, 3) VCS. In the next phase, by using SHA-512 algorithm generate the authentication signal as prescribed in algorithm 1. And finally embed the authentication signal in the alpha channel of each share and distribute the shares in to the participants.
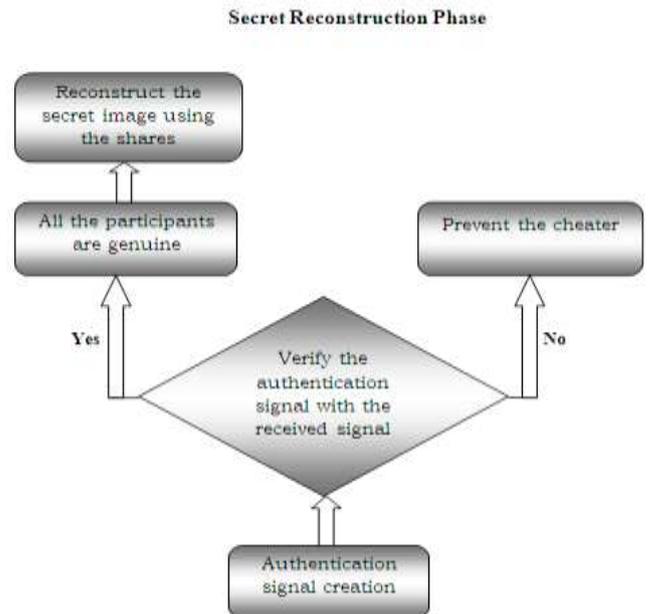
*Algorithm-II*

[Each received share is having a size *row* × 1024;]

*Input:* received shares

*Output*: whether the received share is fake share or not.

**Step1.** Take one row at a time of received shares having size *row* x 1024.

**Step2**. For (*i=0* to *row*) do the steps 3, step 4 and step 5.

**Step3.** Apply SHA- 512 on the bits (*1024*) of the i$^{th}$ row.

**Step4.** Compare the Resulting 512 bit hash value with the 512 bit that is embedded in the i$^{th}$ row of alpha channel.

**Step5**. If both 512 bit hash values are not matching then exit the loop by return the share as fake share and go to the step 7, else continue.

**Step6.** Return the share as the genuine one.

**Step7.** Stop

## 5. SECURITY ANALYSIS:

The security of the proposed method directly depends on the strength of the SHA- 512 algorithm. Suppose a single pixel value changed from 0 to 1 for any of the shares, then the number of bit positions that differ between the authentication signal generated in reconstruction phase and the authentication signal generated in the secret sharing phase is 253, almost the half the bit positions of the authentication signal (of 512 bits), indicating that SHA-512 has a good avalanche effect.

## 6. CONCLUSION:

In this paper, a novel method for cheater identification in the secret sharing schemes is presented. The most important advantage of this method is, it is applicable for all the secret sharing schemes. And the secret analysis shows that even if a single pixel is modified in any of the shares then the authentication signal generated using SHA 1 will be different for that share, and that will lead to the identification of the cheater.

## 7. REFERENCES

[1]M.Naor, A. Shamir, Visual cryptography, in: Proceedings of the Advances in Cryptology, Eurocrypt '94, in: LNCS, vol.950, 1995, pp.1–12.

[2].A. Shamir, _How to share a secret,_ Communications of the ACM, vol. 22, no. 11, pp.612_613, 1979.

[3].P.S.Revenkar, Anisa Anjum, W .Z.Gandhare,_ Survey of Visual Cryptography Schemes _International Journal of Security and Its Applications Vol. 4, No. 2, April, 2010

[4]Duo Jin, WeiQi Yan, and Mohan S. Kankanhalli. Progressive color Visual Cryptography. SPIE journal of Electronic Imaging, 14(3), 2005.

[5] Young-Chang Hou and Zen-Yu Quan _Progressive Visual Cryptography with Unexpanded Shares_ IEEE transactions on circuits and systems for video technology, vol. 21, no. 11,november 2011.

[6]G.Horng,T.H.Chen,D.S.Tsai,Cheating in visual cryptography, Des Codes Cryp-togr.38(2)(2006)219–236

[7]D.S.Tsai ,T.H.Chen ,G.Horng ,A cheating revention scheme for binary visual cryptography with homogeneous secret images ,Pattern Recogn .40(8) (2007) 2356–2366.

[8]C.M.Hu,W.G.Tzeng,Cheating prevention in visual cryptography, IEEE Trans.Image Process.16(1)(2007)36–45.

[9] R.DePrisco,A.DeSantis,Cheating immune threshold visual secret sharing ,Comput. J.53(2010)1485–1496.

[9] Y.C.Chen,G.Horng,D.S.Tsai,Cheating prevention visual cryptography ,Visual Cryptography and Secret Image Sharing,CRCPress,ISBN 9781439837214, 2012

[10] Y.C.Chen,D.S.Tsai,G.Horng,A new authentication based cheating prevention scheme in Naor–Shamir's visual cryptography ,J.Vis.Commun.Image Repre-sent.23(8)(2012)1225–1233.

[11] Y.C.Chen,G.Horng,D.S.Tsai,Comment on " Cheating prevention in visual cryp-tography ",IEEE Trans. Image Process. 21(7)(2012)3319–3323.

[12] Y.C.Chen,D.S.Tsai,G.Horng,visual secrete sharing with cheating prevention revited.,digital signal processing 23 (2013)1496-1504.

[13] Du-Shiau Tsai,Tzung-Her Chen,Gwoboa Horng,Acheating prevention scheme for binary visual cryptography with Homogeneous secret images,pattern rcognition 40 (2007) 2356 – 2366.

[14] C.S Tasai,H.C.Wang,H.C Wu,C.H.M Wang ,"A Cheating –Preventing Visual Cryptography Scheme By Referring The Special Position", International Journal Of

Innovative Computing ,Information And Control volume 7 , N Umber 7(A),July 2011.

[15]Giuseppe Ateniese,carlo Blundo,Alfredo De Santis, and Douglas R. Stnson. Extenede schemes for visual cryptograpgy. Theoretical Computer Science,250:116,June 1996.

[16]V.Rijmen and B.Preneel. Efficient color Visual encryption for colors of Benetton. ERCRYPTO 96, 1996.

[17]Eric R Verheul and Henk C A Van Tilborg. Constructions and properties of k out of n visual secret sharing schemes.Design Codes Cryptography, 11(2):179–196,1997.

[18]ChingNung Yang and Chi Sung Laih.new colored visual secret sharing schemes.Designs,Codes and Cryptography,20(3):325 –336,2000

[19] G. Ateniese, C. Blundo, A. DeSantis, and D. R. Stinson, "Visual cryptography for general access structures", Proc.ICAL96, Springer, Berlin, 1996, pp.416-428.

[20] [17] F. Liu, C.K. Wu, X.J. Lin, "Colour Visual Cryptography Schemes", IET Information Security, vol. 2,No. 4, pp 151-165, 2008.

.

# Blog Quality Assessment Tool (BQAT)

Zuhaira Muhammad Zain
Faculty of Computer and Information Sciences
Princess Nourah Bint Abdulrahman University
Riyadh, KSA

Abdul Azim Abd Ghani
Faculty of Computer Science and Information
Technology
Universiti Putra Malaysia
Serdang, Malaysia

**Abstract**: A blog quality model and guidelines to determine important features of different blog categories have been proposed to determine blog quality and to promote readers' satisfaction. However, no tools have been developed to assist blog readers in the evaluation of their favorite blogs based on their blog satisfaction. This paper discusses each process in the development of the Blog Quality Assessment Tool (BQAT) in detail. The main functions of the BQAT are to calculate the probability of a blog to be of good quality based on blog-reader satisfaction, and to accumulate the results for the assessed blog. Thus, blog-readers can easily assess their favourite blogs and obtain information on the quality of the blogs visited. This study also shows that the more satisfied the blog is, the higher its quality.

**Keywords**: Blog Quality, Blog Quality Model, Rasch Analysis, Blog Quality Assessment Tool

## 1. INTRODUCTION

Scholars have shown increasing interest in blog success by providing blog design advice and checklists [1, 2, 3, 4]. Nevertheless, these criteria are merely based on individual authors' or bloggers' opinions. 49 blog quality criteria have been consolidated by Zain et al. [5, 6, 7] drawn from related studies focused on website design criteria [8], web information quality criteria for different domains including e-commerce [9], data integration [10], decision making [11], organizational networks [12], personal websites [13, 14], web portals [15], criteria extracted from design advice and checklists [16, 3, 17], and design articles extracted from popular blogs [18, 1, 2, 4, 19]. Zain and Ghani [20] provided a relative importance analysis that can help bloggers/blog evaluators/readers focus on the most important criteria during blog category examination. Yet, no tools have been developed to assist blog readers in the evaluation of their favorite blogs based on their blog satisfaction. This study describes the development of a blog quality assessment tool to assist blog readers in the evaluation of their favorite blogs based on their blog satisfaction. This will help maintain blog quality in the blogosphere.

## 2. RELATED WORKS

Quality is an essential factor in the information technology environment. It is an important requirement in information technology-related development (i.e. software, website, and information system domains). It is a composite of many characteristics that operate in particular development domains. Quality might be conceptualized as a quality model/framework that depicts composite characteristics and their relationships. Each model/framework can guide developers/designers during quality product production (e.g. software, data, websites, or information). Alternatively, users can employ a model/framework to evaluate those products. Some commonly accepted software quality models include McCall et al.'s [21], Boehm's [22], Dromey's [23], and the ISO/IEC 9126 quality model [24]. They often serve as foundations for other models in different domains such as website and data development.

Quality is vital to the website development community. Website quality models comprise the Web Quality Evaluation Method (WEBQEM) [25], Web Quality Model (WQM) [26], and a model designed for web-based applications [27]. These models can be applied to evaluate the overall quality of web-based applications. Nevertheless, most concentrate on the usability aspects and lack aesthetic and reputational features. Malak [27] proposed another model to assess quality that highlights on criteria that influence webpage navigational design quality (e.g. information links) and availability of navigational features (e.g. menus and search tools). Even though it attempts to incorporate design features to assess quality, it does not include many important design features, such as multimedia and visual design. A systematically study on important design features of different website domains based on user satisfaction and expectations has been done by Zhang et al. [8]. They described 77 website features and grouped them into 15 feature families. They incorporated aesthetic aspects (e.g. multimedia, visual design, and attractive layout, as well as reputational aspects (e.g. site or company reputation and rewarding experience). These models can be used by Website developers/designers as guidelines during high-quality website development. Furthermore, website users/ evaluators can rely on them to evaluate website quality. For instance, WebQEM has been used to assess websites in different domains including museums [28], academia [29], and e-commerce [30]. Blog characteristics are similar to website characteristics. Therefore, many website quality features are used to determine blog quality. Yet, some features are not relevant to blog quality measurement (e.g. Product and Service Concerns and Security). Hence, we focused on personal blogs because most security criteria solely benefit blog owners, rather than blog readers.

Quality is very essential to the information quality community. Quality begins within the context of management information systems [31, 32] and extends to other contexts, such as cooperative systems [33], data warehouses [34, 35], and electronic commerce [36, 9]. Scholars now focuses on web information quality [37] because of increased awareness of differences between Web applications and traditional information systems. Caro et al. [15] argued that a gap exists among types of information quality specifically developed for web portals. They discovered 33 significant criteria for portal

data quality based on users' perspectives. However, their model failed to include criteria (e.g. search tools and chat rooms). They solely addressed data quality, rather than the entire Web portal. In addition, some criteria are irrelevant and inappropriate for blog quality determination.

Quality is very essential to the blogging community. Blog design advice and checklists include criteria that might influence users' satisfaction (e.g. readability, navigability, clarity, and commentary) [3]. Nevertheless, individual authors defined most criteria. Banks [16] interviewed 30 of the world's top bloggers. He summarized the results and offered suggestions for successful blogging. Yet, the suggested criteria, (e.g. originality, relevant information, and easy navigation), are useful only from bloggers' perspectives. Hopkins [18] conducted a systematic preliminary study focused on ideal blog types. He identified that ideal blogs include comments, photos, and primarily original materials. Ideal blogs feel personal. However, based on our literature review, no empirical evidence confirmed that these criteria are ample and complete. Blog quality includes all blog characteristics that determine a blog's ability to satisfy stated and implied needs [7]. Zain et al. [5, 6, 7] constructed a blog quality model by determining a set of criteria based on a review of relevant studies and blogs. They measured these criteria's acceptability based on questionnaire surveys completed by a sample population of blog readers [7]. The blog quality model comprised of 11 families decomposed into 49 quality criteria that can be used by the blog evaluators to determine blog quality. Bloggers can use it to promote readers' satisfaction. Zain and Ghani [20] provided guidelines that blog designers/evaluators can employ to determine important features of different blog categories. However, no tools have been developed to assist blog readers in the evaluation of their favorite blogs based on their blog satisfaction.

# 3. METHODOLOGY

A prototype of the Blog Quality Assessment Tool (BQAT) was developed in accordance with the processes proposed by Sommerville [38] as follows:

## 3.1 Initial analysis

In this process, basic requirements including the blog quality criteria, desired input and output information were determined. Before specifying the blog quality criteria, we determine the assessors and the assessment process. By reviewing studies on website quality, we determine that quality can be assessed in three different ways: users' view, developers' view, and managers' view [39, 40, 41].

Users are interested in performance quality, primarily an external characteristic, while developers and managers are more concerned with internal quality issues such as maintainability, portability, cost effectiveness, and so on. However, in our case, as our focus is primarily on personal blogs, we can assess these from both blog-readers' and bloggers' viewpoints. Blogs, like websites, focus on users' perspective, an external aspect of quality.

External quality can be defined through both functional and non-functional properties. Apart from functional properties, non-functional properties such as easy to understand, correctness and originality, contribute significantly to blog quality.

In line with suggestions proposed by Evans and King [42] to evaluate Web-based applications, a blog assessment must be

comprehensive, constituting five major components: (i) blog categories (the broad areas to be investigated), (ii) quality factors (specific elements pertaining to each blog category), (iii) weights (relative importance of each blog category and quality factor), (iv) ratings (scores assigned to each category and quality factor), and (v) total score (an overall score based on the weights and ratings).

The first step was to identify the quality factors for a blog. These were determined by Zain et al. [7]. Subsequently, the quality factors were assigned weights; the greater the weight, the more important the quality factor. The weight for each blog category and quality factor was obtained from Zain and Ghani [20]. Subsequently, blog-readers review a blog and rate the quality factors based on their level of satisfaction with the respective factor. The ratings were treated as input in the BQAT prototype. A quality factor total score was then obtained by multiplying the assigned weights with the respective ratings. Finally, the quality factor scores were aggregated to obtain an overall quality score, the BQAT output, for the respective blogs. As the Rasch Measurement Model was used to determine the output, it was referred to as 'the probability of the blog to be a good quality blog'.

## 3.2 Define the prototype objective

The objective of developing the prototype was twofold: (i) to apply the proposed blog quality model, and (ii) to assist readers or bloggers to assess blog quality.

## 3.3 Specify the prototype

All functions relevant to the blog quality assessment were listed, and then each function was either accepted or rejected for inclusion in the prototype system as shown in Table 1 and Table 2.

**Table 1. Functions included in the prototype**

| Functions | Remarks |
|---|---|
| Blog rating | Rating satisfaction using radio buttons. |
| Error handling 1 | If user misses to rate a criteria, a prompt, "Please complete your assessment on the criteria!" will be displayed. |
| Blog quality estimation | Calculates the probability of the blog being of good quality |
| Error handling 2 | If there are less than 30 respondents a prompt, "Insufficient Statistics! Data is not enough to construct a reliable result!", will be displayed. |
| Navigation | User can click on the BQAT banner, Start, Next, and Submit buttons. |

**Table 2. Functions excluded from the prototype**

| Functions | Remarks |
|---|---|
| Registration of blogger | A page where bloggers wishing to participate in the blog assessment project can register their blogs. |
| Submit data to database | Send blogger's data to the BQAT database. |
| Send banner and URL | Automatically email the banner and URL of the blogger's assessment page to the blogger. |

All functions in Table 2 were excluded from the prototype as this study only focuses on the methods to rate a blog and to assess its quality. In order to include these functions, a dummy blog (see Figure 3), with a BQAT banner linking to the blog assessment page, was created. Moreover, the dummy

blog was used to demonstrate how blog readers can attempt the blog assessment and how the BQAT accumulates the results. The details of the dummy blog were manually input in the BQAT database owing to time and cost constraints.

