nigeria computer society

www.ncs.org.ng

26th NATIONAL CONFERENCE

Theme:

Information Technology for National Safety & Security

Conference Proceedings

Volume 27

Edited by:

Professor Adesola ADEROUNMU
Dr. Adesina SODIYA

ISSN: 2141-9663

ACKNOWLEDGEMENT

It is with great pleasure that the Nigeria Computer Society (NCS) acknowledges the immense and revered support received from the following organisaions:

CBC
Chams Plc
Computer Professionals Registration Council of Nigeria (CPN)
Computer Warehouse Group
Data Sciences Nigeria Limites
Galaxy Backbone
Main One Limited
National Identity Management Commission (NIMC)
National Information Technology Development Agency (NITDA)
Nigeria Internet Registration Association (NIRA)
Nigerian Communication Satellite (NICOMASAT)
Nigerian Communications Commission (NCC)
RLG Limited
Sidmach Technologies Limited
Systemspecs Limited
Zinox Technologies

We also recognize the laudable efforts of all others in cash or in kind towards the success of the 12th International Conference

REVIEWERS:

Prof. G. A. Aderounmu FNCS
Dr. A.S. Sodiya FNCS
Dr. A. O. Oluwatope
Dr. S. A. Akinboro
Dr. P. A. Idowu
Prof. 'Dele Oluwade FNCS
Dr. (Mrs.) S. A. Bello
Prof. O. S. Adewale
Dr Mrs M. L. Sanni
Dr. F. T. Ibrahalu
Dr. I. Adeyanju
Dr. (Mrs.) S. A. Onashoga
Dr. A. P. Adewole
Dr. I. K. Ogundoyin
Dr. S. E. Adewumi FNCS
Dr. Wale Akinwunmi
Dr L. A. Akanbi
Dr. B. I. Akhigbe
Dr. F. O. Asahiah
Dr. Mrs I. O. Awoyelu
Dr. A. I. Oluwaranti
Prof.. A. T. Akinwale
Dr. E. O. Olajubu
Dr. O. A. Ojesanmi
Dr. E. Essien
Dr. Mrs I. O. Awoyelu
Dr. (Mrs.) O. R. Vincent
Prof. Olumide B. Longe
Dr. A. A. O. Obiniyi FNCS
Dr. F. T. Ibrahalu
Dr. (Mrs.) O. T. Arogundade
Dr. (Engr.) A. Abayomi-Alli
Dr. R. G. Jimoh
Dr. O. J. Oyelade
Dr. A.A. Adeyelu
Dr. A. O. Ogunde
Dr. G. O. Ogunleye

Publication Office:
Nigeria Computer Society (NCS): Plot 10, Otunba Jobi Fele Way, Central Business District,
Behind MKO Abiola Garden, Alausa, Ikeja – Lagos, Nigeria
P.M.B. 4800 Surulere, Lagos, Nigeria. Phone: +234 (1) 7744600, 4538889, 08097744600, 09038353783
E-mail: ncs@ncs.org.ng Website: www.ncs.org.ng

FORWARD

It is our great pleasure and delight to welcome all to the 26th National Conference of the Nigeria Computer Society which is holding at the beautiful capital of Nigeria, Abuja from July 19 to- 21, 2016. The theme of this year's conference is **"Information Technology for National Safety and Security"**. This year's Conference is intended to provide a forum for policy makers, public and private sector, IT practitioners, academia and Information Security Experts to discuss how Information Technology could be used to enhance national safety and security. The conference will also provide opportunities for the delegates to exchange new ideas, establish business or research relations, and find global partners for future collaborations.

In this year Conference, professionals from government circle, industry, research institutes and academia have submitted insightful papers in the areas of National security and safety, e-government, cloud computing, educational technologies and crime detection. In addition to the above, there are will be Cyber Defense programme, Youth Innovation and Entrepreneurship Platform and Special Keynote sessions. Going by the quality of the papers and the personalities presenting lead papers on well researched and challenging issues, I am persuaded that an extremely rich cross-fertilization of ideas of experts from across the globe is guaranteed.

The organizers of the Conference owe special thanks to our national and international guests and lead paper presenters for accepting to be part of this year's Conference. In particular the Secretary to the Government of the Federation, Engr. Babachir David Lawal; Honourable Minister of Communications, Mr. Adebayo Shittu; Executive Vice-Chairman Nigerian Communications Commission (NCC), Professor Umaru Garba Danbatta; Acting Director General, National Information Technology and Development Agency (NITDA), Dr Vincent Olatunji; Director General, National Identity Management Commission (NIMC), Engr. Aliyu Aziz; Managing Director and Chief Executive Officer Galaxy Backbone, Mr. Yusuf Kazaure; Managing Director and Chief Executive Officer NIGCOMSAT, Ms Abimbola Alale; President of International Federation for Information Processing (IFIP), Professor Mike Hinchey.

We also appreciate our partners – Computer Professionals Registration Council of Nigeria (CPN), Sidmach Technologies, Data Sciences, Systemspecs, Main One, Nigeria Internet Registration Association (NIRA), Nigerian Communications Commission (NCC), Galaxy Backbone, Nigerian Communication Satellite (NICOMASAT), National Information Technology Development Agency (NITDA) and National Identity Management Commission (NIMC). It is our prayer that together, we will move Nigeria to a greater height. I wish you all, very exciting and resourceful deliberations and Journey mercies back to your destinations at the close of this Conference.

We wish to appreciate our editors, Dr. A. S. Sodiya and Dr. A. O. Oluwatope; and all reviewers for their efforts at ensuring quality presentations at this Conference. The Chairman, Local Organizing Committee, Mr Rex Abitogun and all members of his committee who have taken the organization of this Conference as task that must be accomplished, thank you all. The National Executive Council of the Nigeria Computer Society is immensely indebted to the dynamic Conference Planning Committee, which worked assiduously, even against odds, to make this reality.

Thank you and God bless you all.

Professor Adesola Aderounmu *FNCS*
President, Nigeria Computer Society

TABLE OF CONTENTES

SESSION A:

National Security and Biometrics

Full Paper

SECRET SHARING SCHEME FOR SECURING BIOMETRIC TEMPLATE

S. O. Asakpa
Computer Science Department,
Federal Polytechnic, Offa
asakpason@yahoo.com

B. K. Alese
Computer Science Department,
Federal University of Technology,
Akure
kaalfad@yahoo.com

O. S. Adewale
Computer Science Department
Federal University of Technology,
Akure
adewale@futa.edu.ng

A. O. Adetunmbi
Computer Science Department,
Federal University of Technology,
Akure
aoadetunmbi@futa.edu.ng

ABSTRACT

Identity theft is a growing concern in this age of digital technology and applications. Traditional authentication methods such as passwords, passphrases, identity documents etc. are not sufficient to combat this menace. Biometrics lends itself as a better and alternative security measure to fight this problem. However, a major challenge to biometric system is the insecurity of biometric templates stored in the database. As a panacea to the vulnerability of templates stored in the database, we proposed a visual secret sharing scheme for images otherwise known as visual cryptography, such that what is stored is a noise like extracted feature of the template. The subject whose biometric data is being processed also hold a noise like portion of the template. Compromise through attacks on the database becomes difficult without the cooperation of the subject under investigations.

KEYWORDS: database, identity theft, secret sharing , security, template.

1. INTRODUCTION

Keeping secret is an integral part of human society. Important things and documents have always been preserved and protected from possible abuse or loss (Sreekumar, 2009). Computer system serves as a good tool for the storage of vital data and information. However, the emergence of the Internet and its allied technologies has brought with it many threats to information security. Prominent among these threats is identity theft. Identity thefts and identity

frauds have cost several people life fortunes (ISAI, 2013; Siciliano, 2014; Justice, 2015). Trusting user identity becomes a difficult challenge due to the problem of identity thefts. This has necessitated the need for tight security to protect essential data and information from adversaries or malicious use.

Various methods have been proposed to secure stored data and information from adversaries. Traditional/classical methods of protecting valuables include the use of locks and keys, passwords and use of tokens. All these methods have various deficiencies that have necessitated the use of a more perceived secured approach known as biometrics. While lock and keys are not directly linked to the owners, passwords can be forgotten and tokens can be stolen. This is because these methods do not have direct relationship between the security mechanisms, the resources being secured and the owners (Asakpa et al., 2014).

Biometrics is able to identify a person or authenticate the identity of a person based on physiological or behavioural characteristics. Commonly used biometric features are fingerprints, iris, face, voice, written signature, deoxyribonucleic acid (DNA), retina, keystroke dynamics, hand geometry, lip motion, palm prints, gait or posture, body odour.

Although biometrics is better than its precursors but it has its major challenges. It is susceptible to various attacks, either external or internal (Hill, 2001; Revenkar et al., 2010; Jain et al., 2011).

When the biometric template is compromised in the database, it may become difficult to access the biometric data. If the template is altered, authorized user may be denied access to the resources, this will lead to false rejection. Similarly, an adversary may be authenticated as a legitimate user resulting in false acceptance error. While false rejection may lead to frustration and eventual abandonment of the system, a false acceptance may be costly in terms of damages and violations of the system (Claus, 2006). For these reasons various researches have been carried out to protect the biometric data (template) in the system by using different approaches such as cryptography, steganography and watermarking (Raphael & Sundaram, 2010; Kapoor & Neha, 2010).

Liu (1968) proposed a secret sharing scheme as a possible solution for keeping documents safe from compromise. This is typically a case involving a group of mutually suspicious individual with conflicting interest, who must cooperate in order to open a lock. Compromise is still possible in this scenario.

In order to ensure the security of biometric template stored in a database, a 2 out of 2 secret sharing scheme is proposed in this paper.

2. BIOMETRIC SYSTEM

The idea of biometrics was introduced in the late nineteenth century by Alphonse Bertillon, a French policeman. Alphonse developed the first set of tools that are collectively called the Bertillonage system, to identify repeat offenders. Bertillonage involved measurement of certain anatomical traits of a person mainly including head length, head breadth, length of the middle finger, the length of the left foot, and the length of the forearm, etc. Figure 1 shows a typical Bertillonage system (Nagar, 2012).

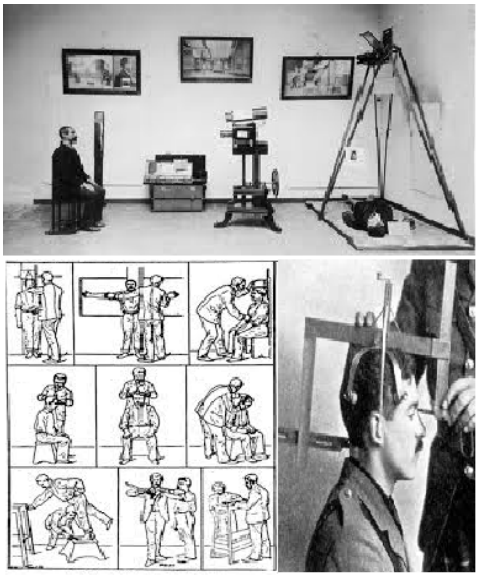


Figure 1: The Bertillonage System

Biometric systems are used in a wide array of applications, which makes a precise definition difficult to establish. Various definitions have been

given but the most acceptable definition of a biometric is (Alese et al., 2012):

"A physiological or behavioural characteristic, which can be used to identify and verify the identity of an individual." According to Mark (2002) there are numerous biometric measures which can be used to help derive an individual's identity. They can be classified into two distinct categories:

Physiological – these are biometrics which are derived from a direct measurement of a part of a human body. The most prominent and successful of these types of measures to date are fingerprints, face recognition, iris-scans and hand scans.

Behavioural – extract characteristics based on an action performed by an individual, they are an indirect measure of the characteristic of the human form. The main feature of a behavioural biometric is the use of time as a metric. Established measures include keystroke-scan and speech patterns.

2.1 BIOMETRIC SYSTEM VULNERABILITIES

The primary reasons for using biometric technology are to arrest criminals, restrict financial fraud, reduce or eliminate identity stealing, protect national borders, or control access to physical facilities and logical resources. When the biometric system fails to meet these objectives, the security of the system is said to be compromised. Such compromise of security can be in the form of denial-of-service to legitimate users, intrusion by unauthorized persons, repudiation claims by authorized persons, or misuse of the biometric data for unintended reasons. Security breakdown can either be due to intrinsic limitations of the biometric system or due to explicit attacks by adversaries, who may be an insider such as an administrator or external attackers. Figure 2 shows the major points of biometric vulnerabilities. However, investigations showed that the major point of compromise is the template stored in the database (Jain et al., 2008).

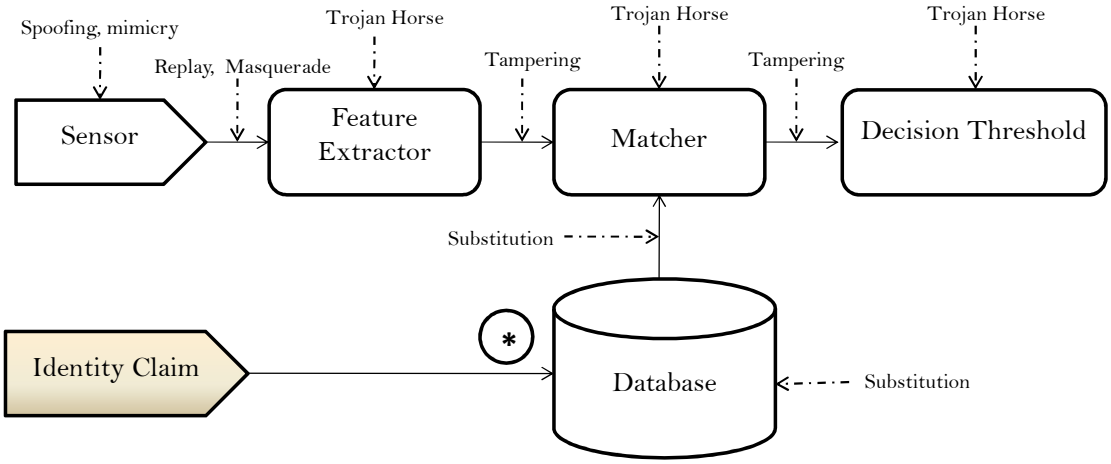


Figure 2: Points of Attack to a Biometric System

3. METHODOLOGY

The methodology employed for this paper is secret sharing scheme. It is a means of sharing a secret message, image, document, etc. among a number of persons such that certain number of legitimate persons can come together in order to

reveal the secret. There are different kinds of secret sharing schemes namely two-party scheme and multiple-party scheme. For this work, we employed a two-party scheme.

Let S be a secret, encoding it as an integer in $\mathbb{Z}/m\mathbb{Z}$. Let $S_1 \in \mathbb{Z}/m\mathbb{Z}$ be generated at random by a trusted party. Then the two shares are defined to be S_1 and $S_2 \equiv (S - S_1)$. The secret is therefore recovered as $S = S_1 + S_2$. This is a simple two-party scheme.

In a multi-party secret sharing scheme, let $S \in \mathbb{Z}/m\mathbb{Z}$ be a secret among n parties. Generate the first $n-1$ shares $S_1, S_2, \cdots, S_{n-1}$ at random and set:

$$S_n = S - \sum_{i=1}^{n-1} S_i \qquad 1.0$$

The secret is recovered using equation 2.0:

$$S_n = \sum_{i=1}^{n} S_i \qquad 2.0$$

The problem with a multi-party scheme is that, a group of corrupt legitimate set may collaborate together in order to reveal the secret. The advantage of our method is that the subject whose biometric data is kept in the database has a share of the secret while the other half is stored in the database. It becomes difficult for the individual to compromise his/her biometric data.

The general idea behind secret sharing scheme is: distribute a secret S to n different participants; any group of t participants can reconstruct the secret; any $t-1$ or fewer participants cannot reveal anything about the secret (Stinson, 2006).

In summary, a scheme could either be a k *out of* k or k *out of* n. Therefore a model of k *out of* k scheme can be represented by equation 3.0:

$$k, k = \begin{cases} m = 2^{k-1} \\ \alpha = \dfrac{1}{2^{k-1}} \equiv \alpha = \dfrac{1}{m} \end{cases} \qquad 3.0$$

However, a model of k *out of* n scheme can be represented by equation 4.0:

$$k, n = \begin{cases} m = \log n \cdot 2^{\Omega(k \log k)} \\ \alpha = \dfrac{1}{2^{\Omega(k)}} \end{cases} \qquad 4.0$$

Where:

- k, is the number of shares;
- m, the number of pixels in a share;
- α, the relative difference in weight between combined shares.

3.1 SHARES DESIGN

A visual secret sharing scheme of a k *out of* n where $k \leq n$ can be constructed as follows.

The basic idea is to construct a polynomial function $f(x)$ of order $(k-1)$ random values $a_0, a_1, \cdots, a_{k-1}$ such that the secret S, equals a_0 as shown in equation 5.0:

$$f(x) = a_0 x^0 + a_1 x^1 + a_2 x^2 + \cdots + a_{k-1} x^{k-1} (mod\ p) \qquad 5.0$$

The polynomial function $f(x)$ is destroyed after each shareholder possesses a pair of value $(x_i, f(x_i))$ so that no single shareholder knows the secret value a_0.

The dealer chooses a prime number p, which is greater than n and the set of possible secret and non-zero distinct elements $x_i \in \mathbb{Z}/_p\mathbb{Z}$, $1 \leq i \leq n$. The share creation process is given thus:

1. The dealer secretly at random chooses elements $x_i \in \mathbb{Z}/_p\mathbb{Z}$, $1 \leq i \leq k-1$ and constructs the polynomial (6.0):

2 $$a(x) = \sum_{i=1}^{k-1} a_i x^i + S \quad 6.0$$ ıares

3. The dealer distribute the share (x_i, y_i) to the i^{th} shareholder, $1 \leq i \leq n$.

3.2 SECRET RECONSTRUCTION

In order to reconstruct the secret S, any k or more shares will be combined together. Suppose the shares are numbered, such that $y_i = ax_i$, $1 \leq i \leq k$ with the polynomial $a(x)$ from equation 6.0. Applying the Lagrange interpolation gives equation 7.0:

$$a(x) = \sum_{i=1}^{k} y_i \prod_{j=1, j\neq i}^{k} \frac{x_j - X}{x_j - x_i} \quad 7.0$$

This polynomial satisfies $a(x_i) = y_i, 1 \leq i \leq k$ and there is exactly one such polynomial of degree $\leq k-1$. Therefore, the shareholders can reconstruct the secret as represented by equation 8.0:

$$S = a(0) = \sum_{i=1}^{k} y_i \prod_{j=1, j\neq i}^{k} \frac{x_j}{x_j - x_i} \quad 8.0$$

4. EXPERIMENT AND RESULT

An application software was developed to simulate the proposed system using MATLAB 7.10.0 (R2010a) and Image Processing Toolbox. The platform for the experiment was Windows 8 Operating System, running on Intel(R) Pentium(R) Duo Processor at 2.40GHZ, and 4.00GB of RAM. The size of each image used is 512 x 512 pixels. Given a secret image represented by a binary string $k = k_1, k_2, \ldots, k_n$, we created two shares, $x = x_1, x_2, \ldots, x_n$ and $y = y_1, y_2, \ldots, y_n$, where x_i is random and $y_i = k_i \text{ XOR } x_i$. We applied equation 6.0 for shares design (encryption) and equations 7.0 and 8.0 for shares reconstruction (decryption). Figures 3 and 4 show samples of two different biometric images captured from two different subjects. The images are of the same dimension. The input image is the captured biometrics while the output shows the reconstruction of the input image from the shares 1 and 2. The subject under investigation holds a share while the other share is stored in the system database. In order to recover the original biometric image, the two shares must be merged together. A single share in isolation is meaningless. It must be combined in order to become meaningful.

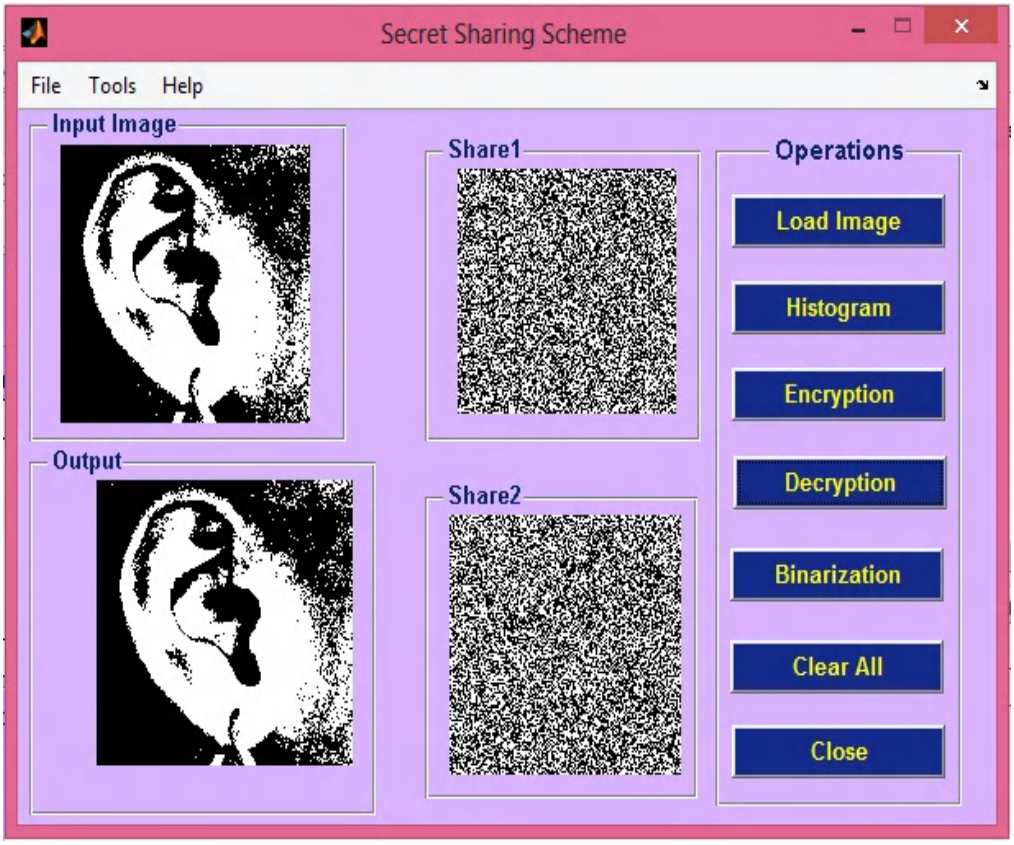


Figure 3: Share Design and Reconstruction for Ear Biometrics

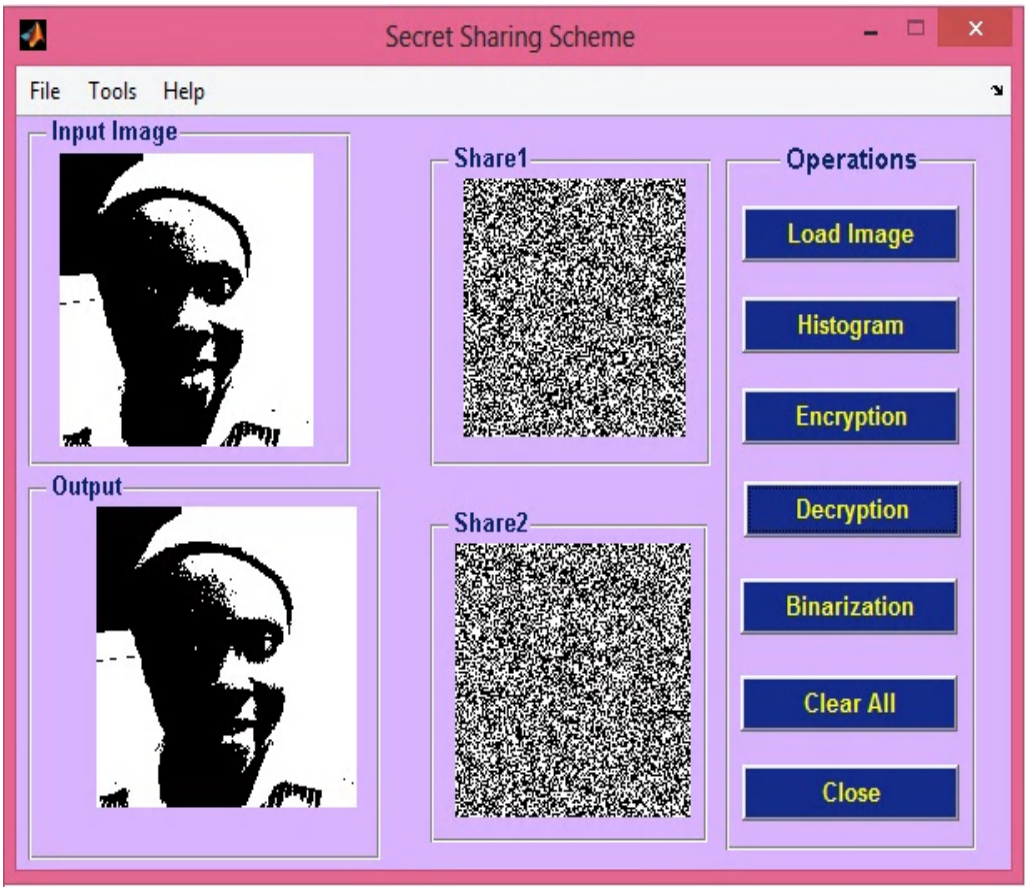


Figure 4: Share Design and Reconstruction for Face Biometrics

5. CONCLUSION

One of the major loopholes to security compromise is the presence of an insider. We have been able to design and implement in this work a secret sharing scheme that is difficult to penetrate by a third party. No individual will want to compromise his/her biometric data because of the implications of such act. This scheme can therefore be applied in the storage of very sensitive information such as biometric template in the database. With this, the issue of identity theft due to biometric manipulations would have been reduced if not totally eliminated.

6. REFERENCES

Alese, B. K., Mogaji, S. A., Adewale, O. S., & Daramola, O. (2012). Design and Implementation of Gait Recognition System. *International Journal of Engineering and Technology, 2* (7), 1102-1110.

Asakpa, S. O., Alese, B. K., Adewale, O. S., & Adetunmbi, A. O. (2014). Secure face authentication using visual cryptography. *In N. C. Society (Ed.). 25*, pp. 61-66. Asaba: NCS.

Claus, V. (2006). Biometric User Authentication

for IT Security, from Fundamentals to Handwriting.

Hill, C. J. (2001). *Risk of masquerade arising from the storage of biometrics*. Australia: Master's thesis, Australian National University.

ISAI. (2013, October 14). *Crazy cases of Identity Theft*. Retrieved May 7, 2016, from International Conference on Information Security and Artificial Intelligence: ttp://www.isai2010.org/10-crazy-cases-of-identity-theft/

Jain, A. K., Nandakumar, K., & Nagar, A. (2008). Biometric template security. *EURASIP Journal on Advances in Signal Processing, Special Issue on Biometric*.

Jain, A. K., Ross, A. A., & Nandakumar, K. (2011). *Introduction to Biometrics*. New York, USA: Springer.

Justice, U. D. (2015, November 2). *Identity Threat*. Retrieved May 8, 2016, from Department of Justice Web site: www.justice.gov/criminal-fraud/identity-theft/identity-theft-and-identity-fraud

Kapoor, S. D., & Neha, b. (2010). Proposed System for Data Hiding Using cryptography and steganography. *International Journal of Computer Applications (0975 – 8887), Volume 8 – No. 9*.

Liu, C. L. (1968). *Introduction to Combinatorial Mathematics*. New York: McGraw-Hill.

Mark, R. D. (2002). *Gait Recognition Master's Thesis*. London: Department of Computing Imperial College of Science.

Nagar, A. (2012). *Biometric Template Security*. Michigan: Michigan State University.

Raphael, J. A., & Sundaram, V. (2010). Cryptography and Steganography – A Survey. *International Journal of Computer Technology and Applications , 2 (3)*, 626-630.

Revenkar, P. S., Anjum, A., & Gandhare, W. Z. (2010). Secure Iris Authentication using Visual Cryptography. *International Journal of Computer Science and Information Security*, 217-221.

Siciliano, R. (2014, June 30). *Unbelievable Identity Theft Cases*. Retrieved May 10, 2016, from Huffpost Crime: http://www.huffingtonpost.com/robert-siciliano/10-unbelievable-identity_b_5239159.html

Sreekumar, A. (2009). *Secret Sharing Schemes using Visual Cryptography*. India: Cochin: PhD Thesis Cochin University of Science and Technology.

Stinson, D. R. (2006). *Cryptography: Theory and Practice*. 2nd ed.: CRC Pres

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

A SECURED VOTING SYSTEM USING FACE BIOMETRIC AND TEXT SEMAGRAM TECHNIQUES

E. A. Salako
Department of Computer Science
FCT College of Education, Zuba-Abuja
Nigeria
kunlesky2@gmail.com,
myadekunle@salakoadekunle.org

ABSTRACT

Election is a formal decision-making process by which a populace chooses an individual to hold public offices. The paper-based voting system is common and insecure from many politicians to manipulate the results of elections; leading to bad governance. Existing biometric voting systems used technical type of steganography to secure election results. However, technical steganography requires more lines of instructions and higher memory capacity to hide information from unauthorized populace. This research focused on a secured voting system using face biometric and text semagram techniques. This research attempts to improve on authentication and confidentiality of electronic voting system. Face feature extraction was to achieve authentication and text semagram technique was used to achieve confidentiality. Developed e-voting system was tested by some selected students of FCT College of Education, Zuba-Abuja, Nigeria. The results of administered questionnaire were evaluated using mean. Developed e-voting system was able to authenticate the voter's identity, prevent multiple registration and multiple voting; as voters were required to use the captured face image for verification. Developed e-voting system was remarked by the respondents to be secured approach of conducting election. Based on these findings, some recommendations were made which include: the voter's face image was required to allow the voter to cast vote.

KEYWORDS: SECURED, VOTING SYSTEM, FACE BIOMETRIC, TEXT SEMAGRAM

1. INTRODUCTION

Election is a formal decision-making process by which a populace chooses an individual to hold public offices. Elections in Nigeria are forms of choosing representatives to the Nigerian federal government and the various states in different capacity. Elections may involve public or private vote depending on the position. Most positions in the local, state, and federal governments are public; where the general populace chooses their representatives with the aim of achieving good governance that is geared towards the populaces' and national development. According to Donald and James (2003), the aims of an electoral system include proportionality of seats to votes, accountability to constituents, durable governments, interethnic and inter-religious conciliation and minority office-holding. The challenges listed by Iwu (2008) included: insecurity, poor funding, attitudes of political class, apathetic and inactive citizenry, delay in amendment to the legal framework, completion of the review of electoral constituencies and polling units and prosecution of election offenders. It is very obvious that traditional methods of voting using papers, ballot boxes and manual counting of votes has made the populaces to lose confidence in the integrity and sincerity of the election in Nigeria.

With the advent of computer technology such as electronic voting (e-voting), the challenges recorded in the past elections can be solved. The technology has the capacity to increase the speed of vote counting; incorporate the broadest assistive technologies for the largest classes of handicapped people, allowing them to vote without forfeiting the anonymity of their vote. Electronic voting (e-voting) is voting using electronic systems to aid casting and counting votes. Electronic voting technology can include punched cards, optical scan voting systems and specialized voting kiosks (direct-recording electronic (DRE) voting systems). It can also involve transmission of ballots and votes via telephones, private computer networks, or the Internet. In Bangor, Kortum and Miller (2011), computer voting system provides the most robust form of immediate feedback to the voter detecting such possible problems as under-voting and over-voting which may result in a spoiled ballot. This immediate feedback can be helpful in successfully determining voter intent.

Five different types of voting systems had been highlighted by Firas and Seifedine (2012). These types are:

i. Paper-based Voting Systems (PVS): This type of voting system records, counts, and produces a tabulation of the vote count from votes that are cast on paper cards or sheets.

ii. Direct-recording Electronic (DRE) voting systems: The system records votes by means of a ballot display provided with mechanical or electronic optical components which could be activated by the voter.

iii. Public network DRE voting systems (PNDRE): This type of voting makes use of electronic ballots and transmit vote data from the polling stations to other locations over a public network.

iv. Precinct Count Voting Systems (PCVS): The voting system puts the ballots in a tabular form at a particular place, say, a polling station.

v. Central count voting systems (CCVS): Voted ballots are safely stored temporarily at the polling station. These ballots are then transported or transmitted to a central counting location. CCVSs may, in some cases, produce printed reports on the vote count.

Existing electronic secured voting systems used technical type of steganography to secure election results. However, technical steganography requires more lines of instructions and higher memory capacity to hide information from unauthorized populace. Hackers in the world have been deploying approaches to detect information embedded using technical steganography. This research attempts an application of different type of steganography called linguistic steganography to hide information from unauthorized populace. Text semagram is a type of linguistic steganography that hides information in modified text. Text semagram requires lesser computational time and lines of instructions in comparison with technical steganography.

2. REVIEW OF RELATED WORKS

In Sanjay and Manpret (2013) work, fingerprint biometric was integrated and

configured with a microcontroller. Secured identification and authentication were achieved in the design. However, problem of data confidentiality was not provided. In this proposed e-voting system, text semagram would be used to achieve confidentiality.

Priyanka, Pooja, Bhagyashri, Rutuja and Priyanka (2013) designed a method of integrating steganography and biometrics to secure online voting system. The voters cast their vote anywhere and the security of the system was preserved by producing cover image for individual voter. The fraudulent acts of multiple registration and voting by a voter could be practiced in the design of Priyanka, et al, (2013). Prevention against multiple registration and voting would be provided in this new proposed e-voting system.

Suresh, Suganthi and Mohanas (2012) designed a multimodal biometrics (fingerprint and hand) was used for authentication. The e-voting system was able to achieve authentication. However, the problem of confidentiality was not addressed. In this new proposed system, text semgram provides confidentiality as part of e-voting requirements.

In Firas, Seifedine and Oussama (2012) voting system, application of fingerprint biometric in order to provide a high performance voting system was provided. The system provided easy way to cast votes. However, the problem of data confidentiality was not provided. The new proposed e-voting system would provide confidentiality.

In Noha, Rabab and Mahmoud (2013), Eigenface filter was used for face verification. The system was able to recognise face of individual voter as part of e-voting requirements. However, the problem of confidential was not addressed. This implies that the results of the election can be manipulated. In this paper, problem of confidential and integrity would be provided to secure the election results from manipulation.

3. STATEMENT OF PROBLEM

For many years in Nigeria, paper-based ballot has been used as a method to vote during elections. This method put an inefficient way of voting process as Nigerians have to queue up to register their names before they can vote. Furthermore, the traditional way of voting would take a long process and time; resulted to stress and ineffective in handling large data. Many politicians (including students) use this opportunity to manipulate the results of elections; leading to bad governance and under-development in the school. How to curb such acts from the politicians is to design a system that is capable of securing electoral data from unauthorized personnel.

In this technological era, there should be a system that provides solutions to problems highlighted above; secure the results (data) of elections using semagram technique. The technique makes the results unreadable by unauthorized individual who may has intension to rig (alter) the results. It is based on these identification and understanding of the problems of election in Nigeria that this research focuses on integration of face biometric and text semagram techniques into voting system for data security.

4. RESEARCH METHODOLOGY

The proposed system can be divided into different units. The figure 7 shows the functional diagram of the proposed e-voting system.

4.1 Registration Unit

The registration process involves in using camera for image acquisition. First, the face image is captured and extracted. The face image template is then registered and stored in a database. The face template is then processed and extracted. It will subsequently match the scanned face image against the stored template. Upon authentication, the voters will have the access to vote for their desired candidates. The unit also acquires the voter's data such as name, gender, date of birth.

4.2 Face recognition

Edge detection analysis can be performed on the face image acquired. Region boundaries and edges are used to describe the image pattern. The figures 1 shows face detection analysis while figure 2 shows edge detection analysis.

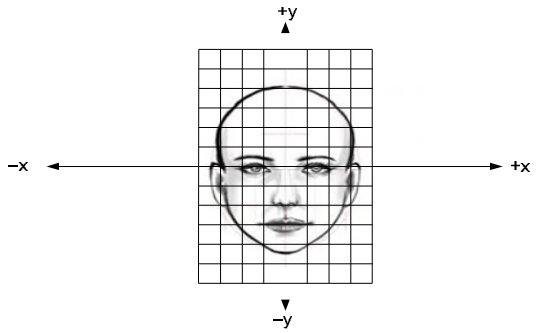


Figure 1: Face detection analysis on face image

i. Determine the centre of the image (origin)
ii. From the centre, locate the edges
iii. Determine the numbers and locations of edges from the centre
iv. The information is stored in the database for matching and future

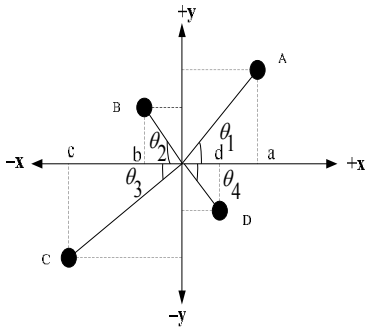


Figure 2 : Edge detection points on x-y axes

The points (edges) of a face image can be expressed with respect to x-coordinate, y-coordinates and the phase (θ).

At point A; the phase (θ_A) with respect to x-axis is:

$$\text{Tan}\ \theta_A = \frac{y_a}{x_a}$$

$$\theta_A = \text{Tan}^{-1}\left[\frac{y_a}{x_a}\right]$$

At point B; the phase (θ_B) with respect to x-axis is:

$$\text{Tan}\ \theta_B = \frac{y_b}{-x_b}$$

$$\theta_B = \text{Tan}^{-1}\left[\frac{y_b}{-x_b}\right]$$

At point C; the phase (θ_C) with respect to x-axis is:

$$\text{Tan}\ \theta_C = \frac{-y_c}{-x_c} = \frac{y_c}{x_c}$$

$$\theta_C = \text{Tan}^{-1}\left[\frac{y_c}{x_c}\right]$$

At point D; the phase (θ_D) with respect to x-axis is:

$$\text{Tan}\ \theta_D = \frac{-y_d}{x_d}$$

$$\theta_D = \text{Tan}^{-1}\left[\frac{-y_d}{x_d}\right]$$

4.3 Authentication Unit

The voter's face is required for authentication. The system denies the access to vote if the voter's face image does not matched with the already registered face template in the database.

4.4 Text Semagram Technique

The semagram technique for the design can be illustrated as it applies to the encoding and decoding stages of the proposed e-voting system.

4.5 Encoding stage

The semagram is a type of steganography using icons (symbols) to carry and hide a message from unauthorized peoples. The stage deals with how to get the "semagotext" from plaintext. The ciphertext is a "semagotext". The plaintext is the original message which is to be transmitted while the ciphertext is the translated or encoded message.

i. Let the key (Matrix A) be represented as:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

ii. Multiply the Matrix A by the Matrix M (plain or original text) to obtain Matrix E.

$$AM = E$$

iii. Multiply each element of Matrix E by the sum of its column position and the product of 5 and its row position to get Matrix E_M (Manipulated Matrix).

Mathematically;

$$E_M = E[5R_p + C_p]$$

Where:

R_p denotes the row position of element E;

C_p denotes the column position of element E.

iv. The result in (iii) above is use to obtain the "semagotext" (Matrix E_S) to be transmitted through the unprotected network, and any number (dividend) in the Matrix E_M that is greater than the total number of symbols in the semago table (64) is divided by 64 and the remainder (if any) is use to locate the character from the table. The expression is written as:

$$E_S = \frac{E_M}{64} = \text{Quotient} + \text{Remainder}$$

4.6 Decoding stage

Decoding is the opposite process of encoding. The decoding is scientific process of converting an encoded message ("semagotext" or ciphertext) back into the original sequence of characters (plaintext). Encoding and decoding are used in data communications, networking, and storage. For this design, the technique is done in such a way that any numerical value in the matrix (E_S) is multiply by the total number of character in the semago table (64); the value of the symbol that followed the numerical value is obtained from the table and is added.

In decoding, to obtain Matrix E from the Matrix E_M, each element of Matrix E_M is obtained by this equation:

$$E = \frac{E_M}{[5R_p + C_p]}$$

Where:

R_p denotes the row position of element E;

C_p denotes the column position of element E

Multiply the inverse of the key matrix A (A^{-1}) to obtain the plaintext (original text) from the unprotected network.

If;

$$AM = E$$

Then;

$$A^{-1}AM = A^{-1}E$$

$$IM = A^{-1}E \quad \text{(I is the unit matrix)}$$

Therefore, the plaintext (Matrix M) can be obtained by:

$$M = A^{-1}E$$

If the key (Matrix A) equals:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

Then there exists an inverse matrix of A, and it is;

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix}$$

4.7 Algorithm for the proposed system

The algorithm for the proposed system is showing below. The algorithm is capable of encrypting and decrypting election results from modifications.

```
Start
        If (open = 1) Run = 1; message
// Welcome screen message
        If (open = 0) Cancel = 0; Quit
        message = "Are you sure you
want to run this software?"
        End if
        m = "..... Text
lines....." // texts
```

```
        texts = "Do you want to encrypt
or decrypt?" // User response
        input = Dialogue (texts,
"Encrypt", "Decrypt", Cancel)
        If (input = 1) Quit  // application
terminates
            If (input = 2); output =
        "Decrypted.txt" // output file
            If (input = 3;  output =
        "Encrypted.txt" // output file
            End if
        Do p = 0, len(m)-1
          character  = extract(m, p) //
        extraction of characters
          ascharacter = ascii(character)
          If as character >= ascii("A")
        and as character <= ascii ("Z")
            result = random(26) // 26
        letters of alphabets
            ascharacter =
        cycle(ascharacter -
        ascii("A")+result,26)
            character = character
        (ascharacter +ascii ("A"))
          Endif
          Show character,  // the result
        show on the user's screen
        Loop p
End
```

4.8 Flowchart

The flowchart of the proposed e-voting system is illustrated in figure 3.

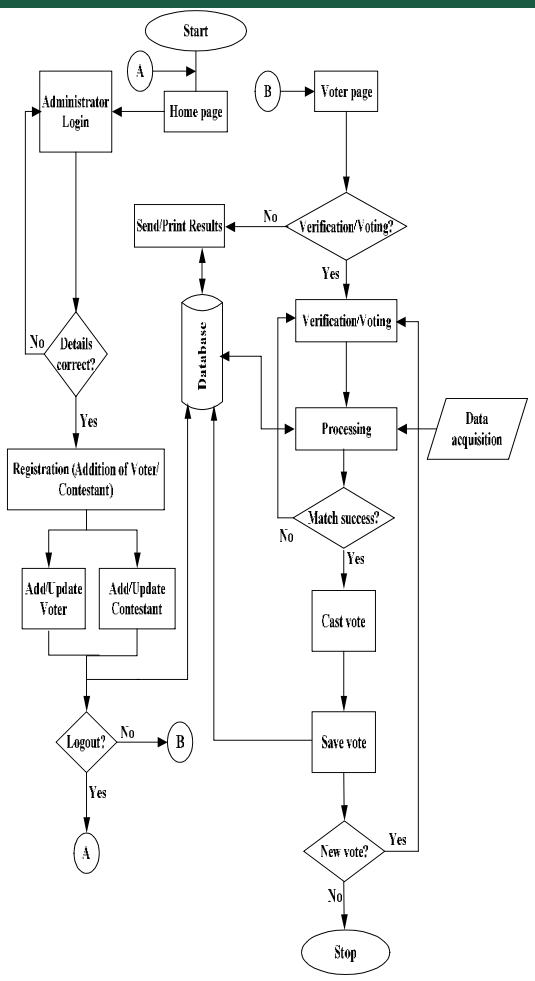


Figure 3: Flowchart

4.9 Source codes

The source code of the proposed e-voting system was developed using visual C# in Microsoft visual studio 2010 ultimate version and Microsoft SQL Database was for the database management system to store all the data used ranging from configuration data such as Candidates, Parties and the election data were all stored on Microsoft SQL database.

5. RESULTS AND DISCUSSION

The figures 4, 5 and 6 show proposed registration page, proposed system operations screenshot and election result screenshots respectively. As part of testing the security requirements of e-voting system, an election was conducted for some selected students of FCT College of Education, Zuba-Abuja, Nigeria. The results of the election was embedded and extracted using text semagram technique. Salako (2015) questionnaire was adopted to gather relevant information about the developed e-voting system from the respondents. The results was evaluated in IBM Statistical Package for Social Sciences (SPSS) 21 Version to obtain the mean scores.