## 3.4 Prototype construction

The BQAT was built using WAMPSERVER technology. It consists of three principal components; Apache web server, MySQL database and PHP scripting language. This package is free and very easy to use enabling easily manipulation of information held in a database and dynamic generation of web pages each time a browser requests for content. PhpMyAdmin program is also included in this package, providing a graphical user interface for the MySQL database manager. The architecture of the BQAT system is depicted in Figure 1.
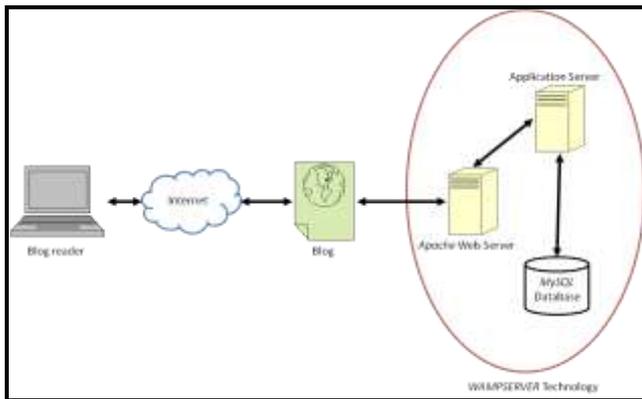


Figure 1.  BQAT architecture

The BQAT prototype was developed as per the flow-chart shown in Figure 2.
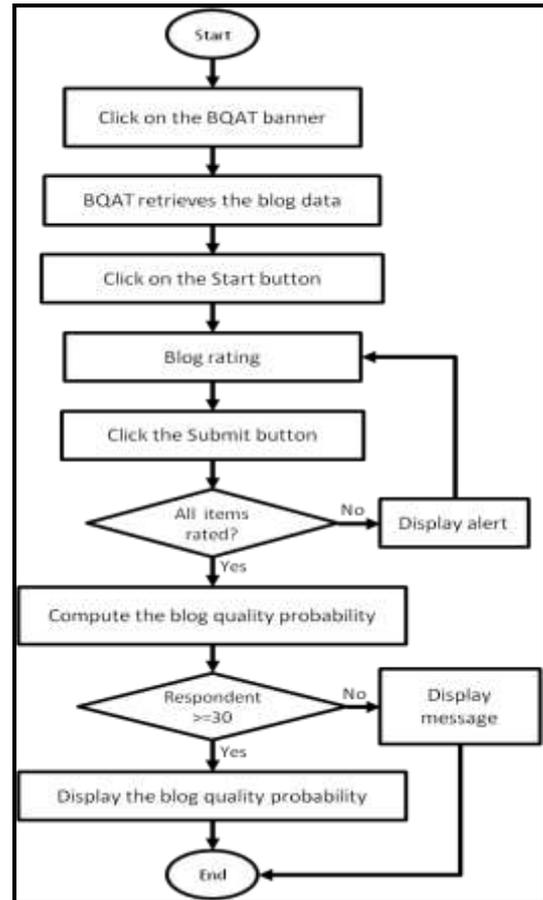


Figure 2.  Flow chart

The process starts with the blog-reader clicking on the BQAT banner in a dummy blog (see Figure 3). This will submit the blog ID parameter to the BQAT system.

Figure 3.  Dummy blog

Using the blog ID, BQAT retrieves the following blog data: blog name, URL, and blog type from the BQAT database. BQAT then displays the data on the introductory page (see Figure 4). Next, the blog-reader clicks the Start button on this page to commence the blog assessment.



Figure 4.  Introductory Page of the Blog Quality Assessment Tool
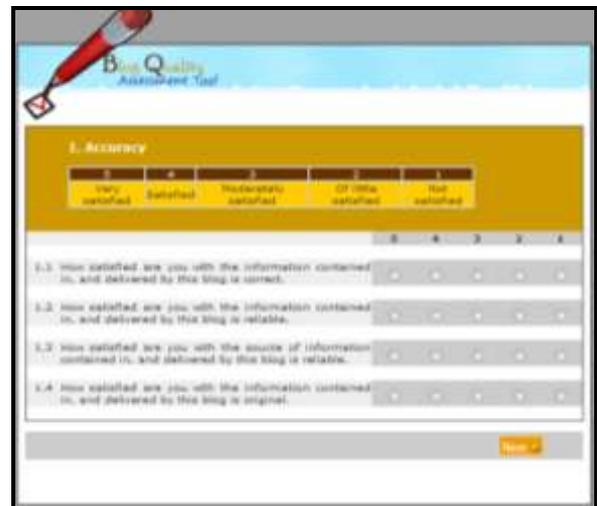


Figure 5.  Accuracy Page

The blog rating consists of eleven pages (see Figure 5 – 15), each representing one of the 11 families of blog quality criteria, respectively. Blog-readers rate the blog by stating their level of satisfaction for the respective criteria in each family on a 5-point Likert scale (1: Not satisfied to 5: Very satisfied) represented by radio buttons. Each page is linked to its following page by a Next button. When a Next button is clicked, all fields in the respective page are verified to confirm whether they have been filled.

Figure 6.  Completeness/Comprehensiveness Page



Figure 7.  Currency Page



Figure 8.  Engaging Page



Figure 9.  Reputation Page



Figure 10.  Info Representation Page

Figure 11. Navigation Page



Figure 12. Visual Design Page



Figure 13. Readability Page
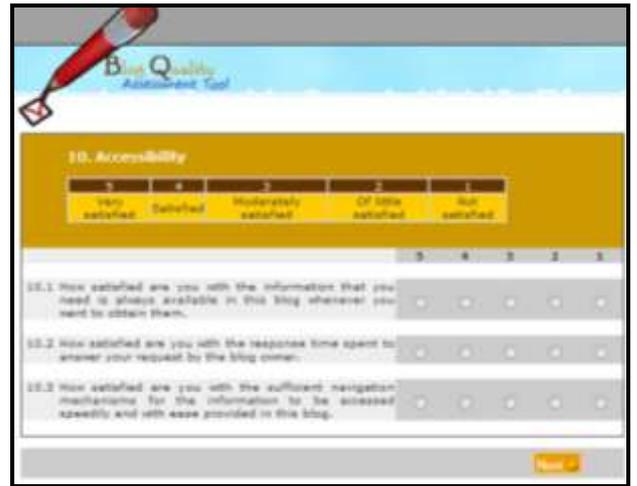


Figure 14. Accessibility Page
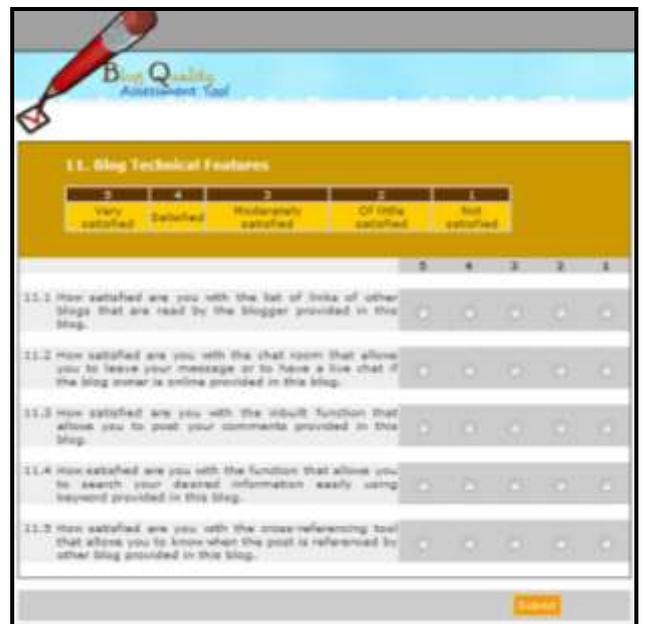


Figure 15. Blog Technical Features Page

If the blog-reader does not rate a particular criterion, an error message will pop-up (see Figure 16). After completing the blog rating, the blog-reader clicks on the Submit button in the Blog Technical Features page (see Figure 15).
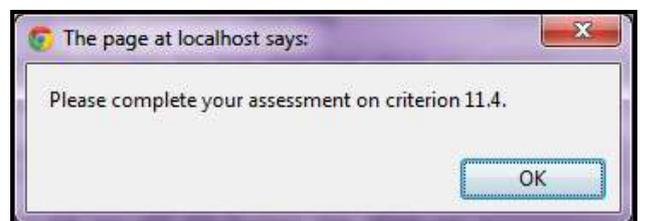


Figure 16. Error message

The rating of the criteria uses Likert's scale to produce ordinal data. Hence, upon submission, the Rasch Model was applied to convert the ordinal data into interval data and then used to estimate the probability of the blog to be of good quality. The system verifies whether the number of respondents is adequate to provide a meaningful result. If the number of

respondents is equal to or greater than 30, then the result will be displayed as follows (see Figure 17).
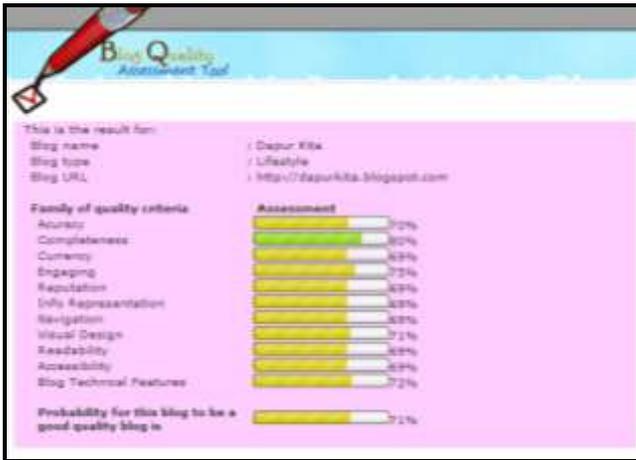


Figure 17.  Result Page

On the contrary, in case of less than 30 respondents, the result is displayed as shown in Figure 18.
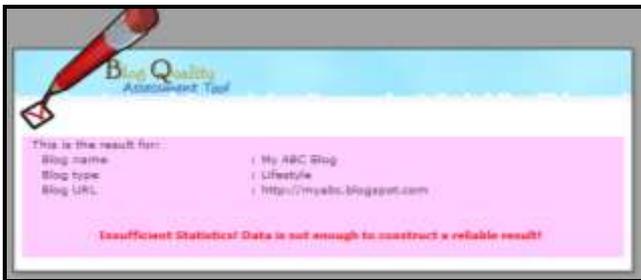


Figure 18.  Insufficient Statistics

There were six steps involved in estimating the probability of a blog to be of good quality based on the Rasch Rating Scale Model [43]. These were as follows:

**Step 1: Estimate the level of satisfaction for item *i***

The raw scores were converted into odds of success by calculating the ratio of the number of people who answered the item on any scale ($x$) to the number of people who did not answer on that scale ($n – x$). For example, if the total number of respondents ($n$) is 30, a raw score ($x$) of 7 on the Very Satisfied (5) scale for item 1 (see Table 3) is divided by the number of people who did not answer Very Satisfied for item 1 ($n – x$), that is, 23, to obtain the ratio 7/23 (see Table 4). The natural logs of these odds (e.g., $\log_{10} 7/23 = -0.52$) are shown in Table 5.

**Table 3. Example of Raw Scores**

| Dimension 1 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|
| Item 1 | 7 | 13 | 5 | 3 | 2 |
| Item 2 | 2 | 18 | 5 | 3 | 2 |
| Item 3 | 3 | 12 | 10 | 3 | 2 |
| Item 4 | 3 | 11 | 11 | 3 | 2 |

Note: 5 – Very Satisfied, 4 – Satisfied, 3 – Moderately Satisfied, 2 – Of Little Satisfied, 1 – Not Satisfied

**Table 4. Example of Ratio (*x/n-x*) for Each Item on Each Scale**

| Dimension 1 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|
| Item 1 | 7/23 | 13/17 | 5/25 | 3/27 | 2/28 |
| Item 2 | 2/28 | 18/12 | 5/25 | 3/27 | 2/28 |
| Item 3 | 3/27 | 12/18 | 10/20 | 3/27 | 2/28 |
| Item 4 | 3/27 | 11/19 | 11/19 | 3/27 | 2/28 |

Note: 5 – Very Satisfied, 4 – Satisfied, 3 – Moderately Satisfied, 2 – Of Little Satisfied, 1 – Not Satisfied

**Step 2: Calculate Item Mean for Dimension 1**

The Item Mean for Dimension 1 was calculated by aggregating the total of the natural logs of the odds for all items, divided by $n$ (30), which gives an Item Mean of -0.48 (see Table 5).

**Table 5. Example of Natural Logs of the Odds ($\log_{10}$ *x/n-x*)**

| Dimension 1 | 5 | 4 | 3 | 2 | 1 | Total |
|---|---|---|---|---|---|---|
| Item 1 | -0.52 | -0.12 | -0.70 | -0.96 | -1.15 | -3.43 |
| Item 2 | -1.15 | 0.18 | -0.70 | -0.96 | -1.15 | -3.77 |
| Item 3 | -0.95 | -0.18 | -0.30 | -0.96 | -1.15 | -3.53 |
| Item 4 | -0.95 | -0.24 | -0.24 | -0.96 | -1.15 | -3.53 |

Note: 5 – Very Satisfied, 4 – Satisfied, 3 – Moderately Satisfied, 2 – Of Little Satisfied, 1 – Not Satisfied

**Step 3: Estimate the Person Ability to Satisfy**

In estimating the Person Ability to Satisfy for person $i$, the raw scores were converted into odds of success, by calculating the ratio of the number of correct items on any scale ($y$) to the number of incorrect items on that scale ($m – y$). For example, if the total number of items ($m$) is 4, the raw data for each item answered by 30 persons are shown in Table 6. The number of correct items answered by Person1 on the Very Satisfied scale is 1, while the number of incorrect items answered on the scale is 3. So, the ratio is 1/3 (see Table 7). If the number of correct items answered by any person on any scale is 0, then the ratio ($y/m-y$) is equal to 0. If the number of incorrect items answered by any person on any scale is 0, then the ratio ($y/m-y$) is equal to the number of correct items. During the development of the BQAT prototype, the Person Ability to Satisfy was only estimated once with 49 items in order to produce an effective estimation.

**Table 6. Example of Raw Data for 4 Items by 30 Persons**

| Person | Item 1 | Item 2 | Item 3 | Item 4 |
|---|---|---|---|---|
| Person1 | 5 | 4 | 3 | 3 |
| Person2 | 5 | 4 | 4 | 4 |
| Person3 | 4 | 3 | 4 | 4 |
| Person4 | 3 | 4 | 4 | 4 |
| Person5 | 5 | 4 | 4 | 4 |
| Person6 | 5 | 5 | 5 | 5 |
| Person7 | 4 | 4 | 4 | 4 |

| Person | Item 1 | Item 2 | Item 3 | Item 4 |
|---|---|---|---|---|
| Person8 | 4 | 3 | 3 | 3 |
| Person9 | 5 | 4 | 3 | 3 |
| Person10 | 5 | 5 | 5 | 5 |
| Person11 | 4 | 4 | 4 | 4 |
| Person12 | 2 | 2 | 2 | 2 |
| Person13 | 2 | 2 | 2 | 2 |
| Person14 | 4 | 4 | 4 | 4 |
| Person15 | 4 | 4 | 4 | 4 |
| Person16 | 3 | 3 | 3 | 3 |
| Person17 | 3 | 3 | 3 | 3 |
| Person18 | 5 | 4 | 4 | 4 |
| Person19 | 4 | 4 | 4 | 4 |
| Person20 | 3 | 4 | 4 | 3 |
| Person21 | 4 | 4 | 4 | 4 |
| Person22 | 4 | 4 | 3 | 3 |
| Person23 | 3 | 3 | 3 | 3 |
| Person24 | 4 | 4 | 3 | 3 |
| Person25 | 2 | 2 | 2 | 2 |
| Person26 | 4 | 4 | 3 | 3 |
| Person27 | 1 | 1 | 1 | 1 |
| Person28 | 4 | 4 | 3 | 3 |
| Person29 | 4 | 4 | 5 | 5 |
| Person30 | 1 | 1 | 1 | 1 |

Note: 5 – Very Satisfied, 4 – Satisfied, 3 – Moderately Satisfied, 2 – Of Little Satisfied, 1 – Not Satisfied

**Table 7. Example of Ratio (*y/m-y*) for Each Person on Each Scale**

| Person | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Person1 | 0 | 0 | 2/2 | 1/3 | 1/3 |
| Person2 | 0 | 0 | 0 | 3/1 | 1/3 |
| Person3 | 0 | 0 | 1/3 | 3/1 | 0 |
| Person4 | 0 | 0 | 1/3 | 3/1 | 0 |
| Person5 | 0 | 0 | 0 | 3/1 | 1/3 |
| Person6 | 0 | 0 | 0 | 0 | 4 |
| Person7 | 0 | 0 | 0 | 4 | 0 |
| Person8 | 0 | 0 | 3/1 | 1/3 | 0 |
| Person9 | 0 | 0 | 2/2 | 1/3 | 1/3 |
| Person10 | 0 | 0 | 0 | 0 | 4 |
| Person11 | 0 | 0 | 0 | 4 | 0 |
| Person12 | 0 | 4 | 0 | 0 | 0 |

| Person | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Person13 | 0 | 4 | 0 | 0 | 0 |
| Person14 | 0 | 0 | 0 | 4 | 0 |
| Person15 | 0 | 0 | 0 | 4 | 0 |
| Person16 | 0 | 0 | 4 | 0 | 0 |
| Person17 | 0 | 0 | 4 | 0 | 0 |
| Person18 | 0 | 0 | 0 | 3/1 | 1/3 |
| Person19 | 0 | 0 | 0 | 4 | 0 |
| Person20 | 0 | 0 | 2/2 | 2/2 | 0 |
| Person21 | 0 | 0 | 0 | 4 | 0 |
| Person22 | 0 | 0 | 2/2 | 2/2 | 0 |
| Person23 | 0 | 0 | 4 | 0 | 0 |
| Person24 | 0 | 0 | 2/2 | 2/2 | 0 |
| Person25 | 0 | 4 | 0 | 0 | 0 |
| Person26 | 0 | 0 | 2/2 | 2/2 | 0 |
| Person27 | 4 | 0 | 0 | 0 | 0 |
| Person28 | 0 | 0 | 2/2 | 2/2 | 0 |
| Person29 | 0 | 0 | 0 | 2/2 | 2/2 |
| Person30 | 4 | 0 | 0 | 0 | 0 |

Note: 5 – Very Satisfied, 4 – Satisfied, 3 – Moderately Satisfied, 2 – Of Little Satisfied, 1 – Not Satisfied

**Step 4: Calculate the Person Mean**
The natural logs of these odds (excluding 0) were calculated and aggregated to obtain an estimate of the Person Ability to Satisfy for each person. The total of Person Ability to Satisfy was calculated by summing up the Person Ability to Satisfy for each person. Its value was then divided by the number of items to get the Person Mean. Based on the above example, the Person Mean is 1.93 logits. Similar to Step 3, the Person Mean was also calculated once.