Table 1 shows mean evaluation metrics on authentication of the developed e-voting system. The findings from the evaluation of the authentication security requirement of the developed e-voting system indicated that the evaluated means were greater than the expected minimum mean of 3.00. This implies that the developed e-voting system satisfied the security requirement of authentication. The respondents' rating of the developed e-voting system on System Degree of Voter's Authentication Index (SDVAI) indicated 3.37 out maximum obtainable value of 4.00.

Table 1: Mean evaluation metrics on authentication of the developed e-voting system

s/n	Items	Observed $\bar{X}$	SDVAI
a.	The developed system provides interface that can be used easily for enrolment and verification.	3.32	
b.	The VIN and fingerprint verifies every individual correctly.	3.34	
c.	The developed system prevents false identity.	3.46	*3.37
d.	The VIN and fingerprint features prevent all unauthorised attempts to cast votes.	3.31	
e.	The developed system provides extremely accurate and secured access to voting procedures.	3.42	

N=120, Expected $\bar{X}$ =3.00

Table 2 shows mean evaluation metrics on confidentiality of the developed e-voting system. The findings from the evaluation of the confidentiality security requirement of the developed e-voting system indicated that the evaluated means were greater than the expected mean of 3.00. This implies that the developed e-voting system satisfied the security requirement of confidentiality. The respondents' rating of the developed e-voting system on System Degree of Confidentiality Index (SDCI) indicated 3.21 out maximum obtainable value of 4.00.

An attempt was made to open an encrypted election file; the figures 8 and 9 were the results. This implies that the election results were secured from fraudulent acts.

Table 2: Mean evaluation metrics on confidentiality of the developed e-voting system

s/n	Items	Observed $\bar{X}$	SDVAI
a.	The election results cannot be viewed by any unauthorized populace	3.12	
b.	The controlling embedding and extraction techniques used protect the secrecy of election result	3.25	
c.	The developed system provides secrecy of election results which cannot be disclosed by unauthorized or government officials	3.08	3.21
d.	The developed system makes the votes casted remain secret	3.17	

e.	The developed system has the potential of restoring confidence for free and fair election	3.43	

N=120, Expected $\bar{X}$ =3.00

6. CONCLUSION

Every voter in an election has right to decide who could occupy elective positions at different levels of government. This fundamental right has been abused by some politicians to forcefully and illegally get positions for themselves. Problems of rigging, false identification, multiple registration and multiple voting had been identified, and different methods to curb these unaccepted acts were reviewed. The study focused on a secured voting system using face biometric and text semagram techniques. The developed system was capable of authenticating the voters correctly, prevent multiple registration, multiple voting using face biometric and secure election results from unauthorized personnel using text semagram techniques, the system was remarked by the respondents to be secured, accurate and convenient ways of conducting voters' registration, voting and collating election results from different polling units.

7. RECOMMENDATIONS

Based on the testing and results obtained from the developed e-voting system, the following recommendations are hereby made: during pre-election stage (enrolment), the voter's fingerprint images is fundamentally required. In the election stage, the acquired face image is required for verification of voter's identity and for voting towards achieving credible elections that is driven by text semagram technique in securing election results from fraudulent acts.

8. REFERENCES

Bangor, A., Kortum, P. T. and Miller, J. T., 2011. An empirical evaluation of the system usability scale (SUS). *International Journal of Human-Computer Interaction*. Retrieved on 23rd October, 2014 at http://chil.rice.edu/research/pdf/everettgreene.pdf.

Donald, L. H. and James, B. D., 2003. Electoral system and their goals: a primer for decision-makers. Department of Law and Political Science, Duke University Durham, North Carolina 27708-0360, U.S.A. Retrieved on 23rd October, 2014 at http://aceproject.org/ero-en/topics/electoral systems/E6ElectoralSystemsHorowitz.pdf

Firas, H. and Seifedine, K., 2012. New system of e- voting using fingerprint. *International Journal of Emerging Technology and Advanced Engineering*. 2 (10), pp. 355-363

Firas, I. H., Seifedine, K. and Oussama, K. Z., 2012. Web-based voting system using fingerprint: design and Implementation. *International Journal of Computer Applications in Engineering Sciences*. 2 (4), 404-409

Iwu, M., 2008. *How we conducted April general elections*. Retrieved August 27, 2014, from http://www.nigeriamuse.com/archives?text=democracy&bt=1

Noha, E. E., Rabab, F. A. and Mahmoud, I. M., 2013. Face recognition as an authentication technique in electronic voting. *International Journal of Advanced Computer Science and Applications, (IJACSA)*, 4(6), pp. 66-71.

Priyanka, C. Pooja, C. Bhagyashri, P. Rutuja, P. and Priyanka, M. (2013). Onvote–secured online voting. *International Journal of Innovative Research in Computer and Communication Engineering*. 1 (9), 2117-2120.

Salako, E. A. (2015). Development of a secured electronic voting system using fingerprint biometric and visual semagram techniques. An unpublished thesis submitted to the Department of Computer Science, Federal University of Technology, Minna,

Nigeria.
Sanjay, K. and Manpret, S., 2013. Design a secure
electronic voting system using fingerprint technique. *International Journal of Computer Science Issues (IJCSI)*, Vol. 10, Issue 4, No 1, July
2013
Suresh, N. Suganthi, S. and Mohanas, D., 2012. Multimodal biometric authentication parameters
on humanbody. *International Journal of Engineering Research and Applications (IJERA)*,Vol. 2, Issue 3, May-Jun 2012, pp. 331-337

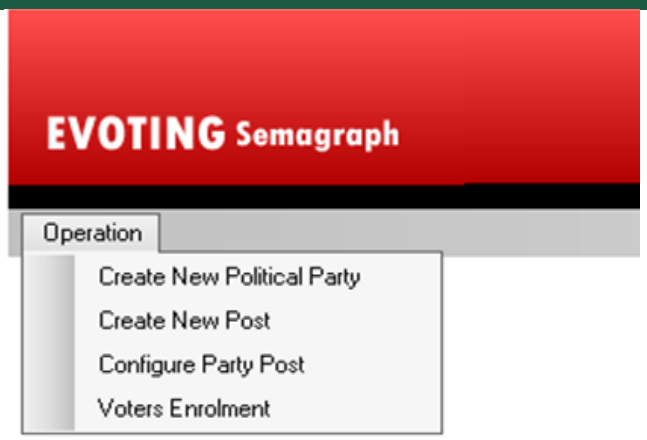


Figure 5: Proposed system operations screenshot

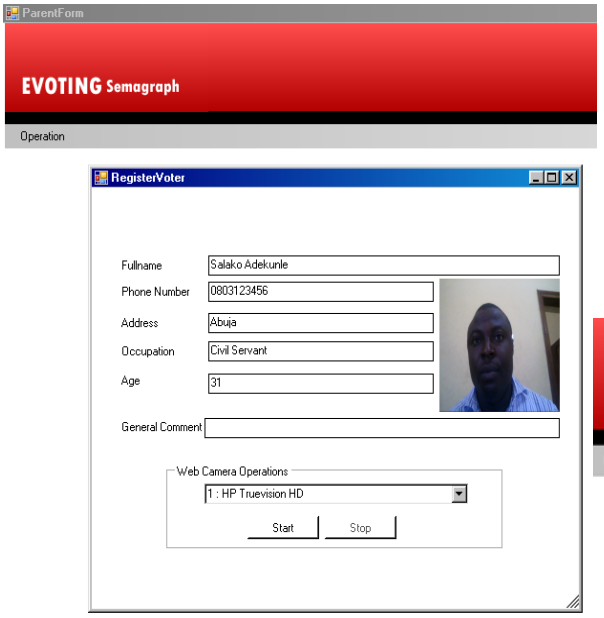


Figure 4: Proposed registration page screenshot

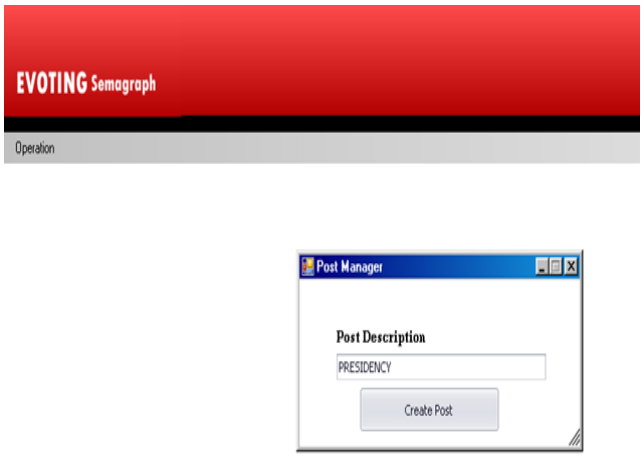


Figure 6: Election result screenshot

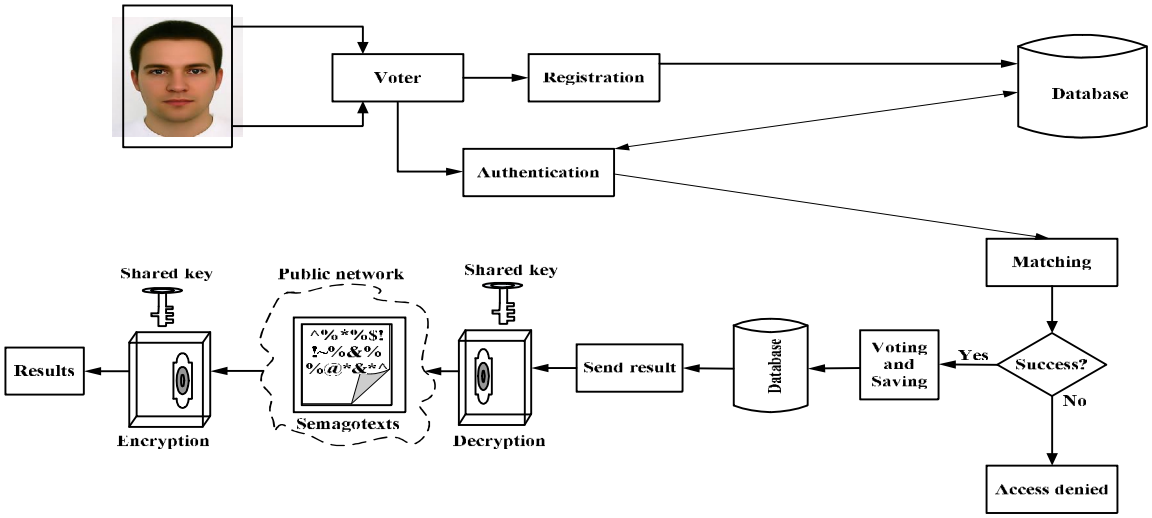


Figure 7: Functional diagram of the proposed e-voting system

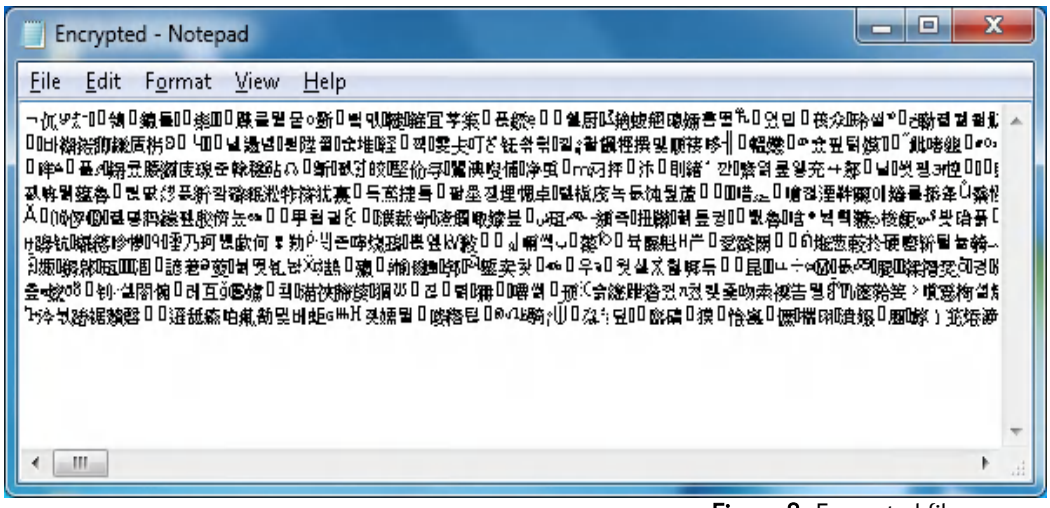


Figure 8: Encrypted file

Figure 9: Encrypted file

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

BIOMETRIC BASED INTEGRITY CONTROL SYSTEM FOR THE NATIONAL PENSION COMMISSION IN NIGERIA

E. S. Alu
Department of Agricultural Economics and Extension,
Nasarawa State University,
Shabu-Lafia Campus, Nasarawa State
Estheralu@gmail.com

D. E. Aniobi
Department of Mathematics, Statistics and Computer Science,
Federal University of Agriculture,
Makurdi, Benue State
david.aniobi@gmail.com

ABSTRACT

The conventional systems in pension administration such as possession of identity card, paper documents to verify pensioners by the respective pension fund administrators is prone to a lot of inadequacies, such as fraud and identity theft. In order to overcome this problem, this research proposes an alternative method in the area of Biometrics technology using advanced computer techniques such as face and fingerprint as a widely adopted front-line security. This is done by developing a biometric system for pensioners which will ensure efficiency and effectiveness in identity verification and control access to pension funds. The method adopted involved study of the existing system and evolutionary prototyping of the new system; C sharp (C#) and MySQL were the technologies used in implementing the system. It is believed that this research will go a long way at eliminating ghost pensioners and restore credibility in Nigeria's pension administration.

Keywords: Biometric technology, Control system, Computer techniques, Pension administration

1.0 INTRODUCTION

The increasing number of older persons has generated a number of challenges that is affecting a number of social institutions such as retirement within the world of work. Nigeria as a country is also affected in addressing the threat posed by increasing number of retirees.

The Federal Government of Nigeria enacted the Pension Reform Act in 2004, a departure from the former retirement scheme that is none contributory (*www.Newage-online.com*). Despite this noble idea, pension administration in Nigeria has been mired by corruption, frequent cases of ghost pensioners, pensioners not receiving their money as at when due, others not even receiving their money at all, among other

anomalies. The emerging trend in organizations is the security of physical, financial, and information assets. Lapses in security such as unauthorized personnel gaining access to an organization's facilities and schemes can have serious consequences that extend beyond the organization. Organizations need to have an absolute trust in the identity of their employees, customers, contractors, and partners. This issue cannot be overemphasized especially when people's money or live savings are involved [1]. It would be disastrous if after working for thirty-five years, all ones' monthly contribution was given to someone who claims to be who he is not or the money developed wings and fly. Or a group of unauthorized persons goes away with all the money as was reported recently where billions of pension funds were embezzled [2]. These are pitiable common trends in the present pension administration which requires urgent attention. If there is a robust biometric pension authentication system, these trends would be curtailed to their barest minimum.

Today, there are many biometric devices based on characteristics that are unique for everyone. Some of these characteristics include but are not limited to fingerprints, face, hand geometry, and voice. These characteristics can be used to positively identify someone. Many biometric devices are based on the capture and matching of biometric characteristics in order to produce a positive identification. By employing a biometric device or system of devices inside the pension system, it will enable organizations to know exactly who is an employee.

Every biometric device or systems of devices include the following three processes: enrollment, live presentation and matching. The time of enrollment is when the user introduces his or her biometric information to the biometric device for the first time. The enrollment data is processed to form the stored biometric template. Later, during the live presentation, the user's biometric information is extracted by the biometric device and processed to form the live biometric template. Lastly, the stored biometric template and the live biometric template are compared to each other at the time of matching to provide the biometric score or result.

A. Problem Statement

Currently, personnel identification for the access to pension funds relies on the use of PIN or password (to login on the pension fund administrator's web site), Identity cards and token. These besides being inconvenient and vulnerable to manipulations and fraud, does not identify the person but simply identify the information that is provided by that person. Biometrics offer automated method of identity verification on the principle of measurable physiological or behavioral characteristics and are believed to be unique to every individual. This type of identification would be more reliable when compared with traditional verification methods such as possession of an object like a key or swipe card, or the knowledge of a password or login to access a scheme, because the person has to be physically present at the time of identification. To achieve a more reliable verification or identification process, this research seeks to use traits that really characterize the given person such as fingerprint and face image. This is done by using biometric authentication method which will eliminate or minimize the problems associated with the conventional system.

B. Aim and Objectives of the Study

The aim of this work is to develop a biometric system for PENCOM which will capture data of both current and would-be pensioners of the Federal Ministry of Transportation which would ensure that only legitimate persons have access to their respective pension funds.

The objectives of the system are as follows:

i. Create a system that ensures efficiency and effectiveness in identity verification and control access to pension funds.
ii. Build a system that provides a platform for pensioners to change from one Pension Fund Administrator to another and model a system such that every pensioner with the Federal Ministry of Transport can have access to their pension funds.

C. Scope and Delimitations

Against the backdrop of the severe lapses observed in pension administration in the country, individuals in the country's pension administration have unauthorized access to pension funds and pensioners cannot be properly identified. The major area of this work lies in the necessity to articulate a new approach for pensioners' identity in pension administration. Its major focus is to design a biometric recognition system using National Pension Commission (PENCOM) as a case study.

D. Motivation/Justification of the study

Previously, Nigerian lives have been lost as a result of the so-called pensioners' verification exercise where they would have to queue up for several hours or days before their identity can be verified. This sometimes lead to loss of pensioners' lives as a result of dehydration and fatigue after waiting several days in a queue just to have their identity verified. This situation is not only alarming but also disgraceful considering the fact that most of these pensioners do not reside in the area where the verification exercise is being conducted. Many had to travel several kilometers before arriving at the verification centre thereby putting their lives at risk of road, air or sea accidents depending on their means of transportation.

Some pensioners were not fortunate enough as those who could no longer locate their identity card and documents were sent back home with nothing. Most at times, pensioners cried and rained abuses on the government for not having a proper means of identifying those who labored for the economic growth of the country. This ugly scenario endeared in me the need to embark on a research that would eliminate ghost pensioners and the use of paper identification. The study introduced an easier, reliable and effective means of identifying pensioners at their respective states and federal levels. The new method will go a long way in eliminating the aforementioned problems associated with the old system of verification.

2.0 RELATED WORK

A. The Need and Use of Biometrics

There is an increasing interest in biometrics technology for crime control and identity credential. This interest brought about the need to reliably verify or confirm people and to control identity fraud. It also monitors online banking and e-commerce. The growing threat of global terrorism makes it an imperative to implement biometrics technology to support identity management [3].

The European Commission supported this argument and stated that the ability of biometrics to increase trust in identity authentication is their greatest advantage. They also stated that ensuring the identity and authenticity of persons is a prerequisite to security and efficiency in modern business operations. Unauthorized intruders can damage physical and logical infrastructure, steal proprietary information, compromise competitiveness and threaten business sustainability.

Biometrics systems are critical in the larger national and homeland security context [4]. It is, therefore, not surprising that national and world "governments" will continue to apply biometrics in their efforts to make society safer.

The Federal Bureau of Investigation (FBI) is embarking on about a billion dollar effort to build the world's largest computer database of people's physical characteristics, a project that would give the government unprecedented abilities to identify individuals in the United States and abroad[3].

Other factors such as these combined and supported the need for biometrics technology: Awareness and global intensification of anti-terrorism; acceleration of identity, Internet, and other forms of frauds; increase in public recognition of the benefits; reduction in errors and improvement in accuracy, and need to control the boarder through identity recognition. A study was conducted in South Africa and noted the opinions of research respondents toward biometrics in the following manner: Biometrics as a possible means of identification will satisfy their security concerns; Biometrics will ensure that only authorized users gain access to certain information; it is a good idea because a user's identity cannot be reproduced by someone else—uniqueness; it is a more workable solution than traditional identification methods because it is easier to use; the use of biometrics as a possible means of identification will provide more confidence in the security of on-line transaction[5].

The "user acceptance of biometrics was the function of three criteria: performance, user satisfaction, and user cost". These criteria are important for biometrics developers and vendors to consider when designing and manufacturing biometrics system. The performance of the system and each user's ability to complete tasks are equally important [6].The perceived usefulness of the technology and each user's satisfaction largely will depend on the assessment of speed and ease of the interaction.

In recent time, there are several biometric characteristics that are in use in various applications. Each biometric has its own strengths, weaknesses, and suitable applications for each biometric methodology. There are no particular biometrics which may successfully meet the requirements of all applications. Depending on the application's usage and the biometric characteristic's features, it was possible to suitably match a particular biometric to an application. The fingerprint- and iris-based techniques are more accurate than the voice-based technique. Nevertheless, in a phone banking application, the voice-based technique might be preferable as the bank could integrate it seamlessly into the existing telephone system [7].

3.0 METHODOLOGY

The programming language that was used at the front-end is Microsoft Visual Studio 2010 (C sharp or C#) While My SQL was used as the back-end database application. The choice of using C# is because it simplifies windows programming, supports any windows operating system, uses object oriented technology, enhances security

of an application, enhances developer productivity through standards and simplifies the deployment of an application. While the choice of My SQL as the back-end database application is due to its database creation simplicity which is relational and has the ability to accommodate large database capacity. The system design only covered the pensioner of federal ministry of transport. However, the system designed can also be used in other ministries, departments, agencies and parastatal with little modification.

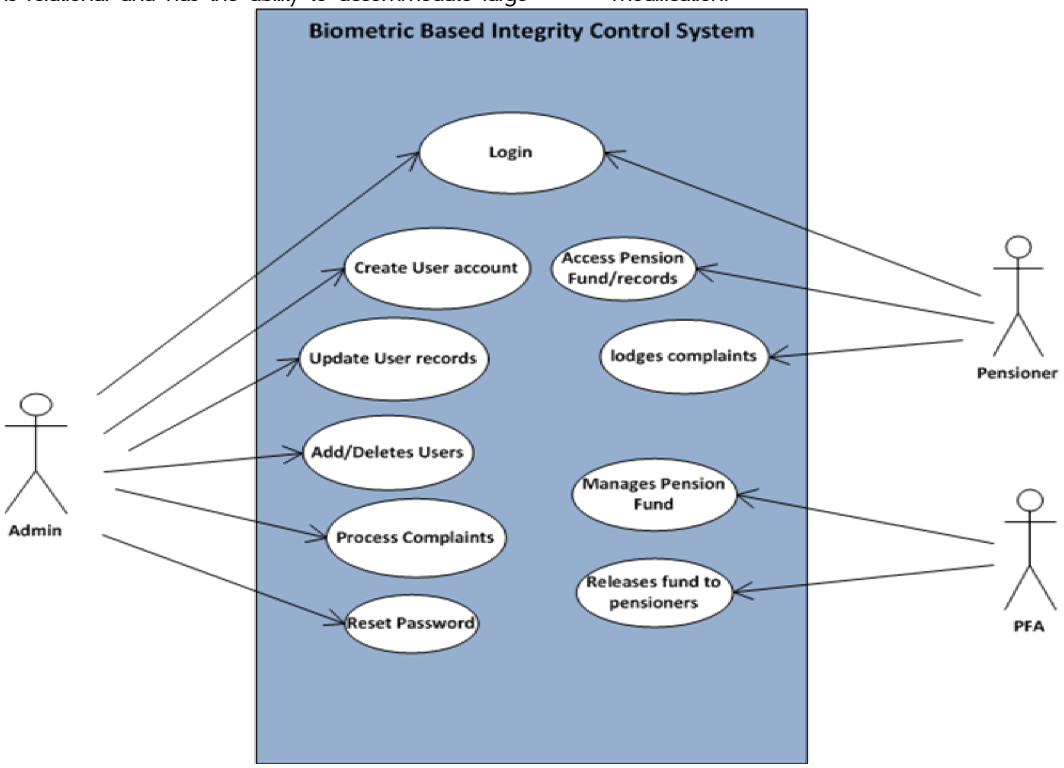


Figure 1: Use case diagram For BBICS

A. Use case modeling

Use case modeling was used to analyze the system from the users' perspective. A use case diagram is part of the Unified Modeling Language (UML) set of diagrams; it shows the important actors and functionality of the system [8]. Use cases are initiated by users or systems called Actors. The use case diagram for the Biometric Based Integrity Control System (BBICS) is shown in figure 1.

Figure 1 shows the relationship between the actors: administrator, pensioner and pension fund administrator (PFA) and the use cases. The diagram shows an outside view of the system. The admin and pensioner initiates the login to perform some instruction like create user account, update user account and lodges complain etc. The PFA is not a major actor because our study is just concerned with how a pensioner accesses his account with the help of an administrator. The PFA just manages and releases funds to valid users.

U

B. Functional Decomposition Diagram

A very important aspect of the interface design is the Functional Decomposition Diagram. A decomposition diagram shows a top-down functional decomposition of a system and exposes the system's structure. The objective of the Functional

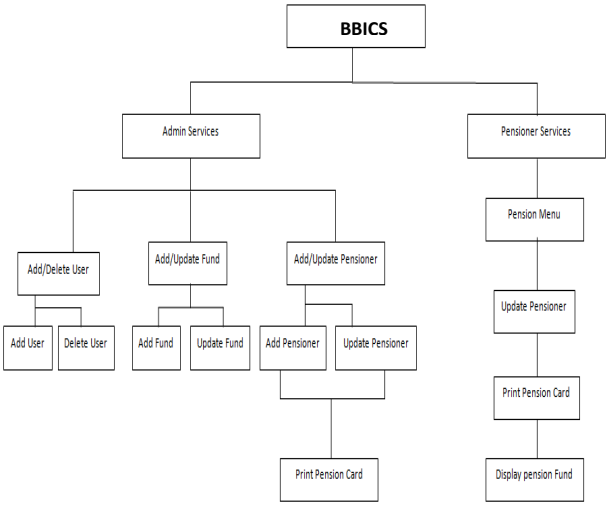


Figure 2: Functional Decomposition Diagram

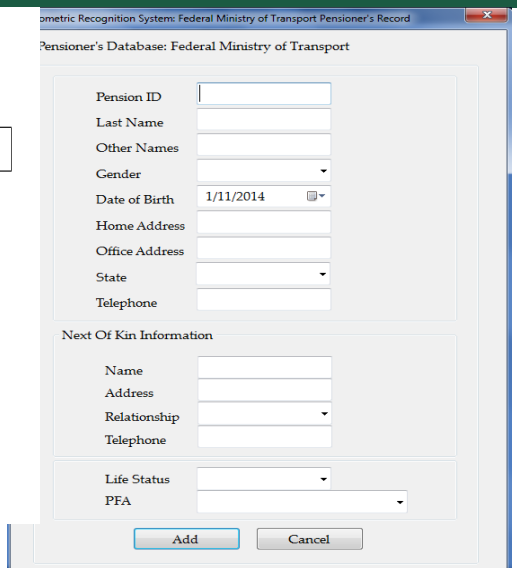


Figure 3: Add Pensioner Form

decomposition is to break down a system step by step, beginning with the main function of a system and continuing with the interim levels down to the level of elementary functions. The diagram is the starting point for more detailed process as shown in figure 2.

4.0 ARCHITECTURAL DESIGN

A. Input Interface Design

Conscious and stringent efforts were taken to ensure that the interface design meets internationally acceptable standard of software development. Microsoft Visual studio 2010 (C#) was used in the interface design because of its simplicity and adherence to best practices. This gave the interface a cool, presentable and acceptable look that will ignite the attention and interest of everyone that uses the program. Figure 3 shows a sample input interface used in this research.

B. Output Interface Design

The output of a system determines the reliability of the system. This form module is used to identify a pensioner using the picture stored in the database. Whenever an image authentication is required, the pensioner stays in front of the camera, while the system searches the image database to find a matching picture. If a match is found, the picture is displayed otherwise; an image not found message will be displayed. Figure 4 and 5 shows the output respectively.

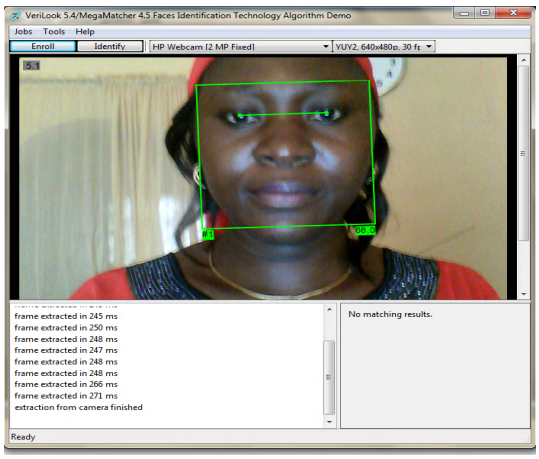


Figure 4: Face recognition output interface

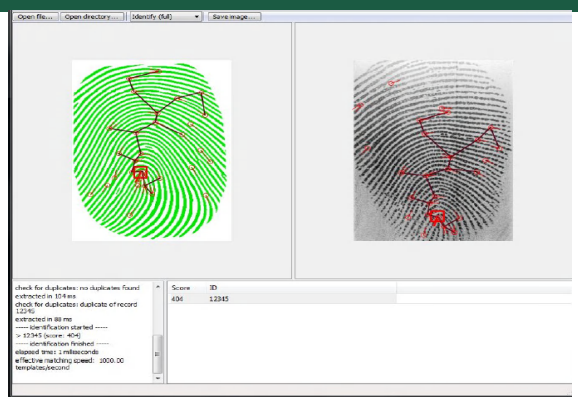


Figure 5: Matching Fingerprint Output Interface

C. Database Design

In order to design a fully comprehensive and functional Biometric Based Integrity Control System (BBICS), the relational database that will hold the entire system data must be designed first. Conceptual design can be divided into two parts: The data model and the process model. The data model focuses on what data should be stored in the database while the process model deals with how the data is processed. To put this in the context of the relational database, the data model is used to design the relational tables.

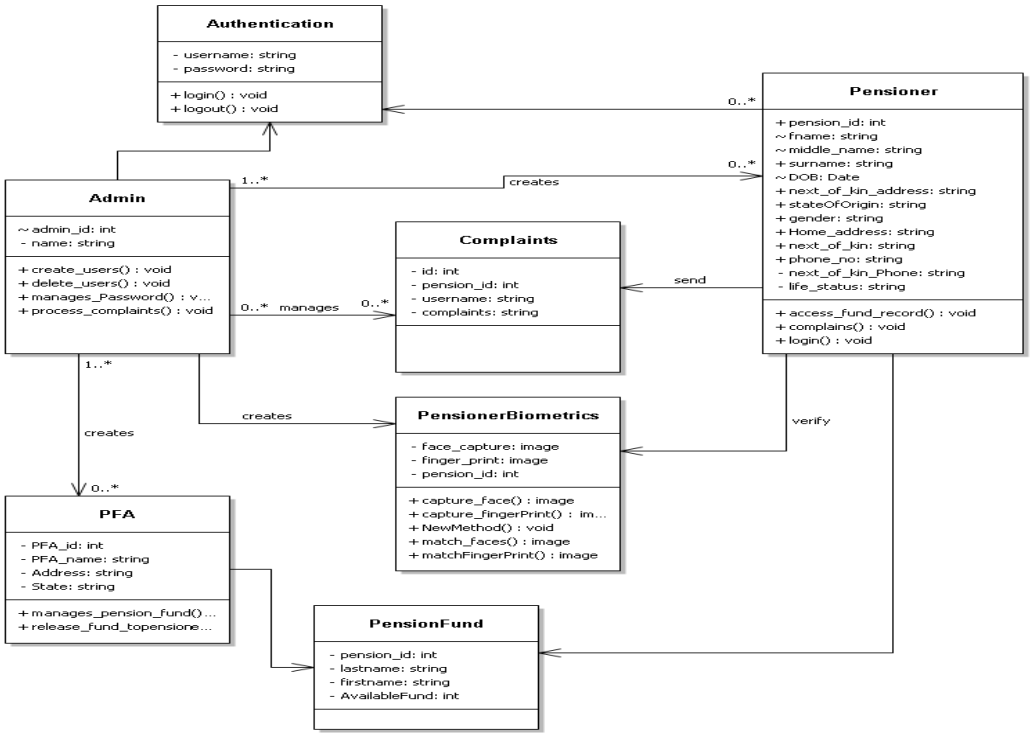


Figure 6: BBICS Entity relationship diagram

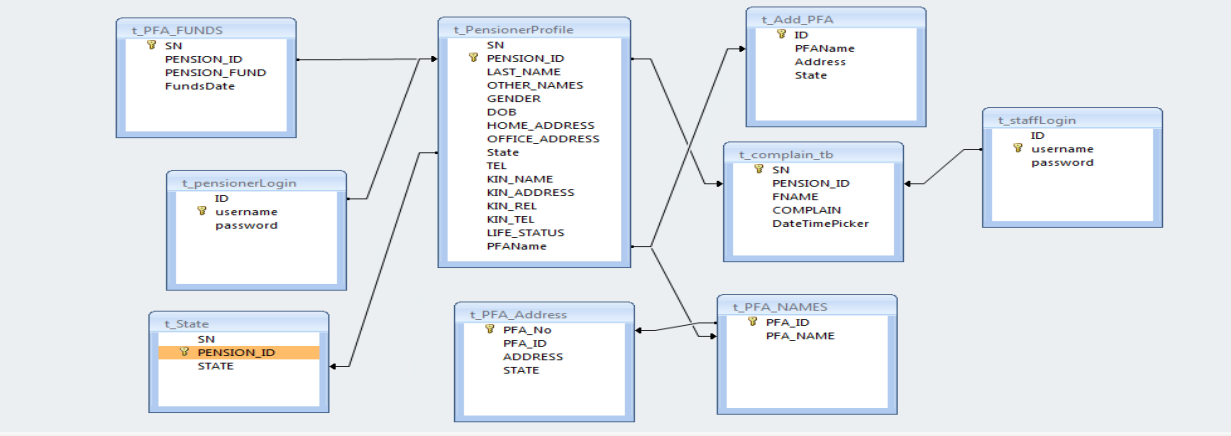


Figure 7: Class diagram for Biometric Based Integrity Control System

A data model is a conceptual representation of the data structures that are required by a database. The first step in designing a database is to develop an Entity-Relationship Diagram (ERD). The ERD serves as a blue print from

which a relational database maybe deduced as shown in figure 6.

D. Program Design

Program design is a method of designing and documenting methods and procedures in software. It could also be seen as the way of designing the internal structure of the entire program.

i. Program Architecture

The program architecture shows the entire design of the program in object

oriented format. Class diagram shows the classes within a model. In an object oriented application, classes have attributes, operations and relationships with other classes.

Figure 7.0 shows the class diagram for BBICS.

ii. Program Algorithm

The output design is with the aid of a data flowchart. Data Flowchart shows the flow of data from external entities into the system and from one process to another within the system. It also contains series of symbols with interconnectivity arrows that shows the overall flow of a program. Figures 8 and 9 are the Flowchart Diagrams for the developed system. Each process within the system is shown as a detailed flowchart.

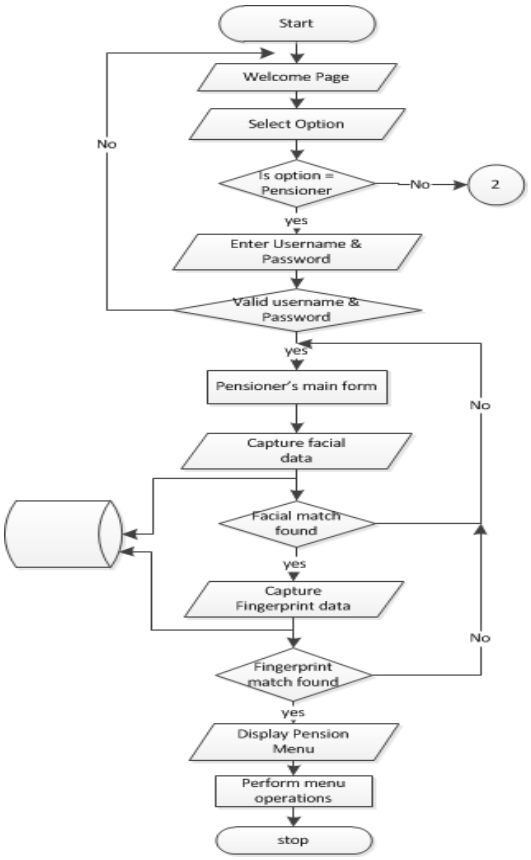


Figure 8: Pensioner Services Flowchart

The detailed flowchart provides a more detailed and comprehensive view of the

interaction among the sub-processes within the system.

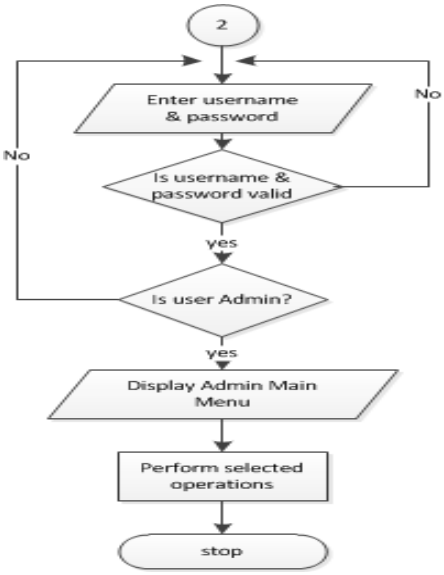


Figure 9: Admin Services Flowchart

iii. System Implementation

The program starts with an on-screen display of a splash screen. A splash screen acts as a welcome page in any application software. It appears briefly on the screen, stays for a few seconds and disappears. It shows the name and version of the software; it also has the name of the software developer. Figure 10 shows the Splash Screen.

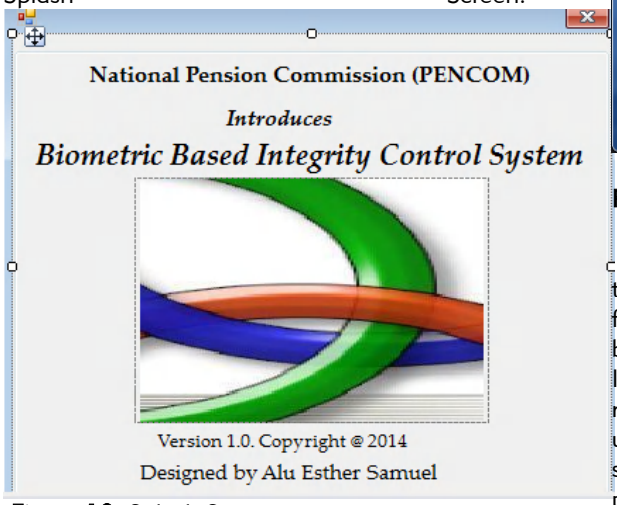


Figure 10: Splash Screen

A timer control was used to control the behavior of the splash screen such that after some seconds, it disappears while the User Type Selection form appears as shown in Figure 11 With this form, a user can select its user type be it administrator or pensioner. If administrator is selected, figure 12 appears where the user is prompted for its login credentials namely the username and password.

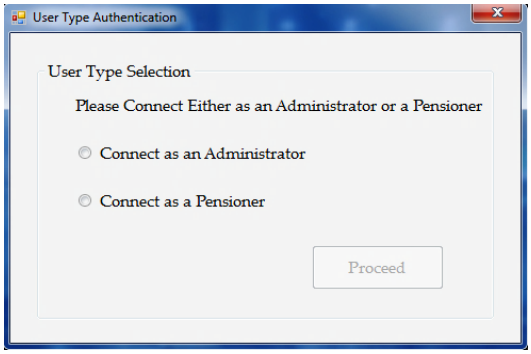


Figure 11: User Type Selection Form

Valid username and password must be entered as shown in figure 12 before access can be granted into the main system.

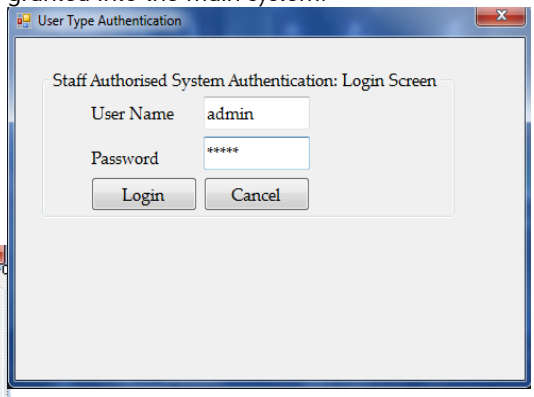


Figure 12: Staff Login Form

After a successful login, the Admin form or the pensioner form will be displayed as shown in figures 13 and 14. The main form is the backbone of the Pencom Biometric Based Integrity Control System since it houses and references most of the procedures or subroutines used in this research work. The main menu has sub-menus which shows the various task performed by the administrator.

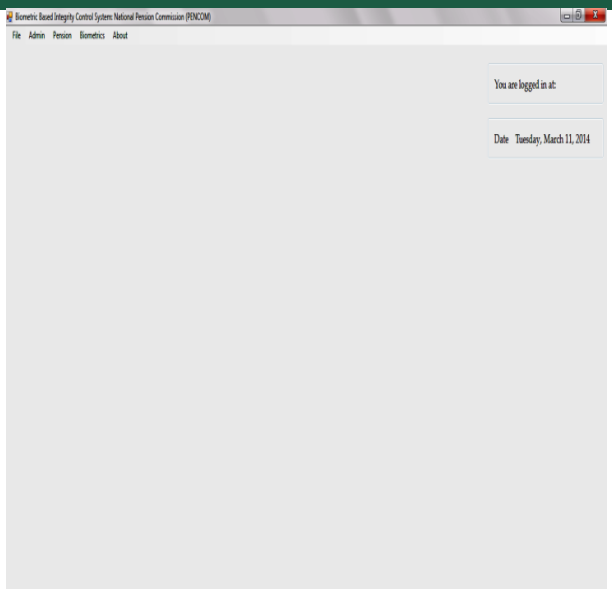


Figure 13: Admin form

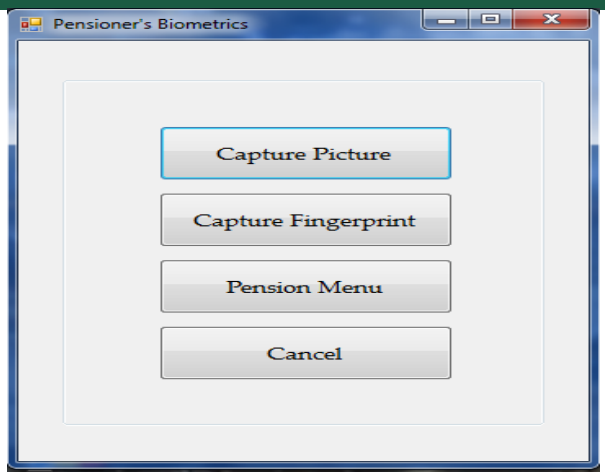


Figure 14: Pensioner form

The pensioner's form is the main interface where authentication is done for a pensioner. The pensioner menu displays after a pensioner is verified, it contains the platform for complaints and fund checking.

iv. System Requirements

System requirement has to do with the kind of computer resources that is needed for an application to perform efficiently. It could be considered in terms of software and hardware.
The software requirements for the implementation of this application are:

- Microsoft Windows 2000/XP/Vista/7.
- .Net Framework 4.0 is required to be installed on the system.
-

- Microsoft GDI+. This library is supplied with Windows XP and Windows .NET Server family. If you are using any other modern Windows platform (Windows 98/Me and Windows NT 4.0/2000) you should download and install it from Microsoft web site.
- The Microsoft® XML Parser (MSXML) 3.0 SP4 is required so if it is not already in the system you should download and install it from Microsoft web site.

While the hardware requirements for the implementations of this application are:

- 128 MB of RAM, 1 GHz CPU, 90MB HDD space for the biometric installation package.
- Secure Gen fingerprint scanner driver or any other finger print scanner
- Camera, Video capture device or (web camera)
- Laptop
- Printer

5.0 RESULTS AND DISCUSSION

A. *Results*

i. System testing

Software testing is the process of executing the software with data to help ensure that the software works correctly. A test case should always include the expected output [8]. The developed system was tested thoroughly using sample test data. Tables 1 & 2 show the different testing carried out on the developed

Table 1: Login Testing

Type of Test:	Unit Testing			
Objective of Test:	Testing to ensure that only registered users have access to the system			
Test Case	**Test Event**	**Description**	**Expected Result**	
			Valid	**Invalid**
Username	type-in	The User is expected to key-in the username.	Gives user access to use the application	Invalid user Login
Password	type-in	The User is expected to key- in his/her Password.	Gives user access to use the application	Invalid user Login
Login	Click	The User have to click on the Login button	Logs User into the system	Invalid user Login
Cancel	Click	The User have to click on the Cancel button	Closes the login interface	login interface visible

application. Types of testing that could be carried out includes but not limited to the following: unit testing, integration testing and system testing.

The aim of the system testing is to check system functionalities and the features in order to find an error, as many as possible and correcting it during the system maintenance. Debugging is an orderly process of finding and reducing the number of bugs, or defects, in a computer program or a piece of electronic hardware, thus making it behave as expected. At the course of the testing, several bugs were detected and fixed.

Table 2:Admin Main Menu Testing

Type of Test:		Integration Testing	
Objective of Test:		Testing to ensure that there is a link between the Admin main menu module and its sub-modules	
Test Case Id	**Test Event**	**Test Case Description**	**Expected Result**
IT1	Click on the Admin Button	To check the link between the Admin main menu Module and Admin module.	Opens the interface containing all task done by the admin: (Admin user, pension user & Password retrieval)
IT2	Click on the Pension Button	To check the link between the Admin main menu Module and Pension module	Opens the interface where you see: add pension fund, add PFA, update pensioner info, view lodged complaints & print pension card.
IT3	Click on the Biometric Button	To check the link between the Admin main menu Module and Biometric module	Opens the interface for Biometric data capture: Face & finger-print capture
IT4	Click on the About button	To check the link between the Admin main menu Module and About tab	Opens a form that gives information about the application.
IT5	Click on the File button	To check the link between the Admin main menu Module and File tab	Gives option on how to exit the program

B. DISCUSSIONS

Based on the validated result of the system, the Biometric-based integrity control system was found to be effective as it will give no room for invalid/unregistered users to access the system. The designed biometric application demonstrated the capabilities of image recognition engine and supported finger print and image acquisition from external sources (such as finger print devices and Web or IP cameras). Discussion on the testing carried out is shown in Table 3.

Table 3: Discussions on Result obtained

Testing Type	Discussion
Login	When correct username and password were supplied to the system the user was successfully logon unto the system, but when an invalid user tries to logon to the system access was denied.
Picture capture & Recognition	The application was able to capture face images and have them saved to the database. It was also able to carry out recognition of existing pensioners by capturing their live face images and matching it against the existing ones. In the event the face is recognized it displays the new and already stored one.
Fingerprint capture & verification	The application was able to capture fingerprints of pensioners by scanning them and have them saved to the database. It was also able to carry out verification of existing pensioners by capturing their live fingerprint data and matching it against the existing ones. In the event the fingerprint is recognized as already existed in the database, it displays the new and already stored one.
Adding Pensioner information	The system allows for the addition and updating of pensioner's information, it takes note of the user's names, date of birth, gender, phone number, pension id, etc.

6.0. SUMMARY AND CONCLUSION

The emerging trend in organizations is the security of physical, financial and information assets. Lapses in security such as unauthorized personnel gaining access to an organization's facilities and schemes can have serious consequences that extend beyond the organization. The biometric-based integrity control system developed could serve as a tool that can be used to cub the enormous challenges envisage by organizations.

The developed application was used in carrying out biometric capture and verification activities which was achieved through the following three processes: enrollment, live presentation, and matching. The time of enrollment is when the user introduces his or her biometric information to the biometric device for the first time. The enrollment data is processed to form the stored biometric template. Later, during the live presentation the user's biometric information is extracted by the biometric device and processed to form the live biometric template. Lastly, the stored biometric template and the live biometric template are compared to each other at the time of matching to provide the biometric score or result. The system met the requirement specification as defined.

This Biometric Based Integrity Control System was developed and duly tested, it proved to be working correctly as specified using parameters like Fingerprint and face for the analysis. When these parameters were used for the verification the system was able to authenticate the users, by comparing the live biometric templates and the stored biometric templates.

By employing a biometric device or system of devices inside the pension system, it will enable the National Pension Commission to tell exactly who is a pensioner.

6.1 RECOMMENDATIONS

It is recommended that the National Pension Commission improves on its current state of pension administration in Nigeria especially in the area of proper identification of pensioners in the federal ministry of transport. Currently, the biometric data of pensioners captured during the registration process is not processed for identification purposes. It is just occupying voluminous space in their expensive database servers.

It is also recommended that the National Pension Commission should and other interest as a matter of urgency, incorporate biometric data capturing and verification of biometric data such as voice, iris and hand geometry in

their pension management software. This is because biometrics offers automated method of identity verification on the principle of measurable physiological or behavioral characteristics such as the use of voice, hand geometry, fingerprint and iris. These characteristics are widely used biometric traits and they are believed to be unique to every individual. This type of identification would be more reliable when compared with traditional verification methods such as possession of an object like an ID card, or the knowledge of a password or login to access a scheme, because the person has to be physically present at the time of identification.

It is equally recommended that other organizations such as the Independent National Electoral Commission (INEC), National Communications Commission (NCC) and the National Population Commission (NPC) introduce biometrics capturing and authentication into their system for effective and efficient data processing and management.

7.0 REFERENCES

[1] Dalang, L.D. (2006): Investment and Risk Management under the New Pension Scheme; *CBN Bullion*, \zApril–June

[2] Kantudu, A.S. (2006): Impact of Pension Reform Act 2004 on Compliance with Accounting Standards on Employee Retirement Benefits in Nigeria, available at http://ssrn.com/absrtact=1106462

[3] Wayman, J.L. and Alyea L.(2000). *Picking the Best Biometric for Your Applications* in *National Biometric Test Center Collected Works*. National Biometric Test Center: San Jose.

[4] Tiwana, A. (1999). *Web Security*: Digital Press An imprint of Butterworth-Heinemann.United States Treasury Department. www.treasury.gov

[5] Pfleeger C.P. (1997). *Security in computing*. Second edition: Prentice Hall.

[6] Woodward, J.D., *et al* (2001). Army *Biometric Applications: Identifying and Addressing Sociocultural Concerns*: RAND.

[7] Prabhakar, S., S. Pankanti, and A.K. Jain (2003). *Biometrics Recognition: Security and Privacy Concerns*. IEEE Security & Privacy.

[8] Gustafson, D. A. (2002). Theory and problems of Software Engineering, the McGraw-Hill Companies, New York, USA.

SESSION B:

National Safety and E-Government

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

AN ANALYSIS OF THE NETWORKED READINESS INDEX DATA OF SOME SUB-SAHARAN AFRICA COUNTRIES

P. K. Oriogun
Lead City University, Ibadan,
Nigeria
p.oriogun@lcu.edu.ng

A. O. Adesanya
Elizade University, Ilara-Mokin,
Nigeria
adelani.adesanya@elizadeuniversity.edu.ng

B. Omolofe
Federal University of Technology Akure,
Nigeria i
bomolofe@futa.edu.ng

P. O. Yara
Lead City University, Ibadan,
Nigeria
yara.po@lcu.edu.ng

R. B. Ogunrinde
Ekiti State University, Ado-Ekiti,
Nigeria
roseline.ogunrinde@eksu.edu.ng

T. O. Akinwumi
Elizade University, Ilara-Mokin,
Nigeria
titilayo.akinwumi@elizadeuniversity.edu.ng

ABSTRACT

African governments and businesses must accept the concept of Internet of Everything (IoE) by being fully digitized with highly robust computer network security in order to embrace modern technologies in the form of cloud, mobile, social and analytics. They must also realize the importance attached to achieving societal and economic transformation by fully understanding the connection between people, process, data and things in order to create the needed opportunities for African citizens. A country's Networked Readiness is an ideal indicator of a country's ability to implement and take a competitive advantage of Information Communication Technologies (ICTs). In this article we investigate 6 sub-Saharan Africa countries (Botswana, Mauritius, Namibia, Nigeria, South Africa and Zimbabwe) in terms of their Networked Readiness Index (NRI) rankings as published in the Global Information Technology Reports (GITR) from 2002 – 2015. We developed a number of statistical models for predicting the NRI for these countries for the next 9 years based on the NRI rankings of these countries in the previous 9 years. Our predictive models for NRI rankings suggest that on average over next the 9 years, the hierarchical ordering is namely, Mauritius (1st), South Africa (2nd), Zimbabwe (3rd), Namibia (4th), Botswana (5th) and Nigeria (6th) respectively. We conclude that in terms of Networked Readiness in Africa during the period of our predictions, data security will be crucial in three areas: confidentiality, integrity and availability. Furthermore, we are of the opinion that programmable security infrastructure (software-based security environment) will have the ability to secure dynamically a particularly sensitive data flow across the network on demand and according to the organization's security policy.

KEYWORDS: INFORMATION COMMUNICATION TECHNOLOGIES (ICTS), NETWORKED READINESS INDEX (NRI), SECURITY, REGRESSION ANALYSIS, SUB-SAHARAN

2. INTRODUCTION

In this article we studied the Networked Readiness Index (NRI) framework as documented in the Global Information Technology Reports over the past 14 years (2002 until 2015) and present statistical analyses of our NRI predictions of six sub-Saharan Africa countries (Botswana, Mauritius, Nigeria, Namibia, South Africa and Zimbabwe) for the next nine years (2016 – 2024) based primarily on the available published data from The World Economic Forum in the past nine years (2007 – 2015) and other sources. Despite the fact that sub-Saharan Africa are always in the bottom half of all the countries covered over the past 14 years, Mauritius, South Africa, Botswana and Namibia have consistently ranked higher than the other countries in the same continent. The rest of the paper is arranged as follows, a brief introduction to the Networked Readiness Framework, followed by background Information on the reporting of sub-Saharan Africa NRI rankings over the past 14 years by the Global Information Technology Reports, this is followed by secure infrastructure for Networked Readiness in sub-Saharan Africa, we then present statistical analysis using Simple Linear Regression to predict the NRI rankings for the 6 countries we have identified based on continuous available NRI data from 2007 until 2015 inclusive. Finally we offer some initial discussion of our results and conclude with our thoughts on possible future of the Networked Readiness Index for sub-Saharan Africa in general.

2. THE NETWORKED READINESS INDEX (NRI) FRAMEWORK

The Global Information Technology Report (GITR) has been updating its readers on the Networked Readiness Index (NRI) Framework since its 2001 – 2002 (first edition in the series) publication. Initially, a country's NRI was defined to be the degree to which a community is *prepared* to participate in the Networked world, however, in the 2001-2002 report, this definition was modified to include the community's potential to participate in the Networked World in the future. In the same article, it was pointed out that a single measure such as the NRI is too restrictive and limited in terms of understanding how a country's national environment affect the adoption of Information Communication Technologies (ICTs). The 2002-2003 report further refined the NRI definition such that individuals, businesses, and governments are stakeholders within the community by including the potential and preparation of a community within its immediate environment.

A much more robust definition was offered in the 2003 – 2004 report, articulating that NRI is a community's degree of preparation to participate in and benefit from Information Communication Technology (ICTs) development. In the 2005 – 2006 report, Mia (2006) suggested that the NRI measures the tendency for a nations/economies to take a competitive advantage of the opportunities offered by ICT and establishes a broad international framework formulating the enabling factors of such capacity. The Networked Readiness Index is therefore a framework that could be regarded as a holistic approach to measure ICT access and impact. According to Bilbao-Osorio (2013), the NRI has provided policy / decision makers with a useful conceptual framework to evaluate the impact of information and communication technologies (ICTs) at a global level, and to benchmark the ICT readiness and the usage of their economies. In order to make any marked impact on ICT readiness, access and usage is of highest priority for developing economies given the need to narrow the so called 'digital divide' Bilbao-Osorio (2014, p.5).

3. GLOBAL INFORMATION TECHNOLOGY REPORTS ANNUAL INCLUSION OF SUB-SAHARAN AFRICA COUNTRIES

The Global Information Technology Report (GITR) started reporting on Networked Readiness Index (NRI) in its 2002 inaugural edition. In this first edition there were 4 sub-Saharan Africa countries included (Mauritius, Nigeria, South Africa and Zimbabwe) out of 75 countries worldwide. In the 2003 report there were 6 sub-Saharan Africa countries in which 2 additional countries were included (Namibia and Botswana) out of 82 countries investigated. The following year, 2004, there were 21 sub-Saharan Africa countries; 15 new countries were added (Angola, Cameroon, Chad, Ethiopia, Gambia, Ghana, Kenya, Madagascar, Malawi, Mali, Mozambique, Senegal,

Tanzania, Uganda, and Zambia) out of 102 countries worldwide. The 2005 report excluded two existing sub-Saharan countries (Senegal and Cameroon), making a total of 19 countries reported out of 104 countries worldwide. The 2006 report excluded 3 existing countries (Angola, Zambia and Malawi) whilst 2 new countries (Benin and Cameron) were added to make 18 reported sub-Saharan Africa out of a total of 115 countries. The 2007 report excluded 2 existing countries whilst 7 (Angola, Burkina Faso, Burundi, Lesotho, Malawi, Mauritania and Zambia) new countries were added, making 23 sub-Saharan countries out of a total of 122 investigated.

There were 23 sub-Saharan countries reported in the 2008, excluding 2 of the existing countries (Angola and Malawi) whilst 2 new countries were added, making 23 sub-Saharan Africa countries out of a total of 127 reported. In the 2009 report 26 sub-Saharan Africa countries were recorded, including 3 new countries (Cote d'Ivoire, Ghana and Malawi) out of 134 countries in total. In 2010 there were no new sub-Saharan countries added, however, one none sub-Saharan country (Moldova) was not covered to make a total number of countries covered to be 133. The 2011 report added three more sub-Saharan countries (Angola, Cape Verde, and Swaziland) to have 29 sub-Saharan Africa countries out of the 138 covered. In the 2012 report, Rwanda was added, making 30 sub-Saharan Africa countries out of the total of 142 worldwide. The 2013 report had recorded NRI scores for 34 sub-Saharan Africa countries, new inclusions are: Gabon, Guinea, Liberia, Seychelles and Sierra Leone. There was no NRI reported for Angola in the 2013 out of the 144 overall countries recorded. The 2014 report recorded NRI scores for 35 sub-Saharan Africa countries (the maximum numbers recorded since the first report in 2002) out of a worldwide total of 148 countries. The latest report in 2015 recorded NRI scores for 32 sub-Saharan Africa countries (Benin, Liberia, Sierra Leone were omitted) out of the 143 countries in the study. Oriogun et al. (2015) were of the opinion that Liberia and Sierra Leone must have been omitted due to the Ebola crisis in West Africa (p.34).

4. REVIEW OF THE GLOBAL INFORMATION TECHNOLOGY REPORTS ON SUB-SAHARAN AFRICA NETWORKED READINESS (2002 – 2015)

The Global Information Technology Report of 2001 – 2002 reported the Networked Readiness Index for 75 countries. Although the report was published in 2002, it actually captured the data for 2001. This has been the case until the reports produced in 2012, 2013, 2014 and 2015 (instead of reporting say, 2011 – 2012 following previous format, it simply reported 2012 etc.). This first report claimed that these countries represents more than 80% of the World's population and more than 90% of its economic output Kirkman et al., 2002 p.10. Furthermore each of the counties included had populations of more than one million. The four sub-Saharan countries included are Mauritius (NRI Ranking 51, NRI Score 3.4, Population 51 million), Nigeria (NRI Ranking 75, NRI Score 2.1, Population 75 million), South Africa (NRI Ranking 40, NRI Score 3.71, Population 40 million) and Zimbabwe (NRI Ranking 70, NRI Score 2.78, Population 70 million). It is interesting to note that the NRI ranking and population of these first four sub-Saharan Africa countries had a very close correlation. Kirkman et al., 2002 p.12 further reported that the NRI is a summary measure and had been designed as a tool for policy makers and global leaders to understand how nations are performing in relation to one another based on their participation in the Networked World. Dutta and Jain (2003) cautioned that it must be noted that the 82 countries considered in the NRI analysis had limitations due to availability of data from reliable sources. It further stressed that ranking other countries in future will possibly pose a serious challenge, and suggested that 'any overall rankings should be done with this taken into consideration'. Dutta et al, 2004 p20 explained that the 102 countries involved in the 2003 – 2004 limits the number of variables that can be considered because the methodology adopted imposed a 65% observation rate for each variable over the 102 countries; consequently, variables with fewer observations had been removed.

The GITR of 2004 – 2005 and that of 2005 – 2006 were not in line with other measurement

protocol adopted the in previous 3 years. The maximum NRI score for the 2004 – 2005 report was 1.73 for Singapore and the lowest score was -1.69 for Chad. In the 2005 – 2006 report USA scored the maximum NRI of 2.02 whilst the lowest NRI was 1.39 by Ethiopia. From the available literature to date, there was no particular reason supplied by the GITR pre and post the 2004 – 2005 and the 2005 – 2006 reports to explain the circumstances surrounding the huge difference in the methodology employed. Consequently, for the purpose of the statistical analysis that we offer in this study on the NRI rankings for sub-Saharan Africa, we have excluded the NRI data for 2004-2005 and 2005-2006. Figure 1 is a graphical depiction of the 35 sub-Saharan Africa countries represented from 2002 to 2015 respectively. Mia and Dutta (2008) concluded that sub-Saharan Africa still lags behind in its Networked Readiness Index due to what they referred to as 'lack of extensive and well-functioning infrastructure, overregulated and inefficient business environments, and poor governance and educational standards are all important hindrances in these countries' (p.16). It was further noted that a 'number of its domestic market that could benefit from networked readiness has thus far been largely ignored' (p16). Dutta e al., 2010 agreed with Dutta et al., 2011 that most of sub-Saharan Africa countries are still lagging behind other economies in terms of its networked readiness. A couple of years later, Bilbao-Osorio et al. (2013) reported that the region has improved its ICT broadband Infrastructure through mobile network coverage, however, only a limited number of its population are poised to take advantage of this improvement. We have concentrated on the statistical analysis of 6 sub-Saharan Africa countries on the basis of continuous availability of the Networked Readiness Index (NRI) data for these countries from 2003 to 2015 inclusive.

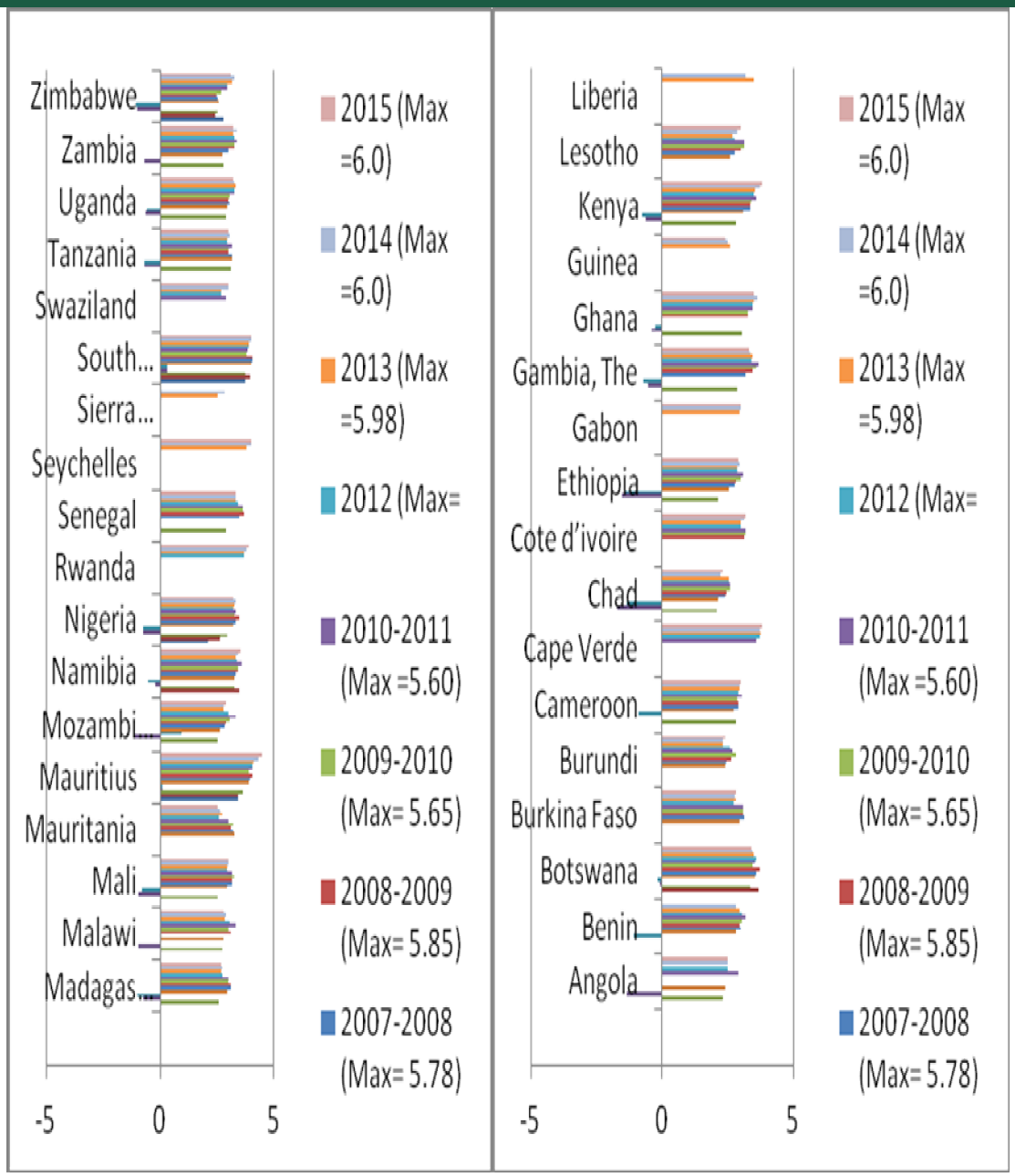


Figure 1: Graphical Representation of the 35 Sub-Saharan Africa Countries NRI Scores (2002 – 2015)

We note the huge difference in the NRI data in the GITR report of 2005 and 2006 respectively compared to the rest of the reports to date; consequently we relied heavily on the NRI data of the past 9 consecutive years (2007 to 2015) as the basis for the prediction of the next 9 consecutive years (2016 to 2024) as documented in this article. Details of our statistical analysis of the available NRI data are as shown and explained in Figures 2 – 8 together with the corresponding regression equations 2 – 8 respectively.

$y = 3.53111 - 0.024167\,x \quad R^2 = 0.470428$ (2)

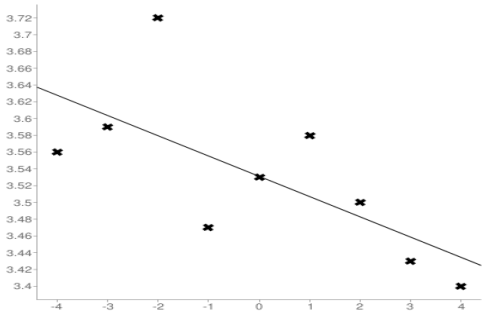


Figure 2: Botswana: Predicted Networked Readiness Index Rankings for 2016 – 2024.

5. EXPLANATION OF REGRESSION LINE

$y = \alpha + \beta\,x$ (1)

According to Oriogun and Gilchrist (2002), the equation above is saying that the expected (mean) y is given by this relation. Note however that this is an estimated relation, with a sampling variability. We could estimate a confidence interval for this mean relation. A given sub-Saharan Africa country's NRI ranking will vary about the true mean value, with a variance which could be estimated. Thus, for example, if $x = 0$, then $y = \alpha$ is the average for such a country's NRI ranking. However, a country with $x = 0$ will not actually have $y = \alpha$. We can estimate the variance about α, but the actual observation is of course unknown. Similarly, if $x = 100$, $y = \alpha + 100\beta$ is the average /expected score for such a country, but the actual NRI ranking will vary about the expectation. Again, we can estimate the variance about the line, although not the actual observation (p.104). Figures 2 – 8 shows Wessa (2015) simple Linear Regression prediction plots of Networked Readiness Index rankings for the next 9 years (2016 – 2024) based on the dataset from the Global Information Technology Reports from the past 9 years (2007 – 2015) for the six Sub-Saharan Africa countries considered for this study. Each of these plots have 95% confidence limit, and F-Test of 6.218224(Botswana), 21.020528 (Mauritius), 1.016126 (Namibia), 0.82249 (Nigeria), 0.420348 (South Africa) and 36.929807 (Zimbabwe) respectively (2007 – 2015) for the six Sub-Saharan Africa countries considered for this study.

$y = 4.09 + 0.064\,x \quad R^2 = 0.750183$ (3)

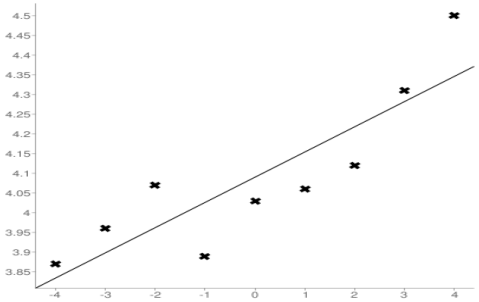


Figure 3: Mauritius: Predicted Networked Readiness Index Rankings for 2016 – 2024.

$y = 3.397778 + 0.012833\,x \quad R^2 = 0.12676$ **(4)**

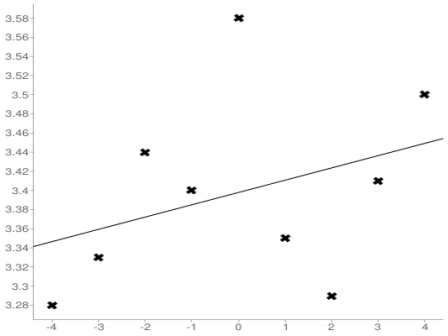


Figure 4: Namibia - Predicted Networked Readiness Index Rankings for 2016 – 2024.

According to Dutta, Geiger and Lanvin (2015, p29-30) Networked Readiness Index has four sub-indexes, namely, Environmental (political and business), Readiness (infrastructure, affordability and skills), Usage (individual business and government) and Impact (economic and social). In this article, we have selected specific pillars of the sub-indexes that directly relate to security within sub-Saharan Africa. The infrastructure must have

appropriate political and business environment to have impact on the population. The infrastructure will include secure internet server per million population, mobile network coverage (% population), international internet broadband (Kb/s per user) and electricity production (KWh/capita). The Government will have laws relating to ICTs, software piracy rate and intellectual property protection. The business environments will then be able to have the latest technologies and procurement of advance technology products. According to Oriogun, Abaye, Forteta and Shorunke (2015, p.108) governments, through ICT regulatory bodies plays a pivotal role in auditing ICT infrastructure projects. This is done through setting national policies, standards, specifications and requirements to govern the execution of projects. Within the context of developing economies, this role cannot be ignored as "best practices" are yet to be developed in many parts of the industry.

$y = 3.285556 - 0.009\,x \quad R^2 = 0.105144$ (5)

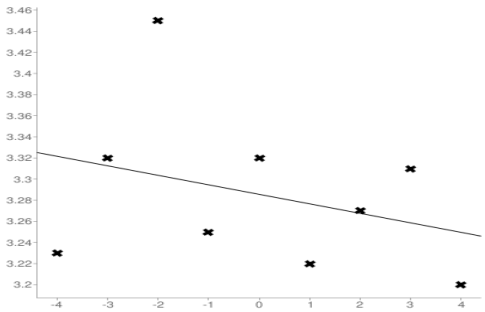


Figure 5: Nigeria - Predicted Networked Readiness Index Rankings for 2016 – 2024.

$y = 3.942222 - 0.008667\,x \quad R^2 = 0.056548$ (6)

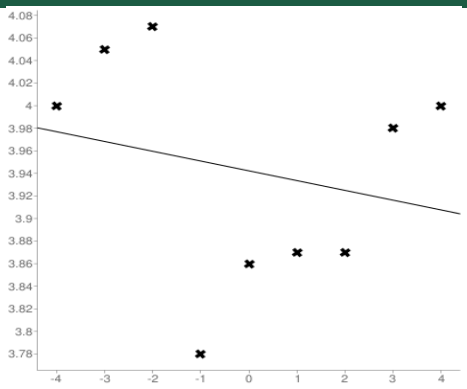


Figure 6: South Africa: Predicted Networked Readiness Index Rankings for 2016 – 2024.

$y = 2.84889 + 0.0975\,x \quad R^2 = 0.840655$ (7)

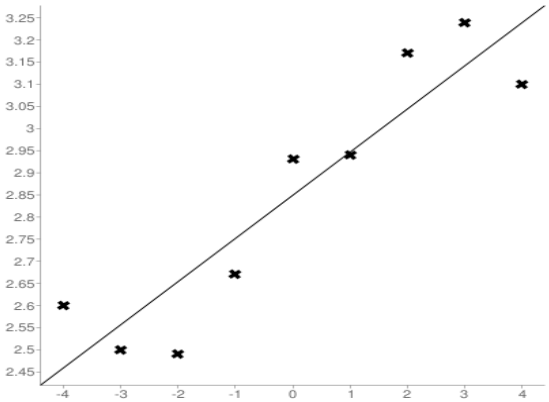


Figure 7: Zimbabwe: Predicted Networked Readiness Index Rankings for 2016 – 2024.

6. SECURE INFRASTRUCTURE FOR NETWORKED READINESS IN SUB-SAHARAN AFRICA

According to Dutta et al. (2015, p29-30) Networked Readiness Index has four sub-indexes, namely, Environmental (political and business), Readiness (infrastructure, affordability and skills), Usage (individual business and government) and Impact (economic and social). In this article, we have selected specific pillars of the sub-indexes which relates directly to security within sub-Saharan Africa. The infrastructure must have appropriate political and business environment to have impact on the population. The infrastructure will include secure

internet server per million population, mobile network coverage (% population), international internet broadband (Kb/s per user) and electricity production (KWh/capita). The Government will have laws relating to ICTs, software piracy rate and intellectual property protection. The business environments will then be able to have the latest technologies and procurement of advance technology products. According to Oriogun, Abaye, Forteta and Shorunke (2015) governments, through ICT regulatory bodies plays a pivotal role in auditing ICT infrastructure projects. This is done through setting national policies, standards, specifications and requirements to govern the execution of projects. Within the context of developing economies, this role cannot be ignored as "best practices", are yet to be developed in many parts of the industry (p.108).

7. DISCUSSION OF RESULTS

Our predictive models for NRI rankings suggest that on average over the next 9 years, the hierarchical ordering is namely, Mauritius (1st), South Africa (2nd), Zimbabwe (3rd), Namibia (4th), Botswana (5th) and Nigeria (6th) respectively. We observe from Figures 2 – 8 three countries (Botswana NRI = 3.314, Nigeria NRI = 3.203 and South Africa NRI = 3.864) had decreases in average predicted NRI while the other three countries (Mauritius NRI = 4.665, Namibia NRI = 3.512 and Zimbabwe NRI = 3.726) in our investigation had increases in their average predicted NRI. The computed $R^2$of the regression equation of each of these countries indicates the percentage increase or decrease in the total population of the data considered. The increase or decrease of the predicted NRI could be due to population (increase or decrease), political climate and other related factors such as security, business environment and ICT infrastructure environment. Consequently we believe that our predictive models are in line with Oriogun et al. (2015) previous findings, as well as other researchers working on NRI, suggesting that the NRI framework is based on political and regulatory framework of each country as well as the business and innovative environment. Furthermore, the affordability of ICT infrastructure has to match the appropriate knowledge and skills acquisition before the environment is deemed to be at the state of *Readiness*. This same environment has to interact with business, government and individuals in order to have meaningful state of usage of the available ICTs resources (p.34).

8. CONCLUSIONS

Three (Botswana, Mauritius and Namibia) out of the six countries investigated are estimated to have average population of less than 3 million each in the next 9 years. Two (South Africa and Zimbabwe) of the countries investigated have a predicted population average of just under 40 million each. Nigeria has the largest economy in the region, based on our prediction, with estimated population of just over 216 million in 9 years' time. It is possible that in general, the political and economic environment together with lack of infrastructure investments underpins the low NRI rankings of sub-Saharan Africa compared to countries in developed economies.

In terms of Networked Readiness in sub-Saharan Africa during the period of our investigation, we believe that *data security* will be crucial in three major areas: confidentiality, integrity and availability. Furthermore, we are of the opinion that programmable security infrastructure (software-based security environment) will have the ability to secure dynamically a particularly sensitive data flow across the network on demand and according to the organization's security policy.

9. ACKNOWLEDGEMENT

We would like to thank the two undergraduate Statistics SIWES (Students' Industrial Work Experience Scheme) students from Federal University of Technology Akure, Miss Funmilayo Akerele and Miss Taiwo Atitebi who were both instrumental in the verification and validation of our data for this research work.

10. REFERENCES

Battista A, Dutta S, Geiger T and Lanvin B (2015). The Global Information Technology Report 2015. The Networked Readiness Index: Taking the Pulse of ICT Revolution.

Dutta S and Jain A(2004).The Global Information technology Report 2003-2004: The Networked Readiness Index 2003-2004: Overview and Analysis framework.

Dutta S, Mia I and Geiger T (2011).The Global Information technology Report 2010-

2011: The networked Readiness Index 2010-2011: Celebrating 10 years of Assessing Networked Readiness 2010-2011.

Dutta S, Mia I, Geiger T and Herrera E.T (2010). The Global Information Technology Report 2009-2010: How networked is the World? Insights from the Networked Readiness Index 2009-2010.

Dutta S, Osorio B.B and Geiger T (2012). The Global Information Technology Report 2012. The Networked Readiness Index 2012: Benchmarking ICT Progress and Impacts for the Next Decade.

Kirkman G, Osorio C and Sachs J(2002). The Global Information Technology Report (2001-2002). The Networked Readiness index: Measuring the preparedness of nations for the networked world.

Mia I and Dutta S (2007).The Global Information Technology Report 2007. Connecting the World to the Networked Economy: A progress Report Based on the Findings of the Networked Readiness index 2006-2007

Mia I and Dutta S (2008).The Global Information Technology Report 2007-2008.Assessing the State of the World's Networked Readiness: Insight from the Network Readiness Index 2007-2008

Mia I, Dutta S, and Geiger T (2009).The Global information technology Report 2008-2009: Gauging the Network Readiness of Nations: Findings from the Networked Readiness Index 2008-2009.

Oriogun P K, Gilchrist R (2002). "A Longitudinal Study on the Impact of Information Systems Analysis and Design Prerequisite on a Software Engineering Module", *Proceedings, UKAIS 2002, Information Systems Research, Teaching and Practice*, Leeds Metropolitan University, United Kingdom, 10th -12th April 2002, pp103-110, Published by Leeds Metropolitan University, ISBN 1 898883 149.

Oriogun Peter, Agbele Kehinde, Aruleba Kehinde, and Agho Adrian (2015), " Introducing a Model to Improve Recent Sub-Saharan Africa Networked Readiness Index", *Proceedings, Nigeria Computer Society Publication*, ISSN: 21419663, Vol 26, 31-38, Akure, Nigeria, 22-24th July 2015

Osorio B.B, Crotti R, Dutta S and Lanvin B (2014). The Global Information Technology Report 2014. The networked Readiness Index: Benchmarking ICT Uptake in a World of Big Data.

Osorio B.B, Dutta S, Geiger T and Lanvin B (2013).The Global Information Technology Report 2013. The Networked Readiness Index 2013: Benchmarking ICT Uptake and Support for Growth and Jobs in a Hyperconnected World.

Wessa, P. (2015), Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7, URL http://www.wessa.net/

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

SMART GOVERNANCE: CONCEPTS AND IT'S APPLICABILITY IN FIGHTING CORRUPTION

T. Balogun
Prototype Engineering Development Institute, Pedi,
Naseni, Fmst, Ilesa,
Osun State, Nigeria
admin.pedi@naseni.org

D. Popoola
Prototype Engineering Development Institute, Pedi,
Naseni, Fmst, Ilesa,
Osun State, Nigeria
admin.pedi@naseni.org

T. Immanuel
Prototype Engineering Development Institute, Pedi,
Naseni, Fmst, Ilesa,
Osun State, Nigeria
admin.pedi@naseni.org

N.F. Efozia
Prototype Engineering Development Institute, Pedi,
Naseni, Fmst, Ilesa,
Osun State, Nigeria
fenngo31@yahoo.com

ABSTRACT

The complete eradication of corruption may be difficult to achieve, however mechanisms must be in place to curb or reduce it in governance. The paper presents a review on the role of Smart governance as a potent tool in reducing corruption in the public sector. It delineates the role of IT as an anti-corruption tool towards achieving this. It also presents a review on how corruption can be reduced in developing parts of the world like African by using technology to lessen the discretion of the ruling elite and thus bringing about transparency in governance. While it is true that ICT eliminates many opportunities for corruption for those who do not understand the new technology fully, however, it opens up new corruption vistas for those who understand the new systems well enough to manipulate them. Therefore proper safeguards are needed. In this paper, we propose a methodology to combat corruption using information and communication technologies (ICT) that entails process restructuring. While e-Governance holds great promise in many developing countries however, substantial challenges are to be tackled to realise optimum benefits that come with it. Many ICT projects fail because of insufficient planning capacity and political instability. Most developing countries are not fully ready to embrace a comprehensive program of e-government, rather than wait for total readiness, an approach of learning by trial and consolidating small gains are recommended.

KEYWORDS: Smart Governance, E-Governance, Corruption, Anti-Corruption, Ict, Good Governance

1.0. INTRODUCTION

In the early 21st century, societies and their governments around the world have been meeting unprecedented challenges, many of which surpass the capacities, capabilities, and reaches of their traditional institutions and their classical processes of governing. Democratic self-governance in 21st century market economies apparently needs to develop new institutional formats and novel mechanisms for staying abreast with the systemic dynamics of a tightly interconnected global society. We claim that actionable and omnipresent information along with its underlying technologies are substantial prerequisites and backbones for developing models of smart (democratic) governance, which foster smart, open, and agile governmental institutions as well as stakeholder participation and collaboration on all levels and in all branches of the governing process. (Hans and Margit, 2014)

In Nigeria many areas have been identified as important economy focus and attention by both the past and present governments of the nation, among them is fighting corruption, making public officers more accountable and keeping government clean. ICT can be utilised here to address these. One very important cord that runs across all the sectors in the networked knowledge economy of today is the Information and Communications Technologies (ICT's.) Until not long ago, ICT was a relatively obscure sector. Today, we live in the digital age and hardly any aspect of human endeavour can be effectively carried out without ICT. It is indispensable in times of National disasters. It considerably reduces the risk and rigours of travel and rural-urban migration." ICT therefore cuts across all aspects of human endeavour and enables us to share knowledge and experiences across ethnic, national and international divides. (Ndukwe, 2004)

Research questions such as "what are the elements of smart governance, smart and open government, and how might they interact?" as well as "what research and practice agenda would logically support the development of smart governance models and the evolution of smart and open government?" are worthy of note.

Technological advancements have been credited for playing a significant role in the globalization of trade, communication, and life styles. Vasarhelyi and Alles (2008) suggest a new business model based on technological advancements. They state that: "*Businesses are taking the lead to adapt and to also accelerate the development of the "now" economy, through the widespread adoption of integrated company software such as enterprise resource planning systems (ERP), modern communication technologies that ensure that workers are on the job 24/7/365, and monitoring systems that give a greater range of managers the ability to track and control key business processes.*"

The term "Governance" according to Wikipedia refers to all the processes of governing. It is the activity of governing. Governance is a set of processes, policies, laws and institutions affecting the way a country, society and organization is directed, administered and monitored. A Governance entity or structure (e.g. a government) need to be in place in making "nuts" and "bolts" in a system to function. The structure provides the necessary direction and specifications of a relevant system. To this end, the governance structure houses policies, institutions and processes to enable governing duty, monitoring and continuously improving the said system. (Governance in a Smart Nation - A Singapore perspective, 2015)

While the term "Smart" in IT normally connote enabling and leveraging the capabilities of modern technology in its various facets to the benefit of its users. Smart Governance is about using technology to facilitate and support better planning and decision making. It is about improving democratic processes and transforming the ways that public services are delivered. Smart Governance is the transformation of the processes of governance using technology. It connotes e-governance and it includes e-government.

So, smart governance is one of the key aspects that define a Smart City. The other 7 key aspects include Smart energy, Smart building, Smart mobility, Smart infrastructure, Smart technology, Smart healthcare, and Smart citizen. (Corruption Perceptions Index, 2016)

The concept E-governance means using technologies at various levels of government and beyond for the purpose of enhancing governance. (Shailendra and Sushil, 2007) It is the ICT enabled route to achieving good governance. E-governance is a notion related to the public sector. By public sector we mean the civil service, state institutions as well as the arms of government. It is a concept which defines how and what the public sector organizations will govern, how they will serve their citizens, how they interact with other stakeholders namely business partners, employees & government departments. (Smart Governance to E –Governance, 2012) It is about the delivery of government services and information by government to its citizens electronically using technology.

E-governance is not just Government web site or e-mail or use of internet for service delivery and electronic payments only but e-Governance allow citizens to communicate with the government, participate in the government policy making. It changes the relationship between the citizen and government as well as among citizen and citizen. It enhances good governance. (Smart Governance to E–Governance, 2012)

Jim Yong Kim (president of the World Bank) opined that Good governance is critical for all countries around the world today. When it does not exist, many governments fail to deliver public services effectively, health and education services are often substandard, and corruption persists in rich and poor countries alike, choking opportunity and growth. (Jim, 2014)

The mandate of any democratically elected government would always include good governance. Attributes of Good governance include provision of services to citizens. These services should be provided in an efficient, convenient, equitable and in effective manner. This can ensure the welfare and wellbeing of it citizens and will in turn facilitate the growth of economic activities (Smart Governance to E–Governance, 2012), which in turn can catalyse development. Accelerated economic growth and development is hinged on good governance. This is a pillar to success.

Good and fair governance promote accountability, integrity and strengthen confidence in government and management administration. Good governance is regarded as anti-corruption whereby authority and its institutions are accountable, effective and efficient, transparent and fair. Lack of good or fair governance is synonymous to corruption. (PricewaterhouseCoopers, 2013) Good and fair governance is tantamount to anti-corruption.

Corruption in the government context is basically a dishonest or illegal behaviour most especially by people who govern. Corruption is a complex subject matter. A vast array of materials abound that define this complex phenomenon. It is basically using public office for private gains.

According to Transparency International, Corruption is the abuse of entrusted power for private gain. It can be classified as grand, petty and political, depending on the amounts of money lost and the sector where it occurs. (What is Corruption? 2015)

Corruption is a symptom that something has gone wrong in the management of the state (geo-political entity). Institutions designed to govern the interrelationships between the citizen and the state is used instead for personal enrichment and the provision of benefits to the corrupt. It undermines the effectiveness of government. (Pathak and Prasad, 2005)

Pathak and Prasad (2005), postulated that the corruption framework can be summarised by the following equation –

Corruption = Monopoly + Discretion – Transparency (in governance)

Corruption thrives when there are opportunities or loop holes in the system of governance which are exploited by corrupt people who govern.

Corruption in any form affects all aspects of societal harmony. It erodes the moral fabric of any society in which it is entrenched. It reflects bad or poor governance. Corruption exacerbates poverty as funds meant for development and the provision of services to citizens are diverted to private and personal purses. It undermines economic growth.

The extent of corruption varies from country to country. The Corruption perception index (CPI)

of Transparency International [the global anti-corruption agency] puts the issue of corruption on the international scene. It uses the CPI to score countries on how corrupt their public sector is seen to be. The scores are on a scale from 0 to 100. 0 indicates a high level of corruption while 100 indicates a country is completely corruption free. Not one single country gets a perfect score and more than two-thirds score below 50. The CPI for 2015 indicated that Denmark and Finland were the top scorers of 91 and 90 respectively, while Somalia and North Korea were the lowest scorers of 8 each. Nigeria scored 26. (Transparency International' Corruption Perception Index 2015, 2015)

Since corruption is a complex societal problem it must be addressed heads on by putting measures in place to reduce or better still eradicate this fraudulent or dishonest activity. This is where anti-corruption comes to play. Anti-corruption is the tool designed to address corruption in all its forms.