**Step 5: Compute the Probability of Dimension 1 to be Satisfied**
The probability of Dimension 1 to be satisfied was computed by substituting $B_n$, $D_i$, and $F_k$ in formula 3.5 (see Chapter 3) with the Person Mean, Item Mean, and 0 respectively. We set the threshold, $F_k$, equal to 0 because it is calculated as a dichotomous 50/50 point [44].

$$P_{nik} = \frac{e^{(B_n - D_i - F_k)}}{1 + e^{(B_n - D_i - F_k)}}$$

$$= \frac{e^{1.93 + 0.48 - 0}}{1 + e^{1.93 + 0.48 - 0}}$$

$$= 0.9176$$

The percentage of the probability of Dimension 1 for the above examples is 92%. Given we have 11 families or

dimensions in this study, Step 1, 2, and 5 were repeated for all 11 families.

**Step 6: Estimate the Probability of the Blog to be of Good Quality**

Finally, the probability of the blog to be of good quality was estimated by aggregating the products of the assigned weights and the probability for each family to be satisfied. For the dummy blog, in this case a Personal Diary blog, the assigned weights were derived from the same blog category. Table 6.8 shows the probability of the family to be satisfied ($P(\theta)_i$), the assigned weights ($w_i$), and the product of the assigned weights and the probability of each family to be satisfied ($P(\theta)_i \times w_i$), for the dummy blog.

**Table 8. The Probability of the Family to be Satisfied ($P(\theta)_i$), Weights ($w_i$), and the Product of the Weights Assigned and the Probability of Each Family to be Satisfied ($P(\theta)_i \times w_i$) for the Dummy Blog**

| Family | Probability, $P(\theta)_i$ | Weight, $w_i$ | $P(\theta)_i \times w_i$ |
|---|---|---|---|
| Accuracy | 70 | 0.0897 | 6.2790 |
| Completeness | 80 | 0.0874 | 6.9920 |
| Currency | 69 | 0.0915 | 6.3135 |
| Engaging | 75 | 0.0851 | 6.3825 |
| Reputation | 69 | 0.0786 | 5.4234 |
| Info Representation | 69 | 0.0999 | 6.8931 |
| Navigation | 69 | 0.0910 | 6.2790 |
| Visual Design | 71 | 0.1013 | 7.1923 |
| Readability | 69 | 0.0964 | 6.6516 |
| Accessibility | 69 | 0.0940 | 6.4860 |
| Blog Technical Features | 72 | 0.0851 | 6.1272 |
| $\sum P(\theta)_i \times w_i$ | | | 71 |

# 4. FINDINGS AND RESULTS OF THE TECHNOLOGY ACCEPTANCE TEST

The Technology Acceptance Test was conducted to gauge the acceptance of the Blog Quality Assessment Tool. This section is divided into two sub-sections; the fit statistics of the Technology Acceptance Test, and the results of the test.

## 4.1 Fit Statistics of the Technology Acceptance Test

The summary statistics for the analysis of the sample of 35 blog-readers on the 9 polytomous scale items comprising the Technology Acceptance Test items are shown in Figure 19. The summary fit statistics for Items and Persons show satisfactory fit to the model. The mean square fit (IMNSQ and OMNSQ) statistics and the $z$ statistics (Infit and Outfit ZSTD) for Items and Persons are close to their expected values, +1 and 0, respectively.

```
+-----------------------------------------------------------------------+
| Persons     35 INPUT     35 MEASURED        INFIT       OUTFIT    |
|        SCORE     COUNT     MEASURE    ERROR   IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN     36.1       9.0       6.08     .91    1.00     .1    .88     .1|
| S.D.      7.1        .0       5.56     .19     .28     .7    .32     .5|
| REAL RMSE   .93 ADJ.SD    5.48  SEPARATION  5.90  Person RELIABILITY  .97|
|-----------------------------------------------------------------------|
| Items      9 INPUT      9 MEASURED          INFIT       OUTFIT    |
| MEAN    140.4      35.0        .00     .44    1.00     .0    .88    -.1|
| S.D.      6.3        .0       1.26     .08     .17     .6    .29     .6|
| REAL RMSE   .45 ADJ.SD    1.18  SEPARATION  2.62  Item   RELIABILITY  .87|
+-----------------------------------------------------------------------+
```

Figure 19. Summary Statistics of Technology Acceptance Test

The Wright map in Figure 20 demonstrates the distribution of blog-readers on the left, represented by r01-r30, and the distribution of item agreement on the right, represented by item ID (refer Table 9). The most easily endorsed item is PEU2 (*I find that the Blog Quality Assessment Tool is easy to use*) located at -2.93 logits (SE .62), while the item that is most difficult to endorse is A2 (*My attitude toward using the Blog Quality Assessment Tool is very favourable*) located at the top of the Item distribution at +1.73 logits (SE .38). The Person distribution confirms the result from the summary statistics. The easiest to endorse blog-readers are r05, r17, r28, r29, and r30 located at +12.53 logits (SE 1.10), while the most difficult to endorse blog-reader is r03 located at the bottom of the Person distribution at -6.63 logits (SE .68). The mean of the Person distribution is higher than the mean of the Item distribution. This indicates that majority of the blog-readers involved in the Technology Acceptance Test have the tendency to agree with most of the items.

```
                    Persons -MAP- Items
              <easy to endorse>|<difficult to be endorsed>
     13                        +
        r05  r17  r28  r29  r30 |
     12                        +
                   r04  r14  r26 S|
     11                         r24  +
                   r07  r27  r35 |
     10             r15  r21  r25 +
                               |
      9                  r01  r02 +
                         r34    |
      8                        +
                         r06   |
      7                        +
                               |
      6  Person Mean = +6.08   M+
                               |
      5                        +
                               |
      4             r09  r16  r33 +
                               |
      3             r11  r20  r22 +
                               |T
      2             r08  r10  r19 +
                               |S                  A2
      1                  r18  +                        ITU1
                         S|  PU1
                             PU2
      0  Item Mean = 0     +M PU3      PEU1
                               |                A1      ITU2
     -1             r12  +
                               |S
     -2        r13  r31  +
                               |T
     -3             r23  +          PEU2
                   r32    |
     -4                        +
                               |
     -5                        T+
                               |
     -6                        +
                   r03    |
     -7                        +
              <difficult to endorse>|<easily to be endorsed>
```
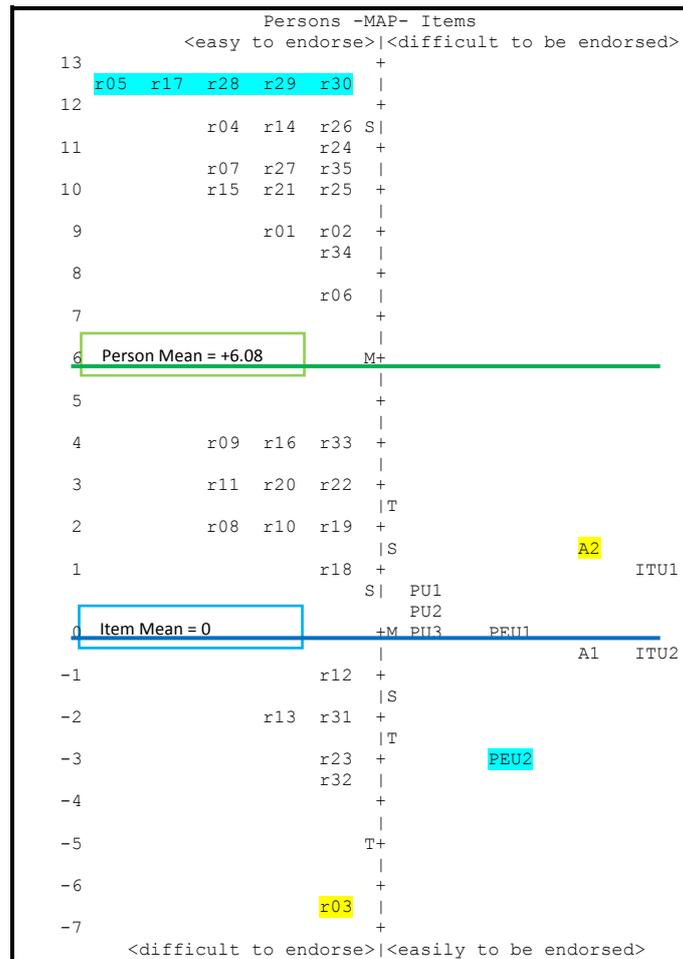
Figure 20. Wright Map of the Technology Acceptance Test

Figure 21 shows the Item statistics in Measure order. The Rasch fit statistics disclose that item A1 behaved more erratically than expected with an Infit MNSQ value > 1.4. However, after confirming that the Infit Z-Std is within the range, it is accepted in this analysis. Other items fit sufficiently to the model, with their Infit and Outfit Mean-square values and Infit and Outfit Z-std values all lying within the acceptable range.

```
+-------------------------------------------------------------+
|ENTRY      RAW              MODEL|  INFIT  |  OUTFIT  |      |
|NUMBER   SCORE   MEASURE    S.E. |MNSQ  ZSTD|MNSQ  ZSTD| Item |
|-----------------------------------+---------+---------+-----|
|    7     130      1.73     .38|1.04    .3| .93   -.1| A2   |
|    8     134      1.15     .38| .90   -.5| .82   -.6| ITU1 |
|    1     138       .55     .39|1.02    .2|1.08    .4| PU1  |
|    2     139       .39     .40| .84   -.8| .70  -1.0| PU2  |
|    3     140       .23     .40| .98    .0|1.15    .6| PU3  |
|    4     140       .23     .40|1.03    .2|1.04    .2| PEU1 |
|    6     145      -.67     .45|1.43   1.5|1.29    .7| A1   |
|    9     145      -.67     .45| .89   -.3| .58   -.9| ITU2 |
|    5     153     -2.93     .62| .86   -.2| .31   -.3| PEU2 |
|-----------------------------------+---------+---------+-----|
+-------------------------------------------------------------+
```

Figure 21. Item Measure of the Technology Acceptance Test
Note: Acceptable range for Infit and Outfit Mean-square is between 0.6 to 1.4 [45] and acceptable range for Infit and Outfit Z-std is between -2 to +2 [44]

The Rasch fit statistics are further inspected by examining the Person statistics. Figure 22 displays the Person statistics in Measure order. There are two possible under-fitting persons;

r34, and r06 having Infit MNSQ values > 1.4. Yet, they are kept in the analysis as their Infit Z-std, Outfit MNSQ, and Outfit Z-Std values are within bounds.

```
+-------------------------------------------------------+
|  RAW           MODEL|  INFIT  |  OUTFIT  |         |
| SCORE  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD| Person|
|---------------------+----------+----------+------|
|   44   12.53   1.10|1.14    .4| .92    .6| r05  |
|   44   12.53   1.10|1.21    .5|1.23    .8| r17  |
|   44   12.53   1.10|1.21    .5|1.23    .8| r28  |
|   44   12.53   1.10|1.21    .5|1.23    .8| r29  |
|   44   12.53   1.10| .68   -.2| .34    .2| r30  |
|   43   11.61    .86|1.00    .2| .73    .2| r04  |
|   43   11.61    .86| .60   -.9| .43   -.2| r14  |
|   43   11.61    .86|1.36    .9|1.17    .5| r26  |
|   42   10.96    .77|1.14    .5|1.15    .5| r24  |
|   41   10.38    .75|1.12    .5|1.02    .2| r07  |
|   41   10.38    .75|1.06    .3|1.02    .2| r27  |
|   41   10.38    .75|1.17    .7|1.06    .3| r35  |
|   40    9.82    .76| .94   -.1| .79   -.2| r15  |
|   40    9.82    .76| .89   -.3| .75   -.3| r21  |
|   40    9.82    .76| .68  -1.1| .59   -.7| r25  |
|   39    9.21    .81|1.17    .5|1.04    .3| r01  |
|   39    9.21    .81| .55  -1.2| .43  -1.0| r02  |
|   38    8.45    .94|1.51    .9|1.05    .3| r34  |
|   37    7.35   1.19|1.58    .9| .98    .5| r06  |
|   35    3.76   1.05| .61   -.4| .32   -.5| r09  |
|   35    3.76   1.05|1.05    .3| .88    .2| r16  |
|   35    3.76   1.05| .86    .1| .53   -.2| r33  |
|   34    2.89    .85|1.00    .2|1.37    .7| r11  |
|   34    2.89    .85| .58   -.9| .41   -.6| r20  |
|   34    2.89    .85| .58   -.9| .41   -.6| r22  |
|   33    2.25    .76| .61  -1.2| .51   -.6| r08  |
|   33    2.25    .76| .87   -.3| .72   -.2| r10  |
|   33    2.25    .76| .65  -1.0| .54   -.5| r19  |
|   31    1.17    .73|1.36   1.0|1.25    .6| r18  |
|   27   -1.20    .77|1.02    .2| .95    .1| r12  |
|   26   -1.77    .74| .78   -.3| .73   -.3| r13  |
|   26   -1.77    .74|1.16    .5|1.16    .5| r31  |
|   23   -3.23    .68|1.15    .5|1.16    .5| r23  |
|   22   -3.69    .68|1.34    .9|1.35    .8| r32  |
|   16   -6.63    .68|1.23    .7|1.28    .7| r03  |
|---------------------+----------+----------+------|
```

Figure 22. Person Measure of the Technology Acceptance Test
Note: Acceptable range for Infit and Outfit Mean-square is between 0.6 to 1.4 [45] and acceptable range for Infit and Outfit Z-std is between -2 to +2 [44]

The principal contrast analysis of the Rasch residual variance is shown in Figure 23. The *variance explained by measures* is significantly good (91.1%). The uni-dimensionality of the technology acceptance test instrument is strongly confirmed by having a good *unexplained variance in the first contrast* (2.0%).

```
          STANDARDIZED RESIDUAL VARIANCE SCREE PLOT
      Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                        Empirical        Modeled
Total variance in observations    =    101.1 100.0%       100.0%
Variance explained by measures    =     92.1  91.1%        89.9%
Unexplained variance (total)      =      9.0   8.9% 100.0%  10.1%
Unexplned variance in 1st contrast =     2.1   2.0%  22.8%
Unexplned variance in 2nd contrast =     1.7   1.7%  18.7%
Unexplned variance in 3rd contrast =     1.4   1.4%  15.4%
Unexplned variance in 4th contrast =     1.1   1.1%  12.2%
Unexplned variance in 5th contrast =      .9    .9%  10.4%
```

Figure 23. Principal Contrast Analysis of the Technology Acceptance Test
Note: Variance explained by measures should be ≥ 50% and unexplained variance in the first contrast should be ≤ 15% [46]