Anti corruption is basically putting mechanisms in place to check mate corruption in public places. IT has a role to play in this task. E-governance can serve as a game changer in the fight against corruption if it is appropriately harnessed.

2.0. LITERATURE REVIEW

The earliest mention of the combined terms of *smart* and *government*, that we were able to find dates back to a short World Bank report on civil service reform. The term was also used without the introduction of a formal definition in a report on the computerization of government operations in the Indian State of Andhra Pradesh. More recently, former US president Bill Clinton utilized the term in the presentation of his views on the future role of government. Last, one of the core conferences in EGR, the Digital Government Society's dgo2013.org conference was held under the motto of "From e-Government to Smart Government" (http://dgo2013.dgsna.org).

In the US, major legislative elements of open government were put in place as early as in the mid-1960s, for example, the Freedom of Information Act of 1966 with its various amendments over the decades including the Open Government Act of 2007. However, the Administration's open government initiative of 2009 marked a radical switch from reactive and lacklustre information provisioning to proactive information sharing by the federal administration. This paradigmatic shift initiated the launch of numerous similar initiatives at local and state levels in the US as well as in other countries around the world. It also reinforced the attention of academic scholarship as evinced by the greatly increased number of published studies on the subject ever since.

The aim of the initiative, which was formally enforced via an Executive Office directive, was to provide transparency to government decision-making, improve accountability, and foster collaboration and stakeholder participation (Orszag, 2009). Practically, the directive required from departments and agencies to make publicly available all unclassified government records in electronic form. However, it also requested from each department a detailed plan for collaboration with and participation of other stakeholders including businesses and citizens. Direct involvement and participation in government service provision and decision making were understood as integral nodes in a feedback loop that safeguarded against falling back into non-open government practices. (Hans and Margit, 2014)

For smart government to thrive there must be the necessary corresponding infrastructure in place. As the Nigerian nation continues to accord priority to the development of necessary infrastructures and access to ICT's for its citizens, sustained policies aimed at encouraging widespread availability of these essential infrastructures must be placed at the front burner just as the case with the more developed nations of the world, which have continued to expand and upgrade their ICT resources. Nigeria has the potential to roll out the most modern of ICT infrastructures in the world by proper planning and forward looking policies by Government. We believe we have made some right moves in the recent past with opening up our ICT market to competition in nearly all sectors.

2.1. WHAT IS SMART GOVERNANCE?

A city is *smart* when investments in human and social capital (smart people), traditional transport (smart mobility), and modern digital infrastructure (ICTs) fuel sustainable economic growth (smart economy) and a high quality of life (smart living), with a wise management of natural resources (smart environment) through participatory governance (smart governance). In order for cities to perform well on the above dimensions, for improvement there is a need for evidence-based planning, which will enable a better identification of problematic sectors (e.g. transport) and areas (e.g. neighbourhoods) and a better allocation of resources. Imperatively smart governance is part of what makes a city or nation smart. (Steenbruggen, et al, 2014)

Smart governance "is an abbreviation for the ensemble of principles, factors, and capacities that constitute a form of governance able to cope with the conditions and exigencies of the knowledge society". It is further acknowledged that smart governance is about "redesigning formal democratic governance" while maintaining the historically developed democratic principles and a free market economy. Smart government, hence, has to cope with (a) complexity and (b) uncertainty, and by so doing, has to (c) build competencies and (d) achieve resilience, the latter two of which have also been referred to as *smart governance infrastructure*, which is seen as an agglomerate of hard and soft elements such as norms, policies, practices, information, technologies, skills, and other resources. When developing smart governance infrastructures, several key factors have been identified such as problem focus, feasibility/ implementability, stakeholders' contributability, continued engagement, coordination, and access to open data and shared information. (Hans and Margit, 2014)

Also, so far the two concepts of smart governance and smart government have only been rudimentarily developed. While the former has recently caught some academic attention along with some foundational theoretical treatment, the latter has not been conceptually developed although component elements such as openness and transparency of government decision-making and actions, open information sharing, stakeholder participation and collaboration, leveraging government operations and services via intelligent and integrated technology use, as well as government's role of facilitator of innovation, sustainability, competitiveness, and liveability seem to converge to a unified concept of smart and open government. Obviously, also, smart government rests on the foundation of smart governance suggesting that both concepts are closely related.

The evolution and active development of smart public governance and smart and open government are interdependent and appear as essential responses when addressing the three challenges to societal wellbeing and liveability in this century.

Smart governance is one way to describe the major institutional adaptations observed in public and international organizations in the face of increasing interdependence. Smart governance, coined by Willke (2007), is "an abbreviation for the ensemble of principles, factors and capacities that constitute a form of governance able to cope with the conditions and exigencies of the knowledge society". Policy decisions in a knowledge society that are based on purely normative considerations lose ground to those based on evidence. At the same time, decision-making requires new methods for coping with, and accounting for, the associated uncertainties that abound when knowledge – always questionable, always revisable –supersedes majority values as the basis for authority.

Smart governance can be understood as the application of so-called smart power, defined as "the combination of the hard power of coercion and payment with the soft power of persuasion and attraction". Hard power (such as use or threat of military intervention or economic sanctions) and soft power (diplomacy, economic assistance and communication, for instance) are wholly descriptive, but smart power is also evaluative. (Ilona and David, 2014)

In summary, traditional setup of government is rather fragmented with each department working in silos. The result of this is lack of coordination which is reflected in the form of poor services to the citizens. Therefore, for

cities to become smart, it is essential that the governance structure is also smart. Therefore effective use of ICTs in public administration to connect and coordinate between various departments is needed. This combined with organizational change and new skills would improve public services and strengthen support to public. This will mean the ability to seek and obtain services in real time through online systems and with rigorous service level agreements with the service providers. (Smart Cities, 2014)

2.2. RELATED CONCEPTS TO SMART GOVERNANCE

E-government (short for electronic government) is a related concept to smart governance. It is also known as e-gov, Internet government, digital government, online government, connected government. It consists of the digital interactions between a Citizen and their government (C2G), between Governments and government agencies (G2G), between Government and citizens (G2C), between Government and employees (G2E), and between Government and businesses/commerce (G2B).

E-government should enable anyone visiting a city website to communicate and interact with city employees via the Internet with Graphical user interfaces (GUI), Instant-messaging (IM), Audio/video presentations, and in any way more sophisticated than a simple email letter to the address provided at the site". Technology can be used to enhance the access to and delivery of government services to benefit citizens, business partners and employees". The focus should be on:

- The use of information and communication technologies, and particularly the Internet, as a tool to achieve better government.
- The use of information and communication technologies in all facets of the operations of a government organization.
- The continuous optimization of service delivery, constituency participation and governance by transforming internal and external relationships through technology, the Internet and new media.

Whilst e-government has traditionally been understood as being centered around the operations of government, e-governance is understood to extend the scope by including citizen engagement and participation in governance. As such, following in line with the OECD definition of e-government, e-governance can be defined as the use of ICTs as a tool to achieve better governance. (Wikipedia, 2016)

In the context of smart city governance, this includes the definition and implementation of the policies aimed at making cities smarter, which requires sharing visions and strategies with the relevant stakeholders. It also includes the management of the implementation of smart city initiatives targeted to making smarter the various city dimensions/components. Finally, it includes the management of the city infrastructures, which also comprises ICT infrastructures and systems that are enabling factors for the development of smart cities and that need to be governed; the management of the resources necessary for the development of smart cities, including the financial resources that are decisive for the prosperity and sustainability of smart cities over time the management of the human asset and of other immaterial capitals (social and relational capital, intellectual capital and innovation, knowledge and information) that are decisive for smart, sustainable and inclusive growth. It assesses the smartness of cities by paying particular attention to city governance and the management of the policy decision-making process. (Walter, et al, 2015)

In smart city, smart governance principles could guide the relatively complex administrative enactment of smart and open government more intelligently than traditional static and inflexible governance approaches could do.

As noted, eight critical factors of smart city initiatives to be analyzed in future research are: management and organization, technology, governance, policy context, people and communities, economy, built infrastructure and natural environment. These factors form the basis of an integrative framework that can be used to examine how local governments are

envisioning smart city initiatives and how they are dealing with these concerns.

By looking at the evolution undergone by the concept of governance over the last fifteen years, it is possible to notice a gradual shift in focus from a mere application of administrative and political authority towards a bidirectional discourse with a diversified constituency who is more and more recognized as an authoritative interlocutor in the process of value creation for society. In this respect, good smart city governance should attempt to achieve two important operational objectives: produce effective decisions - i.e. make the best use of information to optimize decision making - and provide adequate incentives - i.e. given that all individuals act in their own self-interest, provide the incentives that produce the best/desired outcome. But, in order to achieve these results, it is paramount to have developed a clear and strategic vision detailing what value needs to be generated.

Summarily, in this respect, ICT may allow to create decision processes relying on distributed attention, thus enabling a new form of governance, termed "extended governance" whereby the intelligence and the attention of actors residing outside governmental boundaries are harnessed in the management of public resources. (Manuel, 2015)

2.3. DIFFERENCE BETWEEN E-GOVERNMENT AND SMART GOVERNMENT

Despite the fact that both terms are used generally to encompass a lot (and sometimes interchangeably), they are quite different.

- E-Government is about providing services to and engaging with constituents, by leveraging Internet-based technologies.
- Smart government is about leveraging data to make decisions. This can happen on a tactical/operational level, strategic/policy level, or both. The data can be from within government, outside it, or both.

E-Government is about using IT for any government process and project. It's much more understandable and formal. Its quality is measured by international agencies and ratings published annually.

Smart government term is informal. It's much more about government flexibility and adaptability to the changing context. It could be using IT as driver but actually it's still not widely recognized. (Wikipedia, 2016)

2.4. ELEMENTS OF SMART GOVERNANCE CONCEPT

The concept of smart governance and/or government has some elements to enable the governing body tackle complexity and uncertainty, the competencies needed has to be adaptive and capable of serving in a process of coping with such complexity and uncertainty.

There are eight selected areas in focus that are likely candidates for smart governance initiative in line with a nation's goals; the following elements are likely to emerge;

1. *Budgeting/controlling/evaluating*. Example: Under the title "Growth-friendly consolidation" the German Federal Ministry of Finance details a multiyear approach of shrinking government spending while maintaining high levels of governmental investments in growth-related and future-oriented areas.

2. *Electronic government/administrative modernization/process streamlining*. Examples and issues: the German e-government (EGOV) law (eGovG) postulates simplified and reliable administrative processes, needs orientation, economic efficiency, ecological sustainability, modular and adequate ICT support, and a leading role in EGR; however, despite these high aspirations and its economic weight, Germany ranks only 17th in the most recent UN EGOV rankings.

3. *Security and Safety*. Examples: Responding to the sensitivities of the electorate, German governments at all levels have traditionally upheld relatively high standards with regard to data security, privacy, and data parsimony. So far, the focus has been on secure and confidential uses data. However, these practices might need review and reformulation in terms of open data initiatives, with which they may create tensions.

4. *Infrastructure Overhaul and Ubiquitous High-speed Connectivity*. Examples and key points: Germany hosts a number of smart grid projects (in sectors such as energy, traffic, and everywhere gigabit Internet). The latter is badly needed, since the country ranks only 19th worldwide in terms of average Internet bandwidth.

5. *Electric Mobility*. Example: the German Federal government embraced the notion of electric mobility, which would convert individual traffic from fossil fuels to electricity in the long-term, relatively early.

6. *Participation and Collaboration*. Examples and key points: Social media and social networking uses; individual information services; active and individual involvement; fostering individual contributions from citizens and showcasing the effects of contributions.

7. *Open Data / Big Data Provision and Use*. Examples and key points: Provision of accurate, comprehensive, and reliable information (Munchner, 2013) transparency of data uses; accounting for the effectiveness of participatory contribution; currently, open data initiatives in Germany are only partial and selective.

8. *Open Government, Transparency, and Accountability*. Examples, issues, and key points: Although closely related to open data, open government goes beyond the mere provision of government data, it rather has to encompass a proactive involvement of stakeholders in the public decision making processes. Transparency appears as a key to effective administration of the 21st century as well as to the legislative process. An urgent need for significant research on the subject has been identified. (Hans and Margit, 2014)

In summary, figure 1 summarized the context and trajectory as well as the main areas of the smart governance and smart and open government evolution: Emanating from traditional electronic government research, smart governance research will encompass broader fields of interest such as smart administration, smart interaction with stakeholders, smart security and safety, and smart infrastructures, which in turn are enclosed in the larger contexts of 21st century society and environment.

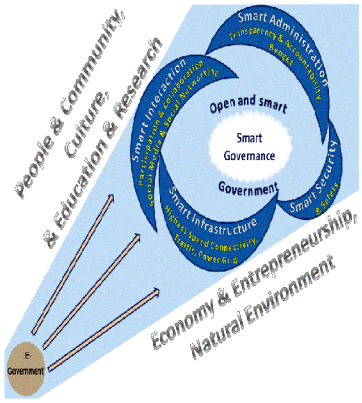


Figure 1: The Trajectory from E-Government Research to Smart and Open Government Research (Source: Hans and Margit, 2014)

2.5. WHAT IS CORRUPTION AND ANTI-CORRUPTION?

Countries are tapping into some of the most potent forces in the world today to bring about good governance, thus improving services and bringing about transparency. These forces include the spread of information technology and its convergence with grassroots movements for transparency, accountability and citizen empowerment. In some places, this convergence is easing the path to better-performing and more accountable governments, therefore reducing the likelihood or extent of corruption. (Jim, 2014)

Corruption can be defined as the use of public resources for private gains. Bribes, corrupt officials and inappropriate procurement undermine health care delivery and compromise its functions through poor governance, for which corruption can be used as a proxy measure. For example, under-5 mortality will improve with additional funding, but only in the presence of good governance and low corruption. While anti-corruption implies the act of fighting corruption.

Corruption benefits a few at the expense of the many and damages three key areas in which the state, citizens and corporations have an interest.

- Corruption undermines the efficiency of government revenue and spending

bribes reduce revenues and limit funding available for services. In recent years, the G20 prioritized tax avoidance as a top issue given the extent of the problem and the impact on public sector fiscal health. In procurement, firms that win contracts may not represent the best value for money. They have incentives to cut costs and have less accountability. This can lead to poor quality infrastructure and services.

- Corruption distorts the distribution of resources within the economy and undermines competitiveness.
- Corruption undermines political fairness, safety and inclusion. It erodes citizens' trust in the government and undermines political legitimacy. It can have the greatest impact on the poor, who are least able to voice their rights. (Nye, 2014)

Attributes of good governance include Simple, Moral, Accountable, Responsive and Transparent government. These can be achieved with electronic intervention and hence reduce corrupt practices.

Citizens expect the Simplification of rules, law, regulations and procedures of Government thus making it user-friendly. Multiplicity of these as well as complex procedures drives citizens to middlemen. This in turn results to delays and corrupt practices.

The attribute of morals in good governance has to do with governance based on morals and ethics. This can be achieved through anti graft agencies and bodies. The role of IT in achieving morals in governance may be limited though.

IT platforms can be engaged in bringing about positive change, thus addressing corruption in areas where it thrives in governance. There are several ways in which IT can be used to identify and address corruption. Some of these are briefly enumerated as follows.

1. Openness regarding government expenditure and disbursements as well as Financial Transactions. IT platforms can be effectively utilised to handle all forms of government expenditure and disbursement of funds. The pilot scheme implementation of the IPPIS public sector workers salaries payment platform saved the Nigeria government over 12 billion Naira between 2007 and 2010. (Idris, et al, 2015) The Foreign Aid Transparency Hub, launched after Typhoon Yolanda (Haiyan) in the Philippines, offered a real-time look at pledges made and money delivered for typhoon recovery. Geo-tagging tools monitored assistance for people affected by the typhoon. (Jim, 2014)
2. Openness in government procurement processes. Manual procurement processes is often characterised with all forms of corruption practices. IT platforms can curb these. In an attempt to bring about a radical change in the government procurement process, the Indian Government introduced an IT based procurement mechanism (an application) to handle government procurements. It sped up the procurement processes of the government and at the same time opened access to more vendors, thus bringing about more transparency. The e-procurement platform made price quotes to drop about 16% in the first year of the pilot scheme. (Judy, 2006)
3. Openness in the award of contracts. IT platforms can also be used to handle and monitor the award of government contracts. The Philippines government opened government data and contract information to enable citizens see how their tax money was being spent.
4. Budgeting should be opened to citizens' scrutiny. The country of Guatemala introduced a budgeting execution IT platform to handle budgeting and financial management processes in government. This was part of a fiscal reform initiative funded by the World Bank. Illegal commitments were eliminated thus saving about S100 million. Also about $2.5 million was saved just by not

issuing paper checks. (Judy Payne, 2006) The Country of Slovakia launched a budget monitoring platform in early 2013 to present budget and expenditure information to its citizens. (Sofia, 2013)

5. Utilizing IT in Tax administration. IT can be used to address corruption is often known to thrive in the government function of tax administration. The Pakistan government departments in Punjab used smart phones to collect real-time data on the activities of government field staff - including photos and geo-tags to help reduce absenteeism and tax performance. (Jim, 2014) Tunisia, Sao Tome, Cape Verde, have opted for electronic tax collection to accelerate the tax processing time and ease the process of paying taxes. (Sofia, 2013)
6. Utilising Technology to report corrupt practices. IT can be used to facilitate the lodgement of complaints as well as to report administrative abuses and corruption. India launched a corruption reporting web portal to enable citizens report corrupt acts that they experienced. Ipaidabribe.com received almost 22,500 reports between 2010 and 2012, some of which were picked up by the media and resulted in arrests and conviction. On the same web portal, citizens can also report on positive experiences they had with honest officers. (Sofia, 2013)
7. Utilizing technological platforms to deliver services to citizens. The biometric e-passport technology introduced by the Nigerian Immigration Service has effectively blocked all loopholes for wastage, inefficiency, corruption and touting. It has brought probity, accountability, transparency and effectiveness in the collection and rendition of government revenue. By the third month of introducing the technology, the Nigeria Immigration Service realised over N2 billion from issuance of e-passports. (Bashir, 2010)

These technological initiatives enumerated above have all helped in one way or the other to promote accountability, transparency, enhance efficiency, orderliness, civil participation thus curbing the likelihood of corruption in public administration processes in the respective aforementioned countries. Figure 2 shows a model of the roles.

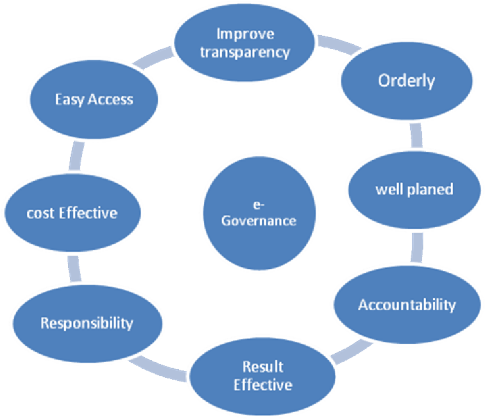


Figure 2: E-Governance role in public delivery system (Source: Authors')

In summary, in spite of IT potential as an anti-corruption tool, it's effectiveness in thoroughly addressing and curbing corruption is not automatic. The realisation of ITs' full potential as an anti-corruption tool or force is dependent on political, economic, social, security as well as infrastructural factors. Significant challenges in terms of internet access, confidentiality, and costs related to the implementation of ICT solutions remain to be addressed. (Sofia, 2013) Also, there must be an enabling political environment that promotes free access to information as well as the freedom of speech.

Security challenges are also associated with the use of technology in addressing corruption. If a system is poorly designed or vulnerable, a whistle blower of a corrupt practice risks being easily identified. In China, for example, the government has allegedly established a SMS monitoring programme to monitor and censor text messages, by setting up SMS surveillance centres around the country. Many governments also require telecom operators to register SIM cards to be able to connect a person to the SIM. This facilitates the ease in identifying a user. Securing confidentiality when sensitive

information about a corrupt practice is being communicated is thus a challenge. (Sofia, 2013)

In spite of this and other challenges, IT initiatives have been successfully utilised and harnessed in many countries around the world to address corrupt practices. Consensus abound that IT has a significant role to play in the fight against corruption in governance. IT helps in eliminating opportunities for corruption in governance.

3.0. METHODOLOGY

Governance is a country's exercise of power in managing its economic and social resources for development. Good governance is generally associated with faster, private sector-led growth and with more pro-poor development outcomes. Poor governance has the opposite effects, providing greater scope for corruption to occur. (Jeffrey, 2009)

Smart governance entails the use of information technology in order to facilitate economic growth in developing countries, as the growing concern is that people, especially those in rural areas have benefited very little from rapid economic growth. This is because the migration of the rural poor to urban areas has helped to cater for urban requirements, it has accentuated urban poverty and migration related social problems. Asymmetric information coupled with poor skill sets are considered the root cause of the inability of the rural poor to take advantage of opportunities in the markets created by technology advancement and policy changes. Addressing the problem of asymmetric information is expected to empower the rural poor to take advantage of the market opportunities as well as overcome the skill set deficits in the long run and therefore enhances inclusiveness. This would also contribute to faster and more balanced growth of the economy. (Gopal, 2011)

3.1. SOME TECHNOLOGICAL INNOVATIONS TO IDENTIFY AND REDUCE CORRUPTION.

There are multiple ways in which ICTs can contribute to identify and reduce corruption and bribery:

1. Technology innovations can be used by governments to improve the efficiency and transparency of public administration and to better communicate with and provide information to citizens;

2. It can also be used by citizens and civil society to raise awareness about the issue of corruption, to report abuses, to collect data and to monitor government activities:

3. The use of ICTs to fight corruption has increasingly served as an avenue to bring the tech community closer to activists and civil society, through the phenomenon of "hackathons". The latest International Anti-Corruption Conference hosted a hackathon focused on finding innovative ways to fight corruption using new technologies. *(U4 Expert Answer. Anti Corruption Resource Centre."technological innovations to identify and reduce corruption" http://www.U4.org)*

3.2. BROAD RANGE OF ICT IMPLEMENTED INITIATIVES IN RELATION TO SMART AND E-GOVERNMENT

More concretely, a broad range of initiatives have been successfully implemented in the last decade throughout the world as reflected by the examples below;

ICTs for Campaigning, Social Mobilisation and Citizen-To Government Interaction

- Citizen mobilization: ICTs can be used for citizen mobilization and awareness raising campaigns. Mobile applications can be designed to reach the majority of mobile subscribers through outreach/publicity campaigns using SMS. Organization running such initiatives need to build a substantial data base of targeted subscribers with active phone numbers, which can prove challenging (Hellström, J., 2010). An example of similar approaches is the campaign ran by #InternetNecesario in Mexico, which used a combination of twitter, blogs posts and media outreach to put

pressure on Mexican legislators to eliminate a 3% tax on internet access which was passed without civil society consultation (Technology for transparency Network, 2010). ICTs can also be used to mobilize people and raise awareness through art. In Tanzania, Chanjo, a collaborative project between musicians, aims to combat corruption through art, mobile phones and social media. The Chanjo project is structured around concerts and tours throughout the country followed by public discussions and debates about corruption. The music tour organized by the artists through Tanzania is coupled with the free distribution, through mobile phones and internet, of songs about corruption issues. The use of internet and social media allowed the project to reach almost 11,000 people between October and December 2011 (Spider, 2011).

- Government-citizen interactions: ICTs can also be used to promote more direct interactions between governments and citizens and empower citizens to influence local governance in their constituency through the use of SMS and the Web. In Kenya, for example, several initiatives enable mobile phone users to pose questions to their local parliamentarians, in order to increase bottom-up communication and citizen-to-government interaction. Bunge SMS, a commercial vendor from South Africa has designed a platform for holding Kenyan Members of Parliament accountable. Citizens can send an SMS to a MP through a designated number which is then routed to the Bunge SMS website (Hellström, 2010).

E-Government Initiatives: ICTs are increasingly used by governments all over the world to deliver government information and services to citizens, to enhance the efficiency and transparency of public administration and to better interact with citizens. E-government plays an increasingly important role in the promotion of participatory and inclusive development and democracy, and has grown in parallel to the rising demand for government transparency and accountability (UNPAN, 2012). Numerous e-government initiatives have been successfully implemented in the last decade and those provided below are just a few examples.

- E-procurement: E-procurement was one of the first applications of ICTs in government activities. E-procurement is the replacement of paper-based procedures with ICTs throughout the procurement processes. E-procurement can reduce administrative costs, speed up the process, increase transparency, facilitate monitoring, encourage cross-border competition and support the development of a centralized procurement administration (OECD, 2011). South Korea adopted its Government e-Procurement System (GePS) in 2002, providing integrated bidding information as a one-stop shop for customers and enabling the electronic processing of the entire procurement process. The bidding system and procurement information are available through mobile phones. According to the OECD, South Korea's e-procurement system has significantly reduced the risks of corruption, through the enhanced transparency made possible by the digitalization of information, and increased competition (OECD, 2005), (UNPAN, 2012).
- E-taxation: Governments also use ICTs for tax collection and payment, with the objective of making the system more transparent and efficient, and to cut out potential corrupt tax collectors. E-taxation has been implemented in 77 countries throughout the world, which is equivalent to 40% of the United Nations' member states. An increasing number of developing countries, such as Tunisia, Sao Tome Principe and Cape Verde, have opted for electronic tax collection to accelerate the tax processing time and

ease the process of paying taxes, (UNPAN, 2012).

- E-judiciary: ICTs offer considerable potential to improve the way the judiciary operates both nationally (filing, archiving, protection of evidence, reporting, traceability) and internationally (international judicial cooperation, training). E-judiciary has helped make workflows more efficient and court proceedings more transparent (Zinnbauer, 2012). In addition, it informs citizens of their rights and can contribute to simplifying procedures (Velicogna, 2007). India, for example, has implemented a number of ICT-based initiatives in its judiciary, like the e-justice process, to provide better access to justice for Indian citizens. Turkey has launched an SMS judicial information system, offering a legal notification service for citizens and lawyers about any development concerning their cases (UNPAN, 2012).
- Electronic identification: New technologies have been used to modernise the process of citizen identification and distribution of social services and benefits. The digitalisation of the procedure to obtain an identity card, E-ID cards and biometric proof of identity captured in electronic authentication mechanisms can have the potential to make the system more accessible, transparent and accountable. Such initiatives can reduce corruption risks in the distribution of social benefits and services, as well as in international aid (Zinnbauer, 2012).

Financial Transactions: In 2009, the Afghan National Police began to test paying salaries through mobiles instead of cash, using a text and interactive voice response system. Most policemen assumed that they had been given a significant raise in salaries, while there were simply receiving their full pay for the first time. The new system revealed that in the past at least 10% of payments had been going to ghost policemen and that middlemen in the police hierarchy commonly pocketed a percentage of other policemen's salaries (Rice and Filippelli, 2010).

ICTs for Reporting: Technology provides effective new channels to report administrative abuses and corruption, and facilitate the lodging of complaints. Reporting can be done via websites, hotlines or phone applications that solicit and aggregate citizens' experience of corruption.

- Reporting bribery and petty corruption: Perhaps the most renowned corruption reporting website is Janaagraha Centre for Citizenship's ipaidabribe.com. Through this website, citizens can report on the nature, number, pattern, types, location, frequency and values of actual corrupt acts that they experienced. Ipaidabribe.com received almost 22,500 reports between 2010 and 2012, some of which were picked up by the media and resulted in arrests and convictions (IACC, 2012). On the same website, citizens can also report on positive experiences they had with honest officers. The initiative started in India but has now been duplicated in Greece, Kenya, Zimbabwe, and Pakistan. New versions of ipaidabribe.com will soon be launched in Azerbaijan, South Africa, Ukraine and Tunisia.

Global Reporting Platforms: Transparency International (TI) has opened over 50 Advocacy and Legal Advice Centres (ALACs) since 2000 to receive citizens' complaints about corruption and engage in strategic advocacy on people's behalf. TI Macedonia has launched an online reporting platform called Draw a Red Line which allows individuals that have experienced or witnessed corruption to report their cases via ONE (Mobile Operator) by sending SMS from their mobile phones, sending an email, using a web form, on twitter by using the hashtag #korupcijaMK or by reporting over the phone. The reports are then verified by TI Macedonia staff and forwarded to the appropriate public institution to solicit follow-up. In 2012, Draw a Red Line received about 200 reports, 60 of which were verified. A number of global

reporting platforms have also been developed in recent years. BRIBEline is a reporting website available in 21 languages that was initiated by TRACE. BRIBEline collects information, through anonymous complaints, about bribes solicited by certain official or quasi-official bodies - governments, international organizations, security forces, state-owned enterprises, etc. - throughout the world. The information gathered is used to take legal or investigative action and the aggregated data is made available to the public to raise awareness about specific corruption challenges.

- Mapping bribery and petty corruption: *Bribe Market* is a similar initiative developed in Romania that allows citizens to share their experiences of bribery when interacting with public services and the amount of money they had to pay. This initiative was developed in 2012 thanks to the support of the Restart Challenges competition financed by TechSoup Global, the Central and Eastern European Trust for Civil Society, US embassies and Microsoft. Within its first four months of existence Bribe Market received nearly 650 reports of corruption. Reports are mapped to help people identify which service providers are the "cheapest" and the least corrupt (IACC, 2012).

- Reporting electoral fraud: Mobile phone reports have also been adapted for citizens election monitoring. In the Philippines for example, during the 2010 presidential elections, the Vote Report PH project encouraged voters to report electoral fraud and irregularities via SMS, email, Twitter and the website, using a collaborative Ushahidi-based platform2. The project has gained much online popularity, attracting around 2,500 unique hits per month (Grönlund, et al, 2010). In Uganda, Uganda watch 2011 is an independent hotline that allows citizens to report problems, fraud and irregularities during the electoral process. The organizations involved then analyze the information and publish reports covering issues such as voter registration issues, money in politics, as well as violence and intimidations (Hellström, 2010).

3.3. VARIOUS CASE MODELS

Smart governance and/or e-governance case models' use of ICTs in the management of delivery of public services in health, education, provision of subsidized food, businesses, land allocation, education, agriculture, etc are highlighted below.

Business Case Model: The India government proposed a sustainable business model for e-governance embedded rural telecenters tagged EGERT. In this model e-governance is an important service to be provided in the centre. Sustainable business models of rural telecentres require high volumes of services to be delivered at low service charges so as to make them affordable to a large number of the rural poor, particularly when cross subsidisation is unlikely to be effective. A high volume of services in a small size population area can come only through provisioning multiple services, which are provided in an integrated fashion and at an affordable cost.

Delivery of government services through telecentres would benefit the government, citizens as well as the telecentres themselves. For telecenters it would mean more services to be provided and therefore more revenue. Many government services such as data collection and recording are also less uncertain and therefore would bring in consistent income and help the telecentres plan their business better. Provisioning certain government services such as health related data gathering would also help in providing many other related public as well as private services. (Gopal, 2011)

To achieve the above, the government can take the help of the private sector to run telecentres through the Public Private Partnership (PPP) model to meet the challenges of investment, technology and manpower management and effective service delivery. Considering the large number of services that can be effectively provided in such

centres, they could be equipped with multiple computers as well as personnel. The private sector would run the telecentre with the revenue that would be generated from the services provided to citizens, the government and business. There has to be perfect clarity on the roles of both the private and the government departments. The government departments have to prepare themselves in terms of backend processes, appropriate systems and more importantly, the mindset to deal effectively with the private sector. The private sector operator needs to have appropriate personnel recruited locally as its employees and provide proper incentive structures. Rigorous training to sensitise the personnel to focus on citizen orientation in service delivery is a pre-requisite to run the telecentres effectively. The viability of the telecentre would depend essentially on how it is able to harness economies of scale and scope. The private sector operator would have contracts with various departments and businesses to provide the needed services. The telecentres act as single points of facilitation for the delivery of various services by the government, business as well as the rural people. (Gopal, 2011)

The Case of Health: Governance reflects how governments and other social organizations interact, how they relate to citizens and how decisions are taken in a complex world. World Health Organisation (WHO, 2007), argued that the main changes in governance at the beginning of the 21st century are also manifesting in relation to health and its governance and will be critical for achieving health gains in the decades to come.

Good governance for health combines financial, material and human resources to deliver timely and good-quality services to citizens, involving them in decision-making, provision and monitoring processes. This requires a system that mobilizes and distributes resources, processes information (and acts upon it) and motivates appropriate behaviour by health care and other workers and administrators. Good governance is a critical factor in making such a system function well.

Smart governance for health is already happening in Europe and other parts of the world. Many governments are approaching governance for health in new and innovative ways, informed by increased understanding of health and transitions in how states and societies work together. Health sector boundaries are being redefined. Also, it is about how governments respond strategically to health challenges, the choices they make regarding which mix of instruments to use, which partners to choose, which levels of government/society to engage, and when. Smart governance for health and well-being means that the state is involved in more complex relationships with other governmental and societal actors, but this does not inevitably reduce its role or power. Indeed, some argue that states (including ministries of health and the health sector) have expanded their power through the new collaborative arrangements. They remain responsible for ensuring that governance arrangements for health are effective, accountable, legitimate and democratic, but many are expanding their regulatory power in relation to a number of health challenges and extending their reach into everyday life and control of markets. (Ilona and David, 2014)

In an experiment in Ethiopia, mobile phone-based tools are being used by community health workers for registration of patients, appointment reminders, and management of inventory. Also, in Akure, the Ondo state government in Nigeria came up with an e-health card for maternal and child care health for pregnant women and children under-5 years of age to get free medical treatment.

Furthermore, improved governance and government leadership across sectors in Slovakia resulted in a modernized drug procurement system featuring a competitive, transparent, online pharmaceutical procurement process, combined with substantially higher patient co-payments. These resulted in the proportion of the cost of pharmaceuticals falling from 38.5% to 32% of total health expenditure between 2003 and 2005. The drug expenditure minimization policy was, however, too aggressive with respect to consumers and had some undesirable side-effects, with the financial

shock derived from patient co-payments hitting the lowest two income quintiles of the population hardest. A new price control policy was implemented in 2008, combining administrative regulation of prices on the production side with regulation of mark-ups on the distribution. (Ilona and David, 2014)

A well-performing health system has the following characteristics;

• delivers qualitative, equitable, efficient and safe medical interventions in a timely and geographically appropriate manner;

• holds a competent, responsive, fair and efficiently working health workforce to achieve the best health outcomes possible (given available resources and circumstances);

• uses an information system that ensures the production, analysis, dissemination and use of reliable and timely information on health determinants, health system performance and health status;

• ensures equitable access to essential medical products, vaccines and technologies of assured quality, safety, efficacy and cost–effectiveness;

• builds on a health financing system that raises adequate funds for health services to ensure access to appropriate health care, while minimizing the risk of financial hardship or impoverishment associated with medical expenses; and

• has a governance and leadership structure that ensures the existence of strategic policy frameworks, combined with effective oversight, coalition building, regulation, attention to system design and accountability.

(Source: Ilona and David, 2014).

In summary, smart governance for health needs to be about better and deeper engagement with a range of societal actors, facilitated by better transparency and held accountable by social values.

Agriculture Case Model: In the case of Nigeria, the distribution of fertilizers was done electronically via mobile phones provided by government for farmers to enable them get access to subsidized fertilizers from government directly without middle-man intervention. This IT enabled electronic delivery initiative helped in reducing corrupt practices in the distribution process.

Also, ICT enables electronic delivery of government services such as caste and income certificates to the rural populations, which in turn provide entitlements to the poor for subsidized food, fertilizer, and health services. Copies of land titles are being e-delivered to facilitate access to farm loans and reduce the cost of land transactions. It is used to share information on development expenditure so that communities get involved in demanding allocation of development expenditure that best meets their needs. Similarly, information on performance of government agencies shared with citizens can promote community audit of project execution. It focuses on enhancing rural incomes by providing information on economic opportunities, knowledge of best practices, current prices of agriculture commodities through websites, call centres, and mobile phones. (Subhash, 2014)

Pensioners Case Model: In many Indian states, pensions of the rural elderly, women, and people with disabilities are already being paid through e-banking. The government applied IT via smart governance to reduce corrupt sharp practices on pensioners by setting up a project tagged ZMF. The project delivers services that offer greater convenience to pensioners and are free of bribes. The government saves nearly 30% of the pension payout every month by weeding out ghost pensioners. ZMF illustrated the potential of transferring cash subsidies to the poor based on biometric identification so that leakage of funds to undeserving claimants can be rooted out. However, collection of biometric data for all poor in any country is a gigantic task that cannot be easily accomplished by an organization like ZMF. Perhaps the answer lies in providing unique identification to every citizen which can be verified easily. A few countries, such as the Republic of Korea and India, have tried to implement such a system. (Subhash, 2014)

Land Allocation Case Model: With computerized management of land records, Smart and/or E-governance can help reduce

corruption in a variety of ways. It takes away discretion from the government functionary, thereby curbing opportunities for arbitrary action, which often results in corruption. For example, in land records computerization in Karnataka (Bhoomi) in India, a first-in first-out (FIFO) discipline is imposed on the order of processing applications for changes in records in the workflow system established to handle the task. Government employees cannot help anyone jump the queue. The date and time are automatically stamped on service requests and they cannot be rejected arbitrarily, as a reason must be recorded if an application is rejected. Biometric log-in by operators and audit trails make it possible to track and link any corrupt operator making illegitimate changes in data through wrongful acts. Through kiosks, websites, and their mobile phones, citizens can check the status of their service request as well as highlight any error. Unlike the traditional system, reduced physical contact with government officials protects the vulnerable classes from bribe seekers. (Subhash, 2014)

Education Case Model: In the case of education an ICT-enabled monitoring and incentives initiative to reduce absenteeism have been tried. In chosen schools (Duflo, et al, 2012) teacher attendance was strictly monitored using cameras and their salaries were made a nonlinear function of attendance. Impact assessment of such initiatives through randomized experiments showed that teacher absenteeism fell 21 percentage points relative to the control group and children's test scores increased by 0.17 standard deviations. While many projects aim to use ICT to improve the "quality of education," it is recommended to go beyond such terms and identify more specific goals so as to help clarify project objectives and better align monitoring and evaluation. It is important, however, to ensure that the specific targets are appropriate in terms of the overall aim of the project and that the targets are achievable within the given time frame, budget, and other constraining factors. (Subhash, 2014)

The highlighted case models are exemplars that the Nigerian government at different levels can take a clue from and leverage the capabilities of ICT in the implementation of Smart governance initiatives to enable the provision and delivery of improved government services to its citizens in both the rural and urban areas alike.

4.0. INFORMATION TECHNOLOGY IN SMART GOVERNANCE & ANTI-CORRUPTION

The importance of IT in smart governance and in the fight against corruption cannot be over emphasized. It is also a fact that it has not been easy to harness this potential as an anti-corruption tool.

There are three main root causes of corruption of which IT can be used to tackle them; information monopolies, concentration of power, and limited accountability. The common thread connecting them is information control. The proliferation of ICT can reduce corruption by "day lighting" activities and strengthening the voice of citizens.

Dismantling monopolies open data dismantles traditional information monopolies by making information available to all. A study by T. B. Anderson47 in 2009 found a strong and direct correlation between implementation of e-government measures and corruption over a 10-year period. The correlation was even stronger than that between corruption and freedom of the press. (Nye, 2014)

Limiting discretion Technology can limit the discretion of public officials by automating processes such as the distribution of payments and benefits. For example, in 2009 the Afghan National Police began to test paying salaries through mobiles instead of cash. Most policemen assumed they received a raise when they were merely receiving their full pay for the first time. In the past, at least 10% of payments went to ghost police officers, and middlemen were skimming off the top.

Technology also provides platforms for engaging citizens in policy formation and enhancing accountability. In enhancing accountability, ICT can enhance the detection of corruption by empowering citizens to hold public service providers to account. A

randomized control trial in 50 communities in Uganda found that publishing basic data on the quality of health services and sharing it at meetings empowered citizens to hold service providers accountable and led to improved health outcomes. (Nye, 2014)

Successful governments are finding ways to reorient their structures by using information technology and policies to address the needs of their citizens and businesses. The concept of Smart Cities is gaining increasingly high importance as a means of making available all the services and applications enabled by ICT to citizens, industries and authorities. It aims to increase citizens' quality of life and improve the efficiency and quality of the services provided by governing entities and businesses. (Kumar, 2013)

Smart Cities gained importance as a means of making ICT enabled services and applications available to the citizens, and authorities that are part of a city's system. It aims at increasing citizens' quality of life, and improving the efficiency and quality of the services provided by governing entities and businesses. Smart City is a type of city that uses new technologies to make them more liveable, functional, competitive and modern through the use of new technologies, the promotion of innovation and knowledge management. Cities today are facing significant challenges including increasing populations, infrastructures, and declining budgets.

Apart from the case model mentioned in the methodology above, IT also promotes citizen engagement through information sharing of which smart and e-governance encourages. Citizens engage with the government on a variety of issues, both at the individual level and at the community level, to file complaints, express their anger, demand services, and influence policy. In the past, governments at various levels have tried to share information to engage with the citizens, but most experiments were not successful. In recent years, several developed countries' governments have created websites to distribute a portion of the data they collect. It is a concept for a collaborative project in municipal government to create and organize a culture of open data or open government data. A program called e-Panchayat is being implemented under the Government of India's National e-Governance Plan, which focuses on computerizing local government functions so that information on development expenditure and performance of executing contractors can be shared with communities. This is expected to empower such communities to demand a fair allocation of development expenditure that best meets their needs.

It also promotes the empowering farmers and fishers via sharing knowledge and price information. Throughout the developing world, mobile phones are playing a significant part in trade and commerce. For example, the Kenya Agricultural Commodity Exchange sends a functionary to visit the Nairobi market, collect prices from the local traders, and then send them back to the office via SMS. The database of prices of fresh produce can be accessed by farmers through their mobile phones through SMS or a call center. Similarly, fishers in Kerala, India, use their mobile phones (operational within 6 kilometers from the shore) to check fish prices, ensuring that they land their catch at the most profitable quayside market. These simple uses of mobile technology create a high economic impact on individual producers by eliminating the intermediaries. (Subhash, 2014)

4.1 KEY PRINCIPLES OF GOOD GOVERNANCE

Transparency: implies openness and visibility, and should apply to almost all aspects of the conduct of governmental affairs. It is the foundation upon which both accountability and participation are built. Information in the public domain is the "currency" of transparency and, together with open and visible decision-making processes, signals that there is really nothing to hide. Transparency facilitates good governance; its absence provides cover for conflicts of interest, self-serving deals, bribery, and other forms of corruption.

Accountability: it has both internal and external dimensions. Internal accountability implies probity in how and why resources are mobilized and used; it involves issues of financial

accountability, efficiency, and effectiveness in the collection of taxes and other revenue, in the creation of public goods, and in the delivery of basic services. External accountability refers to political leaders' responsiveness to citizens' needs and aspirations, including accountability for the overall performance of the economy (sustainable growth and job creation) and for the level and quality of basic services. Such accountability implies that the institutions including the civil service have the capacity to respond to citizens' demands, and that salary levels and other incentives are consistent with those expectations.

Participation, or inclusion: is important not just on principle, but in practical terms. It represents the "demand side" of good governance, and implies that people have rights that need to be recognized; that they should have a voice in the decisions that may affect them; that they should be treated fairly and equally; and that they should benefit from the protection of the rule of law. The benefits of participation are well documented: they are particularly important in decisions on the types of investment projects to be done, their design and implementation, and their operation and maintenance. The involvement of civil society organizations, consumer groups, project beneficiaries, and affected communities in all stages of Bank-financed projects can simultaneously improve development outcomes and reduce the scope for fraud and corruption. (Jeffrey, 2009)

The sample below is a table of design features to improve governance.

Table 1: Examples of Design Features to Improve Governance

Governance Dimension	Explanation	How to Prevent	Examples of Information and Communication Technology Support
Accountability	Traceability of actions and inactions	Intrinsic motivation	Pledge taken by kiosk operators in front of the entire village
		Monitoring	Biometric log-in of operators Cameras in schools Document reasons for actions: delay/denial/rejection
		Client feedback	Easy feedback, action, and escalation of unsettled complaints
Transparency	Data	Put out in public domain without violating privacy	
	Rules and procedures	Formulated in justifiable way, disseminated so clients understand, and simplified and standardized	
	Decisions	Published in public domain and right to information	
Corruption	Abuse of discretion to delay or deny service	Intrinsic values Fear of being caught Consequences	Public pledge of honesty Automate to remove discretion Workflow, so no action is outside the system Traceability through tracking Remove bottleneck in action, i.e., a single authority to sign a document Create multiple service points
	Collusive corruption		Prevent repeated contact between operators and citizens Rotate employees Use unique identification to identify recipient of service

(Source: Subhash, 2014)

4.2 CHALLENGES TO USING DATA AND ICT TO COMBAT CORRUPTION

Data and ICT literacy: Making data available to the public falls short of goals when people are ill-equipped to understand or interpret complex data.