Figure 24 depicts the category probability curves for all items. This corroborates that the 4 thresholds are in order and the probability curves for all categories are not flat. It shows that our 5-point rating scale developed for this questionnaire yields the highest quality measures for the construct of interest.

```
  CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P    ++-------+-------+-------+-------+-------+-------+-------++
R  1.0 +                                4                      +
O     |                            4444 4444                  |
B     |                        44            44            5|
A     |                  333        4            4        5 |
B   .8 +1     222        3   3       4          4       5   +
I     |1    2   2     3    3     4                4    5    |
L     | 1   2    2    3      3    4              4    5     |
I     | 1   2        2  3         3   4              4   5   |
T   .6 +  1 2        2 3          3 4             4   5     +
Y     |   12           23           34                 45    |
    .5 +   *         *          *                 *    +
O     |    21          32          43               54     |
F   .4 +   2 1         3 2         4 3             5    4    +
      |    2 1          3  2         4  3            5   4    |
R     |   2  1         3    2       4    3           5     4   |
E     |2      1        3      2      4      3         5      4   |
S   .2 +2       1   3         2    4        3         5        4 |
P     |       1 3          2  4          3           5      4 |
O     |        *         244           33          55        4|
N     |       333 111      44222           333    555        |
S   .0 +*************************************************+
E     ++-------+-------+-------+-------+-------+-------+-------++
      -9      -6      -3       0       3       6       9      12
         Person [MINUS] Item MEASURE
```
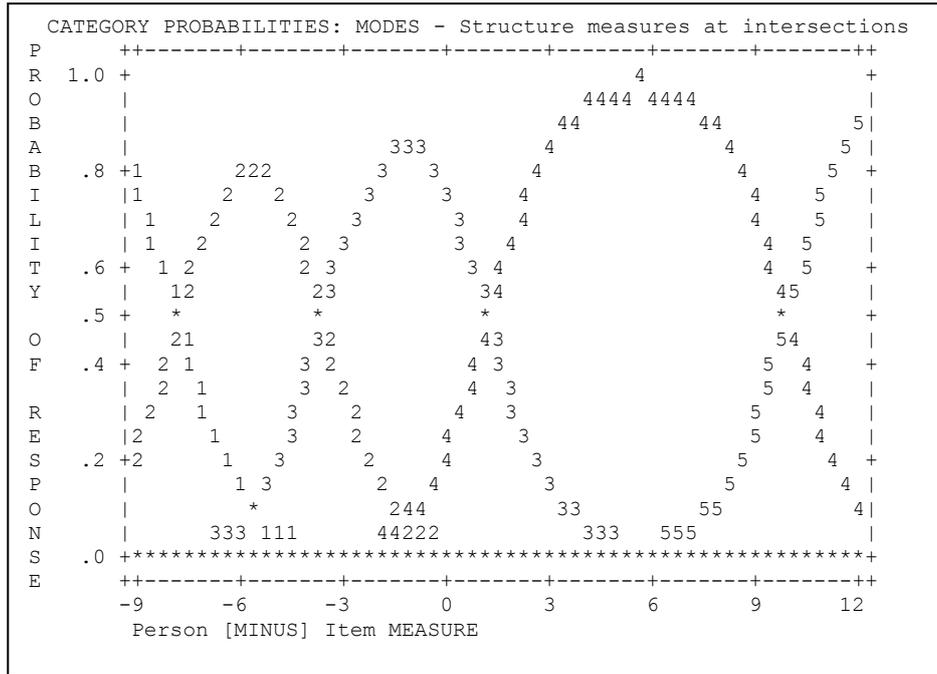
Figure 24. Category Probabilities of the Technology Acceptance Test

## 4.2 Results of the Technology Acceptance Test

After all the data were confirmed to fit to the Rasch model, the probability of each item to be endorsed by blog-readers, on average, was calculated and the results are presented in Table 9. The results show that blog-readers perceived ease of use and usefulness as significant features of the BQAT system. This implies that the BQAT system is easy, effective, and useful to help blog-readers make a good quality assessment. Blog-readers also displayed a significantly positive attitude towards using the tool and intend to use it. Most importantly, blog-readers significantly agree that bloggers should participate in the blog quality assessment project.

**Table 9. The Probability of Factors and Items to be Endorsed**

| Factor | P(Θ) (%) | Item ID and Description | P(Θ) (%) |
|---|---|---|---|
| PEU *Perceived Ease of Use* | 99.85 | PEU1 *Learning to use the Blog Quality Assessment Tool is easy for me.* | 99.71 |
| | | PEU2 *I find that the Blog Quality Assessment Tool is easy to use.* | 99.99 |
| PU *Perceived Usefulness* | 99.66 | PU1 *The Blog Quality Assessment Tool enables me to complete assessing the blog quality quickly.* | 99.60 |
| | | PU2 *Using the Blog Quality Assessment Tool will help me improve my ability to make a good blog quality assessment.* | 99.66 |
| | | PU3 *The Blog Quality Assessment Tool makes the blog quality assessment task more effective.* | 99.71 |
| A *Attitude toward Using* | 99.31 | A1 *Using the Blog Quality Assessment Tool for assessing the quality of my favourite blog is a good idea.* | 99.88 |
| | | A2 *My attitude toward using the Blog Quality Assessment Tool is very favourable.* | 98.73 |
| ITU *Intention to Use* | 99.58 | ITU1 *I intend to use the Blog Quality Assessment Tool when it becomes available on my favourite blog.* | 99.28 |
| | | ITU2 *I think that bloggers should participate in this blog quality assessment project.* | 99.88 |

## 5. CONCLUSION AND FUTURE WORKS

A prototype of Blog Quality Assessment Tool (BQAT) was successfully developed. The main functions of the BQAT are to calculate the probability of a blog to be of good quality, and to accumulate the results for the assessed blog. Thus, blog-readers can easily obtain information on the quality of the blogs visited. This assessment tool can also be used to manage and control a blog's expansion, such that only high quality blogs continue to exist in the blogosphere. The Technology Acceptance Test was conducted to investigate whether or not our prototype of the Blog Quality Assessment Tool was accepted by blog-readers. This study explored the impact of perceived usefulness, perceived ease of use, attitude, and intention to use the system on blog-users' acceptability. Results indicate that blog-readers significantly agree that the Blog Quality Assessment Tool is easy, effective, and useful to them in assessing blog quality. This study also shows that the more satisfied the blog is, the higher its quality.

In future, we will invite bloggers and blog readers to participate assessing actual blogs in different blog categories, by using the Blog Quality Assessment Tool.

## 6. REFERENCES

[1] Lenssen, P. (2006). Good blog writing style. Retrieved January 2, 2009 from http://blogoscoped.com/archive/2006-10-11-n47.html.

[2] Merlin, M. (2008). What makes for a good blog. Retrieved January 2, 2009 from http://www.43folders.com/2008/08/19/good-blogs.

[3] Rowse, D., & Garrett, C. (2008). ProBlogger: Secrets for Blogging Your Way to a Six-Figure Income. Indianapolis, Indiana: Wiley Publishing Inc.

[4] Strauss, L. (2005). Blog design checklist. Retrieved January 3, 2009 from http://www.successful-blog.com/1/blog-design-checklist/

[5] Zain, Z. M., Ghani, A. A. A., Abdullah, R., Atan, R., & Yaakob, R. (2011). Application of Rasch Model in validating the content of measurement instrument for blog quality. In J. M. Zain, W. M. W. Mohd, & E. El-Qawasmeh (Eds.). Software Engineering and Computer Systems (Vol. 180): Springer Publishing Co., Heidelberg/Berlin, 125–136.

[6] Zain, Z. M., Ghani, A. A. A., Abdullah, R., Atan, R., & Yaakob, R. (2012). Blog quality measurement: analysis of criteria using the Rasch Model. Int. J. New Comput. Archit. Their Appl, 1(1), 665–682.

[7] Zain, Z. M., Ghani, A. A. A., Abdullah, R., Atan, R., & Yaakob, R. (2013). Blog quality model. Int. J. Web-based Communities, 9(1), 25–50.

[8] Zhang, P., von Dran, G., Blake, P., & Pipithsuksunt, V. (2001). Important Design Features in Different Web Site Domains. e-Service Journal, 1(1), 77-91.

[9] Katerattanakul, P., & Siau, K. (2001). Information quality in internet commerce design. In M. Piattini, C. Calero, & M. Genero (Eds.), Information and Database Quality: Kluwer Academic Publishers, Norwell, MA, 45-46.

[10] Naumann, F., & Rolker, C. (2000). Assessment methods for information quality criteria. In Proceedings of the 5th International Conference on Information Quality (IQ2000). MIT, Cambridge, MA, (pp. 148-162).

[11] Graefe, G. (2003). Incredible Information on the Internet: Biased information provision and a lack of credibility as a cause of insufficient information quality. In Proceedings of the 8th International Conference on Information Quality (IQ2003). MIT, Cambridge, MA, (pp. 133-146).

[12] Melkas, H. (2004). Analyzing information quality in virtual service networks with qualitative interview data. In Proceedings of the 9th International Conference on Information Quality (IQ2004). MIT, Cambridge, MA, (pp. 74-88).

[13] Eppler, M. (2001). A generic framework for information quality in knowledge-intensive processes. In Proceedings of the 6th International Conference on Information Quality (IQ2001). Cambridge, MA, (pp. 329-346).

[14] Katerattanakul, P., & Siau, K. (1999). Measuring Information Quality of Web Sites: Development of an Instrument. In Proceedings of the 20th International Conference on Information System (1999). Association for Information Systems, Atlanta, GA, (pp. 279-285).

[15] Caro, A., Calero, C., Caballero, I., & Piattini, M. (2008). A proposal for a set of attributes relevant for Web portal data quality. Software Quality J, 16(4), 513-542.

[16] Banks, M. A. (2008). Blogging Heroes: Interviews with 30 of the World's Top Bloggers. Indianapolis, Indiana: Wiley Publishing Inc.

[17] Tan, J.-E., & Ibrahim, Z. (2008). Blogging and Democratization in Malaysia. A New Civil Society in the Making. Petaling Jaya: SIRD.

[18] Hopkins, J. (2009). The "ideal type" blog. Retrieved June 12, 2009 from http://jualianhopkins.net/index.php?/archives/240-The-ideal-type-blog.html

[19] Zarizi, S. S. (2006). 70 weblog paling popular di Malaysia. Retrieved January 2, 2009 from http://syedsyahrul.blogspot.com/2006/05/70-weblog-paling-popular-di-malaysia_28.html

[20] Zain, Z. M., & Abdul Ghani, A. A. (2014). Importance Analysis of a Blog Quality Model for Criteria and Families in Different Blog Categories. International Journal of Virtual Communities and Social Networking, 6(3), 17-56.

[21] McCall, J. A., Richards, P. K., & Walters, G. F. (1977). Factors in Software Quality. New York: General Electric Company.

[22] Boehm, B. (1978). Characteristics of Software Quality. New York: North Holland Publishing Co.

[23] Dromey, R. G. (1995). A Model for Software Product Quality. IEEE Transactions on Software Engineering, 21(2), 146-162.

[24] ISO 9126. (1991). Software product evaluation - Quality characteristics and guidelines for their use. International Organization for Standardization. Retrieved May 18, 2009 from http:www.iso.org

[25] Olsina, L., & Rossi, G. (2002). Measuring Web Application Quality with WebQEM. IEEE Multimedia, 9(4), 20-29.

[26] Calero, C., Ruiz, J., & Piattini, M. (2005). Classifying Web Metrics using The Web Quality Model. Online Information Review, 29(3), 227-248.

[27] Malak, G., Badri, L., Badri, M., & Sahraoui, H. (2004). A Quality Model for Web-based Applications. In Proceedings of the 5th International Conference on Electronic Commerce and Web Technologies (ECWeb2004). Springer, Zaragosa, Spain, (pp. 316-327).

[28] Olsina, L. (1999). Web-site quality evaluation method: a case study on museums. In Proceedings of the 2nd International Conference on Software Engineering Workshop on Web Engineering (ICSE1999). Los Angeles, CA.

[29] Olsina, L., Godoy, D., Lafuente, G., & Rossi, G. (1999). Assessing the quality of academic websites. New Rev. Hypermedia Multimed. J. 5(1), 81–103.

[30] Olsina, L., Lafuente, G., & Rossi, G. (2000). E-commerce site evaluation: a case study. In K. Bauknecht, S. Madria, & G. Pernul (Eds.), Electronic Commerce and Web Technologies (Vol. 1875): Springer Publishers, Heidelberg/Berlin, 239–252.

[31] Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. Information & Management, 40(2), 133-146.

[32] Strong, D., Lee, Y., & Wang, R. (1997). Data Quality in Context. Communications of the ACM, 40(5), 103-110.

[33] Winkler, W. E. (2004). Methods for evaluating and creating data quality. Information Systems Data Quality in Cooperative Information Systems, 29(7), 531-550.

[34] Bouzeghoub, M., & Kedad, Z. (2001). Quality in Data Warehousing. In M. Piattini, C. Calero, & M. Genero (Eds.), Information and Database Quality: Kluwer Academic Publishers.

[35] Zhu, Y., & Buchmann, A. (2002). Evaluating and Selecting Web Sources as external Information Resources of a Data Warehouse. In Proceedings of the 3rd International Conference on Web Information Systems Engineering (2002). Singapore (pp. 149-160).

[36] Aboelmeged, M. (2000). A Soft System Perspective on Information Quality in Electronic Commerce. In Proceedings of 5th Conference on Information Quality (IQ2000). MIT, Cambridge, MA, (pp. 318-319).

[37] Gertz, M., Özsu, M. T., Saake, G., & Sattler, K.-U. (2004). Report on the Dagstuhl Seminar "Data Quality on the Web". SIGMOD Rec., 33(1), 127-132.

[38] Sommerville, I. (2010). Software Engineering (9 ed.). Essex: Pearson Education Limited.

[39] Mich, L., Franch, M., & Gaio, L. (2003). Evaluating and Designing Web Site Quality. IEEE Computer Magazine, 10(1), 34-43.

[40] Olsina, L., Godoy, D., Lafuente, G. J., & Rossi, G. (2001). Specifying Quality Characteristics and Attributes for Websites. In S. Murugesan & Y. Deshpande (Eds.), Web Engineering, 2016 (pp. 266-278): Springer Berlin/Heidelberg.

[41] Signore, O. (2005). A Comprehensive Model for Web Sites Quality. Paper presented at the Seventh IEEE International Symposium on Web Site Evolution (WSE'05), Budapest, Hungary.

[42] Evans, J. R., & King, V. E. (1999). Business-to-business Marketing and the World Wide Web: Planning, Managing and Assesing Web Sites. Industrial Marketing Management, 28(4), 343-358.

[43] Andrich. D. (1978). A rating formulation for ordered response categories. Psychometrika, 43(4), 561–573.

[44] Bond, T. V., & Fox, C. M. (2007). Applying The Rasch Model: Fundamental Measurement in the Human Sciences (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

[45] Wright, B. D., Linacre, M., Gustafsson, J.-E., & Martin-Loff, P. (1994). Reasonable Mean-square Fit Values. Rasch Measurement Transactions, 8(3), 370.

[46] Fisher, W. P., Jr. (2007). Rating Scale Instrument Quality Criteria. Rasch Measurement Transactions, 21(1), 1095.

# Fingerprint Image Compression using Sparse Representation and Enhancement with Wiener2 Filter

Divya Joseph
Amal Jyothi College of Engineering
Kottayam, Kerala, India

Niya Joseph
Amal Jyothi College of Engineering
Kottayam, Kerala, India

**Abstract**: A technique for enhancing decompressed fingerprint image using Wiener2 filter is proposed. First compression is done by sparse representation. Compression of fingerprint is necessary for reducing the memory consumption and efficient transfer of fingerprint images. This is very essential for the application which includes access control and forensics. So the fingerprint image is compressed using sparse representation. In this technique, first dictionary is constructed for patches of fingerprint images. Then a fingerprint is selected and the coefficients are obtained and encoded. Thus the compressed fingerprint is obtained. But when the fingerprint is reconstructed, it is affected by noise. So Wiener2 filter is used to filter the noise in the image. The ridge and bifurcation count is extracted from decompressed and enhanced fingerprints. The experiment result shows that the enhanced fingerprint image preserves more bifurcation than decompressed fingerprint image. The future analysis can be considered for preserving ridges.

**Keywords**: Fingerprint Compression; Sparse Representation; Enhancement; Wiener2 Filter; Minutiae Extraction

## 1. INTRODUCTION

A fingerprint is defined as a mark left by the minutiae of a human finger. Collecting of fingerprints from a crime detection scene is an major method of forensic science. Fingerprint detection refers to the automatic way of finding a match between two human fingerprints. The Federal Bureau of Investigation (FBI) deal with a massive collection of fingerprint cards which contains huge number of cards and is growing at the rate of 30,000-50,000 cards per day [1]. The FBI is digitizing these cards for storage, renewal, and transference. Because of data storage needs and the time needed to transfer a fingerprint card over a modem, these files must be compressed. Fingerprint compression is gaining much attention with the recent widespread transmission of images.

Generally, compression techniques can be classified into lossless compression and lossy compression. Lossless compression allows the original images to be recovered from the compressed data. Lossless compression techniques are used when it is important that the authentic and the decompressed images are similar. Avoiding distortion limits the efficiency of image compression. In image compression slight distortion is acceptable. So lossless compression technologies are usually employed in the outputs of lossy compression. There are several compression techniques available for image compression. Compared with general images, the fingerprint images have simpler format. Fingerprint images only contain ridges and valleys. So compression algorithms are developed by targeting fingerprint images.

## 2. RELATED WORKS

The development of image compression began with several systems from years ago. But a standard image compression method was needed to enable interoperability of equipment from different manufacturer. So JPEG [2] evolved as the first international digital image compression standard used for continuous-tone images (gray scale or color). The JPEG compression has numerous points of interest, for example, effortlessness, comprehensiveness and accessibility. However,

performance is not good at lower bit-rates because of the underlying DCT scheme which is block-based. So in 1995, the JPEG-committee started to build up another wavelet-based compression standard for still pictures, to be specific JPEG 2000[3]. When compared to JPEG, JPEG 2000 has many features that support scalable and intuitive access to huge estimated picture.