Mandates: While it is in the public interest for corporations to be mandated to release some forms of data, accounting standards bodies often object that such disclosures are "corporate social responsibility" and should not be subject to standards.

Scope: Governments must make careful judgements about what data is regarded as privileged. Protecting personal data has an appropriate bias for confidentiality. Corruption could rise if personal information is not properly secured, making people vulnerable to identity theft or bribes.

Political risk: When transparency reveals "inconvenient truths", there can be political risk to individuals and institutions. (Nye, 2014)

4.3 THE NEED FOR CORPORATE SUPPORT TO ICT AND OPEN DATA AGENDA

Corporations and public officials are often two sides of the same "corruption coin" – the former paying the bribes and the latter receiving them. But today, the distinction is blurred. Functions traditionally in the public sector are often outsourced to the private sector and governments have larger stakes in previously private sectors. Corporations can gain in the short and long term from open data and ICT:

• Clearer understanding of the economies in which they operate: Helps to refine the business model to maximize efficiency and impact.

• Level procurement playing field: Protects market competitiveness, limits rent-seeking

• More scientific method to calculate risk and detect and deter fraud: Improves decision-making, minimizes risks and unearths valuable insights that would otherwise remain hidden.

• Synergy when government and corporate bodies work together; Creates new insights to better address specific needs of various segments of the population. (Nye, 2014)

5.0. CONCLUSION

Indeed, no country in history has yet been able to wipe out corruption completely. Corruption, like terrorism, is a global menace. Smart governance through IT may not be able to totally eliminate corruption in governance, however if properly harnessed, it can help in improving the overall scenarios in governance thus reducing the levels of corruption. It is widely believed that good governance needs to be in place to address corruption and e-governance has a role to play in bringing about good governance.

Smart governance along with its administrative enactment of smart and open government, it was argued, can help effectively address the three grand challenges to 21st century societal and individual well-being, which are; the Third Industrial Revolution with the information revolution at its core, the rapidity of change and the lack of timely and effective government intervention, and expansive government spending and exorbitant public debt financing. Although not seen as a panacea, it was also argued that smart governance principles could guide the relatively complex administrative enactment of smart and open government more intelligently than traditional static and inflexible governance approaches could do.

As highlighted in the paper, governance for health requires whole-of-government and whole-of-society approaches and a new positioning and role for ministers and ministries of health. New forms of transitional leadership are beginning to emerge. We believe this is possible to achieve not as a utopian ideal, but is described as good-enough governance, which is characterized by its diversity and adaptability.

The fact that some smart and e-governance projects failed to curb bribery reinforces the need for extensive process reform that will take away the unnecessary discretion that is abused by the corrupt in favouring those who pay bribes. In such cases, new legislation may be necessary, rules and procedures need to be modified, and extensive training is needed to change attitudes. Governments are the largest provider of information and services that are important for the poor. Methods of service delivery have not changed for decades, making them inefficient and corrupt. There is sufficient evidence that well-designed e-governance projects with process reforms that target enhanced transparency and accountability reduce discretion vested with civil servants, enhance efficiency, and can lower corruption. There is a necessity to accelerate the pace of implementation of e-governance and build capacity to reform the process of service delivery.

5.1 RECOMMENDATION AND FURTHER STUDIES

Corruption can be combated by opening up information and decision-making to as many people as possible: Open data is a powerful tool to disrupt the monopoly, discretion, and lack of accountability on which corrupt systems depend.

For better government, the "openness" of data is more important than size: Globally, more data is being generated every day. The volume is overwhelming – HM Revenue & Customs in the United Kingdom reportedly holds over 80 times more data than the British Library. (Chris, 2012) But the size of the data is irrelevant unless it can be used. Governments and corporations need to ensure that the data they publish is accessible, readable, manipulate-able and interoperable.

Data activists are agents for transparency, accountability and change - From citizen app programmers in Silicon Valley to the NGOs and local community activists in Uganda, data activists are vital agents in the open data revolution. They need to be recognized, empowered and protected.

Technology will inevitably lead to policy change, but it needs to be change in the right direction. It should be noted however that there are political risks attached to exposing or insisting on publishing certain data. Commitment to transparency can sometimes come at a price, but policymakers need to position themselves on the right side.

Government and the private sector must work together for mutual gain. Policy-makers should meet with industry representatives to discuss the terms on which such data could be made available to each other. Once a decision is reached, the public will need to be made aware of any initiative and encouraged to see the benefits of being included in the initiative. (Nye, 2014)

REFERENCES

Bashir S. Galadanci (2010) *"e-government: Matters Arising"* A presentation at the Annual Information Technology Professionals Assembly/Conference Retrieved 19/2/16

Chris Yiu, (2012) "*The Big Data Opportunity: making government faster, smarter and more personal*" 2012 Policy exchange report. Retrieved 21/2/16

"*Corruption Perceptions Index*" (2016) https://en.wikipedia.org/wiki/Corruption_Perceptions_Index

Duflo, Esther, Rema Hanna and Stephen P. ryan (2012), "*Incentives work: Getting teacher to come to school*". American Econiomic Review 102(4)1241-78

Ernest C. A. Ndukwe (2004) "*The Imperative Of Accelerating The Deployment Of Information And Communications Technologies (Icts) For Social And Economic Development*" A paper from the chairman Nigerian communications commission (NCC) held at the 11th Herbert Macaulay Memorial Lecture Thursday, July 22, 2004 Retrieved 28/1/2016

Gopal Naik (2011) "*Designing a sustainable business model for e-governance embedded rural telecentres (EGERT) in India*" Economics and Social Sciences, Indian Institute of Management Bangalore, Bangalore, India IIMB Management Review (2011) 23, 110e121 www.elsevier.com/locate/iimb Retrieved 3/2/2016

Grönlund A., et al (2010) "*Increasing transparency and fighting corruption through ICT: empowering people and communities*", SPIDER ICT4D Series No. 3, http://upgraid.files.wordpress.com/2010/11/ict4d_corruption.pdf Retrieved 9/2/16

Hans J. Scholl and Margit C. Scholl (2014) **"***Governance: A Roadmap for Research and Practice***"** In iConference 2014 Proceedings (p. 163–176). doi:10.9776/14060 www.060_ready.pdf Retrieved 28/1/2016

Hellström, J. (2010), "*Mobile technology as a means to fight corruption in East Africa*" http://upgraid.wordpress.com/ Retrieved 9/2/16

http://dgo2013.dgsna.org

IACC, (2012) "*New technologies against petty corruption: Tactics and Lessons from the 2012 IACC*" Retrieved 9/2/16

Idris Haruna, Adaja Joseph and Audu Joel Samsom (2015) **"***Integrated Personnel Payroll and Information System (IPPIS) panacea for ghost workers syndrome in*

Nigerian Public service" Retrieved 14/2/16

Ilona Kickbusch and David Gleicher, (2014) "*Smart Governance for Health and Well-Being: the evidence*" (WHO regional office for Europe,; (http://www.euro.who.int/pubrequest)) Retrieved 29/1/2016

Jeffrey S. Gutman, (2009) "*Dealing with Governance and Corruption Risks in Project Lending Emerging Good Practices*" GAC in Projects Improving Development www.EmergingGoodPracticeNote_8.11.09.pdf Retrieved 1/2/2016

Jim Yong Kim (2014) "*Technology a game-changer in fight against corruption*"

http://www.nationmultimedia.com/opinion/Technology-a-game-changer-in-fight-against-corrupt 30241963.html Retrieved 14/2/16

Jim Yong Kim (2014) "*How Technology is beating corruption*" http://www.weforum.org/agenda/2014/08/jim-yong-kim-corruption-technology-governance/ Retrieved 15/2/16

Jim Yong Kim (2014) "*Corruption Fight Aided by Technology*" http://blogs.worldbank.org/voices/corruption-fight-aided-technology (Accessed 15/2/16)

Judy Payne (2006) "*E-Government: A Critical Anti-Corruption Tool*" http://pdf.usaid.gov/pdf_docs/Pnadm957.pdf Retrieved 15/2/16

Kumar Mishra Mukesh, (2013) "*Role of technology in SMART governance "Smart City, Safe City""* KRITYANAND UNESCO CLUB Jamshedpur, India www.SSRN-id2310465.pdf Retrieved 12/2/2016

Manuel Pedro Rodríguez Bolívar, (2015) "*Smart Cities: Big Cities, Complex Governance?*" © Springer International Publishing Switzerland 2015 M. P. Rodríguez-Bolívar (ed.), *Transforming City Governments for Successful Smart Cities*, Public Administration and Information Technology 8, DOI 10.1007/978-3-319-03167-5_1 Retrieved 12/2/2016

Münchner Kreis, (2013) "Fields of innovation of the digital world: Needs of the day after tomorrow (in German, "Innovationsfelder der digitalen Welt: Bedürfnisse von übermorgen"," Future Study ("Zukunftsstudie'), vol. V, ed. vol. 2013.

Nye Joseph S., Jr., (2014) "*Future of Government Smart Toolbox*" World Economic Forum REF050514 www.WEF_GAC_FutureGovernment_SmartToolbox.pdf Retrieved 28/1/2016

Orszag P. R., (2009) "Open Government Directive: Memorandum for the heads of executive departments and agencies,", Executive Office of the President, Office of Management and Budget, Ed. Washington, DC: The White House, 2009, pp. 1-11.

Pathak R.D. and Prasad R.S. (2005) "*Role of E-Governance in Tackling Corruption and Achieving Societal Harmony: Indian Experience*" Workshop on Innovations in Governance and Public Service to Achieve a Harmonious Society www.napsipag.org/pdf/tackling_corruption.pdf Retrieved 14/2/16

PricewaterhouseCoopers (2013) "*Smart governance and technology*" https://www.pwc.in/assets/pdfs/publications/2013/smart-governance-and-technology.pdf Retrieved 14/2/16

Rice, D and Filippelli, G., (2010) "*One Cell Phone at a Time: Countering Corruption in Afghanistan*" http://commonamericanjournal.com/?p=18685 Retrieved 9/2/16

Shailendra C. Jain Palvia and Sushil S.Sharma (2007) "*E-Government and E-Governance: Definitions/Domain Framework and Status around the World*" http://www.iceg.net/2007/books/1/1_369.pdf Retrieved 16/2/16

"*SMART CITIES*" (2014) Lok Sabha Secretariat Parliament Library and Reference,

Research, Documentation and Information Service (Larrdis) Reference Note. No.28 /RN/Ref./November/2014 www.smartcities.pdf Retrieved 28/1/2016

"*Smart city From Wikipedia - free encyclopedia*" (2016) https://en.wikipedia.org/wiki/Smart_city (Accessed 18/2/16)

"*Smart Governance in a Smart Nation - A Singapore perspective*" (2015) Ltdhttps://www2.deloitte.com/content/dam/Deloitte/sg/Documents/risk/sea-risk-smart-governance-thought-leadership-noexp.pdf, Retrieved 22/2/16

"*Smart Governance to E –Governance*" (2012) http://shodhganga.inflibnet.ac.in/bitstream/10603/3407/9/09_chapter%203.pdf Retrieved 14/2/16

Sofia Wickberg (2013) "*U4 Expert Answer - Technological innovations to identify and reduce corruption*" http://www.u4.no/publications/technological-innovations-to-identify-and-reduce-corruption/ Retrieved 17/2/16)

Spider, (2011) "*Spider Stories*", http://www.spidercenter.org/sites/default/files/SpiderStories2011. pdf Retrieved 9/2/16

Steenbruggen John and Emmanouil Tranos Peter Nijkamp, (2014) "*Data from mobile phone operators: A tool for smarter cities?*" Research Memorandum 2014-1, www.2014-1.pdf Retrieved 3/2/2016

Subhash Bhatnagar, (2014) "*Public Service Delivery: Role of Information and Communication Technology in Improving Governance and Development Impact*" http://digitalcommons.ilr.cornell.edu/intl Retrieved 1/2/2016

Sushil Kumar Singla and Himanshu Aggarwal (2011) "*Combating corruption through e governance in public service delivery system*" http://www.rroij.com/open-access/combating-corruption-through-egovernance-in-public-service-delivery-system-96-100.pdf Retrieved 16/2/16

"*Technology for transparency Network, Technology for transparency: the role of technology and citizen media in promoting transparency, accountability and civic participation*" (2010), http://ifap-is-observatory.ittk.hu/node/498 Retrieved 9/2/2016

"*Transparency International' Corruption Perception Index 2015*" (2015) files.transparency.org/content/download/1955/12832/file/2015_CorruptionPerceptionsIndex_Report_EN.pdf Retrieved 15/2/16

U4 Expert Answer Anti Corruption Resource Centre "*Technological Innovations to Identify and Reduce Corruption*" http://www.U4.org Retrieved 1/4/2015.

UNPAN (2012) "*E-government survey*" http://www.unpan.org/egovkb/global_reports/08report.htm Retrieved 9/2/16

UNPAN (2012) "*Knowledge Base of Innovative E-Government Practices*", http://www.unpan.org/DPADM/EGovernment/KnowledgeBase of E Government Practices/tabid/828/language/en-US/Default.aspx Retrieved 9/2/16

Vasarhelyi, M. and Alles, M. G. (2008), "*The "now" economy and the traditional accounting reporting model:*

Opportunities and challenges for AIS research", International Journal of Accounting Information Systems. vol. 9, n. 4: 227-239. http://dx.doi.org/10.1016/j.accinf.2008.09.002 Retrieved 9/2/16

Walter Castelnovo, Gianluca Misuraca, Alberto Savoldelli, (2015) "*Citizen's engagement and value co-production in smart and sustainable cities*" www.ec.europa.eu/1433973333.pdf Retrieved 1/2/2016

"*What is Corruption?*" (2015) http://www.transparency.org/what-is-corruption/ Retrieved 14/2/16

WHO (2007)

Zinnbauer Dieter (2012) "*False Dawn, Window Dressing or Taking Integrity to the Next Level? Governments Using ICTs for Integrity and Accountability - Some Thoughts on an Emerging Research and Advocacy Agenda*", http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2166277 Retrieved 9/2/16

Full Paper

ASSURING NATIONAL JOB SECURITY THROUGH INFORMATION TECHNOLOGY (IT)

J. O. Rasaki
Computer Science Department, Federal College of Education, Eha Amufu, Enugu State
Rasaq_wale@yahoo.com

V. E. Ejiofor
Department of Computer Science, Nnamdi Azikiwe University, Awka, Anambra State
virguche2004@yahoo.com

ABSTRACT

Information Technology (IT) is very important in the training and development of human resources in any country through the impartation of appropriate skills, capacities, values, knowledge and attitudes which can be used in the transformation of individuals, communities, nations and the world at large. Besides, people get fired or lose their jobs for a lot of reasons. Some of these reasons are genuine like incompetency, corruption, dwindling revenue for the company, company bankruptcy, poor performance and so on. However, some reasons why people are relieved of their jobs cannot be substantiated. A sound knowledge of IT is the right entrepreneurial skill for job security which can be used for wealth creation, poverty reduction, ensuring social-economic empowerment, sustained self and national development. This paper x-rays the meaning of IT, job security or insecurity, unemployment, Entrepreneurial education, IT as a Panacea to Job Security in Nigeria and suggests the way forward.

Keywords: Information Technology, Job Security, Entrepreneurship

1.0 INTRODUCTION

There is a becoming need for people especially those of the working class, to get a constant job. Many of us are beginning to see that part of what distinguishes an employed individual from another, is the type of job and how secured it is. A secured job is that in which the individual would only have a little chance of becoming unemployed. That is to say, the person with a secured job is expected to remain employed for a good period of time probably until retirement age or when the person decides to leave the job. Sometimes the amount of pay is not a determinant to how secured a job would be.

Secured jobs find location across various sectors and segment of labour and productivity. It is not absolutely true that only the public jobs are the secured jobs. The things that make for a secured job may not necessarily be constant factors. We could term them to be variables, because of their possible change. Basic economic theories would support that economic expansion would influence business labor input and output and thus job security; but this too, is subject to change as economic recession may set in at any time (Information Parlour, 2015).

In actual sense and/or professionally, there is no job that is totally secured. What differs is the level of security and may be dependent on performance ability. In Nigeria, the federal government jobs are seen to be much more secured than jobs from capitalist or private individuals. This may or may not be true, as there are known reputable firms and organizations that are not government run, but still ensure a high level of job security for its employees(in as much as a variance has not changed to an adverse negative).

Paul (2014) indicated that lots of jobs are lost in the informal sector but obviously the main focus is on the formal sector. It is always

difficult, especially in a country like Nigeria with very little employment data and statistics, to ascertain the rate of job losses. A lot of private companies lay off staff and that definitely cannot be captured officially in the unemployment or job loss numbers. In 2014, Unity Bank announced the disengagement of 170 staff as part of what it called efforts to reposition for effective service delivery. They also announced hiring 300 new workers mostly entry level. MTN Nigeria laid off 252 worker, 1,110 in Access Bank, 1,185 in Ecobank, 20,000 job losses in to top 10 construction companies, 240 in Zenith Bank, 2,000 workers in Etisalats, Mobil Producing Nigeria unlimited sacked 238, 187 in NCAA, Chevron Nigeria Ltd laid off 154 staff, and Redeemers University sacked 100 workers. People get fired or lose their jobs for a lot of reasons. Some of these reasons are genuine like dwindling revenue for the company, company bankruptcy, poor performance etc. However some reasons why people are relieved of their jobs are cannot be substantiated (Paul, 2014).

2.0 JOB SECURITY

Today unemployment is important phenomenon. Almost every country suffers from job security seems to be decreasing in every part of the world and the most reasons for decreasing job security can be cited as technology, internationalization of capital, demographic change and governmental policies. (Senol, 2011). Employment security as a term is often used interchangeably with work security and job security, job security is the security of a continued employment in the same occupation with the same employer. Conceptually, work security and employment security are broader concepts, including, among other things: self employment, employment security, the confidence of being able to keep, find or create gainful employment, now and in the future, based on the development of your own human capital and in well-functioning institutions (Dekker, 2010). Pearce (1998) in (Mohammad and Shehadeh, 2014) defines job security as a mind state in which the employee sees his job stability with the firm in the near future, and it is the result of the firm's own practices and policies with the employee which make them more secure or insecure towards the job (Javed and Siddiqui, 2012) in (Mohammad and Shehadeh, 2014).

Many factor motivating employees, in fact, job security is one of the most influential means of motivating employees particularly in times of economic clown turn, employee's belief that they will not lose their jobs or they will be employed in the same organization as long as they want is a significant reason for motivation. Therefore job security is one of the most significant variables of the employee's satisfaction which expresses the general attitude of the employee towards his/ her job (Senol, 2011).

Job security plays an important role in both social and working life because it help individuals not to worry about their future, and it contributes to maintaining labor peace, increasing organizations productivity and protecting social balance and values for this employee should not be dismissed from his organization without reasonable grounds, because job security has political and social dimensions (Senol, 2011).

Although, the reason why people would want to doubt the sustainability of non government jobs may include but is not limited to the following reasons (Information Parlour, 2015):

The Size of The Labor Market: there are so many people who are willing and able to work and also have all the requirement to work but still don't have jobs. This makes labor very cheap and rational capitalists would want to reduce the cost of labor to the least minimum. This makes them hire and fire at will, reducing job security. This is quite different in the public sector where the goal is not to make profit; the government may also be weak in monitoring the performance of people that are employed.

Limited Number of People with The Required Technical Know-How: we wouldn't be wrong if we attribute this to the poor state of education in the country (read about fall of education standard in Nigeria). The people may be willing to work but would not meet the requirement. So such people when employed (if employed at all) end up been fired, when someone better qualified comes along. Most unskilled or little skilled jobs are what people in this class end up with. The type of jobs they do

are temporal or even seasonal, thus at a particular time when the demand for their labor is required, then laid off when the assignment is performed.

The Importance of Trade Unions To Secure Employment: labor and trade unions could be very instrumental to ensuring that a worker remains at his work for as long as he wants. Thus workers under the protection of these unions, going on protest and downing their tools, on events that an employer/entrepreneur (although government especially) retrench workers on a basis that is thought to be improper. As such, they are able to establish some security when the employer has it registered that he couldn't just take any such decision.

The way to influence job security would include change of location sometimes. The higher the unemployment rate, the less secured the job would be in an area. If those willing and able to work at a particular time are reduced, then the employer would not have many options of a replacement. Government labor acts and regulations may stipulate that a worker at a particular job must not be removed until at a certain time. With this, the employer is assumed that his/her job cannot be stopped, except if an offense is committed.

As a worker in an organization or public office, it is important that the individual further his or her education as much as possible. He should be readily able to make good use of modern methods, applications, instruments and equipments. These are the essentials of a secured job! An individual's job is secured if he or she climbs to the apex rank as at when due or being able to switch from one viable job to another.

3.0 JOB INSECURITY

Job insecurity is situated between employment and unemployment because it refers to employed people who feel threatened by unemployment. Another definition of job insecurity is a sense of powerlessness to maintain the desired continuity in a threatened job situation (Dachapalli & Paramasurr, 2012). The dimensions of job insecurity include: the importance of job features, the existence of job features, perceived threats to job features, importance of the total job, perceived threat to total job, and the feelings of powerfulness / powerlessness (Dachapalli & Paramasurr, 2012). Job insecurity is actually more than the perceived threat of job loss, it includes thoughts about losing valued job features such as pay, statues, opportunity for promotion. Additionally there are two different forms of job insecurity (Dachapalli & Parumasur, 2012).

- Quantitative job insecurity: worrying about losing the job itself.

- Qualitative job insecurity: worrying about important job features.

Moreover, qualitative job insecurity is defined as the significant deterioration in the working

conditions, reduction in wages and feature, and reduction in opportunities. Qualitative job insecurity reveals behavioral changes related to job.

4.0 UNEMPLOYMENT

Unemployment is a major economic virus militating against the economy and well being of many countries in recent times. It has resulted in increasing agitation from citizens, therefore, increasing insecurity in such countries. Zakaria (2006) and Ajufo (2013) supports

this by stating that the unavailability of job opportunities among youth, especially graduates have been identified as one of the major factors responsible for youth restiveness and other social vices including prostitution, arm robbery, destitution and political thuggery. Armed robbery and stealing are some of the most glaring manifestations of unemployment and poverty in Nigeria

and other developing and underdeveloped nations today.

The frequent lay off of workers which lead to mass unemployment and the resulting poverty have multi-variance consequences on youth, economic, social, political development of a nation leading to youth restiveness and personal society and national insecurity. As noted by (Anho, 2011) and (Nwaosa, Ojohwoh and Jegbefum, 2013), some of the effects

includes; social unrest; school dropout; destruction and vandalization of private and public properties; creation of fear in citizens; threat to life (individual and national); economic wastage and acute reduction in the nation's Gross Domestic Products (GDP) and personal/national income; lack of foreign investment in a country or in particular region; committal of other crimes such as; arm robbery; arson; bombing; cultism; youth exuberance; hostage-taking; human and drug trafficking; gangsterism; kidnapping; thuggery; rape; vandalism of properties; Seizure of facilities; occupation of industrial public and personal site; inter and intra community strife; work stoppage; oil bunkering; fake and illegal drug peddling; and outright will from murder.

5.0 ENTREPRENEURIAL EDUCATION

Entrepreneurial education is a form of education which makes humans to be responsive to their personal, families and national needs and aspirations. Entrepreneurship competencies carry with it, the concept of skills and mental awareness which are needed to understand the functioning of an already existing business. Entrepreneurial education is about developing attitudes, behaviours and capacities at the individual level. It is also about the application of those skills and attitudes that can take many forms during an individual's career, creating a range of long-term benefits to society and the economy. The concept of entrepreneurship education according to (Anho, 2011) is associated with various activities here in stated but not limited to the following: Innovation, creativity, risk taking, initiative, visionary, focus, determination, team spirit, resourcefulness, financial control, self confidence, versatility, knowledgeable, dynamic thinking, optimum disposition, originality, people oriented, flexible in decision, responses to suggestions and criticism, need achievement driven, profit oriented, persistent and persevering, energy for hard work, adjustment to challenges and future looking. Entrepreneurship education is useful for national security by creating career opportunities as identified by (Anho, 2014): agriculture crop production, animal husbandry, barbing, beauty care, coal production and sales, clothes dyeing and tire, driving career (cars, keke and okada), iron and steel production, money collection (daily/monthly, Isuzu), paper and pulp, petroleum/petrochemical production, poultry, tobacco production, soap and detergent, production, wood treatment, sewing and fashion design, petty trading, car wash, waste management technology, information management technology.

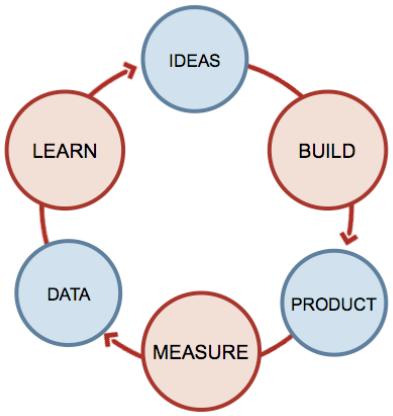


Figure 1: Star Entrepreneurial Education (Cronin, 2014)

Cronin (2014) developed the star entrepreneurial education in figure 1 above. She is of the opinion that an entrepreneur is not a randomly selected person pulled from the crowd. Entrepreneurs are people who are passionate about an idea. They have a problem to solve. A successful entrepreneur follows a validation process, whether that is consciously or unconsciously. Using this process increases the likelihood of getting products to market successfully. People should collaborate to build a system.

6.0 IT AS A PANACEA TO JOB SECURITY IN NIGERIA

Information Technology (IT) is the use of any computers, storage, networking and other physical devices, infrastructure and processes to create, process, store, secure and

exchange all forms of electronic data. The commercial use of IT encompasses both computer technology and telephony.

Certain careers may be doomed, for example, in 1901, 200,000 people washed clothes for a living in Britain. Today, only 35,000 people do that, mostly because families have their own washing machines. Technology creates jobs three ways. The first is directly: people work in the tech sector themselves. The second is that technology increases job demand in knowledge-intensive fields like medicine, business and professional services, marketing, design, and education. And third, because technology reduces the costs of basic goods, there's more disposable income left over for non-essential items. People are resourceful and IT will continue to create jobs in the future (Ben, 2015).

In the IT Profession: m**obile applications developer, big data engineer, wireless network engineer, business intelligence analyst, data security analyst, data architect, lead applications developer, database developer, software engineer, chief security officer, software developer, senior web developer, network security engineer, data modeler, information systems security manager. And host of others are example of job secured position with big salary now and in future all over the continent.** All training in IT can improve your career prospects by getting certification in Cisco Certified Network Associate (CCNA). It is a course that Cisco Systems (a networking equipment manufacturing giant from US) provides to IT professionals who specialise in the field of networking. Database Management System (DBSM) is for database administrators.

Treasury Single Account (TSA) is an electronic transaction adopted by all level of governance to cub corruption by having government money in one treasure to pay all staff and contractors as at when due. Hardware engineering- a sound knowledge of hardware maintenance and repair boost IT sustainability in an organization. Software development- a certification in software engineering in IT gives good programmers in all application and web programming job security along with the high pay. Social media serves as money making medium aside the interaction/advertisement medium.

7.0 THE ECONOMIC BENEFITS OF INFORMATION AND COMMUNICATIONS TECHNOLOGY (ICT)

According to (Robert and Luke, 2013), the followings are the economic benefits of IT:

Create high job paying – 565,000 IT related jobs were created in USA between 2001 and 2011. Which lead to an increase of 22.2 Percent. In 2011, IT workers earned $78,584 a year, 72 percent more than an average worker of $45,230. This happens even in Nigeria.

Comprises a Significant share of GDP – IT industry contribute greatly to the world economy plus the refined output of the IT industry.

Drive Productivity and GDP growth – IT was responsible for the 75 percent productivity growth in USA1995-2002.

Help building high-growth companies – out 5000 fastest growing companies in US, one quarter (1,140) were from the IT industry.

Create New Sectors and Ways of doing Business – US retail sales of goods and services through e-commerce is a new way of doing business between 2002 and 2011, increase sales by 19.8 percent annually.

IT is a key source of competitive advantages – in 2010 US firms held a 26 percent share of the global IT industry and they are the world largest producers of IT goods and services.

Drives Innovation – The probability of a company developing a product or process innovation increase with the firm intensity in using IT.

8.0 CONCLUSION

Computer chips or information and communications technology including hardware, software, telecommunications and the internet has been, is and will likely remain, for the foreseeable future, the domain driver of growth and innovation of the global economy.

This implies Information Technology (IT) is a panacea to national job security in the current Nigeria dispensation.

9.0 RECOMMENDATIONS

The followings are suggested as the way forward in having sustainable job security in Nigeria:

1. There should be personal or corporate constant on the job training to cope with challenges of daily innovation IT industry where new product erupts per second.
2. Nigerian government must promote entrepreneurial education at all level.
3. Proper management and administration of IT policy is required in the industry.
4. All business should be IT compliance to meet up with the competitive market in the world.
5. Zero tolerance to financial misappropriation by the constituted authority for proper implementation of IT policy.

10.0 REFERENCES

Ajufo, B.I. (2013). Challenges of Youth Unemployment in Nigeria: Effective Career Guidance as a Panacea. An International Multidisciplinary Journal7(1), 307-321.

Anho, J.E. (2011). "Impact of Entrepreneurship Education and Training on University Graduates for Sustainable Development" in E.A.Arubayi, N.E. Akpotu and E.P. Oghuvbu (Eds.) A Book of Reading: Education and Training for Entrepreneurship.

Anho, J. E.(2014). "Entrepreneurship Education: A Panacea for Unemployment, Poverty Reduction and National Insecurity in Developing and Underdeveloped Countries". American International Journal of Contemporary Research, Delta State University, Abraka. 4(3). Pp1-13.

(Online – http//www.aijcrnet.com/journals/Vol_4_No_3_March_2014/14.pdf) 19/1/2016.

Ben, S. (2015). Does Technology Boost Jobs Or Kill Them? (online: http:// http://www.fastcoexist.com/3050138/does-technology-boost-jobs-or-kill-them) 22/1/16

Cronin, M. (2010). Star Entrepreneurial Education (Online: http://www.thousandseeds.com/entrepreneurial-education) retrieved 20/05/2016

Dachapalli, I. and Parumasur, S. (2012). Employee susceptibility to experiencing job security, SAJEMS NS 15,NO1.

Dekker, R.(2010). Employment security: a conceptual exploration, working document for the programme "employment security. new security for charging labour market .

Information Parlour (2015) Job Security in Nigeria. (Online: http://www. http://informationparlour.com/article-job-occupation-job-security-nigeria 22/1/16

Mohammad, T and Shehadeh, M.A. (2014). The Impact of Job Security Elements on the Work Alienation at Private Universities in Jordan (A Field Study from Employees perspective). European Journal of Business and Management. 6(24) pp 61-62

Nwaosa, I.P.; Ojohwoh, R.; and Jegbefom F.M. (2013). "Curbing Youth Restiveness in Nigeria Through Entrepreneurship Opportunities in Business Education Programme" -A Paper Presented at the Annual Conference of the Institute of Education, Delta State University, Abraka 11th -17th June.

Paul, E. (2014). Do You Know How Many People Lost Their Jobs in 2014? https://blog.ngcareers.com/4463/do-you-know-how-many-people-have-lost-their-jobs-in-2014/

Robert, D. A. and Luke, A. S. (2013). *The Economic Benefits of Information and Communications Technology*. (online: http://www2.itif.org/2013-tech-economy-memo.pdf) retrieved 01/03/2016.

Senol, F.I (2011). The effect of job security on the perception of external motivational tools: A study in Hotel

Businesses. Journal of economic and social studies, (1)2.

Zakaria, Y. (2006). Youth, Conflict and Development. (Online: http://www.realityofoid.org/reareport.php) retrieved 01/03/2016.

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

GENERIC PREDICTION OF MALARIA TREATMENT OUTCOMES USING BIG DATA ANALYTICS

A.S Sodiya
Federal University of Agriculture
Abeokuta, Ogun State
sodiyaas@funaab.edu.ng

S.O Olusoga
Babcock University,
Ilishan-Remo,
Ogun State
aderonkeolusoga@gmail.com

A.O Akande
Babcock University,
Ilishan-Remo,
Ogun State
vikbola@yahoo.com

O.O Ebiesuwa
Babcock University,
Ilishan-Remo,
Ogun State
seunebi@gmail.com

E.E Onuiri
Babcock University,
Ilishan-Remo,
Ogun State
ernestonuiri@gmail.com

O.K Amodu
University College Hospital,
Ibadan
Oyo State

ABSTRACT

Research has shown that the synergy between big data analytics and healthcare enhances the quality of care on a patient by patient basis as well as a massive reduction in expenditure made on account of healthcare-related problems. This study addresses malaria disease, a perennial problem that has plagued the vast majority of people in Africa, especially as it pertains to ascertaining the best drug combination and progression that should be followed by physicians in the treatment so as to provide get the best outcomes and provide wholesome healthcare for those suffering from malaria. The study implemented big data analytics presented in the National Prediction Framework, in order to predict best effective anti-malarial drug combination to countries plagued with the disease. Large malaria data was sourced from the Institute of Child Health, University College Hospital, Ibadan – Nigeria. The dataset was imported into WEKA (Waikato Environment for Knowledge Analysis) and pre-processed; the relevant attributes in the dataset used for this study are 22 in number. The Hadoop MapReduce framework was employed for this research because of the enormity of the data captured in terms of volume. The association data mining technique was applied in order to relate patients' symptoms with the medication; this association between the patients' symptoms and the medication was done using the APRIORI algorithm for mining the data and for generating the best ten rules during the rule selection phase of mining. The rule extraction phase succeeded the rule selection phase in which the most pertinent rules needed for prediction of malaria treatment outcomes were extracted and subsequently, inferences were generated as regards the better progression of drug use that should be adopted by a patient infected with malaria. Using WEKA software, the best ten rules were generated and results showed that the adopted rules give the best progression to be followed in the treatment of malaria using combination treatment approach. Consequently, findings in this study lends credence to the fact that existing big data mining techniques can generate reliable results which can be very instructive for healthcare practitioners in general to help salvage third world countries particularly in Sub-Saharan Africa from the malaria menace that claims many innocent lives daily.

KEYWORDS: Big Data, Big Data Framework, Clinical Analytic Systems, Malaria, Mining

1.0 INTRODUCTION

Big data refers to a collection of extremely large volume of data or complex data format in data collections. In computing, big data analysis is used for reliable predictions derived from studying hidden patterns in large and/or complex data. Big data explores the opportunity of leveraging on the vast amount of data in order to discover associations, and hidden patterns within the pool of data (Raghupathi & Raghupathi, 2014). One of the most commonly recognized applications of Big Data is social media data analysis (IBM Institute for Business Value & Saïd Business School at the University of Oxford, 2012)

.

The application of big data analytics in healthcare is believed to have to have potential to improve care, save lives and lower costs (Institute for Health Technology Transformation, 2013). Big data concepts are currently being integrated into the healthcare system as a panacea for a myriad of healthcare related problems. Big data analytics provides the requisite intelligence needed by electronic health systems, thereby equipping such systems with the ability to connect clinical analytic systems and support for evidence-based healthcare. Insights can be derived to aid relevant and informed decisions in healthcare through observed relationships amongst symptoms and antimalarial drugs, amongst others.

In mapping out big data analytics solution, it is important establish a structure for rapid and seamless health data acquisition. Without this, big data analytics may not be successful in medical predictions (Andreu-Perez, Poon, Merrifield, & Wong, 2015). "In big data research, the data are usually stored and organised in order to maximize the efficiency of the data analytics process" (Viceconti, Hunter, & Hose, 2015).

Many on-going research work, now introduce big data concept to the healthcare system as a solution to a variety of healthcare related problems. Big data analytics provides intelligence to electronic health systems, providing the ability to connect clinical analytic systems and support for evidence-based healthcare (helping decision makers through systematic review of past clinical data) (Hermon & Williams, 2014).

As a matter of fact, IBM established partnership with researchers, to work on using big data and analytics to predict the outbreak of deadly diseases such as dengue fever and malaria (Takahashi, 2013).

The rate of mortality due to malaria is known to be high in Africa. Humans are also known to often build resistance to drugs over time. The World Health Organisation documents that "Malaria is caused by parasites that are transmitted to people through the bites of infected female mosquitoes - P. falciparum". The first symptoms of malaria (fever, headache, and vomiting) usually appear between 10 and 15 days after the mosquito bite. This may lead into severe sickness and eventual death if P. falciparum is not treated adequately (World Health Organisation, 2016). Malaria control measures include the use of insecticide treated nets, effective malaria drugs, effective insecticides spray, effective drainage system, advancement in malaria research and the discovery of effective drugs can help drastically to reduce the malaria incidences (Olugbenga & Clarence , 2012). Recent trends in the practice of medicine promotes the use of Artemisinin-Based Combination Therapy (ACT) for effective treatment of the malaria scourge.

The study aims to apply big data algorithms on large malaria data from Nigeria to accentuate the predictability of malaria medication administration. This is especially important in the wake of drug resistant malaria strains which warrants the need for combination treatment. Hence, this study makes use of a predictive framework to prescribe the best sequence for the combination treatment used to combat drug resistant malaria. This will give rise to relevant future expectations in the prevalence of malaria, resistance to generic antimalarial drugs.

2.0 RELATED WORK

Based on reviewed literature, it was found that some predictor models made use of environmental data and historical malaria incidence data. Environmental data is known to be less labour intensive and less expensive. The disadvantages of using the second include: the possibility of introducing bias in the estimation of predictor effects and under-estimation of standard errors due to inability to naturally account for serial autocorrelation; the likelihood of high correlation with the data used for building it but this does not

indicate future performance with data not used in the model development.; and lastly, the model may require data from the previous month which may not be readily available.

The different models found in literature are:

1. Event prediction model
2. Event detection model
3. Spatial model
4. Dynamic model
5. Risk assessment model
6. Mathematical forecasting model

In order to achieve prediction accuracy, separate data should be used in building the data and another set of reserved data should be used for testing. This approach will also help to avoid a biased result (Buczak, et al., 2015).

A malaria prediction model was developed by Wangdi, Singhasivanon, Silawan, Lawpoolsri, White, and Kaewkungwal. This was done in order to forecast malaria incidence in Bhutan The model uses the ARIMAX modelling to determine malaria predictors monthly. Time series models were derived from retrospective monthly malaria data was gathered from 1994 to 2006, and the best-fit model was identified. This was used to forecast the monthly trends from January 2009 to December 2010 in Bhutan. The modelling resulted into two common (ARIMA) models which are the (2,1,1) (0,1,1)12 and (1,1,1) (0,1,1)12. Temperature was found to be a strong positive predictor for malaria. The forecast varied from 15 to 82 cases in 2009, and 67 to 149 cases in 2010 (Wangdi, et al., 2010).

With a cursory look at the need to use targets (other than the plasmodium sequence in 18s rRNA) which can detect malaria in cases of multiple or single infection. *Demas et al* used data mining to detect hidden genomes that indicate malaria apart from 18s rRNA genome. The study was carried out with the objective of mining genomes of P. falciparum and P. Vivax to identify species-specific, repetitive sequences that serve as new PCR targets for the detection of malaria. Genome sequence data for P. falciparum (3D7 strain) and P. Vivax (Sal-1 strain) were obtained from PlasmoDB (release 5.5). P.Falciparum is more stable and advanced, compared to P.Vivax. The identification of consensus repeat sequences (CRS) through the use of Repeat-Scout. CRS is the related and similar sequences which appear in the same position on the sequence alignment. Repeat sequences that can interfere with PCR amplification were removed through the use of Tandem Repeat Finder Program. Repeats containing vector sequences introduced during genome sequencing were identified by a comparison with the NCBI UniVec database with an E-value cut-off of 1E-10. Screening was done in parallel and any sequence failing the screen test was removed from further consideration. All P. falciparum and P. vivax CRS were compared (WU-BLAST) to all available Plasmodium sequence data, and the results were manually inspected to ensure species specificity (Demas, et al., 2011).
A neuro-fuzzy prediction framework was designed and implemented, in which the prediction system has the capability to reason and to make logical decisions based on varying values of corresponding diseases' controlling factors. The framework of the malaria incidence predicting system is made up of 4 components; the user, the GUI, the application, and the artificial intelligence components. It was observed that the combination of the different control measures gave rise to a reduction in malaria incidences. The predicting system generated results for 30%, 60% and 90% combination respectively. The result showed that malaria incidences can be drastically reduced through a combination of controllable factors. (Olugbenga & Clarence , 2012).

In 2012, Pitale and Ambhaikar also developed a prototype Sensitive Region Prediction System (SRPS), using the Linear Regression data mining technique on past data sets, collected from various surveys to predict the number of cases of Malaria in future. Datasets were collected and saved into Attribute-Relation File Format (ARFF) which is used in the Weka machine learning software. After the dataset preparation, the Weka software created a model using linear Regression. At last the model was used to predict the future values and analysis was done on predicted values for detecting the sensitive regions for a country. The prototype was implemented in java Technology and JFreeChart library was used for generating plots and charts. The SRPS uses data mining technique to produce periodical forecasts about malaria disease. The technique complement proven numeric forecasting method used regression analysis with technology taking as input the malaria disease information. It was observed

that the accuracy for the prediction is dependent on data sets. If data sets are of huge numbers high level of model training is achieved and in case of small data sets a low level of model training is achieved. It was also observed that the accuracy of the prediction depends upon the level of training. If model is highly trained, then the prediction will be most accurate.

Furthermore, another study carried out in Maputo to investigate the prediction models of malaria. The study employed data with administrative districts, malaria cases, indoor residual spray and climatic variables temperature, rainfall and humidity as attributes. Regression trees and random forest were employed on 900 trees to develop models using the R statistical tool, and applied to predict the number of malaria cases during one year, based on observations from preceding years. Models were generated using the number of malaria cases, names of administrative districts, month of the year, level of indoor residual spray applied, measurements of temperature (maximum and minimal), relative humidity and rainfall as attributes, considered in different time frames. The subdivision of the basic data set into several sub-training sets allowed for the analysis and determination of the time frame that best predicts future malaria cases and incidence. The model allowed for recognizing the most important variables in the study. The models were compared with respect to the mean squared error (MSE) and correlation coefficient. From the results, Indoor Residual Spray (IRS), month of January, minimal temperature and rainfall variables were found to be the most important factors when predicting the number of malaria cases, with some districts showing high malaria incidence. It was also discovered that by reducing the time window for what historical data to take into account, predictive performance can be increased substantially. Furthermore, the study showed that a consistent application of indoor residual spray may lead to a stable decrease of malaria cases in most of the districts in the study, as the results indicate (Zacarias & Bostrom, 2013).

A predictive model was also developed for public health professional's use. The model helps public health professionals to predict outbreak of malaria in Republic of Korea or South Korea. This study made use of data mining techniques alongside complex spatial and temporal (which previous correlation methodology cannot handle). The methodology used for the prediction model entailed 1. the selection of predictor variables: epidemiological, environmental, and socioeconomic data, transmission-reducing interventions like-DPRK mosquito net data, external funding for mosquito control sent to DPRK, and yearly malaria data for DPRK as predictor variables; 2. data pre-processing and data mining using fuzzy association rule; 3. rule extraction; 4. rule selection; 5. prediction generator (Buczak, et al., 2015)

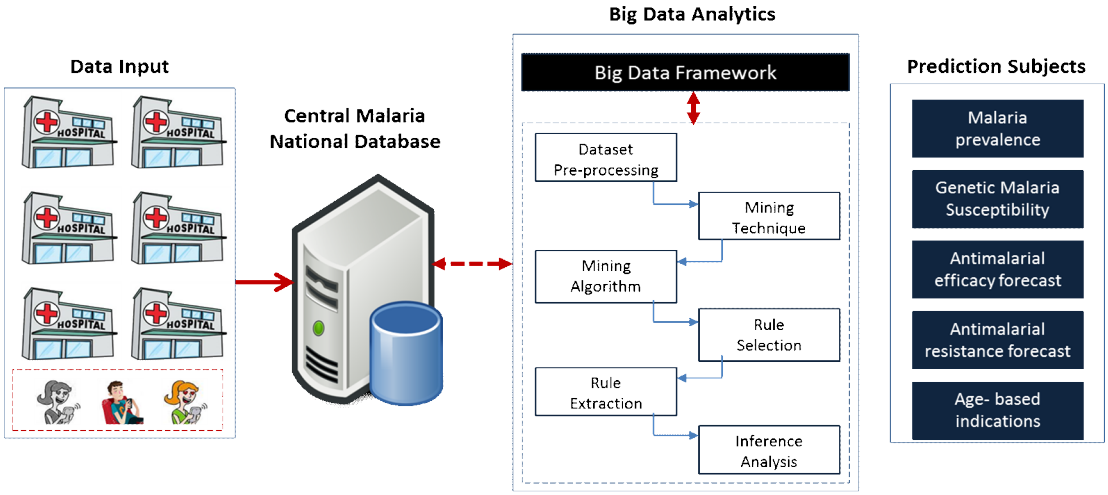


Figure 1: Generic prediction framework for malaria treatment outcomes using big data analytics

NATIONAL MALARIA PREDICTION FRAMEWORK

The establishment of a concrete national effort to predicting malaria outcomes can adopt the prediction framework presented in Figure 1. The business logic fashioned by the framework separates the prediction effort into 4 areas of concern: the data input, the national malaria database, the Big-data analytics, and the prediction interpretation.

a. ***Data Input:*** there is a great need to put a structure in place for continuous capturing of real time data. This is to ensure that the data set used for big data analytics largely captures true malaria incidences in the nation or region under study. The proposed framework is structured to receive malaria data from hospitals as they occur, as well as from individual patients. The medium through which these data is transmitted can vary from a web access to a mobile access or a user-driven GSM interaction.

b. ***Central Malaria National Database:*** a dedicated database server is required for housing the real time data received from the hospitals and the individual patients. There must be a way to uniquely identify and authenticate user access at the data-link layer.

c. ***Big Data Analytics:*** this serves as the core of the proposed National framework. It starts with a decision on the big data framework to be used for the analytics e.g. MapReduce.

   Following the big data framework, the analytics covers activities of pre-processing data that has been captured over a period. Based on a selected mining technique (association, classification, or clustering), an appropriate mining algorithm (K-Means, K-Nearest Neighbour, Naïve Bayes, Support Vector Machines, or Decision Trees) is applied for the generation of best rules. The best rules are selected and extracted for the generation of required prediction model. These stages require the use of mining tools such as WEKA, Rapid Miner, R-Programming, KNIME or Orange.

d. ***Possible Predictions:*** It is important to have the end in mind. The dynamics of the desired prediction pattern informs the mining techniques to be used and the details to look out for during inference analysis. Predictions that can be made from implementing a framework such as this includes: in-depth forecast on the prevalence of malaria, information on the genetic implications on future susceptibility to malaria, efficacy of antimalarial and resistance to effective antimalarial over a period, age-based, predictions on identified pattern in specific age group concerns, and many more.

METHODOLOGY

This study implements the big data analytics presented in the Nation Prediction Framework in order to predict best effective antimalarial drug combination to countries plagued with malaria disease. Large malaria data was sourced from the Institute of Child Health, University College Hospital, Ibadan – Nigeria. The dataset (which was made up of several attributes) was imported into WEKA (Waikato Environment for Knowledge Analysis) and pre-processed; the pre-processing phase involved the removal of certain attributes in the dataset that can be regarded as extraneous for this study hence, the required attributes reduced to 22. The 22 attributes analysed include *age, sex, fever, loss of appetite, vomiting, diarrhoea, cough, body pain, paracetamol, chloroquine, sulphadoxine and pyrimethamine, quinine, halofantrin, artemeter, mefloquine, weight, height, temperature, outcome, blood group, and genotype*.

In order to relate patients' symptoms with the medication, the association data mining technique was applied. This made use of the APRIORI algorithm for mining the data and for generating the best ten rules. This phase is called the rule selection phase. This phase is followed by the rule extraction phase where the rules that are most germane for this study were extracted and inferences made as regards the better drug use progression that should be adopted by a patient infected with malaria.

Results and Discussion

After applying the Apriori algorithm on the malaria dataset imported into the WEKA software, 10 rules were generated by the algorithm. These rules are the best 10 generated when the malaria

symptoms are associated with the medication for malaria. Figure 2 shows a screenshot of the best 10 rules generated by the WEKA software. The rules give the best progression to be followed in the treatment of malaria using medications.

```
1. qn=2.0 artemete=2.0 3073 ==> mefloq=2.0 3072    <conf:(1)> lift:(1.71) lev:(0.24) [1
2. qn=2.0 halofan=2.0 artemete=2.0 3066 ==> mefloq=2.0 3065    <conf:(1)> lift:(1.71) l
3. sp=2.0 qn=2.0 artemete=2.0 3044 ==> mefloq=2.0 3043    <conf:(1)> lift:(1.71) lev:(0
4. sp=2.0 qn=2.0 halofan=2.0 artemete=2.0 3037 ==> mefloq=2.0 3036    <conf:(1)> lift:(
5. aq=2.0 qn=2.0 artemete=2.0 2961 ==> mefloq=2.0 2960    <conf:(1)> lift:(1.71) lev:(0
6. fever=1.0 qn=2.0 artemete=2.0 2960 ==> mefloq=2.0 2959    <conf:(1)> lift:(1.71) lev
7. aq=2.0 qn=2.0 halofan=2.0 artemete=2.0 2954 ==> mefloq=2.0 2953    <conf:(1)> lift:(
8. fever=1.0 qn=2.0 halofan=2.0 artemete=2.0 2953 ==> mefloq=2.0 2952    <conf:(1)> lif
9. aq=2.0 sp=2.0 qn=2.0 artemete=2.0 2934 ==> mefloq=2.0 2933    <conf:(1)> lift:(1.71)
0. fever=1.0 sp=2.0 qn=2.0 artemete=2.0 2932 ==> mefloq=2.0 2931    <conf:(1)> lift:(1.
```

Figure 2: Screenshot of best 10 rules generated

Interpretation of Rules Generated

The first rule shows a progression of malaria treatment in which quinine is first prescribed and then if there is no much improvement, the prescription is changed to a combined therapy of artequin (artemeter and mefloquine) giving a cumulative mean potency value of 3072 for the two drugs.

The second rule shows a progression of malaria treatment in which quinine is first prescribed and then if there is no much improvement, the prescription is changed to halofantrin with the prescription changed again to a combined therapy of artequin (artemeter and mefloquine) now giving a cumulative mean potency value of 3065 for the combination of the three drugs.

The third rule shows a progression of malaria treatment in which sulphadoxine is first prescribed and then if there is no much improvement, the prescription is changed to quinine and changed again to a combined therapy of artequin (artemeter and mefloquine) giving a cumulative mean potency value of 3043 for the combination of the three drugs.

The fourth rule shows a progression of malaria treatment in which sulphadoxine is first prescribed and then if there is no much improvement, the prescription is changed to quinine with the prescription changed again to halofantrin if no major improvement is recorded, and then changed to a combined therapy of artequin (artemeter and mefloquine) now giving a cumulative mean potency value of 3036 for the combination of the four drugs.

The fifth rule shows a progression of malaria treatment in which aq is first prescribed and then if there is no much improvement, the prescription is changed to quinine with the prescription changed again to a combined therapy of artemeter and mefloquine (artequin) if no major improvement is recorded now giving a cumulative mean potency value of 2960 for the combination of the three drugs.

The sixth rule shows a progression of malaria treatment for a patient who has been confirmed to have fever as a symptom. In this scenario, quinine is first prescribed and then if there is no much improvement, the prescription is changed to a combined therapy of artequin (artemeter and mefloquine) now giving a cumulative mean potency value of 2959 for the combination of the two drugs.

The seventh rule shows a progression of malaria treatment in which aq is first prescribed and then if there is no much improvement, the prescription is changed to quinine with the prescription changed again to halofantrin if no major improvement is recorded and the prescription is

Table 1: Table of Results

again changed to a combined therapy of artequin (artemeter and mefloquine) in the event that the

patient is still showing symptoms of malaria now giving a cumulative mean potency value of 2953 for the combination of the four drugs.

The eighth rule shows a progression of malaria treatment for a patient who has been confirmed to have fever as a symptom. In this scenario, quinine is first prescribed and then if there is no much improvement, the prescription is changed to halofantrin and the prescription is changed again to a combined therapy of artequin (artemeter and mefloquine) now giving a cumulative mean potency value of 2952 for the combination of the three drugs.

The ninth rule shows a progression of malaria treatment in which aq is first prescribed and then if there is no much improvement, the prescription is changed to sulphadoxine with the prescription changed again to a combined therapy of artequin (artemeter and mefloquine) in the event that the patient is still showing symptoms of malaria now giving a cumulative mean potency value of 2933 for the combination of the three drugs.

The tenth rule shows a progression of malaria treatment for a patient who has been confirmed to have fever as a symptom. In this scenario, sulphadoxine is first prescribed and then if there is no much improvement, the prescription is changed to quinine with the prescription changed again to a combined therapy of artequin (artemeter and mefloquine) now giving a cumulative mean potency value of 2931 for the combination of the three drugs.

Rule Number	Drug 1	Drug 2	Drug 3	Drug 4	Cumulative Mean Potency Value
1	quinine			artequin	3072
2	quinine	halofantrin		artequin	3065
3	sulphadoxine	quinine		artequin	3043
4	sulphadoxine	quinine	halofantrin	artequin	3036
5	aq	quinine		artequin	2960
6	quinine	artemeter		artequin	2959
7	aq	quinine	halofantrin	artequin	2953
8	quinine	halofantrin		artequin	2952
9	aq	sulphadoxine	quinine	artequin	2933
10	sulphadoxine	quinine		artequin	2931

The rules 6, 8 and 10 were selected because of the inclusion of fever as a relevant symptom. In other words, all the three rules select fever as the dominant malaria symptom. The progression of the 3 rules selected is explained below:

i. **Rule 6:**
Patients are advised to periodically take prophylactic antimalarial (such as quinine)
IF clinical malaria symptoms emerge, THEN artequin (a combination of artemeter and mefloquin) should be administered next.

ii. **Rule 8**:
Patients are advised to periodically take prophylactic antimalarial (such as quinine)
IF clinical malaria symptoms emerge, THEN halofantrin should be administered.
IF clinical malaria symptoms persist, then artequin (a combination of artemeter and mefloquin) should be administered.

iii. **Rule 10:**
Patients are advised to periodically take prophylactic antimalarial (such as quinine)

IF clinical malaria symptoms emerge, THEN quinine should be administered.
IF clinical malaria symptoms persist, THEN artequin (a combination of artemeter and mefloquin) should be administered.

CONTRIBUTION TO KNOWLEDGE

The data captured is fundamental to the effectiveness of the proposed framework for malaria treatment prediction. Bearing in mind that existing mining techniques can generate reliable results, there is still need for developmental research in custom malaria mining techniques. This is to ensure that the algorithms applied are not just generic but they consider specific analytical concerns that are relevant to malaria disease prevention and eradication in countries. This study therefore proposes a generic prediction framework for malaria treatment outcomes using big data analytics.

CONCLUSION

The study involved an evaluation of the potency of a number of antimalarial medication commonly used in treating malaria. The results obtained after Association Rule Mining showed the best ten rules that represent the best sequence possible. The results obtained validates the practice of prophylactic administration of antimalarial medicine, with Quinine or Fansidar as the most efficacious. In cases where clinical malaria symptoms eventually emerge, the use of Artequin as an ACT which gave high potency values is recommended. In cases where malaria symptoms persist after administering prophylactic antimalarial, the use of halofantrin is recommended which is then succeeded with ACT if there is no significant improvement.

REFERENCES

Andreu-Perez, J., Poon, C. C., Merrifield, R. D., & Wong, S. T. (2015, July). Big Data for Health. *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, 19*(4). doi:10.1109/JBHI.2015.2450362

Buczak, A. L., Baugher, B., Guven, E., Ramac-Thomas, L. C., Elbert, Y., Babin, S. M., & Lewis, S. H. (2015). Fuzzy association rule mining and classification for the prediction of malaria in South Korea. *BMC Medical Informatics and Decision Making, 15*(47), 1-17.

Demas, A., Oberstaller , J., DeBarry, J., Lucchi, N. W., Srinivasamoorthy, G., Sumari, D., . . . Kissinger, J. C. (2011, July). Applied Genomics: Data Mining Reveals Species-Specific Malaria Diagnostic Targets More Sensitive than 18S rRNA. *Journal of Clinical Microbiology, 49*(7), 2411–2418. doi:10.1128/JCM.02603-10

Hermon, R., & Williams, P. A. (2014). Big data in healthcare: what is it used for? *Australian eHealth Informatics and Security Conference*. Edith. Retrieved from http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1021&context=aeis

IBM Institute for Business Value & Saïd Business School at the University of Oxford. (2012). *Analytics: The real-world use of big data*. New York: IBM Corporation. Retrieved from https://www.ibm.com/smarterplanet/global/files/se__sv_se__intelligence__Analytics_-_The_real-world_use_of_big_data.pdf

Institute for Health Technology Transformation. (2013). *Transforming Health Care Through Big Data*. New York: Institute for Heatlh Technology Transformation.

Olugbenga, O., & Clarence , S. Y. (2012). Computational Predictive Framework towards the Control and Reduction. *Egyptian Computer Science Journal ,ECS, 36*(2). Retrieved from http://eprints.covenantuniversity.edu.ng/740/1/2012-Gbenga.pdf

Pitale, P., & Ambhaikar, A. (2012). Sensitive region prediction using data mining technique. *International Journal of Engineering and Advanced Technology (IJEAT), 1*(1).

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems, 2*(3). doi:10.1186/2047-2501-2-3

Takahashi, D. (2013). *IBM uses big data to predict outbreaks of dengue fever and malaria*. VentureBeat. Retrieved from http://venturebeat.com/2013/09/29/ibm-uses-big-data-to-predict-outbreaks-of-dengue-fever-and-malaria/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Venturebeat+%28VentureBeat%29

Viceconti, M., Hunter, P., & Hose, R. (2015, July). Data, Big Knowledge: Big Data. *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, 19*(4). doi:10.1109/JBHI.2015.2406883

Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N. J., & Kaewkungwal, J. (2010). Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: A case study in endemic districts of Bhutan. *Malaria Journal*, 9(25). Retrieved from http://www.malariajournal.com/content/9/1/251

World Health Organisation. (2016). Retrieved from http://www.who.int/topics/malaria/en/

Zacarias, O. P., & Bostrom, H. (2013). Predicting the incidence of malaria cases in Mozambique using regression tress and forests. *International Journal of Computer Science and Electronics Engineering (IJCSEE)*, *1*(1).

SESSION C:

Cloud Computing and Applications

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

A MONITORING SYSTEM FOR PROVISION OF RESOURCE SERVICES IN THE CLOUD

O. F. Otusile
Computer Science Department, School of Computing and Engineering Sciences
Babcock University, Ilishan Remo, Ogun State, Nigeria
buhkieotusile@yahoo.com

O. Awodele
Computer Science Department, School of Computing and Engineering Sciences
Babcock University, Ilishan Remo, Ogun State, Nigeria
delealways@yahoo.com

A.C. Ogbonna
Computer Science Department, School of Computing and Engineering Sciences
Babcock University, Ilishan Remo, Ogun State, Nigeria
acogbonna@yahoo.com

S.O. Okolie
Computer Science Department, School of Computing and Engineering Sciences
Babcock University, Ilishan Remo, Ogun State, Nigeria
okolieso@babcock,edu.ng

A.O. Ajayi
Computer Science Department, School of Computing and Engineering Sciences
Babcock University, Ilishan Remo, Ogun State, Nigeria
deboxyl@gmail.com

ABSTRACT

There is a growing awareness among entrepreneurs of the need to be IT compliant. It is imperative to ensure that agreed Service Level Agreements (SLAs) are respected as customer satisfaction remains a vital focus of Service Providers (SPs). Appraisal of the minimum quantity of resources that a service provider has to provide to guarantee stipulated SLAs deserves a lot of attention and remains a viable research area. Hence, this study was set to implement a monitoring system for managing cloud service disposition based on acceptable SLAs.

A purposive design for determining in advance the needed Virtual Machines (VMs) to satisfy specified Quality of Service (QOS) in the SLAs was setup. This design proposed techniques for optimal provisioning of Cloud resources with the aim of maximizing profit by handling the dynamism associated with SLAs and heterogeneous resources. The proposed technique were evaluated using the CloudSim simulator with workloads from CloudMinder (a service product from Computer Associates (CA) Technologies). Data centres with physical machines which configuration resembled public Cloud such as Amazon EC2 large image, were simulated. The VMs of different types were mapped to physical machines and the general scheduling policy was time shared. The negotiation framework performs an adaptive and intelligent bilateral bargaining of SLAs between SaaS brokers and SaaS providers necessary for effective decisions for negotiation.

A prototype system called Service Level Agreement Management System (SLAMS) was implemented using SharePoint platform and Java programming language to validate and demonstrate the usefulness and feasibility of the proposed technique.

In conclusion, the negotiation framework provided more flexibility in terms of services needed to cater for variations associated with an individual customer. The use of the proposed framework for resource management by enterprise software as service providers was therefore recommended.

Keywords: Cloud Computing, Resources, Service Level Agreement, Software Service, Quality

3. INTRODUCTION

Over recent years, computer and Internet technologies have been incorporated into many everyday activities such as aircraft control, shopping, banking and so on (Linlin & Buyya, 2014). This rapid growth in interconnected computer networks has allowed companies to offer their services electronically (Murugesan 2011). Before the advent of Cloud technology (computing), the ICT administrative tasks were easy because the important objective of resource provisioning was performance (Mensce & Almeida, 2002). Eventually, the complexity of applications grew, thereby increasing the difficulties in their administration.

Consequently, enterprises realized that it is more efficient to outsource some of their applications to third-party Software as a Service (SaaS) providers enabled by Cloud computing due to some of the following reasons (Yang, et al., 2011): to reduce the maintenance cost, because as complexity increases the level of sophistication required to maintain the system also increases and enterprises need not to invest in expensive software licenses and hardware before knowing the business value of the solution.

A service is a software system used to perform a specific task for its customers using request-response messages (Linlin & Buyya, 2014). A service customer may choose a specific service from among similar ones that offer the same business. For this reason, it is a challenge for a service provider to maintain the running of the service at an adequate level in order to keep attracting potential customers (Linlin & Buyya, 2014; Mensce & Almeida, 2002). Customers' interest regarding the level of service offered may vary and this can be related to different dependability, performance and performability metrics such as response time, availability, throughput, reliability, exception handling, and security (Katerina & Kishor, 2000; Keller & Ludwig, 2003;). In this context, and in order to give customers the ability to choose which service is best suited to them, the term Quality of Service (QoS) has evolved to denote the quality of the non-functional properties of a service (Keller & Ludwig, 2003). Service providers and customers choose QoS metrics and specify guarantees of their values over a certain period of time; these are called Service Level Objectives (SLOs) (Keller & Ludwig, 2003). Owing to their importance in attracting customers, SLOs have become a crucial part of a larger legal document called a Service Level Agreement (SLA) (Keller & Ludwig, 2003).

The cloud computing technology is an efficient and cost effective way for managing and delivering services over the Internet. It allows customers to acquire resources fast on a pay-per-use basis, allowing minimized start-up costs and to rapidly scale up or down resources avoiding performance degradation in case of peak load and over-provisioning in case of scarce demand. SLAs were first developed in the 1980s by telecommunications firms and their importance was strengthened later by the Grid computing community. The most important reason why an SLA is used in service provision is to clarify and formalise relationships between the contractual parties regarding the overall quality of the service offered (Jin & Machiraju, 2002 Linlin & Buyya, 2014).

The problem of service level provisioning in clouds essentially consists of determining the minimum quantity of resources that a service provider has to use to guarantee stipulated SLAs in a highly dynamic environment.

The main challenge is to determine in advance the needed resources (Virtual Machines) since the incoming workload is extremely variable and difficult to predict. The service provider has to find a good trade-off between the risk of overestimating and underestimating the incoming traffic. Indeed, while in the first case there is an economic loss, in the second case the economic damage for the provider is determined by the penalty that has to be paid to customers if SLAs are violated. To perform online service level provisioning in a dynamic environment subject to highly changing workload conditions, autonomic solutions are required.

This study automates the determination of required Virtual Machines necessary for a SaaS provider to meet with established and the acceptable SLA (response time) in a dynamic environment with variable workload condition and providing a detailed implementation of SLA based management to demonstrate the usefulness.

2. RELATED REVIEW

The ability to monitor at application layer in Clouds provides the opportunity for efficient and cost-effective Cloud management. Carol &

Karsten (1991) presented ChaosMon, an application for monitoring and displaying performance information for parallel and distributed systems. ChaosMon supports application developers in specifying performance metrics and to monitor these metrics visually to detect, analyze, and understand factors that affects performance. This tool was a distributed monitor with a central control. It included local monitors that reside on the target machines and communicate the monitored information to the central control. However, this tool has not been applied in Cloud environments and it does not support SLA violation detection.

Zoltan, et al. (2001) propose application monitoring in Grid with GRM and PROVE, which were originally developed as part of the P-GRADE graphical program development environment running on Clusters. In their work, they showed how they transformed GRM and PROVE into a standalone Grid monitoring tool. However, their approach did not consider automatic finding of optimal measurement intervals.

Bartosz, et al. (2002) proposed an infrastructure for Grid application monitoring. Their approach was based on OCM-G, which is a distributed monitoring system for obtaining information and manipulating applications running on the Grid. They aimed to consider Grid-specific requirement and design a suitable monitoring architecture to be integrated into the OCM-G system. However, their approach considered only Grid specific applications.

Zoltan & Gabor (2003) discussed resource and job monitoring in the Grid. They presented a monitoring architecture with advanced functions like actuators and guaranteed data delivery. Their motivations toward application monitoring was to understand its internal operations and detect failure situations. They did not consider the monitoring of application resource consumption behaviours.

Bartosz et al. (2004) discussed the monitoring of Grid applications with Grid-Enabled OMIS monitor, which provided a standardized interface for accessing services. In their approach, they described the architecture of the system and provided some design details for the monitoring system to fit well in the Grid environment and support monitoring of interactive applications. Their monitoring goal was focused toward application development and they did not consider detecting application SLA violations.

Jan, et al (2007) proposed the building, deploying, and monitoring of distributed applications with Eclipse. In their approach, they first analysed applications using Eclipse to determine the best way to deploy them in a distributed manner. After deploying the applications, they applied a tool to visualize the distributed execution of the applications and identify factors affecting performance and failures. With this information they enforced the performance goals of the applications. However, they did not describe the usage of their approach in a large scale Cloud environment and moreover, their approach depended heavily on Eclipse framework.

Xu, et al (2009) proposed an architecture for monitoring of multi-tenant systems whereby they aimed to monitor QoS at tenant level to detect aggressive tenants consuming more resources as agreed. However, their architecture was theoretical. It was not implemented and there are no explanations of how to realize monitoring of resources consumed by a single tenant. Shicong, et al. (2009) present REMO - a resource-aware application state monitoring for large-scale distributed system, which produced a forest of optimized monitoring trees through iterations of two procedures. The first procedure explores the chances of sharing per message processing overhead based on performance estimation while the second procedure refines the monitoring plan produced by the first procedure. The authors argued that careful planning of multiple application state monitoring task, by jointly considering multi-task optimization and resource-constrained monitoring tree construction, can facilitate much gain in scalability and performance. However, their approach did not consider automatic finding of optimal measurement interval for efficient application monitoring.

Wang, et al. (2010) discussed a scalable run-time correlation engine for monitoring in a Cloud computing environment. Their approach was based on the use of log files to determine the behaviour of distributed applications. Thus, they developed a framework for run-time correlation of distributed log files in a scalable manner for enterprise applications in a Cloud environment. The correlation engine was capable of analyzing and performing symptom matching with large volume of log data. But, it did not consider automatic determination of intervals for

measuring/logging the application behaviours. Stuart, et al (2010) presented Lattice framework for Cloud service monitoring in the RESERVOIR EU project. It was capable of monitoring physical resources, virtual machines and customized applications embedded with probes. The Lattice framework is not generic since its application monitoring capabilities are restricted to applications preconfigured with probes and it does not consider measurement intervals in its operation.

Jin & Wang (2011) presented a performance guarantee for Cloud applications based on monitoring. The authors extracted performance model from runtime monitored data using data mining techniques, which was then used to adjust the provisioning strategy to achieve a certain performance goals. They did not consider finding optimal measurement intervals in their approach.

Massimiliano, et al (2011) proposed Cloud application monitoring using the mOSAIC approach. In a first step, the authors described the development of customized applications using mOSAIC API to be deployed on Cloud environments. For these applications, they proposed in a second step some monitoring techniques. Their interest was only to gather information that can be used to perform manual or automatic load-balancing, increase/decrease the number of virtual machines or calculate the total cost of application execution. Their approach did not consider the detection of SLA violations to avoid SLA penalty cost and moreover, it was not generic since it monitors only applications developed using the mOSAIC API.

Jin & Wang (2011) discussed a performance guarantee approach based on a performance model, which was extracted from actual runtime monitoring data using data mining techniques. It considered two QoS metrics: availability and response time. To build the performance model, they analyzed several attributes including number of CPU, number of application deployed on the same virtual machine, resource consumption, etc. However, their approach was not implemented and there are no evaluation results.

Ana (2015) presented the fundamentals of a toolkit for service platform architectures, which enable flexible and dynamic provisioning of Cloud services within the OPTIMIS EU project. The focus of the toolkit was aimed at optimizing the whole service lifecycle including service construction, deployment, and operation. It does neither detail the application monitoring strategy nor consider the determination of optimal measurement intervals.

However, these approaches did not consider the monitoring of SLA violations to avoid SLA penalty cost in SaaS cloud computing environment.

3. METHODOLOGY

The negotiation framework performs an adaptive and intelligent bilateral bargaining of SLAs between SaaS brokers and SaaS providers including negotiation policies, protocols, and strategies. The negotiation framework was evaluated and compared with a baseline.

In sophisticated markets, the negotiation objective is not only price but also other elements such as quality, reliability of supply, or the creation of long-term relationships. Multiple objectives were considered including cost, refresh time, process time and availability. The main objectives for a customer, a SaaS broker and a provider are:

- **Customer**: minimize price and guaranteed QoS within expected timeline.
- **SaaS Broker**: maximize profit from the margin between the customer's budget and the providers' negotiated price.
- **SaaS Provider**: maximize profit by accepting as many requests as possible to enlarge market share.

The negotiation protocol refers to a set of rules, steps or sequences during the negotiation process, aiming at SLA establishment. It covers the negotiation states (e.g. propose offer, accept/reject offer, and terminate negotiation). It is common to characterize negotiations by their settings: bilateral, one-to-many, or many-to-many. This work focuses on the one-to-many bargaining setting, where three types of agents are considered (CA, BCA and PA). A BCA negotiates with many PAs in a bilateral fashion.

During the negotiation process, the negotiation status is updated using negotiation states described in Table 1.

STATES	DESCRIPTION
PROPOSE	THE AGENT PROPOSE INITIAL OR COUNTER OFFER TO THE OPPONENT AGENT.
REJECT	THE AGENT DOES NOT ACCEPT THE OFFER PROPOSED BY THE OPPONENT AGENT.
ACCEPT	THE AGENT ACCEPTS THE OFFER PROPOSED BY THE OPPONENT AGENT.
FAILURE	SYSTEM FAILURE, TRIGGER RENEGOTIATION.
TERMINATE	NEGOTIATION IS TERMINATED DUE TO TIMEOUT OR NO MUTUAL AGREEMENT.

Table 1: The Negotiation States and Description Summary

The sequential negotiation process for this framework is described as follows and depicted in Figure 1:

Phase 1: *CA submits requests*: CA requests services on behalf of the customer to the Broker.

Phase 2: The BCA requests initial proposals from all providers, who are registered in the Directory. The values sent from BCA to PAs are expected values.

Phase 3: *PAs propose initial offer:* All PAs propose initial offers based on their current capabilities and availability to fulfil BCA's requirements.

Phase 4: *Negotiation Process with PAs:*

a). If there are providers who can fulfil all requirements, then the BCA selects the best vendor.

b). If there is no provider that can fulfil all requirements, then the BCA starts the negotiation process with PAs.

Step 1: BCA selects the best initial offer from all offers that are proposed by all providers according to the objective.

Step 2: BCA adjusts its initial offer according to the offer selected in **Step 1** to generate new counter offer and propose it to all providers.

Step 3: A PA evaluates BCA's counter proposal.

Step 4: If the counter offer proposed by BCA cannot be accepted, PA proposes a counter offer.

Step 5: Terminate negotiation. There are three termination conditions: First, when negotiation deadline expires. Second, when the offer is mutual agreed by both the CA and the PA. Third, when BCA is not able to accept any counter offer proposed by all providers within the negotiation deadline.

Phase 5: *SLA Generation:* Initiate SLA creator to generate SLA for customer and provider respectively using SLA templates stored in KB.

Phase 6: *Send SLA to all participants:* The generated SLA will be sent to the customer and provider respectively by the SLA creator.

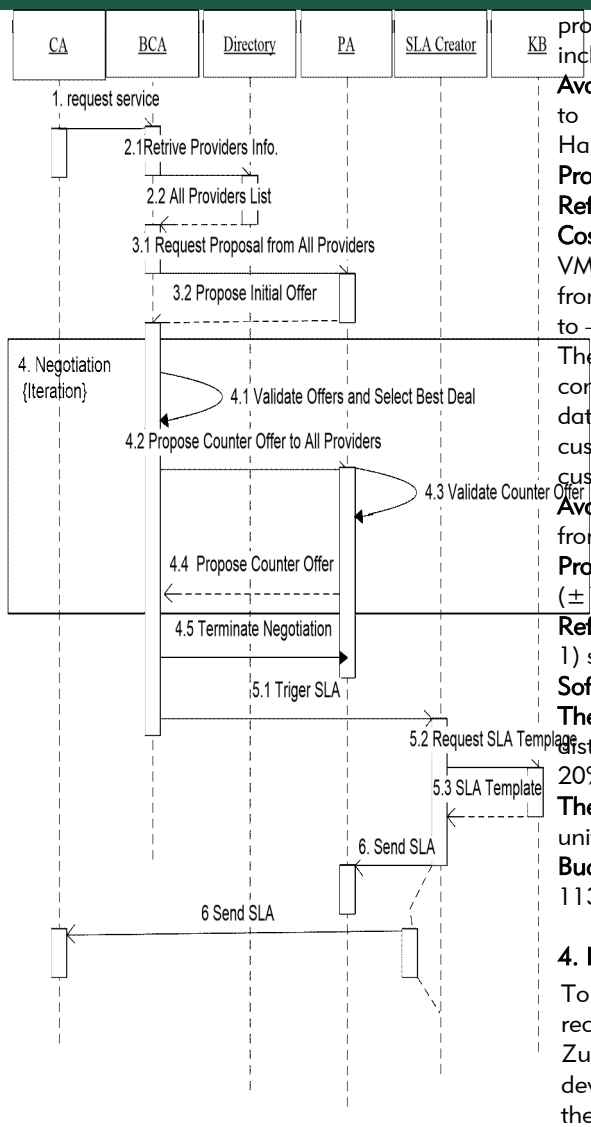


Figure 1: The Interaction between Components during Negotiation Process

A prototype of the framework considering both time and market factors using real data shared by cloud provider CA Technologies was implemented. CA Technologies offers a number of enterprise software solutions to customers delivered as SaaS. The data provided included the response, refresh and processing times of an enterprise solution hosted on VMs, as measured by the quality assurance team. Availability data was collected from CloudHarmony benchmarking system (2015), which provides real data from Cloud providers. These data were collected over 4 days including weekdays, weekends.

Availability: Varies from 98.654% (Colosseum) to 100% (Amazon EC2) as derived from Cloud Harmony.

Process Time: The mean 5.243 (± 2.043) s.

Refresh Time: The mean 1.581 (± 1.383) s.

Cost: Cost is considered similar to Windows VMs from 3rd party IaaS providers, which varies from -N-96.39 kobo per hour (VCloud Express) to –N-130.41kobo per hour (Amazon EC2).

The experiments conducted considered 50 concurrent users based on the CA provided data, which is designed according to their customer historic data. The summary of customer data is:

Availability: uniformly distributed and varies from 99.95% to 100%.

Process Time: normally distributed mean 1.5 (±1) s.

Refresh Time: normally distributed mean 2 (± 1) s.

Software service set: consists of 3 editions.

The expected discount percentage: normally distributed with mean value 30% (variation ± 20%).

The preference level of each QoS parameter: uniformly distributed between 0 and 1.

Budget: normally distributed with mean -N-11339.96 kobo (± -N-2834.99 kobo).

4. RESULTS

To compare the proposed heuristics, the most recent work related on negotiation proposed by Zulkernine and Martin (2011) was used, who developed a time-based Sigmond function in their negotiation process for generating counter offers. However both time and market functions in Clouds was considered. To compare negotiation strategy, their heuristics and Sigmond function was used for implementation with the objective of cost minimization.

The following performance metrics were considered for evaluation based on the objectives of the negotiating parties:

Average broker's profit: The broker's average profit from accepted customers.

CSL improvement: The average CSL improvement over base.

Average provider's profit: The average provider's profit for accepting customers.

Average round of negotiation: The average number of negotiations conducted during the negotiation process to reach mutual agreement.

1) Variation of Negotiation Deadline

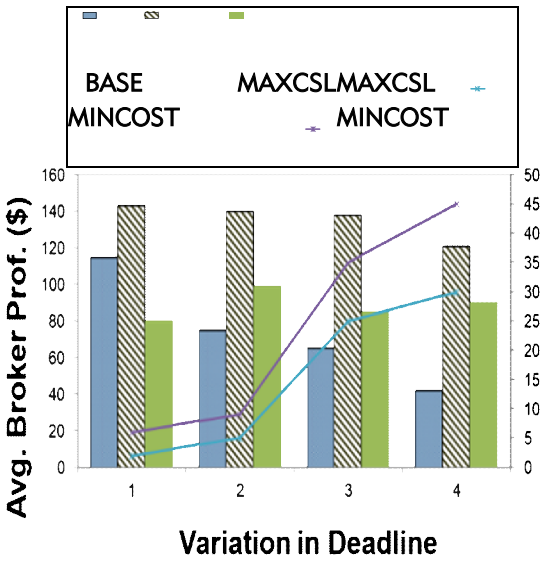
The experiment was designed to evaluate mincost and maxcsl during negotiation deadline variations.

The bar chart in Figure 2a represents average broker profit while the line chart represents the CLS improvement over base heuristic. For all the negotiation deadline variations, mincost generates the highest profit (up to 400%) for the broker over maxcsl and base. The reason for such a trend is that the broker concedes less or bargains harder for more profit. In terms of CSL improvement, maxcsl results in the highest improvement (up to 15%) over base, since it is designed to sacrifice profit for a higher CSL.

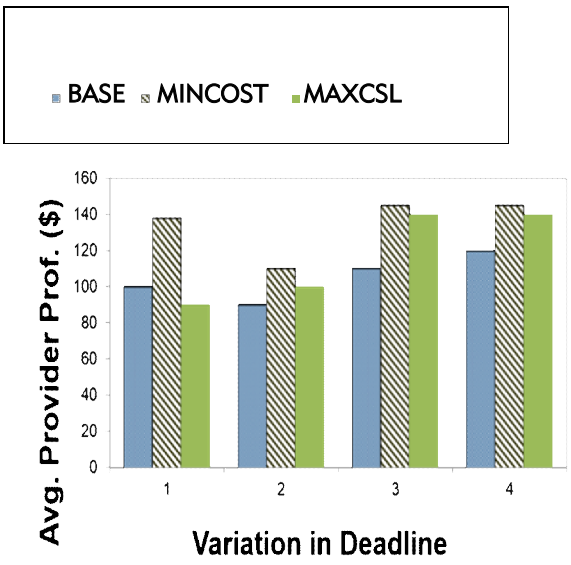
From the providers' perspective (Figure 2b), on average maxcsl generates more profit for providers, because the maxcsl aims at satisfying all Issues within the broker's budget, which leaves more profit for providers.

Figure 2c shows the average negotiation round for base increases dramatically when deadlines were varied (as base is only time dependent), whereas the proposed heuristics increases slightly (less than 2 rounds), as market factors also impact on the negotiation process. In terms of the number of successful negotiations (Figure 2d), when the deadline becomes trivial, the proposed heuristic performs better and increases in trend, as there is more bargaining time.

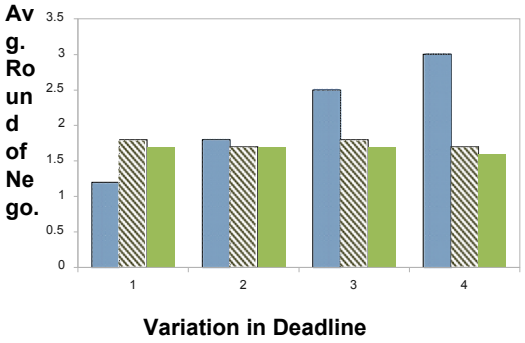
In summary, mincost generates more broker profit while maxcsl generates improved CSL and increased provider profit by increasing the number of successful negotiations with similar negotiation rounds.



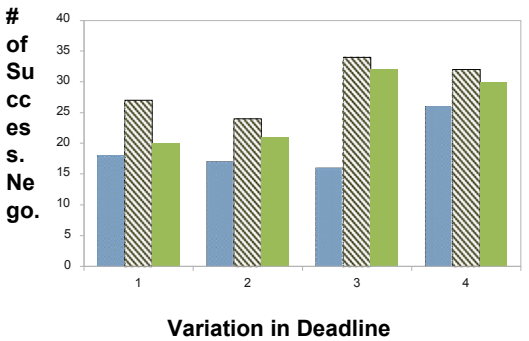
(a). Average Broker Prof.($)



(b). Avg Provider Prof.($)



(c). Avg Round of Nego.



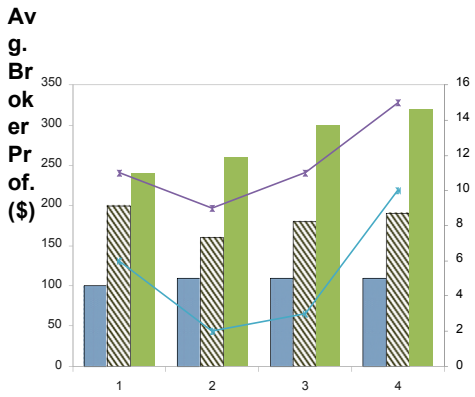
(d). # of Success. Neg.

Figure 2: Impact of Deadline Variation

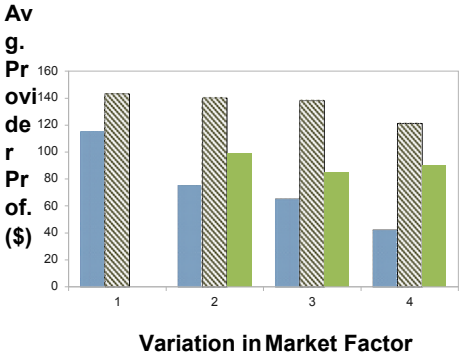
2) Variation of the Market Factor

The experiment was conducted to evaluate the proposed heuristics during the variation of market factors. When market factors vary from 1 to 4, which represents an increase in market competition, the mincost generates up to twice the profit than the base (Figure 3a bar chart) and the maxcsl improves up to 4 times more CSL compare to mincost (Figure 3a line chart). The broker's profit generated by base only changes slightly during market factor variations, as base does not consider market conditions.

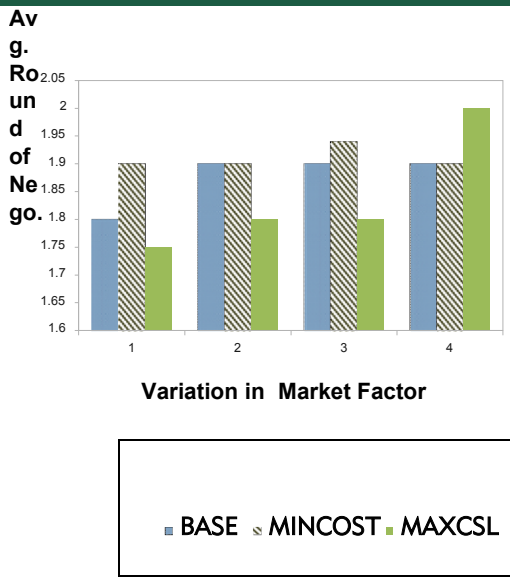
Figure 3b illustrates that the provider's profit decreases due to an increase in market competition. The maxcsl generates more profit for providers than mincost and base, as maxcsl considers the CSL as the highest priority, which leaves more profit for providers. When competition increases, more negotiation rounds are required to reach agreement (Figure 3c), as participants bargain harder and the number of opportunities to reach agreement increases (Figure 3d). To conclude, the experiment demonstrates that mincost produces more profit while the maxcsl achieves better CSL for the broker and more profit for providers.



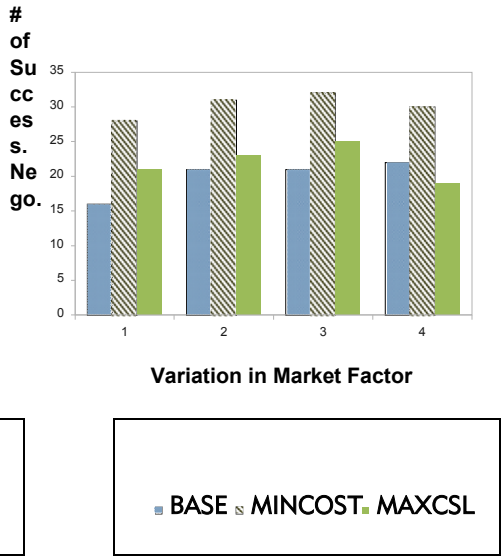
(a). Average Broker Prof.($)



(b). Avg Provider Prof.($)



(c). Avg Round of Nego.



(d). # of Success. Neg.

Figure 3: Impact of Market Factor Variation

5. CONCLUSION

As SaaS providers want to enlarge market share, they need to provide more flexibility in terms of services to cater to variations associated with an individual customer. This is generally done by a negotiation process between customers and service providers. However, while undertaking this negotiation process, the service provider needs to take into consideration not only what they can provide to customers but also the competition with other SaaS providers. Thus, the new negotiation frameworks proposed are needed for the SaaS provider that considers dynamism in Cloud environment with time and market factors to make the best possible decisions for negotiation. The proposed negotiation framework can be used for the SaaS provider and the SaaS broker model.

6. ACKNOWLEDGEMENT

I wish to acknowledge the Association of African Universities (AAU) for the grant which assisted me in pursue of my doctoral degree.

7. REFERENCES

Ana J. (2015). Optimis: a holistic approach to cloud service provisioning. http://www.optimis-project.eu/.

Bartosz B., et al (2004). Monitoring grid applications with grid-enabled omis monitor. In Grid Computing, volume 2970, pages 230–239.Xu C., Yuliang S., and Qingzhong L. (2009). A multi-tenant oriented performance monitoring, detecting and scheduling architecture based on sla. In 2009 Joint Conferences on Pervasive Computing (JCPC), pg 599 – 604.

Bartosz B., et al (2002). An infrastructure for grid application monitoring. In Proceedings of the 9th European PVM/MPI Users' Group Meeting on Recent Advances in Parallel Virtual Machine and Message Passing Interface, pages 41–49.

Carol K., and Karsten S. (1991). Chaosmon - application-specific monitoring and display of performance information for parallel and distributed systems. In Proceedings of the 1991 ACM/ONR workshop on Parallel and distributed debugging, PADD '91, pg 57–67.

Faratin, P., et al (1998). Negotiation Decision Functions for Autonomous Agents, Robotics and Autonomous System, 24(3-4), (pp. 159-182).