The above algorithms are for compressing regular images. Focusing on fingerprint images, there are certain compression algorithms. Hopper proposed the most common fingerprint image compression algorithm called as Wavelet Scalar Quantization (WSQ) [4]. It is used by the FBI (Federal Bureau of Investigation) for the compression of 500 dpi fingerprint images. However due to the drawback of quantizer in this method, it cannot cope with fuzzy images. Thus an enhanced WSQ fingerprint image compression algorithm is made by Tang et al [5]. A generic fingerprint image compression technique depend on wave atoms decomposition was proposed by Mohammed et al [6] which is better than the standard method for FBI fingerprint image compression, the wavelet scalar quantization (WSQ). The proposed compression scheme is based upon linear vector quantization of dissolved wave atoms illustration of fingerprints. Then coefficients which are quantized are encoded with arithmetic entropy scheme. Application of wave atoms based decomposition for fingerprint image compression results into a significant improvement in peak signal to noise ratio (PSNR) compared to FBI's WSQ fingerprint compression standard. However, wave atoms cannot deal with the fact that the edges found in fingerprints are sharp curves. This problem is solved by transforms known as contourlets[7], preserves edges.

The contourlet transform is an extension of wavelet transform in 2D using directional filter banks. The requirement of storage in multistage vector quantization (MSVQ) is lesser when compared with full search vector quantization. Multistage vector quantization is used for quantizing coefficients of contourlet transform. Huffman coding is used to encode quantized coefficients. As a representation of directional multiresolution image, the contourlet transform

can surely grab curved and aligned geometrical formats in images. However, the contourlet transform has the shortcoming of a repetition in its oversampling ratio. Zhang and Moloney have found a nonredundant version of the contourlet transform, known as the nonredundant contourlet transform (NRCT) [8] and have demonstrated that this redundancy can be eliminated by the NRCT. With the advantages of sampling and reconstruction, the NRCT can be good enough for tracing and well coding of oriented structure in images, such as the structure of ridges in fingerprint images. Moreover, as an extension of the wavelet transform, the NRCT is easily adaptable with the wavelet transform. A new transform which associate the NRCT with the wavelet transform is called the hybrid-NRCT.

The main idea behind this paper is to compress the fingerprint based on sparse representation and later enhancement is done using Wiener2 filter. The sparse representation is efficient to preserve the minutiae contents which are required for fingerprint applications. However it is suffered by noise when decompressed and the accuracy is very low. In order to avoid that Wiener2 filter is used for removing noise from decompressed fingerprints.

The rest of our paper is organized as follows. The main principles of the techniques mentioned in the proposed method are introduced in the next section. Section 4 sums up the performance analysis of our scheme in terms of number of ridges and bifurcation count on different fingerprint images. Conclusions are provided in Section 5.

# 3. PROPOSED SYSTEM

The methods we propose consist of construction of dictionary, compression of fingerprint, its decompression and enhancement. Finally the minutiae are extracted from both the decompressed and enhanced fingerprint images.

## 3.1 Construction of Dictionary

First construction of training set is done. Then, from the training set the dictionary is obtained. For constructing the training set, select all the fingerprint images from the database and slash them into square patches of fixed size. When we get the patches after initial screening, the training samples can be constructed by applying a greedy algorithm.

- The dictionary is initially empty and the first patch is added to it.

- Then the similarity between this patch and the next patch is tested. If they are sufficiently similar, the other patch is tested; if not, the patch is appended into the dictionary. Here, the optimization problem is solved to calculate the similarity measure between two patches (eq.1).

  $$S(P1, P2) = \min \|P1/\|P1\|2 F - t*P1/\|P1\|2F \quad (1)$$

  where $\|\cdot\|2F$ is the Frobenius norm. P1 and P2 are the matrices that are correspondent of two patches. t, a parameter of the optimization problem (eq.1), is a scaling factor.

- Repeat the 2nd step and make sure that all patches have been tested.

Before constructing the dictionary, calculate the mean value of each patch and subtract from the corresponding patch. Construction of the dictionary is done by using a training method called K-SVD [9], [10]. An optimization problem is iteratively solved to obtain the dictionary (eq.2). Y is consisted of the training patches, A is the dictionary, X represents the coefficients and Xi is the ith column of X. MP method [11] is used to compute the coefficients matrix X in the sparse solving stage. It guarantees that the coefficient vector Xi has non-zero elements which is limited to T. Then, singular value decomposition (SVD) is used to update each element in the dictionary.

$$\text{Min } A,X \ \|Y-AX\|2 F \ \text{ s.t for all i, } \|Xi\|0 < T \quad (2)$$

## 3.2 Compression of Fingerprint

We selected a new fingerprint for compression, cut it into square patches. It has the same size with the training patches. The size of the patches has a affects the compression efficiency. The algorithm becomes more efficient as the size increases. However, there can be rapid increase in computation complexity and the dictionary size. In order to fit the patches in the dictionary, calculate the mean of each patch and subtract from the patch. Then l0 problem is solved to compute the sparse representation for each patch. If the values of coefficients are less than a given threshold, it is treated as zero [12].

Record four kinds of information for each patch. They are the mean value, number of atoms to use the coefficients and their locations. Arithmetic encoding of the mean value, number of atoms, coefficients and their location is done and compressed value is obtained.

## 3.3 Decompression of Fingerprint

The values obtained after compression is given to the arithmetic decoder. Thus the mean value, number about how many atoms to use, the coefficients and their locations of patches are regenerated. The patches are converted into images and the decompressed image is obtained.

## 3.4 Enhancement of Fingerprint

Wiener2 filter is applied to the decompressed image to remove the noise and to preserve more minutiae content. A class of filters, referred to as Wiener filters, use interrelationship data between signal and noise to enhance image or reduce distortion. Wiener2 is a 2D adaptive noise removal filtering. Wiener2 lowpass fiters filter an image which is vulgarized by additive noise. Wiener2 uses a pixel based Wiener method based on data collected from each pixel. Wiener2 filter is thus used to remove the noise from the image and thus preserve more minutiae than decompressed images.

## 3.5 Minutiae Extraction

Minutiae Extraction is applied to decompressed image and enhanced image. The image is converted to binary image. This process consists of changing the gray image into binary image, i.e, the intensity has two values: black is representing the minutiae, and white is representing the valleys and the background. A method to binarize is to use a global threshold value; however, it is not suitable for images with noise content, a more flexible method that uses mask which is rectangular in shape, and rotate it according to the direction of the ridges.
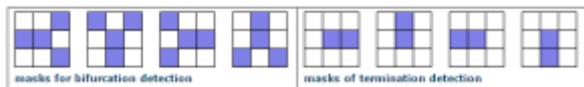
Figure 1. Pattern Masks

After the image is binarized, thinning is applied on to the image. The thinning process is to find the minutiae of one pixel width. The process consists of carrying out consecutive destruction until a set of joined lines of unit width is attained. These lines are also called skeletons. One of the characteristics of thinning is the preservation of the association and topology which can lead to creation of bifurcation artifacts and detection of false minutiae. From the binary thinned image, the minutiae are founded by using 3 * 3 pattern masks. Samples of masks used for identifying the ridge ending and bifurcations point are shown in the Figure 1.

## 4. EXPERIMENTAL ANALYSIS

In this section, we present experiments on fingerprint database for fingerprint compression, decompression and its enhancement. Minutiae are extracted from both the decompression and its enhancement. Minutiae are extracted from both the decompressed fingerprint and its enhanced fingerprint. Then the minutiae content is analyzed in both the compressed and its enhanced fingerprint. Finally, through the experiment it is shown that the minutiae content of decompressed fingerprint. The implementation is done in MATLAB.

### 4.1 Databases

To construct a dictionary of fingerprint patches, we used a set of training fingerprints. The fingerprint images can be downloaded from [13]. The size of the fingerprints in the database is 640 * 640. The quality of fingerprints in the database is good.

### 4.2 Implementation

First the fingerprint images in the database are sliced into square patches. The dictionary is obtained from the patches by solving the optimization problem, which is described in previous section. The obtained dictionary of the images in the database is shown in Figure 2. A new fingerprint is selected and cut into square patches which have the same size with the training patches. The size of the patches has a great effect on the efficiency of compression.



Figure 2. Dictionary of Fingerprint

The patches should be fit in the dictionary. For that, the mean of each patch needs to be calculated and subtracted from the patch. After that compute the sparse representation for each patch by solving l0 problem. Those coefficients whose absolute values are less than a given threshold are treated as zero.

For each patch, four kinds of information need to be recorded. They are the mean value, the number about how many atoms

to use, the coefficients and their locations. Arithmetic encoding of the mean value, number of atoms, coefficients and their location is done and compressed value is obtained. The compression ratio is calculated and obtained as 16%. When the compression ratio is too high, the minutiae contents cannot be preserved and when the compression ratio is too low, the data is not well compressed. So a moderate compression ration should be maintained.

The values obtained after compression is given to the arithmetic decoder. Thus the mean value, number about how many atoms to use, the coefficients and their locations of patches are regenerated. The patches are converted into images and the decompressed image is obtained. The obtained decompressed image is shown in Figure 3 (a). Wiener2 filter is applied to the decompressed image to remove the noise and to preserve more minutiae content. A class of filters, referred to as Wiener filters, exploit correlation information between signal and noise to enhance image or reduce distortion. Wiener2 is a 2D adaptive noise removal filtering. The obtained image is shown in Figure 4(a). The output image after enhancement is usually blurred visually. But it contains more minutiae content.
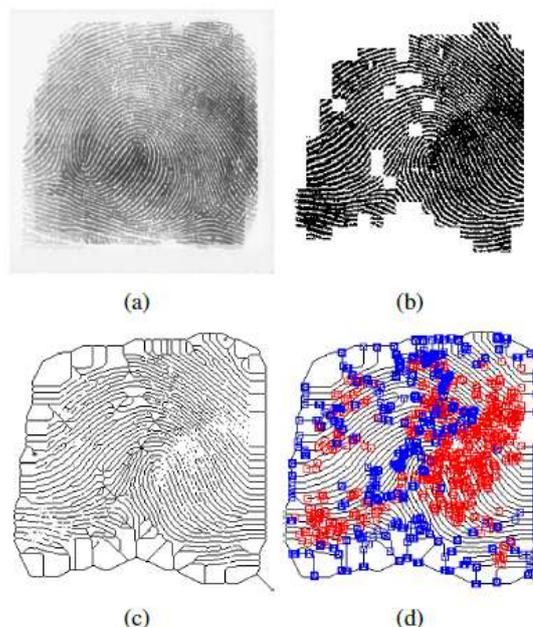


Figure 3. (a) Decompressed Fingerprint, (b) its binarization, (c) thinning and (d) Minutiae extraction.

Minutiae Extraction is applied to decompressed image and enhanced images. The image is converted to binary image. This process consist in converting the gray scale into binary image, i.e, the intensity of the image has only two values: black, representing the ridges, and white, representing the valleys and the background. A simple method to binarize is to use a global threshold value; however, it is not suitable for noisy images. So a more robust method consist of using some rectangular mask, rotate according the orientation of the ridges. The binarized images of decompressed and enhanced fingerprint are shown in Figure 3(b) and 4 (b).
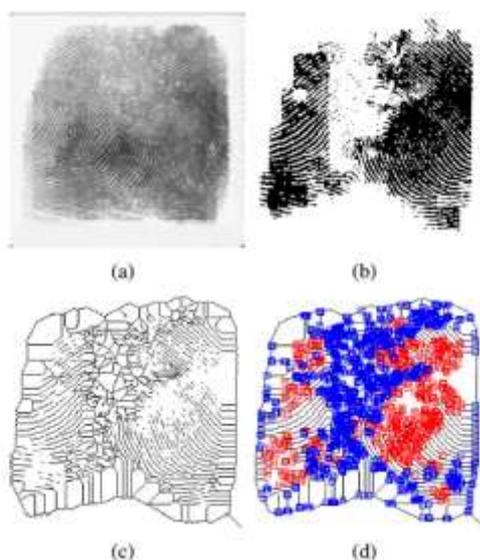
Figure 4. (a) Enhanced Fingerprint, (b) its

binarization, (c) thinning and (d) Minutiae

extraction.

After the image is binarized, thinning is applied on to the image. The objective of thinning is to find the ridges of one pixel width. The process consists in performing consecutive destruction until a set of connected lines of unit-width is reached. These lines are also called skeletons. Thinned images of decompressed and enhanced fingerprints are shown in Figure 3(c) and 4(c). An important property of thinning is the preservation of the association and topology which lead to creation of small bifurcation artifacts and detection of false minutiae. From the binary thinned image, the minutia is detected by using 3*3 pattern masks. The minutiae extracted from decompressed and enhanced fingerprint is shown in Figure 3(d) and 4(d). A pixel is then classified as a ridge if it has only one neighboring ridge pixel in the window, and classified as a bifurcation if it has three neighboring ridge pixels. For display purpose, the ridge is marked by a red window square and bifurcation is marked by a blue square window.

When the fingerprint images are experimented, we reached a conclusion that the enhancement technique preserves bifurcation content. However the broken ridges cannot be preserved. The experimental results obtained are tabulated to evaluate the preservation of minutiae content in various fingerprint images. Figure 5 shows the bifurcation content in each fingerprint images after decompression and enhancement. Table I shows the performance values of the proposed method.
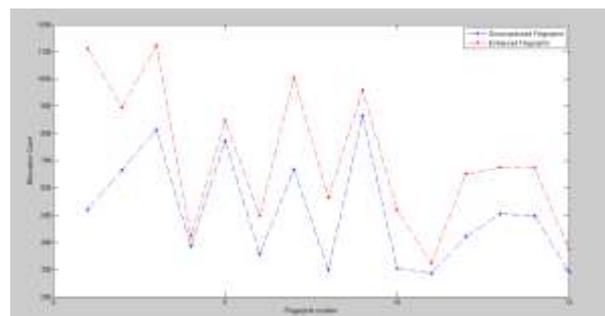


Figure 5. Performance Analysis

Table I
Experimental Results for Fingerprint Images

| Fingerprint Image | After Decompression | | After Enhancement | |
|---|---|---|---|---|
| | Ridge | Bifurcation | Ridge | Bifurcation |
| 1.bmp | 554 | 521 | 510 | 1112 |
| 2.bmp | 338 | 666 | 276 | 895 |
| 3.bmp | 411 | 814 | 232 | 1122 |
| 4.bmp | 650 | 385 | 426 | 423 |
| 5.bmp | 543 | 772 | 427 | 848 |
| 6.bmp | 548 | 358 | 410 | 497 |
| 7.bmp | 571 | 668 | 413 | 1003 |
| 8.bmp | 622 | 299 | 425 | 565 |
| 9.bmp | 146 | 864 | 85 | 960 |
| 10.bmp | 130 | 305 | 147 | 519 |
| 11.bmp | 86 | 287 | 84 | 325 |
| 12.bmp | 33 | 422 | 22 | 650 |
| 13.bmp | 680 | 504 | 581 | 676 |
| 14.bmp | 214 | 498 | 203 | 673 |
| 15.bmp | 177 | 291 | 168 | 373 |

## 5. CONCLUSION

In this paper, we propose an enhancement technique that is applied on compressed fingerprint images to make it better of fingerprint matching and so on. When the compressed fingerprint is decompressed, the image preserve only very less minutiae due to the noise in the image. So we applied Wiener2 filter to overcome the noise and thus to preserve more minutiae. Experimental results show that the enhancement technique preserves bifurcation content. However the broken ridges cannot be preserved. The method can be further improved by preserving ridges also. Latent fingerprints that are obtained from crime scenes can also consider for future works.

## 6. ACKNOWLEDGMENTS

# REFERENCES

[1] D. Maltoni, D. Miao, A. K. Jain, and S. Prabhakar, ``Handbook of Fingerprint Recognition,'' 2nd ed. London, U.K.: Springer-Verlag, 2009.

[2] W. Pennebaker and J. Mitchell, ``JPEG-Still Image Compression Standard,'' New York, NY, USA: Van Nostrand Reinhold, 1993.

[3] A. Skodras, C. Christopoulos, and T. Ebrahimi, ``The JPEG 2000 still image compression standard,'' IEEE Signal Process. Mag., vol. 11, no. 5, pp. 36–58, Sep. 2001.

[4] T. Hopper, C. Brislawn, and J. Bradley,`` WSQ gray-scale fingerprint image compression specification,'' Federal Bureau of Investigation, Criminal Justice Information Services, Washington, DC, USA, Tech. Rep. IAFIS-IC-0110-V2, Feb. 1993.

[5] Jinshan Tang and Xiaoming Liu, ``An Improved WSQ Fingerprint Image Compression Algorithm,'' IEEE Conference,2008.

[6] Abdul Adeel Mohammed, Rashid Minhas, Q.M. Jonathan Wu, Maher A. Sid-Ahmed, ``A Generic Fingerprint Image Compression Technique Based On Wave Atoms Decomposition,'' IEEE Conference 2009.