Jin S., and Wang Q. (2011). A performance guarantee approach for cloud applications based on monitoring. In 2011 IEEE 35th Annual Computer Software and Applications Conference Workshops (COMPSACW), pg 25 – 30.Massimiliano, Salvatore, Tam, Gorka & Gorka (2011)

Jin, L. J., and Machiraju, V. A. (June 2002). Analysis on Service Level Agreement of Web Services. Technical Report HPL-2002-180, Software Technology Laboratories, HP Laboratories.

Keller, A., Ludwig, H. (2003). The wsla framework: Specifying and monitoring service level agreements for web services. J. Netw. Syst. Manage. 11(1) (2003) 57-81.

Linlin W., Buyya R. (2014). Service Level Agreement (SLA) in Utility Computing Systems, Cloud Computing and Distributed Systems (CLOUDS) Laboratory.

Mensce, D., and Almeida, V. (2002). Capacity Planning for Web Performance: Metrics, Models and Methods. Prentice-Hall, Upper Sadale River, NJ.

Murugesan M. (2011). Cloud computing concepts. In 2011 3rd International Conference on Electronics Computer Technology (ICECT), volume 6, pages 236 –239.

Parkhill, D. (1966). The Challenge of the Computer Utility, Addison-Wesley, USA.

Peter M. and Tim G. (2009). The nist definition of cloud computing. National Institute of Standards and Technology, 53(6):50.

Stuart C., et al (2010). Monitoring future internet service clouds. In Towards the Future Internet - A European Research Perspective book.

Wang M., et al (2010). Scalable run-time correlation engine for monitoring in a cloud computing environment. In 2010 17th IEEE International Conference and Workshops on Engineering of Computer Based Systems (ECBS), pg 29 –38.

Yang, E. F., et al (2011). A Hybrid Approach to Placement of Tenants for Service-Based Multi-tenant SaaS Application. Proceedings of the 6th IEEE Asia-Pacific Services Computing Conference, Korea.

Zoltán B., and Gábor G. (2003). Resource and job monitoring in the grid. In Proceedings of the International Conference on Parallel and Distributed Computing (Euro-Par'03), pg 404–411.

Zoltán B., et al (2001). Application monitoring in the grid with grm and prove. In Proceedings of the International Conference on Computational Sciences-Part I, ICCS '01, pages 253–262.

Zukernine, F., and Martin, P. (2011). An Adaptive and Intelligent SLA Negotiation System for Web Services. IEEE Transactions of Service Computing, 4(1), (pp. 31-43).

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

ADOPTION OF CLOUD COMPUTING SERVICES BY NIGERIAN ENTERPRISES; ISSUES AND CHALLENGES

C.C. Chigozie-Okwum
National Open University of Nigeria, Lagos
chiomaokwum@gmail.com

S.G. Ugboaja
Michael Okpara University of Agriculture, Umudike
sammuelugboaja@gmail.com

D.O. Michael
Alvan Ikoku College of Education, Owerri
hanniboff@gmail.com

ABSTRACT

The study aimed at showcasing issues and challenges associated with the adoption of cloud computing services by enterprises in Nigeria. The study was a survey type, 20 companies were purposively sampled. Questionnaires and interviews were used to gather primary data. Primary data collected were analyzed using frequency distribution tables and percentages. The study revealed that a great percentage of the respondents had various degree of awareness of cloud computing services, and despite the awareness they enterprises seldom adopted cloud computing services as was represented by 60% of respondents pointing to this fact. Efficiency in service delivery to their customers (100%), and on- demand self-service (100%) among other benefits were enumerated by enterprises as what they intended to enjoy should they move their services to cloud. The study also pin pointed challenges that faced adoption of cloud computing by enterprises; and Cloud data security, Privacy breech of users' data and cost implications stood as with high percentages as major challenges facing cloud adoption by Nigerian enterprises among other challenges. The respondents identified factors that would improve adoption of cloud computing services by enterprises and these include; ensuring cloud data security (100%), decrease in privacy breech issues (100%), and increase in staff integrity of cloud providers (100%), amongst others. The study recommended that cloud service providers should strive to maintain high level of trust with their customers by ensuring cloud data security, reducing privacy breech issues and lock-in, ensure reliability in other to win user confidence and attract more enterprises to move their IT services to cloud.

KEYWORDS: ADOPTION, CHALLENGES, CLOUD COMPUTING, DATA SECURITY, ENTERPRISES.

1.1 INTRODUCTION.

Cloud computing is a set of IT services that are provided to the customer over a network on a leased basis, and with the ability to scale up or down their services requirements. Cloud computing is a form of distributed computing that provides information technology (IT) services as a commodity and enable users to benefit from the efficient use of computer resources, (Radack, 2012). Users are able to control the computing service they access while sharing the investment in the underlying IT resources with other consumers. When computing resources are provided by another organization over wide area network, cloud computing becomes similar to an electric power utility platform. Users of cloud computing services do not have knowledge of the physical location of the server or how the processing of their data is configured. To this effect, users consume service without information about the processes involved. According to Chigozie-Okwum, (2015), Cloud Computing is not a very new concept in IT; in fact Cloud Computing is a more advanced version of the Data Processing Service Bureaus that we had 40 years ago. Nevertheless, the best known companies in the IT field offer or will shortly offer Cloud Computing services to a range of customers from organizations of all sizes to individuals. The biggest and best known Cloud Computing providers include Amazon with EC2, Microsoft with Azure and Google with Google apps (e.g. Gmail, Google Docs, and Google Calendar). The paradigm of Cloud Computing can be described in simple terms as offering particular IT services that are hosted on the internet, (Chigozie-Okwum, 2015). Data in the cloud are easy to manipulate but also easy to lose control of. Despite the benefits enterprises stand to enjoy should they move their IT services to cloud the rate of cloud adoption by Nigerian Enterprises is still low hence prompting the need to brainstorm on the issues and challenges facing adoption of cloud computing services by enterprises in Nigeria. With the market growth of cloud computing in developing countries which Nigeria is not an exception, the cloud service technology would houses application with wide area in electronics, health, commerce, education and business which are critical sector of national development, (Afshari et al, 2014). Because most business and service providers in Nigeria are looking for way to equate their business with their counter part in other countries, and that could only be possible by adopting the current trail in cloud technology.

1.2 CLOUD COMPUTING: A CONCEPTUAL OVERIEW.

According to Kuyoro S.O et al, (2011), cloud computing is defined "as a set of IT service that are provided to a customer over a network on leased basis with the ability of scaling up or down their service requirement". The National Institute of Standard and Technology, (NIST, 2012), defined cloud computing as " a model for enabling convenient on-demand network access to a shared pool of configurable computing resources (e.g. network , servers, storage, application and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction . Usually, cloud computing service are delivered by a third party provider who owns the infrastructure. Cloud computing is a model for enabling convenient, on demand network access to a pool of configurable computing resource that can be rapidly provisioned and released with minimal management effort or service provider interaction. It reduces the capital expenditure and operational expenditure involved in the IT infrastructure of an organization. Cloud computing offers an innovative business model for organization to adopt IT services without upfront investment. Cloud computing encompasses activities such as use of social networking sites and other forms of interpersonal computing.

Cloud computing is an emerging technology that can help organizations become more efficient and agile and respond more quickly and reliably to their customer needs. Many government and private sector organization are currently using or considering the use of cloud computing to increase the performance and efficiency of their information system operation, and to improve the delivery of

services (Radack , 2012). ITU – T technology watch, (2012), states that cloud computing refers to the ability to access and manipulate Information stored on servers using an internet enabled platform, including Smart phones. Computing facilities and applications will increasingly be delivered as a service over the internet. Furthermore, it states that we are already making use of cloud computing when for example we use application such as Google mail, Microsoft office 365 which is the software as a service commercial offering of Microsoft office or Google docs. In the future they went on to emphasize , that governments companies and individual will increasingly turn to cloud. The cloud computing paradigm changes the way in which information is managed, especially where personal data processing in concerned. End – user can access cloud service without the need for any expert knowledge of underlying technology, without knowledge of the physical location of the server or of how the processing of personal data is configured end user consume cloud service without any information about the process involved.

CLOUD DEPLOYMENT MODELS

Cloud computing can be deployed in three different models namely;

I. Private cloud -enterprise owned or leased
Ii. Public cloud - sold to the public mega scale infrastructures
Iii. Hybrid cloud – the combination of two or more cloud types

Private Cloud :- It is set up within an organization's internal enterprise data center (Kuyoro et al , 2011). In private cloud, scalable resources and virtual application provided by the cloud vendor are pooled together and available for cloud user to share and use. It differs from public cloud in the sense that all the cloud resources and application are managed by the organization itself, similar to intranet functionality. Utilization on the private cloud can be much more secure than that of public cloud because of its specified internet exposure. Only the organization and designated stakeholder may have access to operate on a specific private cloud (Arnold 2009).

Public Cloud: Public cloud describes cloud computing in the traditional mainstream sense, whereby resources are dynamically provisioned on a fine – grained basis over the internet via web application / web services , from an offsite third provider who share resources on a bills on a time –grained utility computing basis (Kayoro et al, 2011) .public clouds are less secure than other cloud models because it places additional burden of ensuring all applications and data accessed on the public cloud are not subjected to malicious attacks.

Hybrids Cloud:- Hybrid cloud is a private cloud linked to one or more external cloud services, centrally managed, provisioned as a single network unit and circumscribed by a secure network. It provides virtual IT solutions through a mix of both public and private clouds. Hybrid cloud provides more secure control of the data and applications and allows various parties to access information over the internet.

CLOUD SERVICE MODELS

The three cloud service delivery models are

i. Software as a Service (SaaS)
ii. Platform as a Service (PaaS)
iii. Infrastructure as Service (IaaS)

These three classic cloud service ,model have different divisions of responsibility with respect to personal data protection. The risk and benefit associated with each model will also differ, and need to be determined on a case – by – case basis and in relation to the nature of the cloud service in question.

Software as a Service (SaaS):

Software as a Service is a software distribution model in which application are hosted by a vendor or service provider and made available to customer over a network, typically the internet, (Kuyoro et al,2011). SaaS enables the consumer to use the providers applications running on a cloud infrastructure. The applications are accessible

from various client devices through a client interface such as a web browser (e.g. web-based Email such as Gmail or CRM from Saleforce) . With the SaaS models, the consumer has little or no influence on how input data is processed, but should be able to have confidence in the cloud providers' responsibility and compliance or can control which input he gives to a SaaS.

Platform as a Service (PaaS)
According to Kuyoro et al 2011, platform as a service is a set of software and development tools hosted on the provider servers. This offers an integrated set of developer environment that a developer can tap in to build their application without having any clue about what is going on underneath the service. It offers developers a service that provides a complete software development life cycle management from planning to design to building application, to deployment, to testing, to maintenance. PaaS provides tools supported by a cloud provider, that enable developers to deploy applications (e.g.Salesforces' force .com, Google APP Engine, Mozilla Bespin, Zoho creator etc.) (ITU-T Technology Watch 2012).

Infrastructure as a Service (IaaS):
Infrastructure as a service is a single tenant cloud layer where the cloud computing vendor's dedicated resources are only shared with contracted client at a pay – per use fee . This greatly minimize the need for huge initial investment in computing hardware such as serves, networking devices and processing power.
They also allow varying degrees of financial and functional flexibility not found in internal data centre or with collocation services, because computing resource can be added or released much more quickly and cost – effectively than in an internal data centre or with model a collocation services (Brodkin , 2008).

CHARACTERISTICS OF CLOUD COMPUTING.

The National Institute of Standards and Technology's definition of cloud computing identifies "five essential characteristics" of cloud computing which include;

1. **On-demand self-service**. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

2. **Broad network access**. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).
3. **Resource pooling**. The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.
4. **Rapid elasticity**. Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear unlimited and can be appropriated in any quantity at any time.

5. **Measured service**. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service. (National Institute of Standards and Technology, 2011).

1.3 CLOUD COMPUTING CHALLENGES

Cloud Computing is often marketed as an efficient and cheap solution that will replace the client-server paradigm. The paradigm shift involves/results in the loss of control over data as well as new security and privacy issues. For

this reason caution is advised when deploying and using Cloud computing in enterprises. With Cloud Computing rapidly gaining popularity, it is important to highlight the resulting risks. As security and privacy issues are most important, they should be addressed before Cloud Computing establishes an important market share. Many IT and important research agencies are aware of these risks and have produced reports and analyses to document them (Brodkin and Gartner), (Microsoft whitepaper, 2010), (ENISA report, 2009), (CSA report, 2009).

The current adoption of cloud computing is associated with numerous challenges causing users to be skeptical about its authenticity. Despite all the hype surrounding the cloud, customers are reluctant to deploy their businesses to the cloud. Security issues in the cloud has played a major role in slowing down its acceptance. In fact security ranked highest and first as the greatest challenge to cloud computing (Gens, 2009).

Based on a survey conducted by IDC in 2008, the major challenges that prevent cloud computing from being adopted as recognized by organizations are as follows:

1. **Security:-** It is clear that security issue has played the most important role in hindering cloud computing acceptance. Well – know security issue such as data loss, phising , botnet (running remotely on a collection of machine) pose serious threats to organization data and software.

2. **Costing model**: Cloud consumers, must consider the tradeoffs amongst computation, communication and integration. While migrating to the cloud can significantly reduce the infrastructure cost, it does raise the cost of data communication i.e. The cost of transferring an organization data to and from the public and community cloud and the per unit of computing resources used is likely to be higher.

3. **Charging Model:-** The elastic resource pool has made the cost analysis a lot more complicated than regular data centers, which calculate their cost based on consumptions, of static computing. Moreover an instantiated virtual machine has become the unit of cost analysis rather than the under lying physical server.

4. **Service Level Agreement (SLA)**; Although cloud consumers do not have control over the underlying computing resources; they do need to ensure the quality, availability, reliability and performance of these resources when consumers have migrated their core business functions onto their entrusted cloud. In other words, it is vital for consumers to obtain guarantees from providers on service delivery.

5. **What to Migrate:** based on a survey (sample size = 244) conducted by IDC in 2008, the seven IT systems/application being migrated to the cloud are
   a. IT management applications (26.21%)
   b. Collaborate applications (25.4%)
   c. Personal applications (25%)
   d. Business applications (23.4%)
   e. Applications development and deployment (16.8%)
   f. Server capacity (15.6%)
   g. Storage capacity (15.5%)

The survey shows that organizations still have security/privacy concerns in moving their data cloud. The survey also shows that in three years time 31.5% of the organization will move their storage capacity to the cloud.

Cloud Interoperability Issues: currently, each cloud offering has its own way on how cloud clients/applications/users interact with the cloud, leading to the "**Mazy Cloud"** phenomenon. This severely hinders the development of cloud ecosystems by forcing vendor locking which prohibits the ability of users to choose from alterative vendors/offering simultaneously in order to optimize resources at different levels within an organization.

Furthermore, Gartner, (2008), in his survey identified seven security Issues that needed to be address before enterprises consider switching to the cloud computing model. They are as follows:

1. **Privileged user access:** information transmitted from the client through to internet poses a certain degree of risk , because of issue of data ownership. Enterprises should spend time getting to know their providers and their regulations as much as possible before assigning some trivial applications first to test the waters.

2. **Regulatory compliance:** clients are accountable for the security of their solution as they can choose between provider that allow to be audited by 3rd party organization that check level of security and provider that don't .

3. **Data Location:** depending on contracts some client might never know what country or what jurisdiction their data is located.

4. **Data segregation:** Encrypted information from multiple companies may be stored on the same hard disk, so a mechanism to separate data should be deploy by the provider
5. **Recovery:** Every provider should have a disaster recovery protocol to protect user data.
6. **Investigation support:** if a client suspects faulty activity from the provider, it may not have many legal ways to pursue an investigation.
7. **Long-term viability:** this refers to the ability to retract a contract and all data if the current provider is bought out by another firm.

A recent survey by cloud security alliance (CSA) and IEEE indicates that enterprises across sector are eager to adopt cloud computing but that overcoming challenges amongst which is ensuring security are needed both to accelerate cloud adoption on a wide scale and respond to regulatory drivers. It also details that cloud computing is shaping the future of IT but the absence of a compliance environment is having dramatic impact on cloud computing growth (CSA, 2010).

The Implementation of any of the cloud models, either the private or public model or combination of both needs adequate expertise to maximize benefits of adopting cloud services. The challenges encountered by most enterprises in Nigeria are the lack of knowledge of critical requirement for cloud service infrastructure. These drawbacks would make them not to be confident enough to adopt such cloud service. Awareness of the kind of business enterprises model being run by Nigerian enterprises would give a greater insight of the cloud infrastructure. The fair of whether the existing cloud infrastructure could handle or influence latest virtualization, hardware software solution and accommodates or integrate traditional IT system requirement to meet the standard of current data center, Liang and Ulander, (2010). The lack of standard by regulating computing body to providing tool that would determine the defining and metering of the cloud service provided in order for the services to be quantified could be a challenge of adopting cloud computer service. This is where the service management comes into play when the Nigeria enterprise can have confidence in the service management provided by cloud service providers.

2.1 PROBLEM STATEMENT

Cloud computing has a lot of beneficial properties which enterprises enjoy should they deploy their IT services to cloud. However, it is observed that the rate at which enterprises in Nigeria are adopting cloud computing services is at its lowest ebb. This problem now challenged the researcher to design a survey aimed at attempting to find out the issues and challenges that face the adoption of cloud computing services by enterprises in Nigeria today.

2.2 RESEARCH QUESTIONS

The study aimed at providing answers to the following research questions;

1. How aware are Nigerian enterprises on cloud computing services?
2. Do Nigerian enterprises often adopt cloud computing services?
3. What benefits do enterprises intend to enjoy should they choose to adopt cloud computing services?
4. What challenges face the adoption of cloud computing services by Nigerian enterprises?
5. What factors could improve cloud adoption by Nigerian Enterprises?

2.3 METHODOLOGY

The study was a survey type of research carried out between November and December 0f 2015. The broad objective of the study was finding out the issues and challenges confronting adoption of cloud computing services by enterprises in Nigeria. The specific objectives include;

1. To determine the level of awareness of Nigerian enterprises on the concept and deployment of cloud computing technologies.
2. To determine how often Nigerian Enterprises adopt cloud computing technologies.
3. To establish benefits Enterprises were to enjoy should they adopt cloud computing technologies.
4. To find out challenges facing cloud adoption by Nigerian enterprises.
5. To determine factors that if put in place could improve cloud adoption by Nigerian enterprises.

Twenty (20) Information and Communication Technology compliant companies in Owerri Municipal Area Council of Imo State were purposively sampled. The instruments for primary data collection included a 10-item structured questionnaire, distributed to the 20 sampled enterprises and interview of respondents. The 20 distributed questionnaires were properly answered and returned hence making the response rate to be **n = 20.** A four-point modified Likert scale was used in rating the questionnaire responses. The questionnaires were validated by 5 experts in the field of study. The questionnaire responses were analyzed using frequency distribution tables and percentages. Secondary data were collected from review of related literature.

3.1 RESULTS

The results of the survey as extracted from the questionnaires are presented and interpreted below:

Responses	Frequency	Percentages (%)
Very aware	6	30
Aware	10	50
Fairly aware	4	20
Unaware	-	-
Total	20	100

Table 1: Level of awareness of Nigerian enterprises on cloud computing services.
Source; field data 2015

Table 1 above shows that of the 20 sampled companies, 6 were very aware of the existence and services of cloud computing techniques, 10 companies representing 50% of the sampled population were aware while 4 of the sampled companies stated they were fairly aware of cloud computing services. However worthy of note is that none of the companies (0%) accepted they were unaware of cloud computing services.

This finding illustrates the fact that Nigerians enterprises were not unaware of cloud computing services, hence answering the first research question that Nigerian enterprises were aware of cloud computing services as represented by a high percentage.

Table 2: How often Nigerian Enterprises adopt Cloud Computing Services.
Source; field data 2015

Benefits	Frequency	Percentages (%)
Saving of their over-head costs	14	70
Conservation of their computing resources	18	90
Bridging the distance gap	16	80
Making their operations easier	14	70
Safety of their data in case of system crash, loss or hack	18	90
On –demand self service	20	100
Efficiency in service delivery to their customers	20	100
Flexibility	12	60

Results from table 2 above shows that 15% of the sampled population very often utilize cloud computing services, 25% often use cloud computing services while a higher percentage(60%) of the respondents seldom make use of these cloud computing services. Worthy of note is that all sampled companies make use of cloud computing services in one way or the other as none of the companies indicated they never (0%) use cloud computing services. On further oral interview the researcher discovered that the enterprises were yet to make use of cloud infrastructures, platforms and servers rather they only mode of use of cloud computing were in backing up their data in cloud. This answers the second research question by showing that even though enterprises were aware of cloud computing services, they seldom used them in their operations.

Table 3: Benefits enterprises intend to enjoy should they adopt fully cloud computing services.
Source; field data 2015

Responses	Frequency	Percentages (%)
Very often	3	15
Often	5	25
Seldom	12	60
Never	-	-
Total	20	100

Data as presented on table 3 above answers the third research question by displaying that amongst so many benefits enterprises intend to enjoy should they move their IT services to cloud include, on-demand self-service (100%), efficiency in service delivery to their customers (100%), safety of their data in event of system crash, loss or hack (90%), and conservation of their computing resources (90%) were top ranking. The researcher went on to interview the enterprises representatives on why they still seldom integrated cloud computing services in their IT operations despite their knowledge of these numerous benefits they would benefit if they adopt cloud computing services. The enterprises raised some issues and challenges to this effect and this is presented in table 4 below.

Table 4: Challenges facing adoption of cloud computing services by enterprises in Nigeria

Challenges	Frequency	Percentages (%)
Cloud data security issues	20	100
Breech in privacy of users' data	20	100
Lack of information from providers to users	12	60
Lack of integrity by staff of cloud –provider firms	16	80
Cost implications	18	90
Vulnerability to competitors	18	90
Trans –border flow	10	50

Source; field data 2015

Enterprises elucidated some challenges that posed limitation factors to their effective adoption of cloud computing services. These challenges included the following, cloud data security issues (100%), breech in privacy of

users' data (100%), cost implications (90%), and vulnerability to competitors (90%), among other challenges. This answers the fourth research question and elucidates the challenges facing adoption of cloud computing services by Nigerian enterprises.

Table 5: Factors that would improve cloud computing services adoption by enterprises in Nigeria

Factors	Frequency	Percentages (%)
Ensure cloud data security	20	100
Reduction in privacy issues	20	100
Reduced cost of leasing cloud services	16	80
Increase in the integrity of cloud providers	20	100
Bridging communication gap between providers and users.	12	60
Promotion of cloud usage education by providers to users.	14	70
Reducing lock-down.	10	50

Source; field data 2015

The study not only attempted at finding out the challenges facing adoption of cloud computing by enterprises in Nigeria but also went further to prompt enterprises to proffer factors that would improve adoption of cloud computing services by enterprises in Nigeria. The enterprises stated that based on challenges facing cloud computing services adoption by enterprises in Nigeria, the following factors would help improve cloud adoption; Ensured cloud data security (100%), reduction in privacy issues (100%), increase in the integrity of cloud providers (100%), and reduced cost of leasing cloud services (80%), amongst others. This finding answers the fifth research question.

3.2 SUMMARY OF FINDINGS

The findings from the research show that;

1. Nigerian enterprises were aware of cloud computing services, as the sampled companies stated varying degrees and percentages of awareness of cloud computing services.
2. Enterprises in Nigeria seldom used cloud computing services and the major service they used was storage services for backup of their data.
3. There were several benefits enterprises intend to enjoy if they were to adopt cloud computing services and this included; conservation of computing resources, efficiency in service delivery to their customers amongst other benefits
4. Enterprises identified key challenges to adoption of cloud computing services and these included cloud data security, privacy breech of users' data, cost implications amongst others.
5. Ensuring cloud data security, decrease in privacy issues, and reduction in the cost of leasing cloud computing services among others were some factors enumerated that would improve adoption of cloud computing services by enterprises in Nigeria.

3.3 RECOMMENDATIONS.

The study recommends that;

1. Enlightenment campaigns should be intensified to improve awareness of enterprises on availability and benefits of cloud computing services.
2. Enterprises in Nigeria should move their IT services to cloud to save overhead cost, and increase efficiency and reliability.
3. Cloud providers should assure customers of their commitment to ensure cloud data security, reduce privacy issues and uphold high integrity of their staff in handling users' data.

4.0 CONCLUSION.

The research shows that cloud computing is an aspect of information technology that affects our lives as individuals and corporations on a day to day basis. Cloud computing provides a wide range of advantages to its users and enterprises that deploy them as well. Even though a number of pitfalls have been pointed

out as reasons why users and enterprises are slow in deploying cloud computing especially in Nigeria, the benefits are numerous and hence identifying and implementing lasting solutions to these issues and challenges will ensure that enterprises' confidence in cloud computing are boosted hence leading to overall promotion of the adoption of cloud computing by enterprises in Nigeria.

5.0 REFERENCES

Afshari, M., Andersen, S., Thayer, R., Welder, L., Malizia, M. and Van Eijk, P.H. (2014). *Cloud computing adoption in developing countries*. Available at: http://cloudtweaks.com/2014/06/cloud-computing-adoption-developing-countries/ (Accessed: 20 May 2016).

Arnold.S, (2009). *Cloud computing and issue of privacy*. Retrieved from http:// www.kmworld.com.

Brodkin. J. (2008). *Gartner: Seven Cloud–Computing Security Risks*. Available atwww.infoworld.com/d/security-central/gaterner-seven-cl-computing security- risk-853

Chigozie-Okwum.C.C. (2015). *Ensuring Cloud Data Security; a Panacea for Promotion of the Adoption of Cloud Computing Techniques by Enterprises*. Thesis submitted to the National Open University of Nigeria for the Award of Master of Science in Information Technology.

Cloud Security Alliance (2009). Security Guidance for Critical Areas of Focus in Cloud Computing. Vol.2.1. Available at www.cloudsecurityalliance.org/guidance/csaguide.v2.1pdf.

Cloud Security Alliance (2010). *Top Threats to Cloud Computing*. Vol.1.0. Available at http://cloudsecurityalliance.org/research/top.Threats.v1.0pdf.

ENISA Report (2009). *Cloud computing Benefits, Risks, and Recommendations for Information Security*. Available at www.enisa.europa.eu/act/rm/files/deliverables/cloudcomputingriskassesment.

Gens. F. (2009). *New to Cloud Services: Top Benefits and Challenges*. Available at http://blogs.idc.com/ie/?.

IDC. (2008). *IT Cloud Service Forecast*. Retrieved from http://blogs.idc.com/ie/?p=224.

ITU-Technology Watch Report,(2012). *Distributed computing utilities*. Retrieved from http://www.itu.int/dms. Extracted 12/12/15.

Kuyoro,S.O, Ibekwe,F, Awodele, D, (2011). *Cloud computing security issues and challenges*. International Journal of Computer Networks, Volume (3); Issue (50).

Liang, S. and Ulander, P. (2010) 7 *requirements for building your cloud infrastructure*. Available at: http://www.cio.com/article/2412506/cloud-computing/7-requirements-for-building-your-cloud- infrastructure.html (Accessed: 20 May 2016).

Microsoft Whitepaper (2010). *Cloud Computing Security Consideration, a Microsoft Perspective*. Available at http://www.microsoft.com.malaysia\ea\whitepaper.aspx.

National Institute of Standard and Technology (2011).*The NIST definition of cloud computing*. Available at:http//www.gni.com

Radack Shirley, (2012). *Guidelines for improving security and privacy in public cloud computing*.ITL bulletin, Volume 5, March 2012. Available at http//www.nist.gov/itl/cloud-def-v15. Pdf, retrieved on July 2011.

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

PROSPECTS OF CLOUD COMPUTING AS SAFE HAVEN FOR IMPROVING MATHEMATICS EDUCATION IN NIGERIA TERTIARY INSTITUTIONS

C. O. Iji
Department of Science Education
University of Agriculture, Makurdi,
Benue State
ijiclements07@yahoo.com

J. A Abah
Department of Science Education
University of Agriculture, Makurdi,
Benue State
abahjoshua@uam.edu.ng

ABSTRACT

Abstract

Historically, mathematics education has been bedeviled by the deployment of instructional strategies that seriously stunt the growth of students. Methodologies and approaches of instructional delivery in tertiary institutions have raised the need for technological augmentation for both students and mathematics educators. Cloud computing yield itself to this quest by strengthening individualized learning via unrestricted access to infrastructure, platforms, content, and powerful web-based tools from anywhere within the private cloud architecture adopted by tertiary institutions in Nigeria. This paper considers broad safety concerns of this technological intervention and the prospects of improving mathematics education in tertiary institutions in Nigeria.

KEYWORDS: MATHEMATICS EDUCATION, CLOUD COMPUTING, DATA SECURITY

1.0 INTRODUCTION

Mathematics education is an umbrella term that encompasses all aspects of learning and teaching mathematics in schools and in other settings. Mathematics itself is an aid to representing and attempting to resolve problem situations in all disciplines. According to Odili (2012) it is a powerful means of communication, an intellectual endeavour and a mode of thinking. Mathematics is a discipline through which students can develop their ability to appreciate the beauty of nature, think logically and make sound judgement. Mathematics education is considered as an intersection with the nature of mathematics as a discipline. With focus on teacher education, mathematics education considers the design, implementation and effects of curriculum and instructional interventions; and contemporary developments in learning theories and technologies.

Technology, broadly understood, has been transforming human life in one way or another for thousands of years. But in the computer age, the pace of technological change is very rapid, altering schooling, work and social lives in ways that have significant consequences for young people (Craig, 2009). In rethinking education to cope with these changes at the threshold of the twenty-first century, innovation and research are indispensable tools. Failure to innovate, by and large, means repeating yesterday's educational programmes and strategies tomorrow (Raja, 2002). The society which education is meant to sustain is becoming transformed by trend such as automation, globalization, workplace culture and personal responsibility.

Within the demands of the time, both the education system and the educational process must be amenable. This calls for a fundamental qualitative transformation of

education in terms of its content, methods and outcomes. Education should seek to inculcate skills that are aimed at accelerating technological change, rapidly accumulating knowledge, increasing global competition and rising workforce capabilities (Partnership for 21st Century Skills, 2002). Schools must equip students who will ultimately spend their adult lives in a multitasking, multifaceted, technology-driven, diverse and vibrant world. The reality on ground has made it imperative for the education system to be more strategic, aggressive and effective in preparing students to succeed and prosper. Educational institutions must rethink what, but even more important, how and where we learn (Innovation Unit, 2014).

Although it is clear that technology is not the solution to present day education (Lokesh, 2013), utilizing emerging technologies to provide expanded learning opportunities is critical to the success of future generations. The level of penetration of ICT among students signals more than a change in pedagogy; it suggests a change in the very meaning and nature of mathematics education itself (Italiano, 2014). Schools all over the world are becoming an integral part of the broadband and technological transformation, harnessing the potentials of technology to drive and empower more personalized mathematics learning.

One of the specific ways technology is enhancing present day mathematics teaching and learning is through the utilization of the cloud. The cloud is a set of hardware, networks, storage, services, and interfaces that enable the delivery of computing as a service (Hurwitz, Bloor, Kaufman, & Halper, 2010). Cloud services include the delivery of software, infrastructure and storage over the internet, reducing cost and providing flexibility and mobility (Kovachev, Cao, & Klamma, 2011). These services are delivered via the internet from high-specification data centres in locations remote from the end user.

Broadly, the cloud can be seen as an on-demand access to computer services, applications, network and data anywhere (Powell, 2009). In an educational institution, the cloud provides students with standard internet access that promotes the use of heterogeneous thin or thick client platforms (e.g. mobile phones, laptops, and tablets). Students make use of several self-services by connecting to wireless access points spread across their school. This has become a modern tool, a way of fact-based learning which allows students to do a lot of research using the web and various tools (Lokesh, 2014). In the process, students' critical and literacy skills are enhanced.

Active utilization of cloud services provided by educational institutions has grown in importance as a result of a new genre of students with learning needs vastly different from their predecessors (Thomas, 2011). Present day students require increase network access to sustain their culture of learning, leisure and social interaction. The computing power provided by the cloud avails the opportunity to extend students' mathematics learning beyond the walls of the classroom, thereby offering the learner greater participation and control of the learning process. Much flexibility, as provided by the availability of cloud services in institutions of higher learning, especially universities, is needed particularly in the teaching and learning of mathematics.

2.0 LITERATURE REVIEW

The review of related literature is done according to the following sub-headings, namely, the educational cloud, essential characteristics, cloud service models, cloud deployment models, the wireless network technologies in tertiary institutions, and cloud architecture for tertiary institutions.

The Educational Cloud

According to the National Institute of Standards and Technology (NIST) (as cited in GTSI, 2009), cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be provided with minimal service provider effort. Typically, cloud services run in a web browser requiring the user to have only basic components while enjoying high speed, bandwidth and computing power. This

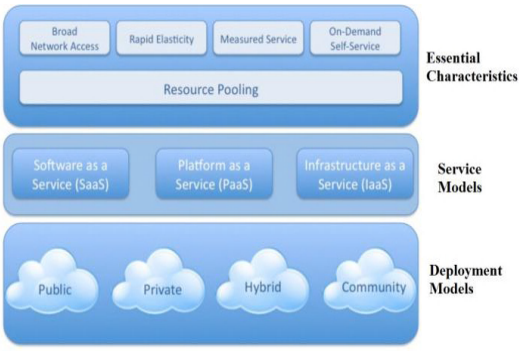
simplicity is why the emergence of cloud services is fundamentally shifting the economics of IT-based businesses (Harms & Yamartino, 2010).

Education has not remained unaware of this trend in migration to the cloud (Niharika, Lavanya, Murthy & Satya Sai Kumar, 2012). Presently, virtualized resources are being provided to educational institution over the internet, without users having knowledge of, expertise in, or control over the technology infrastructure. More students are enriching their educational experience daily through network access provided by private cloud services.

To grasp the extent of utilization of cloud services in education, it is important to briefly study the acceptable model of cloud computing.

Figure 1: NIST Visual Model of Cloud Computing Definition (*Source: Niharika et al, 2012)*

Essential Characteristics



Most commentators on cloud computing agree on five key characteristics

i. Broad Network Access: Services are provided over the network and accessed through standard mechanism.
ii. Rapid Elasticity: The cloud gives the user the impression that the services are infinitely scalable. The service needs to be available all the time and it has to be designed to scale upward for high periods and downward for lighter ones (Hurwitz et al, 2010).
iii. Measured Service: A cloud environment has built-in system that bills users. In educational institutions such as the public universities in Benue State, students are given a number of hours daily to access the cloud, logging in with their user accounts as created on the university portal after payment of tuition fees.
iv. On-Demand Self-Service: The cloud allows the user to request an amount of computing facility needed automatically, without requiring direct human interaction with a service provider.
v. Resource Pooling: Computing services such as storage, processing, network, bandwidth, and virtual machines are dynamically assigned and reassigned according to the user's demand.

Cloud Service Models

Cloud service delivery is divided into three models. They are infrastructure as a service, platform as a service, and software as a service. Infrastructure as a Service (IaaS) is the delivery of computer hardware for customized needs of the user. Such computer hardware include resources like servers, networking technology, storage, and data centre space. Educational institutions benefit maximally from networking technology (wireless network) and access to servers.

Platform as a Service (PaaS) is the capability provided to the user to deploy onto the cloud user-created or acquired applications created using programming languages and tools supported by the provider.

Software as a Service (SaaS) implies the provision of applications which are accessible to users from various client devices through a thin interface such as a web browser. The private cloud services of the tertiary educational institutions offer web-based email, virtual library, among other application services.

Cloud Deployment Models

The NIST recognizes four deployment models for cloud services.

i. Public Cloud: The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services. Popular examples are the Amazon Cloud and Google Cloud.

ii. Private Cloud: This cloud infrastructure is operated solely for a single organization. It is managed by the organization or a third party, and may exist on-premises or off-premises. The ICT Directorate of the Universities power the private cloud service on-premises, and serves the entire university community.
iii. Community Cloud: This cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns.
iv. Hybrid Cloud: This cloud infrastructure is a composition of two or more clouds that remains unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g. Cloud bursting for load-balancing between clouds).

The Wireless Network Technologies in Tertiary Institutions

Considering cost and technical factors, higher educational institutions are seeking private cloud services to provide a common interface, common identity infrastructure, and common service attributes (Katz, Goldstein, & Yanosky, 2009). Universities are turning into network hubs, as mobile devices are carried by students and staff, and these devices are communicating with the world around them (Steijaert, Boyle, Leinen, Melve, & Mitsos, 2012). Educational institutions are responding to the availability of cloud services by enforcing the use of a limited set of services such as official e-mail, internal institutional portal, and e-library services. Users are increasingly connecting to the wireless network on campus.

One of the primary requirements for benefiting from the wide range of services rendered by the university cloud is access to the school's wireless network. The term "wireless network" refers to two or more computers communicating using standard network rules or protocols, but without the use of cabling to connect the computers together (Bakardjieva, 2014).

For a telecommunications network to work, connectivity needs to be ensured at different levels by the elements present (Neto, 2004). The wireless network is basically a system of radio technologies deployed in the 2.4GHz and 5GHz bands.

Best (2003) presented a hypothetical network installation as depicted in Figure 2.

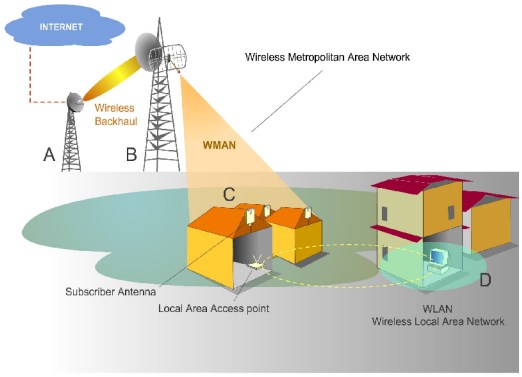


Figure 2: Connectivity in Wireless Networks (*Source: Best, 2003)*

This schematic diagram shows two radio towers (A and B), houses and other buildings (C), and a personal computer inside a building (D). Radio tower A is connected through a wired link to an Internet Point of Presence owned by an Internet Service Provider (ISP). The ISP radio tower could belong to any of the telecommunications companies with presence in the university (e.g. Airtel, Glo, Etisalat, or MTN). So, the PC shown at point D is ultimately connected to the Internet by several wireless links.

To start with, a point-to-point connection is used between radio towers A and B, with only one antenna (i.e. one receiver/transmitter) in both extremities. According to Neto (2004), the purpose of this connection is typically to transmit over long distances (in the order of tens of kilometres). Several of these links can be used, one after the other; in this way the signal will be transmitted, in "hops", to a potentially remote location. This is normally referred to as wireless backhaul.

The connection from B to C is a point-to-multipoint connection. This means that radio tower B is now radiating to and receiving from several stations of type C – i.e. several buildings with base stations, or access points. This is normally called a Wireless Metropolitan Area Network (WMAN) (Neto, 2004).

Finally, there is a radio connection between the subscriber equipment mounted on the side of the building (point C) and the individual

personal computer inside the building (point D). This is what is normally called a Wireless Local Area Network (WLAN). An outdoor repeater may be required to redistribute the signals from the access point in a situation where there are blockages in the direct line of site (LOS) between the base station (access point) and the personal computer (PC).

For a PC to access the wireless network, it must possess a network interface card (NIC) or a network adapter card. Most modern laptops and mobile devices come with in-built network cards. Such Wi-Fi enabled systems can track signals from base stations available at the offices of the Deans and Heads of the various colleges/faculties and departments of the institution. The usual transmitting proximity ranges from 100 metres indoors to 350 metres outdoors (Bakardjieva, 2014).

3.0 CLOUD ARCHITECTURE FOR TERTIARY INSTITUTIONS

According to Mircea and Andreescu (2011), the architectural pattern of using cloud computing in tertiary educational institutions may be described starting from the development and supply of services and resources offered to the schools' communities. This may be illustrated in figure 3.

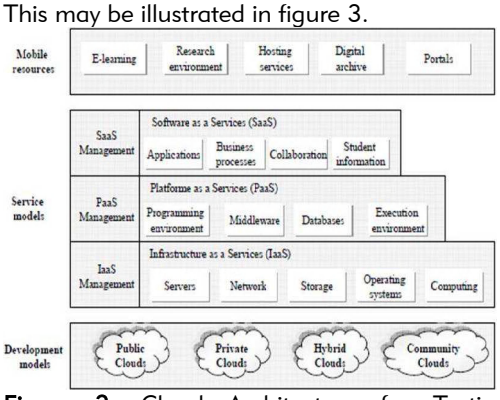


Figure 3: Cloud Architecture for Tertiary Institution (*Source: Mircea & Andreescu, 2011)*

As shown in figure 3, students and staff benefit from mobile resources as e-learning, expanded research environment, e-mail hosting services (for instance @uam.edu.ng), digital archive and student portal services. The services models indicate areas of direct impact of the three service models of cloud computing. The widest area of impact is the availability of wireless Internet network as a service.

Benefits of the Cloud

Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and services) which can be dynamically re-configured to adjust to a variable load (scale), allowing also for optimum resource utilization (Vaquero, Merino, Caceres, & Lindner, 2009). This pool of resources is typically exploited by a pay-per-use model which the infrastructure provider offers through customized Service Level Agreement. At the tertiary education level, the infrastructure provider is the ICT Directorate of the institutions.

The goal of utilizing the cloud as a tool is the achievement of virtual communities of educators, researchers and practitioners working in collaborative groups to advance their practices (Thomas, 2011). Students stand to gain from online services from anywhere within the hotspot and anytime access to powerful web-based tools. It lets both the teachers and the students to access, share and publish documents, class calendars or web pages (Miseviciene, Budnikas, & Ambraziene, 2011).

There is also 24 hours access to infrastructure and content. Resource sharing, network speed and flexible access are all accompanying benefits of the cloud to students. The cloud services of the universities allow offline usage with further synchronization opportunities. Uncountable research materials can be sourced from the National Virtual Library and other linked library services. This functionality of the cloud is gradually changing the way students do research in the school.

4.0 A SAFE HAVEN FOR MATHEMATICS EDUCATION

Mathematicians work on the important practical issues of their era, which have always required both the development of the subject together with the tools to support this. Pascal's and Leibniz's mechanical calculating machines, Napier's logarithms, Babbage's difference engine, Newman's Colossus and Turing's Bombe for crypto-analysis at Blethcley Park are just few a examples of computational tools which have been fundamental to the evolution of digital technologies to support mathematical developments. In this respect

mathematics is, and has always been, a dynamic problem solving activity for which humans have continued to develop and exploit new tools (Clark-Wilson, Oldknow, & Sutherland, 2011).

The level of penetration of cloud computing in Nigerian tertiary educational institutions can be best explained by the Unified Theory of Acceptance and Use of Technology (UTAUT). The UTAUT as propounded by Venkatesh, Morris, Davis and Davies (2003) provides a useful tool for educational administrators needing to assess the likelihood of success of new technology interventions.

Venkatesh et al (2003) theorize that four constructs will play a significant role as direct determinants of user acceptance and usage behaviour: performance expectancy, effort expectancy, social influence, and facilitating conditions. These determinants are in turn influenced by key moderators such as gender, age, voluntariness, and experience. Thus, the degree to which cloud computing helps both teachers and students of mathematics education attain gains in performance, the ease associated with the cloud technology, the availability of organizational and technical support infrastructure, and the social benefits that are derivable, all contributed to the appeal of the cloud to tertiary educational institution in Nigeria.

Despite this appeal, a primary concern that cloud computing adopters have is the security of enterprise information (Tout, Sverdlik, & Lawver, 2009). Large amount of student, teacher, and institution data are placed in the hands of third party service providers, particularly in the public cloud model. Where such applies, it is critical that teachers, school administrators, and educational regulatory bodies take steps to ensure that the cloud services that are used in the tertiary institutions comply with all applicable laws and otherwise protect data from improper use or transfer (Mutkoski, 2014). A seamless integration of cloud security controls with campus-wide departments and their various applications will be necessary for maintaining high level of information assurance of such applications, including their confidentiality, integrity, and availability.