[7] S. Esakkirajan, T. Veerakumar, V. Senthil Murugan and R. Sudhakar, ``Fingerprint Compression Using Contourlet Transform and Multistage Vector Quantization,'' International Journal of Computer, Information, Systems and Control Engineering, 2007.

[8] Shenqiu Zhang, and Cecilia Moloney, ``Gray-Scale Fingerprint Image Compression Based on the Hybrid-NRCT,'' IEEE Conference 2009.

[9] M. Aharon, M. Elad, and A. M. Bruckstein, ``On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them,'' J. Linear Algebra Appl., vol. 416, no. 1, pp. 48-67, 2006.

[10] M. Aharon, M. Elad, and A. M. Bruckstein, ``The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation,'' IEEE Trans. Signal Process., vol. 54, pp. 4311-4322, 2006.

[11] S. Mallat and Z. Zhang, ``Matching pursuits with time-frequency dictionaries,'' IEEE Trans. Signal Process., vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[12] Guangqi Shao, Yanping Wu, Yong A, Xiao Liu, and Tiande Guo, ``Fingerprint Compression Based on Sparse Representation,'' IEEE Transactions on Image processing, Vol. 23, No. 2, February 2014.

[13] (2013, Jun.) Fingerprint Images [Online]. Available:http://pan.baidu.com/share/link?shareid= 3872475108 & uk=2301599420

# A framework for Performance Prediction of Service-Oriented Architecture

Haitham A.Moniem,
Department of Information Technology,
College of Information Technology,
Future University
Khartoum, Sudan

Hany H Ammar,
Lane Department of Computer Science and
Electrical Engineering,
College of Engineering and Mineral Resources,
West Virginia University
Morgantown, USA

**Abstract:** Service-Oriented Architecture (SOA) considered as one of the important architectural styles to build future applications. This architecture consists of a group of homogenous and autonomous components that interact with each other to accomplish a task. However, performance prediction of SOA based applications is regarded one of the complex tasks that face software developers and engineers. This paper presents a novel approach for SOA performance prediction at early stages of SDLC by using Machine Learning technique. Firstly, annotated UML diagrams are presented. Secondly, translate the UML diagrams into (QNM) model in order to extract performance indices such as Response Time, Throughput, and Utilization. Finally, machine learning technique used to predict the application model performance. The prediction result "Risk" means design does not meet customer requirement and "No Risk" means the design satisfies the customer requirements. Machine learning technique predicts the performance based on the training set against the extracted test set of the application model performance indices. The new method has many advantages, such as reducing time, scale with large system size, and avoiding problems before the service put into the production environment. To illustrate our approach, we present the results of a simple practical example.

**Keywords:** SOA; Queuing Network Model; Machine Learning; Performance; Prediction; Architecture

## 1. INTRODUCTION

Service-Oriented Architecture (SOA) provides excellent features for designing distributed internet applications include reusability, flexible configuration, and easy implementation. The Service defined as functional logical part that performs a unique business task. This task could be for consumer or for another service connected to it. Many challenges come with SOA as a new architecture style. Firstly, challenges related to finding methods to analyze and predict the Quality of Services (QoS) such as security, performance, service availability, and standardization during designing time. Secondly, challenges related to applying QoS standers on SOA based applications such as centralization and service integrity in environment exhibits autonomous, encapsulation, and privacy [1].

The proposed approach considers the report "Risk" if the performance indices extracted from the architectural design does not meet the customer requirements and "No Risk" if the performance indices satisfy the customer requirements. We assume that customer has non-functional requirements such as response time, resource utilization, and throughout. According to ISO 9126 all the previous requirements called Performance [9].

Generally, there are two types of software performance prediction, at the design time (Model-based) and at running time. Our proposed approach concerns about model-based performance prediction from the initial stages of Software Development Life Cycle (SDLC). Model based performance prediction has many advantages such as low cost, easy, and practical.

The proposed framework consists of three major steps as presented in Fig.1.

**First step:** annotated UML diagrams will be used to describe software system as follows:

- Use Case diagrams represent workloads applied to the system [7].
- Deployment diagrams describe available physical resources where computations take place.
- Activity diagrams describe both the order in which resources are used, and corresponding service demand.

**Second step**: mapping annotated UML model into Queuing Network Model (QNM) to simulate the SOA application and generates performance indices.

**Final step:** take the output of the Queuing Network Model (QNM) as a new instance and provides it to the machine learning, then based on the training set the machine learning algorithm will predict whether there will be a risk or no risk if we implement the SOA application by current performance indices.

The reminder of this paper structured as follows section 2 presents the related work. Section 3 explains the term SOA. Section 4 demonstrates SPT UML profile (Schedule, Performance, and Timing). Section 5 states the translation from UML to QNM. Section 6 defines QNM. Section 7 presents the machine learning technique. Finally, section 8 presents the case study.

Figurer 1. The Proposed Framework

Service Oriented Architecture Protocol (SOAP) and Web Service Description Language (WSDL).



Figure 2. Service-Oriented Architecture
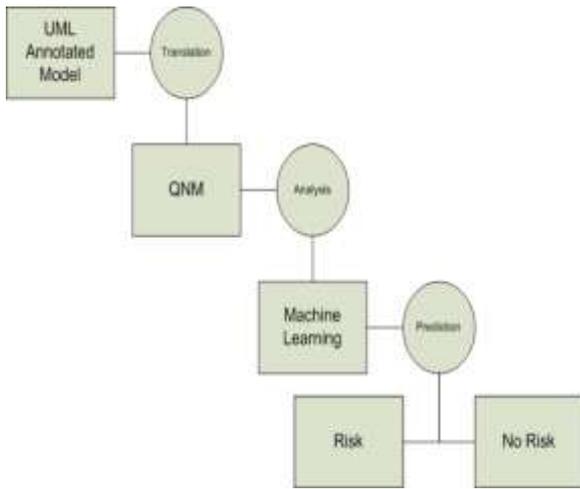
## 2. RELATED WORK

[Ganapathi, 2009] proposed a statistical machine learning technique to predict and optimize multi components and parallel system utilization and performance. The proposed technique extracts correlation between a workload's pre-execution characteristics or configuration parameters, and post-execution performance observations [2]. The correlation has been used for performance prediction and optimization.

[Dubach, 2009] used machine-learning to efficiently explore the compiler architecture design. The researcher firstly, developed two performance models and used them to increase efficiency of searching the design space of micro-architecture [3]. These models accurately predict performance metrics such as cycles or energy, or a tradeoff of the two.

[Malhotra et. al, 2012] have employed machine learning to measure the maintainability, number of CHANGE is observed over a period of three years on dataset [4]. Change can be defined as the number of lines of code which were added, deleted or modified.

[Mohanty et. al, 2012] have employed machine learning technique to classify and rank web services [5]. The researchers proved that by using Naïve based Bayesian network the classification performs better than other techniques.

[Ipek et. al, 2005] used multilayer neural networks trained on input data from executions on target platform. The approach is useful for predicting many aspects of performance, and it capture full system complexity. The study focuses on the high performance, parallel application SMG2000 [6]. The model has predicted performance within error 5% - 7% error across a large, multidimensional parameter space.

## 3. SERVICE-ORIENTED ARCHITECURE (SOA)

Service-oriented architecture (SOA) is an architectural style where a system comprises of three major components: First, Service Provider is the service or entity that accepts and executes request from service user and service registry. Second, Service User is an application or service that requires a service. Third, Service Registry is a network based directory that contains available services [8, 9]. An architectural style characterizes the types of components, connectors, and configuration. Fig 2 explores SOA architecture.

Web service technology is a type of the implementations of SOA. Web Service consists of many published standards such as
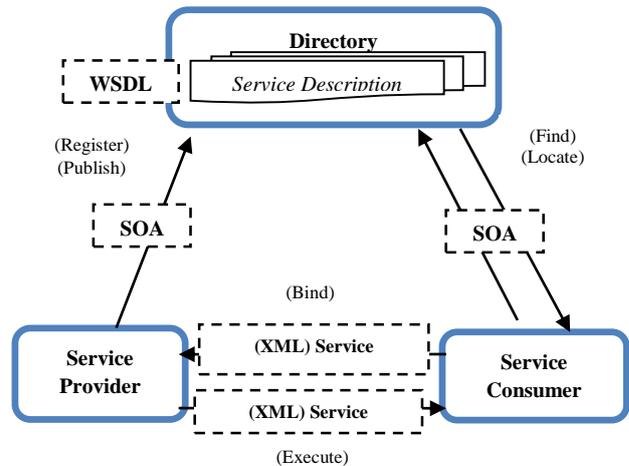
## 4. UML SPT PROFILE

UML profile for scheduling, performance, and Time has been released by OMG standard. The basic objectives of the UML profile are to declare the requirements for system performance and a tool to model physical time, timing specifications, timing services, logical and physical resources, concurrency and scheduling, software and hardware infrastructure, and their mapping [10].

UML profile provides an extension of the UML standard with special modeling components and their semantics. The main components of profile are new Stereotypes. Stereotype provides a style of extending UML by declaring simple terms and using them to explain UML components and cooperation in a system. Stereotypes are implemented to existing UML entities, such as class and association, and increase semantics of these elements with newly predefined meaning.

## 5. UML TRANSLATION INTO QNM

Based on the roles stated by (Simonetta et. al, 2004), we use algorithm named UML-QNE to translate an annotated UML diagrams specification into QNM. UML model components are translated into the corresponding QNM model elements as follows. From actors in Use Case diagrams we identify the type of QNM model. From activity diagrams we derive the network topology that is the behavior of the classes of users circulating through the system. Finally, each node in the Deployment diagrams defines a service center. The mapping between UML and performance model element is illustrated in Fig.3.
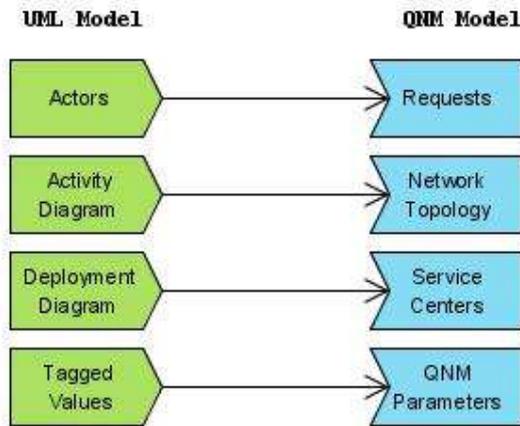
Figure 3: Mapping UML into QNM

# 6. QUEUINGNETWORK MODEL (QNM)

Queuing network model is a mathematical model used in computer systems performance analysis to predict quality attributes of a system and its parts [11]. The basic parts of QNM are Queues and Service Center. A queue is a buffer, which similar to any queuing system. A service center provides services to the queue's customer. Each service center has an associated queue containing jobs to be processed by that service center.

To get performance characteristics for any individual service center, two kinds of information must be presented. Firstly, the average rate $R$ at which new jobs arrive in service center. Secondly, the average time taken by service center $S$ to perform one job. Based on this information the following quality of services could be calculated:

- *Utilization of each Service Center* $(U_i) = R * S_i$
- *Average Response Time* $(R_i) = S_i/(1-u_i)$
- *Average Number of users at each Service Center* $(P_i) = u_i/(1 - u_i)$

From above equations it is possible to make calculations for latency, throughput, and highly utilized service centers.

# 7. MACHINE LEARNING

Machine learning is the capability of the computer program to acquire or develop new knowledge or skills from existing or non existing examples for the sake of optimizing performance criteria [12]. Software engineers and researchers have been started using machine learning techniques in the area of quality of service classification and prediction. Moreover, machine learning has proved it is efficiency to asset and optimizes model based performance prediction.

Machine learning can be categorized into two groups that are, supervised and unsupervised machine learning. These two learning categories are associated with different machine learning algorithms which represent how the learning method behaves [12].

# 7.1 Supervised Learning

Supervised learning comprises of algorithms that reason from externally supplied instances to produce general hypothesis which then make predictions about unseen instances. Moreover, with supervised learning there is presence of the outcome variable to orient the learning process. There are many machine learning algorithms for supervised learning such as Support Vector Machine (SVM), K-Nearest Neighbor, and Random Forests.

# 7.2 Unsupervised Learning

Opposite to supervised learning where there is presence of the outcome variable to orient the learning process, unsupervised learning builds models from data without predefined example [12]. This means no guidance is available and learning must perform heuristically by the algorithm examining different training data.

# 8. A CASE STUDY
## Step 1: UML Annotated Diagrams

In order to make our proposed approach of a model based performance prediction more understandable, an example is represented describing a commerce system.

The application scenario starts as the client browses the system to obtain information about products and prices via commercial server. The commercial server connects the database server in order to get the information required and complete the task. Moreover, the client can use the system to purchase products via the commercial server. In this step, the commercial server connects the online shopping mall server to make the purchase operation and accomplish the job.
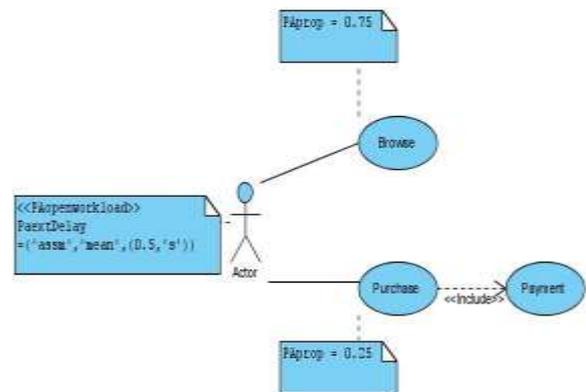


Figure 4. The system described by Use Case Diagram
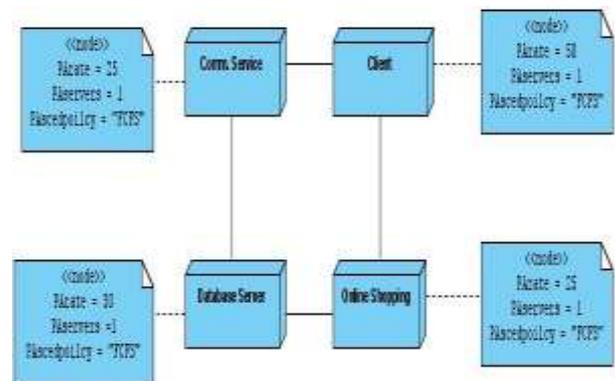


Figure 5. The System described by Deployment Diagram
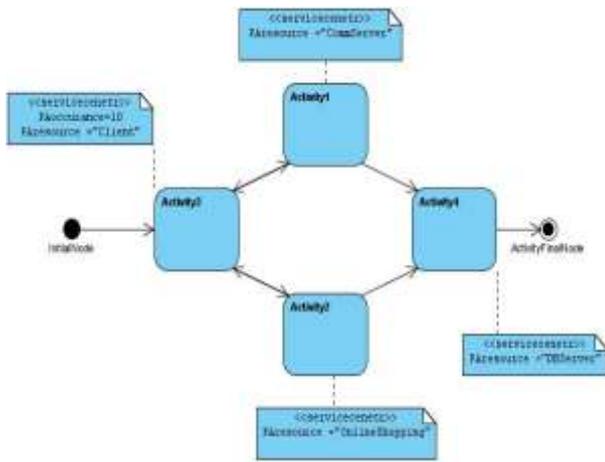
Figure 6.The System described by Activity Diagram

Fig.4 shows the annotated Use Case diagram. Each actor in a use case diagram may represent a stream of requests arriving at the system. There may be an unlimited sequence of requests (open workload) stereotyped as <<OpenUser>>, or a fixed population of users requiring service from the system (closed workload) stereotyped as <<ClosedUser>>. The tagged value can be linked with use case diagrams:

**PApopulation** specifies the total number of requests circulating in the system.

Fig.5 represents the annotated deployment diagram. Deployment diagrams are used to model the physical resources available in the system. Each resource is represented by a node in the deployment diagrams. Each node which must be stereotyped as <<node>> represents a processor with a given scheduling policy. The tagged values can be link with deployment diagrams:

**PAschedpolicy** denotes the scheduling policy of the processor, which can be one of: First in First out (FIFO), or Last in First out (LIFO), or Processor Sharing (PS).

**PArate** denotes the processing rate of the processor or time taken by the processor to complete a task.

**PAservers** denotes the number of processors simultaneously executing the requests.

Fig.6 shows the annotated Activity diagram. Activity diagram describes the workloads and the physical resources available in the system, also it's specifies how the resources are used in the meaning of computations are performed in the system.

Each action state of an activity diagram, stereotyped as <<ServiceCenter>>, represents a computation which requires service to one resource. Tagged values below can be specified to convoy information for making up the performance model:

**PAresource** denotes the name of the resource from which service is requested, which called node in Deployment diagram.

**PAoccurrence** denotes the inter-arrival time of this service request. It can be used with Queuing Network Model to specify open queuing model.

## Step 2: Transformation to QNM

Based on Table 1 notation algorithm 1 is used to translate annotated UML elements into the corresponding QNM components. Each node in the Deployment diagram defines a service center. Use case diagram used to categories the type of model (open or closed), and from Activity diagrams network topology is derived.

---

**Algorithm 1: QNM Generation Algorithm 1**

---

**for all** Deployment diagram node $R_i, i = 1 .. N$ **do**
   $S_i := New\ Service\ Center$
   $\mu_i := 1/PAserviceTime(R_i)$
   $NS_i := PAservers(R_i)$
**end for**
Let $C \leftarrow 0$
**for all** Actor $A$ **do**
 Initialize routing matrix $\mathbf{P^c}$ for class $C$ to zero
 $AD := Active\ diagram\ associated\ to\ A$
 **for all** Transition t from action state $a_i$ to $a_j$ of
 Activity diagram $AD$ **do**
   $R_k := PAhost(a_i)$
   $R_l := PAhost(a_j)$
   $P^c_{k,l} := PAprob(t)$
 **end for**
**if** $A$ is a ClosedUser **then**
   Set class $C$ as a closed chain with *PApopulation (A)*
   requests
**else**
 Set class $C$ as an open chain
 **for all** Action state $a$ of Activity diagram $AD$ **do**
   **if** PAoccurrence(a) is defined **then**
    $R_i := PAhost(a)$
    $\lambda^c_i := PAoccurrence(a)$
   **end if**
 **end for**
 **end if**
 Let $C \leftarrow C + 1$ {New Customer Class}
**end for**

---

**Table1. Notation used for Algorithm1**

| | |
|---|---|
| $N$ | Number of nodes in the UML deployment diagram. |
| $S_i$ | $i^{th}$ service center |
| $\mu_i$ | Service rate of service center $S_i$ |
| $\lambda^c_i$ | Arrival rate of class $C$ customers at service center $S_i$ |
| $NS_i$ | Number of servers in service center $S_i$ |
| $P^C$ | $N \times N$ routing matrix of class $C$ customers |

## Step 3: QNM Analysis

The translation algorithm works on the annotated UML diagrams, and starts mapping the UML into QNM model. Fig7 presents the derived QNM from the Use case, Deployment, and Activity diagrams of respectively, assuming that the Activity diagram is associated to an actor stereotyped <<OpenUser>>.

We used a tool for drawing and calculating performance indices from queuing networks called **P**erformance **E**valuation and **P**rediction **SY**stem for Windows platforms (**WinPEPSY-QNS**).
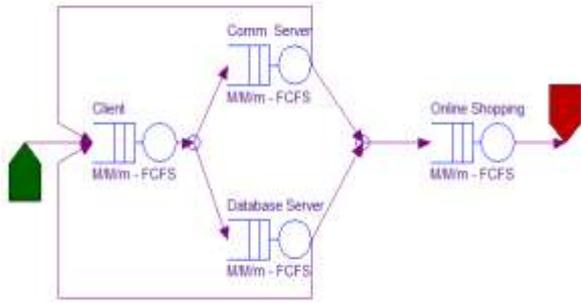
Figure 7. QNM - Output of the Transformation Algorithm

WinPEPSY-QNS analyzes the QNM and gets the result by using Mean Value Analysis (MVA) algorithm.

The analysis of the QNM of Fig. 7 provides a set of average performance indices that include utilization, throughput, response time, and latency. Table 2 shows numerical results for the set of parameters value for the whole system. The performance analysis can provide indication for possible system modification and further software performance evaluation can be iterated by choosing a different set of parameters value, to be inserted in the UML annotation.

**Table 2. Performance Result for the QNM Fig. 7**

| | |
|---|---|
| Utilization ($U_i$) | 0.4 |
| Throughput ($X_i$) | 10.0 |
| Response Time ($R_i$) | 1.42 |

### Step 4: Machine Learning Prediction

According to above performance indices presented on table 2 we will consider the results as a test set. To complete the test set we assumed availability, success-ability, reliability, and number of operations performed by the web service the complete test set will be as Table 3. We used **W**aikato **E**nvironment for **K**nowledge **A**nalysis (**WEKA**) as a tool for machine learning. In order to classify the test set as "**Risk**" or "**No Risk**".

**Table 3. The Complete Test Set**

| | |
|---|---|
| Utilization ($U_i$) | 0.4 |
| Throughput ($X_i$) | 10.0 |
| Response Time ($R_i$) | 1.42 |
| Availability ($Av_i$) | 100 |
| Success-ability ($Su_i$) | 100 |
| Reliability ($Re_i$) | 100 |
| Operation ($Op_i$) | 9 |
| Class | No risk |

Utilization ($U_i$) means total amount of resources required in order to complete a task, Throughput ($X_i$) is a total number of invocations for a given period of time, Response time ($R_i$) is the time taken to send a request and receive a response, Success-ability ($Su_i$) measured as number of response / number of request messages, Operation ($Op_i$) is the number of operations performed by the web service.

### Step 5: The Result

We have used a training set contains 166 web services and their measurements collected by (Al-Masri, 2007) to train our model. After feeding the test set on the machine learning the result

confirms our claim that there will be "No Risk" if we apply the current web service architecture. The result is promising as we get relative absolute error 15.4 % fig.8.
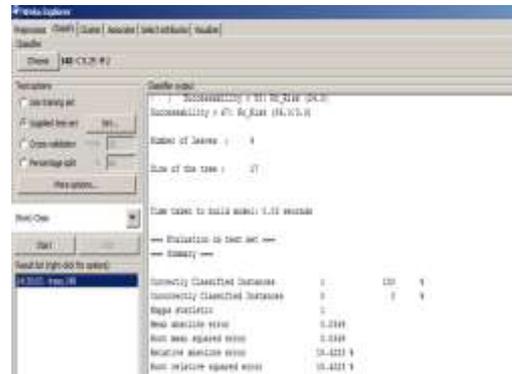


Figure8. Result

## 9. CONCLUSION

In this paper, we proposed a new technique to predict the performance of service oriented architecture based applications. The study focuses on web service as implementation of SOA. This approach avoid the needs of transforming UML model into queuing network at each time we want to predict the performance at early stages of development process.

We extract utilization, response time, throughput, and number of operations from the architectural description after mapping application architecture into QNM model. Machine learning technique has been used to predict is there will be risk if we implement the web service with same indices or no risk.

As a result the prediction model gives error percentage 15.4 % which considered as promising result.

## REFERENCES

[1] Mani, N., Petriu, D. C., & Woodside, M. (2011, March). Towards studying the performance effects of design patterns for service oriented architecture. In*Proceedings of the 2nd ACM/SPEC International Conference on Performance engineering* (pp. 499-504). ACM.

[2] Ganapathi, A. (2009). Predicting and optimizing system utilization and performance via statistical machine learning.

[3] Dubach, C. (2009). Using machine-learning to efficiently explore the architecture/compiler co-design space.

[4] Malhotra[1], R., & Chug, A. (2012). Software maintainability prediction using machine learning algorithms. *Software Engineering: An International Journal (SEIJ)*, *2*(2).

[5] Mohanty, R., Ravi, V., & Patra, M. R. (2012). Classification of web services using Bayesian network.

[6] Ipek, E., De Supinski, B. R., Schulz, M., & McKee, S. A. (2005). An approach to performance prediction for parallel applications. In *Euro-Par 2005 Parallel Processing* (pp. 196-205). Springer Berlin Heidelberg.

[7] Balsamo, S., Mamprin, R., & Marzolla, M. (2004). Performance evaluation of software architectures with queuing network models. Proc. ESMc, 4.

[8] Punitha, S., & Babu, C. (2008, September). Performance prediction model for service oriented applications. In *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on* (pp. 995-1000). IEEE.

[9] Moniem, H. A., & Ammar, H. H. (2014) Performance Prediction of Service-Oriented Architecture-Asurvey.

[10] Youn, H., Jang, S., & Lee, E. (2007, August). Deriving queuing network model for UML for software performance prediction. In *Software Engineering Research, Management & Applications, 2007. SERA 2007. 5th ACIS International Conference on* (pp. 125-131). IEEE.

[11] Punitha, S., & Babu, C. (2008, September). Performance prediction model for service oriented applications. In *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on* (pp. 995-1000). IEEE.

[12] Omary, Z., & Mtenzi, F. (2010). Machine learning approach to identifying the dataset threshold for the performance estimators in supervised learning.*International Journal for Infonomics (IJI)*, *3*, 314-325.

**Haitham A.Moniem** is a lecturer of Computer Science and Information Technology in College of Information Technology at Future University, Sudan. He is a PhD student at Sudan University of Science and Technology. He received B.Sc. and M.Sc. from Gizera University, Sudan; (2002, 2006). His research interest includes Software Engineering, Service-Oriented Architecture, Machine Learning, Cloud Computing, and Software Architecture.

**Dr.Ammar** is a Professor of Computer Engineering in the Department of Computer Science and Electrical Engineering at West Virginia University. He has published over 170 articles in prestigious international journals and conference proceedings. Dr. Ammar is currently Editor in Chief of the Journal of Computer Science and Engineering, in Arabic. Dr. Ammar has been teaching in the areas of Software Engineering and Computer Architecture since 1987. He has been Principal Investigator on a number of research projects on Software Risk Assessment and Software Architecture Metrics funded by the NASA IV&VFacility and NSF, and projects on Automated Identification Systems funded by NIJ and NSF. Dr. Ammar is a member of the IEEE and the ACM professional organizations. He has served as Chairman and member of Steering Committees and Program Committees of several International Conferences and Symposia. He previously served as the Chairman of the Upper Monongahela Subsection of the Pittsburgh section of the IEEE, a Director of the Pittsburgh section of the IEEE, and the student section advisor of the IEEE Computer Society at WVU.

# Enhanced IHMAC with Dead Line Aware Packet Scheduling Scheme for Wireless Sensor Networks

Y. Sravya Lakshmi,
VBIT, Hyderabad
India

Bhaskar Gugulothu,
VBIT, Hyderabad,
India

**Abstract:** - Wireless detector networks have become very democratic in Holocene years. For detector network, S-MAC and T-MAC are used to improve the energy benefits by applying sleep control. Our ultimate aim is to achieve high power benefits under extended range of traffic and also promises shorter delay for this type of emergency and latency-sensitive messages and removed collision. In this project, we present the light weight Intelligent Hybrid MAC, a newer low power with QoS promising medium access control protocol for wireless detector networks. Our proposed Enhanced ACTIVITY CONTROL (Enhanced-IHMAC) is combines TDMA and CSMA with buffering technique. There are many applications of wireless detector network where it is really wanted to ensure the priority services for the emergency information. And also we have not used synchronization messages unlike our base model. Our proposed ACTIVITY CONTROL (Enhanced-IHMAC) protocol guarantees shorter latency for this type of emergency and latency-sensitive messages with low overhead and emergency scheduling.

**Key words:** WDN, power saving, sleep mode, priority, MAC

## 1. INTRODUCTION

A wireless detector network (WDN) consists of detector devices capable of collecting information from the place and sharing the information with each other via wireless devices. The collected data will be delivered to one or more destinations, very often through multi-hop information sharing. The detector devices are very often expected to operate with power units and are often deployed to not-easily-accessible or remote place, sometimes in prominent numbers. It can be unmanageable or impossible to replace the power units of the detector devices. On the other hand, the destination is very often rich in power. Since the detector power is the most precious resource in the WDN, effective utilization of the energy to prolong the lifetime of network has been the main focus of the research on the WDN.



**Fig.1 wireless detector network**

The information sharing's in the WDN has the many-to-one mode in that data from a prominent number of detector devices tend to be concentrated into a few destinations. Since multi-hop routing is very often wanted for distant detector devices from the destinations to save power, the devices near a destination can be burdened with relaying a prominent amount of traffic from other devices. Communication in wireless detector networks will, like most network communication, be divided into many layers. one in all those is that the Medium Access control (MAC) layer. This layer is delineated by a Mac protocol, that tries to make sure that no two nodes are officious with every other's transmissions,

and deals with things once they do. Wireless detector networks have a further aspect: as detector nodes are typically battery-operated, energy consumption is incredibly necessary. The radio on a detector node is sometimes the part that uses most energy. Not solely transmission prices energy; receiving, or just scanning the ether for communication, will expand to half the maximum amount, reckoning on the sort of radio [8]. Whereas ancient mackintosh protocols are designed to maximize packet output, minimize latency and supply fairness, protocol style for wireless detector networks focuses on minimizing energy consumption. the appliance determines the wants for the (modest) minimum output and most latency. Fairness is sometimes not a problem, since the nodes in a very wireless detector network are generally a part of one application and work along for a standard purpose

## 2. RELATED WORK

This paper [1] presents RAP, new real-time information sharing architecture for prominent-scale detector networks. In some of previous work the Authors proposed Velocity Monotonic Scheduling. VMS assigns the priority to a message bundle based on its requested speed. A message bundle with a higher requested speed is assigned a higher priority. VMS improves the number of messages that meet their deadlines because it assigns the "right" priorities to messages based on their emergency on the current node. But there is no detail for message bundle receiving distance in that paper. And in this paper, when the queue is filled, most priority incoming messages overwrite the lower priority messages at ones. It may trouble to be the low priority information.

In paper [2], Authors proposed an Adjustive Staggered SLEEP Protocol for effective energy management in wireless detector networks targeted to periodic data collection. This algorithm lively adjusts the sleep schedules of devices to match the network needs, even in time-changing operating conditions. It utilizes the CSMA scheme for process the info, but it may be not efficient in fixed WDN network and there is no detail to data handling. In this paper [2], Authors presented how to place detectors by use of a minimal number to maximize the coverage area when the information sharing radius of the SN is not less than the detecting radius, which results in the application of regular topology to WDNS placement.

In this paper Authors discussed the details of detector placement. Due to optimal coverage detector placement, it reduces the no of detectors usage and also step-ups the lifetime of detectors. But till lifetime of detector need to step-up. In paper [3], Authors proposed a grouping method with coverage and power aware TDMA scheduling method. And the group creation is done by the base station according to the current residual power, and the coverage area of group Member is reduced to avoid the congestion and power management. In that paper also there is no discussion on the real time and non-real time message bundle transmission. Authors developed scheme by deploying the network with different in size and fixed grids while taking into account the arbitrary shaped area sensed by the detector devices.

In paper [4], Authors considers the different initial power level of detectors, and placing that detector according to that power range. So power reduction was for fended. But measuring different initial power level and placing the device according to that power level is unmanageable in real time.

In paper [5] Authors presented a group based routing algorithm. One of main goal is to design the power efficient routing protocol. This algorithm makes the best use of device with low number of group leader know as super device. Here Authors divided the full area in equal zones and the center area of the area is used to select for special device. Every area is considered separately and the zone may be or not divided further that's depending upon the tightness of devices in that zone and capability of the special device. In this paper Authors considered, group leader changes when the group leader is failed. It may be the trouble to detecting in that area.

## 2.1 Existing system & demerits

The Existing Z-MAC has the initial phase. In initial phase there are neighbor node identification, timing assignment, neighbor message exchange and global time synchronization steps has been done. Q-MAC method gives QoS by differentiating network services based on priority values. The priority values reflect the emergency of messages sectioning from different detector devices.

## 3. PROPOSED SYSTEM

The ACTIVITY CONTROL (Enhanced-IHMAC) does it by using the strength of carrier sensing and time division approach with intelligence. The innovative idea behind the ACTIVITY CONTROL (Enhanced-IHMAC) is that it uses both the broadcast scheduling and link scheduling. ACTIVITY CONTROL (Enhanced-IHMAC) classifies messages according to their emergencies and stored the messages into the appropriate queue. The source device knows the degree of importance of the sensed information and accordingly the application layer assigns the priority.

### 3.1. Techniques

- o Route Discovery
    - Timer
    - Sectioning Message Generator
    - Route establishment
- o Virtual Grouping
    - Allocation of Sections
- o Data Transmission/Buffering
- o Power Saving

### 3.1.1 Route Discovery

Here we are going to enable the timer to send sectioning message in specific intervals. Sectioning message is to know about the neighbor devices sections, based on the reply; device will store the neighbor table information. The Request message will flood to the each device in the network to update the information of base station availability.

### 3.1.2 Power Saving

The ACTIVITY CONTROL uses Request-To-Send, Clear-To-send messages to handshake with the neighbor device. And this method is used to adjust the transmitting power to the minimum level necessary to reach the intended neighbor. Thus, the ACTIVITY CONTROL (Enhanced-IHMAC) reduces power loss by suitably changing the transmit power. And this handshaking method used to synchronize the sleep scheduling also.

### 3.1.3 Data Transmission

According to the energy saving method and in the allocated sections based on handshaking method, detectors are going to transmit the data and similarly they will change to sleep mode when there is no transmission.

In our proposed system, we have implemented ACTIVITY CONTROL with reduced over leader model. As like as in TDMA method, we also divided time into time sections, but unlike TDMA method each device can use the other time section when the time section is not in use. To check whether time section is 'not in use' or use, we have connected CSMA/CA method. In our project, first we divided time into time sections and then further we divided time section into further three sections (priority section, own section, other section). If device has any priority data then the device can transfer the data at beginning of time section which may be own or others. If device not has any priority data then it will check for section is for me or not. If section is own section then it can send data in second section of main section.
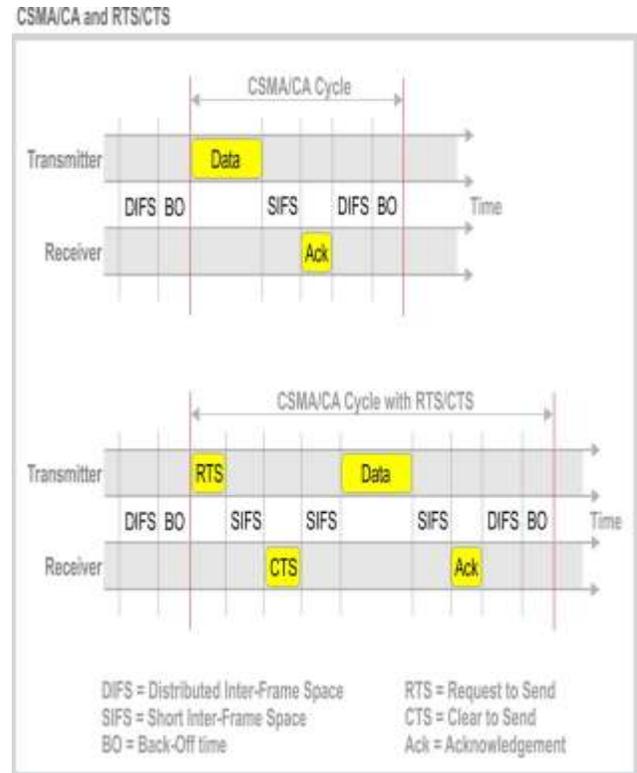


**Fig.2 CSMA/CA method**

If section is others section then it will wait for third section in main section with small random interval if device not detecting information sharing then device can transfer the data in that section. Fig.7 shows the result of data transmission through other section

### 3.1.4) Priority message bundle scheduling

Among large number of  network design problems, such as routing protocols and data collection, that reduce sensor energy loss and data transmission delay, message bundle scheduling (interchangeably use as task scheduling) at sensor devices is highly important since it ensures delivery of different types of messages based on their priority and fairness with a lesser delay. Indeed, most existing Wireless Detector Network use FCFS schedulers that process messages in the order of their receiving time and, thus, need a large amount of time to be delivered to a relevant base station. However, to be useful, detected information has to reach the BS within a certain period or before the deadline. Additionally, emergency info should be delivered to base station with the lesser possible end-to-end latency. Hence, intermediate devices require changing the delivery order of data messages in their ready queue based on their importance and delivery priority. In our enhanced work, we have considered the buffering problem; the message bundle may chances to loss due to scheduling at a time. So in our enhancement work the device can remember the previous scheduling message bundle information. We are

concentrating on reducing the overhead and avoiding the packet loss due to inefficient packet scheduling. In our base work, the researcher used the synch packet to make synchronization. Once the node gets the slot to transfer the data then it has to synchronize with neighbor to deliver the data to next node. To synchronize, the sender node broadcasts the synch message to neighbors with the name of indented receiver. If the neighbor received then it has to check whether own id is present in found in that or not. If own id found then it can know that it should be in active or else it can go to in sleep. In our enhancement work, we have proposed the RTS synchronization. If any node received the RTS/CTS signal then it has to check whether own id is found or not, if own id is not found then the device has to go to sleep otherwise node will be in active. So we can reduce unnecessary overhead in the network. RTS/CTS message is enough to synchronize the network.

In our base work, there are no details to schedule the multiple priority packets at one node. In case, the node generated one packet with priority and received the one priority packet at same time then the node has to schedule both at same time. If both packets are scheduled at same time then any one of the message bundle will be successfully scheduled and another one will be deleted. In our enhanced work we have considered this problem. So in our enhancement, the node checks the scheduling info whether any packet is already scheduled or not, if priority packet is already scheduled then another priority packet will be added into the queue. So that message bundle can send in next time slot without loss. We can improve the packet delivery by queuing technique.

## 3.2) Algorithm

### 3.2.1) Algorithm:

Inputs: N –number of sensor nodes, 1-base station node. At beginning pre-defined sensor will generate slot req, after some predefined duration base station will share origination message.

1) Sensor "i" sends the slot req to immediate neighbors (1,2…m) (m<15) and set the timer to confirm the slot (confirmation timer)
2) If slot req received in any sensor
   a. Checks the waiting list and confirmation list
     i. If No slot is matching with requested slot
       1. Add the requested slot into waiting list with node id, time and hop cont
     ii. If requested slot already available in confirmed list
       1. Generates the rejection message with
         a. Node id
         b. Confirmed slot info
     iii. If request not found in confirmed list but available in waiting list
       1. Check the time of slot req generation

    a. If new slot_req_time <already available pending req_time
      i. Add new req in waiting list
      ii. Reject already available pending slot_req
    b. If new slot_req_time >already available pending req_time
      i. Reject new req, and send rejection mesg
    c. If new slot_req_time >already available pending req_time
      i. Reject both and send rejection mesg

3) If rejection messg received in any sensor node
   a. If rej_mesg is for me
     i. Then reset the own slot
     ii. Then send cancelation message to all neighbor node
     iii. And checks confirmed slot info from pkt and updates the confirmed list
   b. If not for me
     i. checks confirmed slot info from pkt and updates the confirmed list
4) if cancelation message received then
   a. check the canceller info in waiting list
     i. if found
       1. delete the info
     ii. else
       1. ignore
5) if in any sensor confirmation timer is expire after sending the slot req
   a. checks it own slot info in own waiting list (slot requested sensor)
     i. if info found with requested slot
       1. confirm the slot for own use
       2. send the confirmation message to all neighbors
     ii. if info not found
       1. set the timer to re-allocate the slot
6) if confirmation message received
   a. update the confirm list
   b. checks the own confirm list for own slot
     i. if node has own lot
       1. ignore the pkt
     ii. if node not has any slot for own use then
       1. set the limited random time to generate slot req

## 3.2.2) Data transfer

If synchronization is completed

I) if data generated
   a. set the mode (priority or non-priority)
   b. checks the packet priority
     i. if priority- packet
       1. then set the packet sending in next section starting
     ii. if non-priority- packet
       1. then add into buffer

II) if data recv
   a. continue the same process as *I-b*
III) if current time is for next section
   a. count the current section number
   b. if section number matching with own section
      i. set the timer for own section access
   c. if section number is others
      i. set the timer for other section access
   d. schedule next section
     IV) if current time to access own section
   a. check the mac
      i. if not in use
         1. send the data
      ii. if not
         1. re-buffer the packet
V) if current time to access other section (continue the process of (*IV-a*))
VI) if current time is to emergency packet
   a. Go to buffer scheduling algorithm

## 3.2.3) Buffer scheduling algorithm:

Each node has to use CSMA/CA technique to transfer the data in allocated TDMA slot. Let consider $f$ forwarder of data and $n$ is receiver

1) If $slot\ initiated$
   a. If $packet.data$ available in f
      i. If can use Sub slot
         1. Generate RTS packet
         2. $(f,n) \cup packet.addrs$
         3. $broadcast\ RTS$
      ii. Else
         1. Wait for next sub slo$t$
   b. Else
      i. Switch to idle mode
2) If $packet.RTS$ received in $n$
   a. If $n \ni packet.addrs$
      i. Switch to Active state
      ii. Send CTS $\ni (f,n)$
   b. Else
      i. Set NAV
      ii. Switch to sleep till next slot initiation
3) If $packet.CTS$ recv in $n$
   a. If n $\ni$ packet.addrs
      i. Switch to Active state
      ii. Send data
   b. Else
      i. Set **NAV**
      ii. Switch to sleep till next $slot$ initiation
4) If $packet.data$ recv in n
   a. If $packet.data_{emergency}$
      i. Get interval$_{nslot}$

      ii. If $\exists schedule$
         1. En-queue $packet$
      iii. Else
         1. schedule packet

## 3.2.4) power saving algorithm

(I)    Power control technique
   A) set the maximum transmission power
   B) while sending RTS/CTS set maximum power as tx power
   C) if CTS pkt is received
      a. checks the recv power info
      b. set new tx power   $P_{tx} = \dfrac{P_{max}}{P_{rec}} * P_{rth} * C$
(II)    Sleep
      a. If node ready to send data
         i. Send synch message with (addrs(F,N))
      b. If Synch recv
         i. If node is not in synch
            1. Make node to sleep

## 4. RESULT ANALYSIS

We have tested our proposed system with the Network Simulator 2 tool. Our proposed system the delay is less than all other slandered systems (see fig.3)
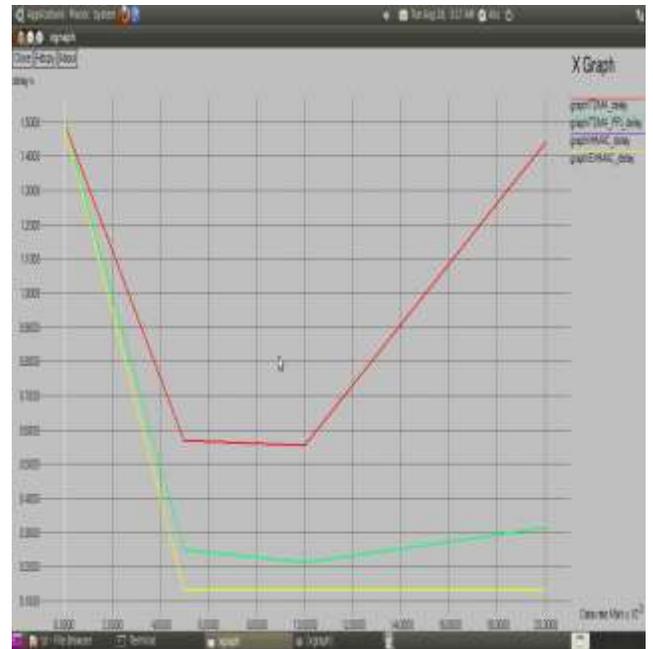


**Fig.3 delay comparison**

Our proposed system provides less energy loss compare than all other standard methods (see fig.4)
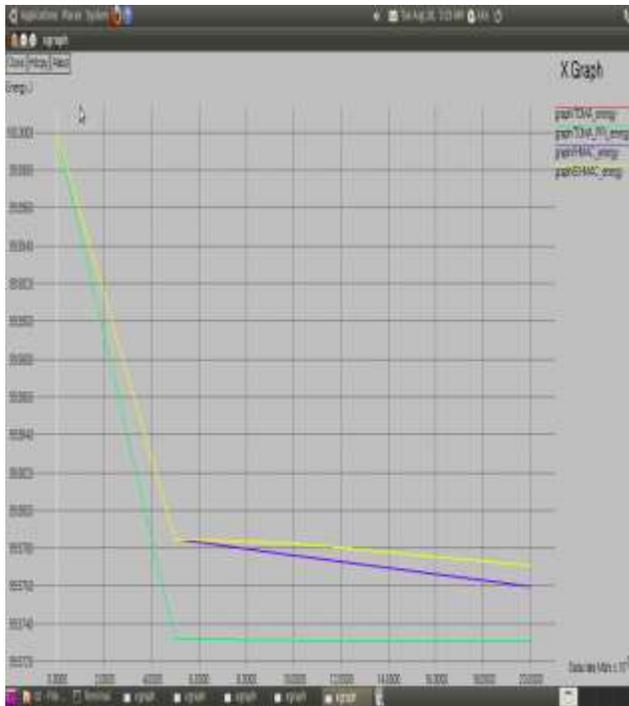
**Fig.4 Energy comparison**

In our enhanced method, we have not used the synch packets so our enhancement method provides less overhead compare than all other standard methods (see fig.5)
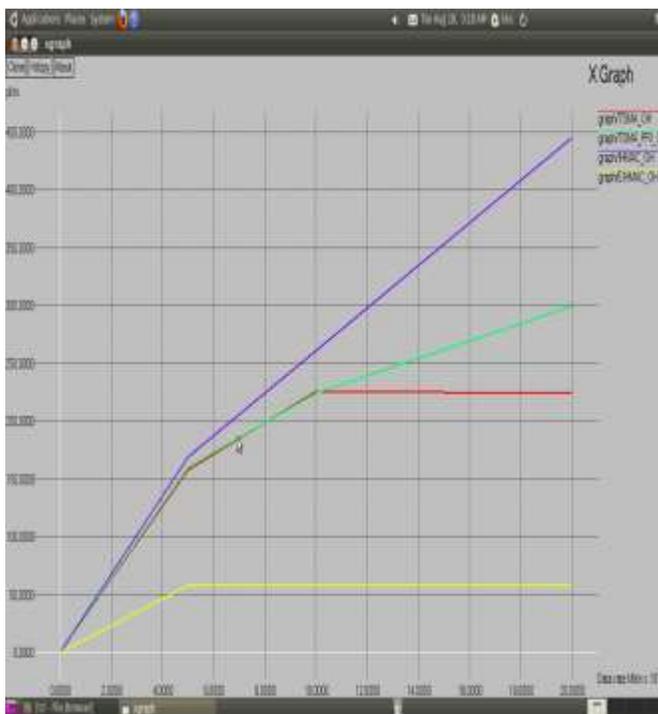


**Fig.5 overhead comparison**

Our proposed system provides higher stable packet delivery compare than all other standard methods (see fig.6)
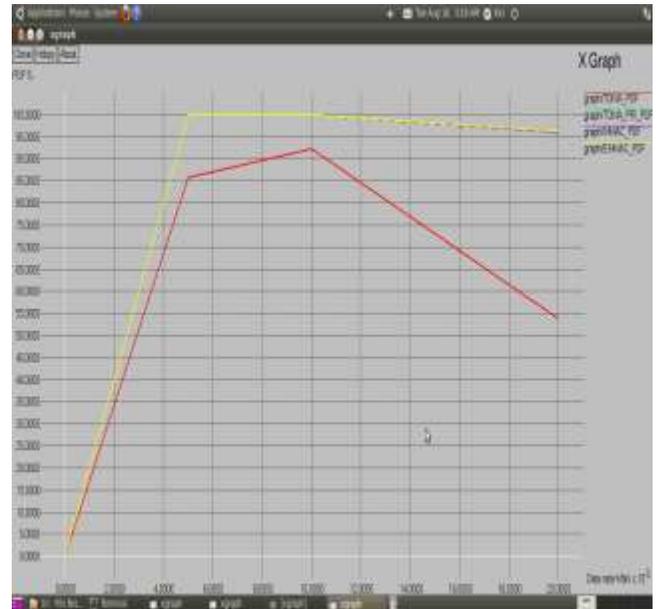


**Fig.6 PDF comparison**

## 5. CONCLUSION

In this paper, we have presented the ACTIVITY CONTROL (Enhanced-IHMAC), a newer low energy with QoS promising medium access control protocol for wireless detector networks. Our proposed ACTIVITY CONTROL is combines Time division and carrier sensing techniques. There are many applications of wireless detector network really wanted to ensure the priority services for the emergency data. Our proposed ACTIVITY CONTROL protocol promises the shorter latency for this type of emergency and latency-sensitive messages. And we have proved our method is lightweight power saving method. In our future work, we will study the details of mobile detector devices.

## 6. REFERENCES

1)      Chenyang Lu, Brian M. Blum, Tarek F. Abdelzaher, John A. Stankovic, Tian He, "RAP: A Real-Time Information sharing Architecture for Prominent-Scale Wireless Detector Networks", 2002.
2)      Giuseppe Anastasi, Marco Conti, Mario Di Francesco, "Extending the Lifetime of Wireless Detector Networks through Adjustive Sleep", 2007

3) *"Maximizing Networking Lifetime in Wireless Detector Networks with Regular Topologies",* Hui Tian, Hong Shen and Matthew Roughan.

4) Ms.Namrata S. Baghele, "An Adjustive Power Aware Device Scheduling Scheme In Grouping for Wireless Detector Networks", 2012

5) *"A Strategic Placement and Group-Leaderer Selection for Wireless Detector Networks",* Tarandeep Kaur, and Jinsuk Baek, Member, IEEE.

6) Ashim Kumar Ghosh1, Anupam Kumar Bairagi2, Dr. M. Abul Kashem3, Md. Rezwan-ul-Islam1, A J M Asraf Uddin1 "Power Efficient Zone Division Multihop Hierarchical Grouping Algorithm for Load Balancing in Wireless Detector Network. ", 2011.

7) *"Power-Efficient Group Formation for Prominent Detector Networks using a Minimum Separation Distance",* Ewa Hansen, Jonas Neander, Mikael Nolin and Mats Björkman.

8) Sinem Coleri Ergen and Pravin Varaiya, "TDMA Scheduling Algorithms for Wireless Detector Networks", 2009

9) Renjie Huang, Wen-Zhan Song, Mingsen Xu, Behrooz Shirazi, Guoliang Xing, "Localized QoS-Aware MAC Protocol for Real-time Dynamic Detector Networks"

10) Y. Liu, I. Elhanany, and H. Qi, "An power-efficient QoS-aware media access control protocol for wireless detector networks," Nov. 2005.