A wide range of educational institutions in Nigeria are presently operating the private cloud model hosted by the ICT directorates of the schools (Iji, Abah, & Anyor, 2014). The choice of the model over popular commercial cloud enterprises reduces the risk of security breaches at data centres. In a campus-hosted cloud ecosystem, Trusted Virtual Data (TVD) technologies are often adopted to address the need for strong isolation and integrity guarantees (Berger et al, 2009). The TVD implements controlled access to networked storage based on security labels and enforces isolation constraints and integrity checking. With such efforts in place, mathematics educators and students in tertiary institutions in Nigeria can focus on leveraging on the power of cloud computing to achieve the goals of their respective programmes of study.

The methodology and approach of instructional delivery at the tertiary educational level call for technological augmentation on the part of both students and instructors. The utilization of cloud technology offer students the opportunity to appropriately use ubiquitous wireless network to access, manage, integrate and evaluate information, construct new knowledge, and communicate with others in order to participate effectively in the society (Partnership for 21st Century Skills, 2002). With the aid of available cloud services, students of mathematics education can drive more personalized learning, forming a positive image of their discipline (Artigue, 2012).

The cloud by its very nature allows mathematics education students endless opportunities to engage in advanced researches and even in online entrepreneurial outfits. Uncountable research materials can be sourced from synchronized virtual libraries and other linked instructional content repositories. Such readily available support to conventional classroom instruction stands to improve the students' mathematical reasoning and problem solving skills. It opens up more frontiers for the expansion of knowledge and enriched learning experience.

CONCLUSION

The cloud has come to stay in tertiary education in Nigeria as a tool for hooking students to the information grid. Though there are safety concerns in cloud adoption, campuses across the country are already

leveraging on the dividends of mobile and wireless technology, considering the ubiquity of smartphones and other computer devices among present-day student population. Mathematics education is being improved by transforming opportunities offered by cloud computing into active learning strategies that put students in charge of their own learning.

REFERENCES

Artigue, M. (2012). *The Challenges of Change in Teaching Practice*. UNESCO: Paris

Bakardjieva, T. (2014). *Introduction to computer networking*. Varna: Institute of Technology- Varna Free University. 1-22

Berger, S., Caceres, R., Goldman, K., Pendarakis, D., Perez, R., Rao, J. R., … (2009). Security for the cloud infrastructure: Trusted virtual data center implementation. *IBM Journal of Research & Development*, 53(4), 6:2-6:12

Best, M. L. (2003). The wireless revolution and universal access. *Trends in Telecommunications Reform, 1-24*.

Clark-Wilson, A., Oldknow, A., & Sutherland, R. (2011). *Digital technologies and mathematics education*. A report from the working group of the Joint Mathematical Council of the United Kingdom, pp 1-32.

Craig, D. J. (2009). *Defining a 21st century education*. Alexandria, VA: The Centre for Public Education, 1-79.

GTSI (2009). *Cloud computing: Building a framework for successful transition*. Herndon, VA: GTSI Corp.

Harms, R. & Yamartino, M. (2010). *The economics of the cloud*. Microsoft, 1-21.

Hurwitz, J., Bloor, R., Kaufman, M. & Halper, F. (2010). *Cloud computing for dummies*. Hoboken, NJ: Wiley Publishung Inc.

Iji, C. O., Abah, J. A., and Anyor, J. W. (2014). Enhancing national competitiveness through the utilization of cloud services to improve the learning of school mathematics. Proceedings of September 2014 Annual National Conference of the Mathematical Association of Nigeria (MAN), 230-243

Innovation Unit (2014). *21st century learning*. Retrieved on 18th March, 2014 from http://www.innovationunit.org/knowledge/our-ideas/21st-century-learning

Italiano, E. (2014). *Community, contemplation, and computers: The role of technology in education*. Retrieved on 18th March, 2014 from http://www.thepublicdiscourse.com/2014/02/11789/

Katz, R., Goldstein, P., & Yanosky, R. (2009). *Cloud computing in higher education*. EDUCAUSE. Retrieved on 5th April, 2013 from http://net.educause.edu/section_params/conf/CCW10/highered.pdf

Kovachev, D., Cao, Y. & Kamma, R. (2011). *Mobile cloud computing: A comparison of application models*. Aachen: RWTH Aachen University, 1-8.

Lokesh, U. (2013). *Technology and its role in 21st century education*. Retrieved on 18th March, 2014 from http://www.edtechreview.in/trends-insights/insights/277-role-of-technology-in-21st-century

Mircea, M. & Andreescu, A. I. (2011). Using cloud computing in higher education: A strategy to improve agility in the current financial crisis. *Communications of the IBIMA, 2011*. Retrieved on 25th February, 2014 from http://www.ibimapublishing.com/journals/CIBIMA/cbima.html DOI:10.5171/2011.875547

Miseviciene, R., Budnikas, G., & Ambraziene, D. (2011). Application of cloud computing at KTU: MS Live @Edu Case. *Informatics in Education, 10(2), 259-270*.

Mutkoski, S. (2014). Cloud computing, regulatory compliance, and student privacy: A guide for school administrators

and legal counsel. *The John Marshall Journal of Information Technology & Privacy Law*, 30(3), 510-534

Neto, M. I. A. S. (2014). *Wireless networks for the developing world: The regulation and use of license-exempt radio bands in Africa*. An M.Sc Thesis submitted to the Engineering Systems Division at Massachusetts Institute of Technology. 1-226

Niharika, K., Lavanya, G., Murthy, G. V., & Satya Sai Kumar, A. (2012). Educational cloud: Utilization of Iaas versus Paas Services. *International Journal of Scientific & Engineering Research, 3(1)*. ISSN 2229-5518. Retrieved on 29th March, 2014 from http://www.ijser.org

Odili, G. O. (2012). Towards a new paradigm of teaching mathematics in Nigerian universities: The role of mathematics educators. *Online Journal of Science Teachers Association of Nigerian (STAN), 47(1)*.

Partnership for 21st Century Skills (2002). *Learning for the 21st century: A report and mile guide for 21st century skills*. Washington: Partnership for 21st Century Skills, 1-5.

Powell, J. (2009). *Cloud Computing- What is it and What does it Mean for Education?* Retrieved on 10th September, 2012 from http://erevolution.jiscinvolve.org/wp/files/2009/07/clouds-johnpowell.pdf

Raja, R. S. (2002). *Education for the twenty-first century: Asia-Pacific Perspective*. Bangkok: UNESCO, 1-111.

Steijaert, A., Boyle, B., Leinen, S., Melve, I., & Mitsos, Y. (2012). The Adoption of Cloud Services. ASPIRE. 1-37.

Thomas, P. Y. (2011). Cloud computing: A potential paradigm for practicing the scholarship of teaching and learning. *Electronic Library, 29(2), 214-224*.

Tout, S., Sverdlik, W., & Lawver, G. (2009). Cloud computing and its security in higher education. *Proc ISECON*, 26, 1-5.

Vaquero, L., Merino, L., Caceres, J. & Lindner, M. (2009). A Break in the clouds: Towards a cloud definition. *SIGCOMM Computing Community Revolution, 39(1), 50-55*.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478

SESSION D:

Approaches for Enhancing National Security

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

A REVIEW OF CONTEXT- AWARE SURVEILLANCE AND DETECTIVE SYSTEM

A.P. Adegbiji
Automation Unit, Main Library, University of Lagos
adebayo@unilag.edu.ng

N.A. Azeez
Department of Computer Sciences, University of Lagos
nazeez@unilag.edu.ng

ABSTRACT

Abstract

This papers provides a thorough review of context-aware surveillance and detective system and examined its aplication to the Nigeria System considering recent Government efforts in harmonizing various national databases such as National Identification Number (NIN), Bank Verification Number (BVN) and other data collected from the Federal Road Safety Commission, Joint Admissions and Matriculations Board, The Nigeria Immigration Service etc.

Constant improvement and innovations in mobile computing has made ubiquitous computing more real to us than it has ever been and the evolution of Cloud Computing and the Internet of Things has further complimented the reality that several activities can take place on our mobile devices anywhere anytime.

The effective relationship of required databases and the application of context-awareness tools and technologies for effective surveillance and detective systems are explained in this paper.

KEYWORDS: Context-Awareness, Human-Computer Interaction, Mobile Computing, Pattern Recognition, Ubiquitous Surveillance

1. INTRODUCTION

The constant request to make computing services available anywhere anytime has developed study areas such as ubiquitous computing, Human Computer Interaction and Mobile computing.

According to (Dey Anind, 2000), (Yılmaz & Erdur, 2012); "***Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves***".

The word ***context*** in a generic form refers ***to the environment where an object is, how it behaves and how the behaviors of other objects affect it.*** (Johnson, Phillips, & Stein, 2002)

According to (Schilit, Adams, & Want, 1994) who are the early writers on context-awareness, context is referred to as location, identities of nearby people and objects, and changes to those objects.

Surveillance and detective systems are major areas of applications of context-awareness because for every context- aware project the type of context required needs to be explicitly stated; in this case,

Surveillance means to keep a close watch on something or someone (Wright, et al., 2010) (Subodhani, Khalil, & Atiquzzaman, 2015). Surveillance helps in monitoring people, processes and objects without being physically present to observe or take the data (Subodhani, Khalil, & Atiquzzaman, 2015). By putting adequate surveillance measures in place, defined surveillance data becomes readily available for use to detect operations pattern and activities, prevent crime and safeguard people and infrastructure. In this paper, context-awareness is explicitly discussed, it's application to security, and surveillance was demonstrated.

1.1 Motivation

We are motivated to embark on this study and other related studies so as to make recommendations and provide solutions to the

alarming insecurity in our indigenous environment and the global village through the applications of tested and demonstrable computing methods and discoveries.

2. *Mobile Computing and Context-Awareness*

Mobile Computing has evolved as a means of resolving issues relating to size, continuous availability of quality service irrespective of location and time, availability of various basic applications like calendar, camera, browsers, games and others, the need for power all the time, reduced booting time, flexibility and a host of others (Stojmenovic, 2002).

Mobile Computing is a technology that allows transmission of data, via a computer or any other computer-related devices, without having to be connected to a fixed physical link.

Mobile computing technology enables the mobile worker to create, access, process, store and communicate information without being constrained to a single location. By extending the reach of an organization's fixed information system, mobile computing enables interaction with organizational personnel that were previously disconnected.

Mobile computing is the discipline for creating an information management platform, which is free from spatial and temporal constraints. The freedom from these constraints allows its users to access and process desired information from anywhere in the space. (Deepak & Pradeep, 2012)

Mobile computing devices combined the basic functions of always-available communication facilities and that of a computer for performing basic computing activities.

Mobile Computing is synonymous with, computing–on-the-move, computing anywhere anytime, roaming computing etc.

Mobile Computing and wireless communication is being applied in various ways not limited to e-commerce, activities monitoring and control, personal communications, telecommunications, monitoring remote or dangerous environments, national defense, security, emergency and disaster operations, remote operations of appliances, wireless Internet access etc.

Mobile computing facilities as embedded facilities in cars, medical equipment, aircrafts etc. has made mobile so many traditional stationary operations (Stojmenovic, 2002).

Mobile computing has undoubtedly become a very important new paradigm in today's world of networked computing systems. Ranging from wireless laptops to mobile phones of different sizes and types, Wi-Fi, Bluetooth- enabled Personal Digital Assistants (PDA's) to wireless sensor networks. Mobile computing has become ubiquitous in its impact on our daily lives. (Deepak & Pradeep, 2012)

2.1 Context-Awareness functionality in mobile devices and Applications

Context awareness simply can be said to be adaptation of the behavior of an application as a function of its current environment, which can be characterized as a physical location, an orientation, or a user profile.

Context can be broadly categorized into four;

a. Computing Context: This has to do with network connectivity, network bandwidth, communication cost, nearby resource etc.
b. Time Context: This has to do with time of the day, week, month or season of the year.
c. User Context: This comprises of user location, user profile, current user situation etc.
d. Physical Context: This has to do with temperature, humidity, darkness or lighting state, noise etc. (Yılmaz & Erdur, 2012)

A context-aware application can sense the environment and interpret the events that occur within it.

Context-aware components can sense who you are, where you are, and what you are doing and use that information to adapt their services to your needs. Mobility and services on demand are greatly impacted by the location of the devices and the requested services. Examples range from relatively rudimentary device following services such as phone call forwarding to the location of the device, to more complex issues of detecting locations of available services and selecting the optimal location for obtaining the services, such as printing services. (Stojmenovic, 2002), (Brown P. , 1996), (Fortier, Rossi, Gordillo, & Challiol, 2010)

Mobile applications are now contextualizing proximity, location, weather, time, etc. to deliver hyper- specialized, dynamic, rich content to users through context-aware applications. Previously, web applications would often provide contextualized content based on time, detected location and language.

The context-aware system development has the following layers of hierarchical architecture, which reflects the complex functionality of Context-aware systems.

a. Sensor Layer → This is the lowest level of the location management architecture and it represents the variety of physical and logical location sensor agents producing sensor specific location information.
b. Context Reasoning Layer → This layer of the hierarchy receives the sensor-specific location information and other contextual information related data to the user and transforms it into standard format.
c. Web Services/Application Layer → This layer interacts with the variety of users of the context aware system and therefore needs to address several issues including access rights to location information (who can access the information and to what degree of accuracy), privacy of location information (how the location information can be used) and security of interactions between users and the context-aware system.

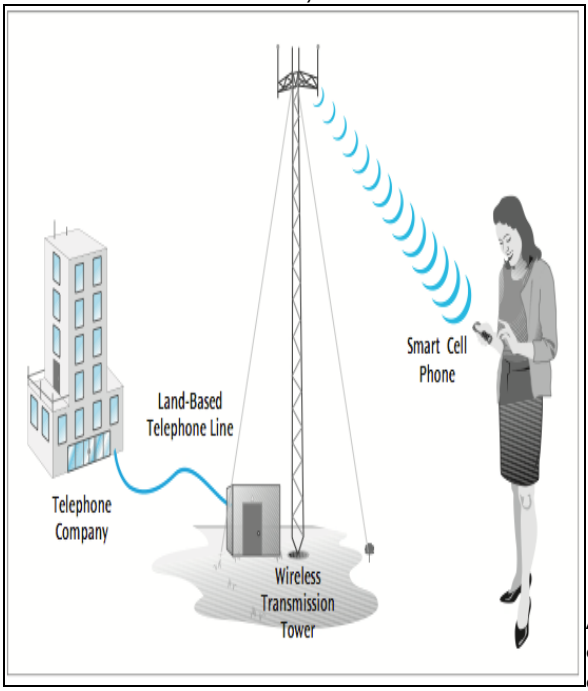


Figure 1: A mobile phone collecting context-data and transmitting it through a wireless transmission tower. Source: (White, 2013).

Our mobile devices cannot only found other devices, services and activities; it can also be found. Its various activities per time at different locations can also be monitored, reported and controlled by other remote devices and applications.

Some of the value – added services that can be provided using the Context awareness facilities embedded in mobile devices are but not limited to:

1. Network Signal strength tracking
2. Location based service i.e. (supermarkets, clubs, churches, universities, vulcanizers, gas stations, medical centers available within a named range or location) (Corral, Janes, & Remencius, 2012)
3. Troop monitoring
4. Vehicle tracking and Fleet management
5. Disaster/hazards detection (Chang, Kang, Ahn, Jang, & Choi, 2012)
6. Activities monitoring
7. Remote Locations control
8. Weather prediction
9. Financial activities reporting
10. Exchange rate monitoring
11. Road maps and directions (Sirichaia, Kaviyab, Fujiic, & Yupapind, 2010)
12. Facilities monitoring and reporting
13. Ships movement tracking and reporting (Katina & Roger, 2013)
14. Pipeline layout monitoring, tracking and security
15. Intelligence Surveillance (Blasch & Aved, 2015)
16. Point-of-sale applications
17. Mobile medicine (Al-Bashayreh, Hashim, & Khorma, 2013)
18. Quick Accident reporting and response
19. Product distribution Monitoring
20. Insurgence/terrorism activities monitoring
21. Emergency services management
22. Election Management (Voting locations surveillance, activities and materials movement monitoring, results collation etc.) (Blasch & Aved, 2015)

According to (Xue, Pung, & Sen, 2013) In context-aware computing, applications automatically adapt their operations or behaviors based on the real- time context data acquired from context sources installed in diverse operating spaces.

The ultimate goal of context-aware computing is to provide information and assistance for the applications to make appropriate decisions in the Right manner, at the Right time and in the Right place (3R). A context source can be a hardware sensor or a software process (e.g. web service or legacy database) that generates context data, or a virtual source of a collection of such sensors and processes whose context data is acquired through a well-defined access point.

An operating space in this view is referred to as any entity (object or area) in the real or virtual world having multiple context sources to provide context data for context- aware applications. Example of operating spaces can be persons, shops, homes and offices.

A context-aware application may run on the mobile device worn or carried by a person, such as a smart phone, a PDA or a laptop. With mobility, the application should be able to acquire context data from different operating spaces forming its dynamic operating environment at different run times, rather than restricting the data access to a few pre- defined spaces at the build time.

Programmers and Software developers have ample opportunities in developing indigenous mobile applications that will use the context-awareness facilities to provide localized solutions to their immediate environment.

Researchers alike could also explore the application of context-aware services in deciding the best model to expose opportunities and threats within their local domain.

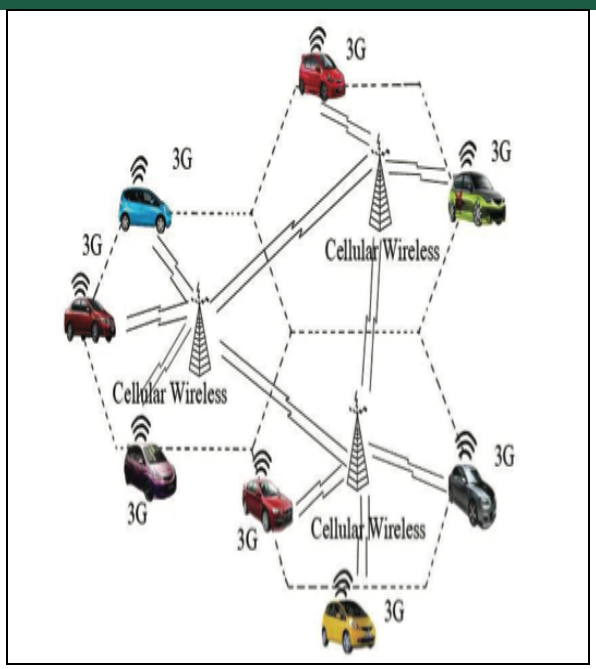


Figure 2: Context-aware sensors system built into cars for receiving and transmitting context-aware data. Source: (Sirichaia, Kaviyab, Fujiic, & Yupapind, 2010)

3. *Issues in developing context-aware Applications*

According to (Garg, Lather, & Dhurandher, 2012), every location – conscious system should be able to provide these three (3) basic functionalities:

i. **Sensing** – The ability to find and present information and services to users (devices and applications). Sensors can be implemented as hardware of software sensors, hardware sensors can sense, record and communicate issues relating to temperature, humidity, location, lightning etc. while software sensors can discover user's personal information, the status of a service and so on.

ii. **Reasoning** – The ability to tag context information in readiness for later retrieval by users

iii. **Acting** – The ability to execute a service to the user

Protocols and users' preferences should guide and govern the use and invocation of the context services.

Context computing aims at designing applications that automatically adapt their behaviors to the

available location information and the available nearby sensors and devices.

The ability of the sensor to collect various context related information accurately and made the same available to different applications at varying time without distortion in forms and content is a major issue in ensuring accuracy of context information. (Al-Bashayreh, Hashim, & Khorma, 2013)

Accuracy becomes an important issue to programmers and users alike because decisions are made using such data received from non-accurate devices like sensors.

An agent-based service where middle-agent took cognizance of the user's locations using location-ontology and determines best-matched services per time was used to coordinate various information services by heterogeneous devices.

Also, in tackling the issue of accuracy, an important design decision approach was to decouple those independently changing blocks of the system to support scalability, which is a way of anticipating future changes.

Also, sensing devices should be abstracted, so that the application logic does not have to get involved with the burden of connecting to hardware devices and sensing information, which is also a way of thinking of variability at the design and implementation level of the application (Fortier, Rossi, Gordillo, & Challiol, 2010).

Another germane issue in context-awareness software development is security and privacy of information during acquisition, generation and dissemination.

The issue of privacy in computing and software development in general is a strong issue of interest because an unsecured system poses more danger to automation and solution development.

The upward looking data of mobile users and the sensitivity of data accessed and transmitted by these arrays of users makes the issue if security a much more important area of concern in Context – aware computing.

Due to large amount of context information and devices in dynamic and heterogeneous environments, creating a total solution for managing security and privacy policies is known to be a very challenging open issue in ubiquitous environments (Yılmaz & Erdur, 2012)

The traditional approaches to security take into cognizance the static nature of users computing environment, and therefore base security-related decisions on such static attributes like identity or role.

Todays dynamic computing environment of mobile users may: 1) use a variety of mobile computing devices with varying configurations; 2) connect over various networks; and 3) be in varying physical settings when requesting access to remote resources. To achieve effective security in this dynamic computing environment, security decisions must consider the user's context (e.g., co-location, network characteristics, and device characteristics etc.), which can change frequently and rapidly. (Johnson, Shakariana, Gupta, & Agrawala, 2011)

4. Understanding the relationship between Context-awareness, Internet of Things and Cloud Computing

One thing is to be well grounded on context, context-awareness in computing devices and another is to know how to write, correct and evaluate context-aware applications.

This section looks deeply into:

i. The Nitty-gritty of developing context-aware applications
ii. Context –aware computing & the Internet of Things (IoT)
iii. Context- aware computing & the Cloud
iv. The combination of practical understanding of i→iii above in delivering context-aware service

4.1 Understanding Context – aware Applications

The purpose/type of a context-aware application must be clearly stated whether the application is meant to:

i. Detect activities or/and location at a particular given time?
ii. Store data from detected activities or/and location?
iii. Interpret and/or make decision with data from detected activities or/and location?
iv. Trigger a reaction/application or/and service as a result of the detected activities and/or location?
v. Combination of any of points i, ii, iii and iv above?

Another important point to note after establishing the function(s) of a context-aware application is to determine how the context data is gathered and how the logical instructions are executed which could be manually, semi-manually or automatically. (Dey Anind, 2000). Whatever

context an application programmer wants to capture must be clearly stated and must have necessary sensors support to capture the context adequately.

According to (Dey Anind, 2000), (Brown P. , 1996) and (Sánchez-Pi, Carbó, & Molina, 2012) there are several context-aware applications using selected type of context to provide solutions to various category of users i.e. medical doctors, tourist, conference participants, students, marketers etc.

Considering the built in sensors and Global Positioning System in Mobile phones, it has become a ready platform for building and implementing context-aware applications.

(Brown, Bovey, & Chen, 2000), Classify context-aware applications into ***discrete*** and ***continuous*** ones. In a ***continuous application***, the information presented to users is rapidly changing according to movement of the device from one place to another; a good example of this is the ***here*** (www.here.com) map. The here map shows the movement of the mobile phone along its current route and location.

Discrete application separate different likely occurring situations and attach different behaviour to the likely occurring situation; when the predicted situation occurs then the logic behaviour built for it will be automatically triggered.

4.2 Context-aware computing and Internet of Things

Internet of things (IoT) is a platform where different computing approaches such as ubiquitous computing, mobile computing, pervasive computing, embedded devices, sensing technologies and communication technologies are merged together in order to produce a symbiotic digital experience (Borgia, 2014), (Gubbi, Buyya, Marusic, & Palaniswami, 2013).

The unprecedented number of devices, data and technologies being combined together in Internet of things (IoT) has made it the next big thing to happen to the technology world.

Figure 3: A Prototype of the emerging IoT Scene. **Source:** (Borgia, 2014)

The application of the concept of Internet of Things is such that will hugely affect every human in different areas of endeavours with the cloud at the Centre providing storage and data exchange facilities; it will also bring tangible benefits to individuals, the environment and the society with the creation of smart and intelligent services and products while ensuring the protection and privacy of information and the exchanged data.

As with other powerful and emerging technologies, for the actualization of the benefits to be derived from the Internet of Things, a whole lot of issues have to be resolved. Some of such issues include hardware, architecture, data processing, network management, communication, power, storage, security etc.

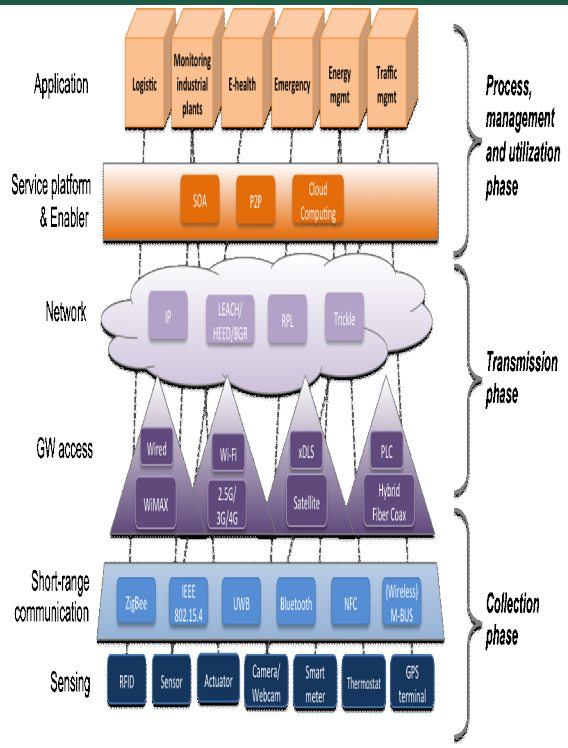


Figure 4: A typical horizontal representation of IoT Applications. **Source:** (Borgia, 2014)

Government organizations, industries and research institutions are all making efforts to research into IoT and leverage on its colossal potential to their personal and collective advantage, there are wide number of funded projects globally aiming at researching into the challenges of implementing IoT and how to proffer solutions. A wide range of European Union research and applications projects have been launched to research into context-awareness and Internet of Things (IoT). Since 2010 the United State have funded about four (4) projects since 2010 which aim at designing and validating comprehensive new architectures for the next-generation Internet as part of the "National Science Foundation's Future Internet Architecture" (Borgia, 2014), (Gubbi, Buyya, Marusic, & Palaniswami, 2013).

Context-aware computing and Internet of Things (IoT) has a strong interrelationship because:

i. Context-aware applications will provide a great support to process and store big data and it will also make the interpretation of the data easier.
ii. Context-awareness will ensure the implementation of efficient services
iii. Combining the capabilities of context-awareness with that of IoT; information about object's features, its status, its geographical location, and security data may be exploited to enrich the knowledge on services and refine for instance the choice of the most suitable provider

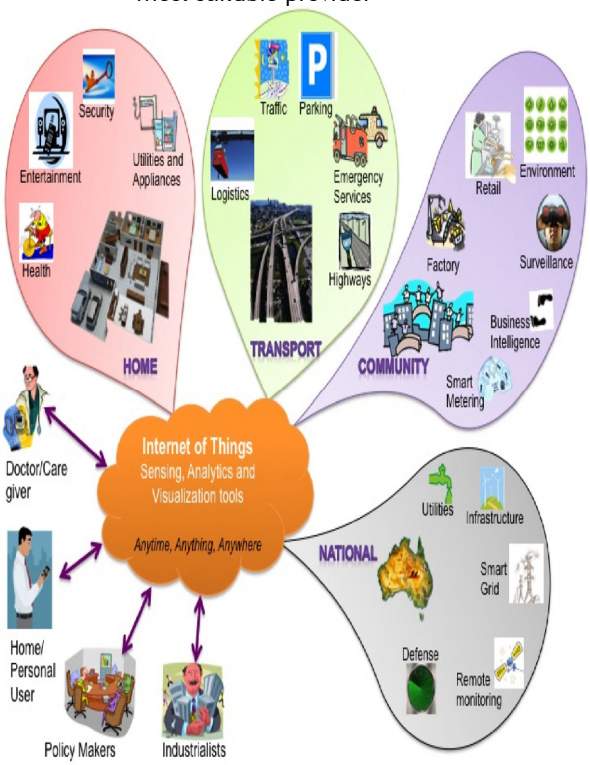


Figure 5: IOT's Schematic showing the end users and application areas based on data. **Source:** (Borgia, 2014)

4.3 Context- aware computing and cloud computing

Since the emergence of cloud computing, researchers and practitioners alike are eager to look deep into it's structure, operations and the benefits they tend to derive by embracing it as one of the new contributions to the field and practice of computing.

At the Centre of context-awareness and Internet of Things (IoT) is the cloud, which connects both the application, data, platform and the services being rendered. Sensing service providers can join the network and offer their data using a storage cloud; analytic tool developers can provide their software tools; artificial intelligence experts can provide their data mining and machine learning tools useful in

converting information to knowledge and finally computer graphics designers can offer a variety of visualization tools. Cloud computing can offer these services as Infrastructures, Platforms or Software where the full potential of human creativity can be tapped using them as services (Gubbi, Buyya, Marusic, & Palaniswami, 2013).

Cloud computing provides a common platform for all users to be their own IT person. It has reduced the cost of It services through it's three (3) cardinal service provision i.e. Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and Software as a Service (SaaS); all these are made possible because Cloud Computing allows users to use resources without owning the computing resources.

IT Users lend IT resources (storage, servers, network, software) as the need arises and then pays as he/she uses any of the resources; this increasing acceptance of the new IT service provision has made mobile cloud computing another evolving concepts in the field of computer science and service provision.

Mobile cloud computing combines the superiority and economic benefits of cloud computing to meet the convenience and mobility of mobile devices and this now makes available "anytime anywhere computing service using platform, software and infrastructure as a service". Mobile devices computing capacity has been limited because of size, battery life, storage size etc.; Cloud Computing has come as a means of making mobile devices an access route to the unlimited services and computing capabilities provided by the cloud.

5. Harmonized National Database and Context-aware data

Surveillance and tracking can be applied to individuals, groups of people or other objects such as vehicles, jets, motorbikes, ships and even drowns. Any organisation (Company, State or Country) with a harmonized database of such humans or objects has laid a good foundation for context data to be interpreted and used accurately especially where such context refers to the unique objects in the database.

Case Study:

Assuming Mr. Cole Ahmad is duly registered with the National Identity Management Commission with a valid (NIN) and also has his bank accounts linked to his NIN through the Bank Verification Number (BVN), his records with the Federal Road Safety Commission also contains his NIN and BVN and his vehicle also has similar details at the point of registration, other databases such as his International Passport with Nigeria Immigration Service, his drivers license and his SIM registration data can access any of these databases for the purpose of unique identification verification then the first level of individual uniqueness records can be seen to be in order.

In relating context data to harmonized database; it means the GPS on his phone and that of his car can record his movement, position at any point in time and such records can be communicated to any of the authorized aforementioned databases.

The results will then be that:

a. His location at any point in time can be known without asking him of his whereabouts
b. The places he has been in the last three (3) months can be known and how long he stays in each location can also be known i.e. movement pattern can be established
c. Relationship between different classes of people in each of the databases can then be studied and it could be of help to government agencies, businesses, security agencies and the community at large in terms of where they go, what they do and with whom they associate.
d. Trends of other GPS Co-ordinates of other phones or communication devices that in close range with that of Mr. Cole Ahmad can be studied to know those he relates with often and create a multiplier surveillance and detective effect.

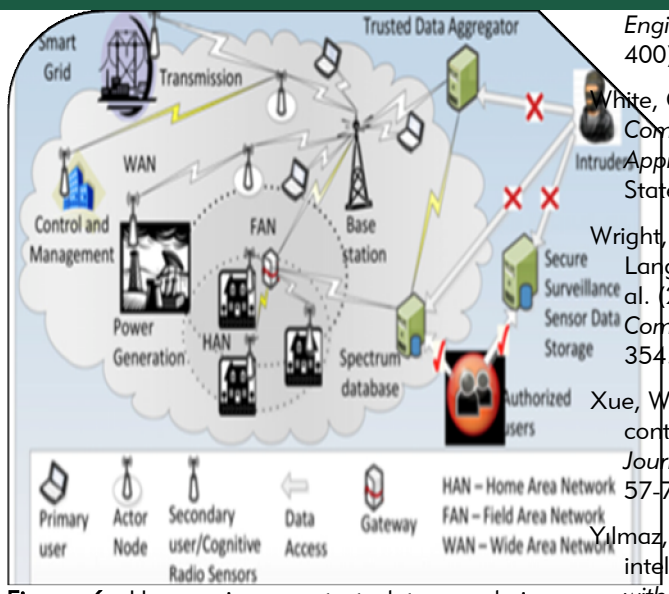


Figure 6: How various context data are being transmitted. Source: (Chang, Kang, Ahn, Jang, & Choi, 2012)

Such surveillance and detection performed on Mr. Cole Ahmad can also be performed on objects with a valid SIM card or other wearable objects.

6. *Conclusion*

Context-awareness in ubiquitous computing has provided no hidden place for both human and objects whose actions and/or inactions are important to our co-existence and unity.

This paper has discussed its application as it relates to security and surveillance, harnessing the power of context-awareness into different areas of our operations can help us build indigenous solutions to our problems.

References

Stojmenovic, I.´. (2002). *HANDBOOK OF WIRELESS NETWORKS AND MOBILE COMPUTING.* MEXICO: John Wiley & Sons, Inc.

Katina, M. A., & Roger, C. (2013). Location and tracking of mobile devices: U¨berveillance stalks the streets. *Computer Law and Security Review* , 216-228.

Wasserman, A. (2010). Software Engineering issues for mobile application development. *FSE/SDP Workshop on Future of Software Engineering Research (FoSER '10)* (pp. 397-400). ACM, Ney York: Sciverse ScienceDirect.

White, C. M. (2013). *Data Communications and Computer Networks (A Business User's Approach)* (7th Edition ed.). Boston, United States of America: COURSE TECHNOLOGY.

Wright, D., Friedewald, M., Gutwirth, S., Langheinrich, M., Mordini, E., Bellanova, R., et al. (2010). Sorting out smart surveillance. *Computer Law & Security Review* , *26*, 343-354.

Xue, W., Pung, H., & Sen, S. (2013). Managing context data for diverse operating spaces. *Journal of Pervasive and Mobile Computing* , 57-75.

Yılmaz, O., & Erdur, R. (2012). iConAwa – An intelligent context-aware system. *Expert Systems with Applications* , 2907-2918.

Al-Bashayreh, M., Hashim, N., & Khorma, O. (2013). Context-Aware Mobile Patient Monitoring Framework Development: A Detailed Design. *2013 International Conference on Electronic Engineering and Computer Science* (pp. 155-167). Malaysia: SciVerse ScienceDirect.

Blasch, E., & Aved, A. (2015). Dynamic Data-Driven Application System (DDDAS) for Video Surveillance User Support. *ICCS 2015 International Conference On Computational Science* . *51*, pp. 2503-2517. ELSEVIER.

Borgia, E. (2014). The Internet of Things vision: Key Features, Applications and Open Issues. *Computer Communications* .

Brown, P. J., Bovey, J. D., & Chen, X. (2000). Context-aware Applications: from the Laboratory to the Marketplace.

Brown, P. (1996, September). The stick-e document: a framework for creating context-aware applications. *Electronic Publishing* , 1-14.

Chang, H., Kang, Y., Ahn, H., Jang, C., & Choi, E. (2012). Context-aware Mobile Platform for Intellectual Disaster Alerts System. *2012 International Conference on Future Energy, Environment, and Materials* (pp. 1318-1321). Korea: SciVerse ScienceDirect.

Corral, L., Janes, A., & Remencius, T. (2012). Potential advantages and disadvantages of multiplatform development frameworks – A

vision on mobile environments. *International Workshop on Service Discovery and Composition in Ubiquitous and Pervasive Environments (SUPE)* (pp. 1202-1207). Italy: SciVerse ScienceDirect.

Dey Anind, K. (2000). Providing Architectural Support for Building Context-Aware Applications. *Providing Architectural Support for Building Context-Aware Applications* . Atlanta, Georgia: Unpublished Doctoral Thesis.

Deepak, G., & Pradeep, B. (2012). Challenging Issues and Limitations of Mobile Computing. *International Journal of Computer Technology & Applications* , *3*, 177-181.

Fortier, A., Rossi, G., Gordillo, E., & Challiol, C. (2010). Dealing with variability in context-aware mobile software. *The Journal of Systems and Software* , 915-936.

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems* , 1645-1660.

Garg, N., Lather, S., & Dhurandher, S. (2012). Smart applications of Context services using Automatic adaptive module and making Users Profiles. *2nd International Conference on Communication, Computing & Security [ICCCS-2012]* (pp. 324-333). India: Sciverse ScienceDirect.

Johnson, G., Shakariana, P., Gupta, N., & Agrawala, A. (2011). Towards Shrink-Wrapped Security: Practically Incorporating Context Into Security Services. *International Symposium on Frontiers in Ambient and Mobile Systems (FAMS)* (pp. 782-787). New York: ELSEVIER ScienceDirect.

Johnson, M., Phillips, S., & Stein, R. (2002). Contextual data and the study of elections and voting behavior: connecting individuals to environments. *Electoral Studies* , 219-233.

Leach, M. J., Sparks, E., & Robertson, N. (2014). Contextual anomaly detection in crowded surveillance scenes . *Pattern Recognition Letters* , *44*, 71-79.

Subodhani, P. U., Khalil, I., & Atiquzzaman, M. (2015). Secure and reliable surveillance over cognitive radio sensor networks in smart grid. *Pervasive and Mobile Computing* , *22*, 3-15.

Sánchez-Pi, N., Carbó, J., & Molina, J. (2012). A knowledge-based system approach for a context-aware system. *Knowledge-Based Systems* , 1-17.

Schilit, B., Adams, N., & Want, R. (1994). Context-aware computing applications. *Workshop on mobile computing systems and applications, IEEE Computer Society* (pp. 85-90). London: IEEE.

Sirichaia, P., Kaviyab, S., Fujiic, Y., & Yupapind, P. (2010). Smart Car with Security Camera for Road Accidence Monitoring. *2nd International Science, Social Science, Engineering and Energy Conference 2010: Engineering Science and Management* (pp. 308-312). Thailand: ELSEVIER.

Ramadath, S., & Collins, M. (2012, 12 01). Mobile Application Development: Challenges and Best Practices. North America, USA.

Full Paper

AUTOMATED VEHICLE SCRUTINY THROUGH MOBILE-BASED SYSTEM USING SHORTCODE

J. B. Awotunde
Department of Computer Science,
University of Ilorin, Ilorin
jabonnetbylinks@gmail.com

A. O. Umar
School of Science,
Department of Computer Science,
FCT College of Education,
Zuba, Abuja
Ojo_raheem@yahoo.com

O.S. Isiaka
Institute of Information & Communication Technology,
Kwara State Polytechnic
Ilorin
isiakaosalman2@gmail.com

M.B. Akanbi
Institute of Information & Communication Technology,
Kwara State Polytechnic
Ilorin
bolatwo@yahoo.com

ABSTRACT

Crimes rate have been of increase in our society, and vehicles are being used for committing most of these crimes. Research shows that perpetrator of these acts make use of stolen vehicles snatched at gun points which is causing damages to live and properties. The Nigeria Police, Road safety Corps, Roadworthy Officers and other security agents adopted manual method for vehicle inspection system that involves the use of showing papers for proof of ownership which is prone to error and very easy to manipulate. There are many ways to scheme the inspectors since lots of tricks can be done on paper. The proposed system try to design an automated GSM-based vehicle inspection that will send a short message to a particular designated shortcode, which in turns bring back all necessary information needed by the vehicle inspection team to verify the authenticity of the vehicle ownership. The system is designed using a descriptive conceptual approach Unified Modelling language (UML) tools while PHP and MYSQL server which is used in storing data being captured are used for implementation of the proposed system. The proposed system uses the staff identification number as primary identification (authentication) and all the staff phone number will be register in the database. If the system is implemented, the rate at which vehicles will be stolen at gun-points or crime rate will be reduced if not totally eradicated and also shows that GSM is a practicable platform for managing the process of inspection of vehicles. Future work is to consider how the system can be further secured from authorized users, and further research work should be carried out on the use of Short Message Service (SMS) and Multimedia Messaging Service (MMS) and its application to accommodate the picture of the vehicle owners.

Keyword: Automated, Mobile-based, Vehicles, Scrutiny, Shortcode

Information technology for national safety and security is essential particularly now that there are many crimes in our nation where people cannot sleep with relaxed mind (concentration). The need for good record-keeping and information sharing practices has taken on added importance in today's global environment. Not only do good records keeping provide crucial internal information (that is, business operations and case management support not to mention the official memory of an agency's investigations), law enforcement agencies now need to communicate agency-to-agency and across continents in order to protect the Nation's citizens. Nothing is more important to accomplishing that mission than having accessibility to accurate and timely records. Calls for service records and investigation, arrest, criminal identification, detention, and even civil records hold information that by themselves mean little; however, when pieced together with information from other jurisdictions, the result can help with all levels of investigations and aid in safeguarding the Nation.

Crime is a human experience and it has to be controlled (Awotunde et al, 2015). In order to control crime, it has to be properly managed and analyzed. Crime has no boundaries and criminal activities are becoming more sophisticated. Crime analysis is attainable with the aid of an effective information system. To fight crime, criminal investigations have to be conducted using technology solutions that facilitate communication, enhance investigative work and prevent crime.

The new age of technology has redefined communication. Most people nowadays have access to mobile phones and therefore the world indeed has become a global village. At any given moment, any particular individual can be contacted with the mobile phone. But the application of mobile phone cannot just be restricted to sending SMS or starting conversations. New innovations and ideas can be generated from it that can further enhance its capabilities. Technologies such as Infra-red, Bluetooth, and so on which has developed in recent years goes to show the very fact that improvements are in fact possible and these improvements have eased our life and the way we live. Remote management of several home and office appliances is a subject of growing interest and in recent years we have seen many systems providing such controls.

Global System for Mobile Communications (GSM): It is a cellular communication standard. GSM is the most popular and accepted standard for mobile phones in the world established in 1982 and it operates in 900 MHz frequency. Over one billion people use GSM service across the world. The utility of the GSM standard makes international roaming very common between mobile phone operators, enabling subscribers to use their phones in many parts of the world. GSM differs significantly from its predecessors in both signaling and speech clarity, as its channels is digitized. It means that the GSM system is now considered as a third generation (3G) mobile communication system.

Today's second-generation GSM networks deliver high quality and secure mobile voice and data services (such as SMS/ Text Messaging) with full roaming capabilities across the world.

SMS stands for Short Message Service. It is a technology that enables the sending and receiving of message between mobile phones. SMS first appeared in Europe in 1992. It was included in the GSM Later it was ported to wireless technologies like Code Division Multiple Access (CDMA) and Time Division Multiple Access (TDMA). GSM standards right at beginning, and later it was ported to wireless technologies like (CDMA) and (TDMA). The GSM and SMS standards were originally developed by ETSI. ETSI is the abbreviation for European Telecommunication Standard Institute. Now the 3GPP (Third Generation Partnership Project) is responsible for the development and maintenance of the GSM and SMS standards (Rappaport, 2000).

One SMS message can contain at most 140 bytes (1120 bits) of data, so one SMS message can contain up to:

- 160 characters if 7-bit character encoding is used. (7-bit character encoding is suitable for encoding Latin characters like English alphabets.)
- 70 characters if 16-bit Unicode UCS2 character encoding is used. (SMS text messages containing non-Latin characters like Chinese character should use 16- bit character encoding.)

Shortcodes (also known as short numbers) are special telephone numbers, significantly shorter than full telephone numbers that can be used to address Short Message Service (SMS) and Multimedia Messaging Service (MMS) messages

from certain service provider's mobile phones or fixed phones.

Short codes are designed to be easier to read and remember than normal telephone numbers. Like telephone numbers, short codes are unique to each operator at the technological level. Even so, providers generally have agreements to avoid overlaps. In some countries, some classes of numbers are inter-operator meaning they cut across several operators (Phanerus Technology, 2011).

Shortcodes are widely used for value-added services such as television program voting, ordering ringtones, charity donations and mobile services. Messages sent to a short code can be billed at a higher rate than a standard SMS and may even subscribe a customer to a recurring monthly service that will be added to the customer's mobile phone bill until the user texts, for example, the word "STOP" to terminate the service.

Short codes are often associated with automated services. An automated program can handle the response and typically requires the sender to start the message with a command word or prefix. The service then responds to the command appropriately.

In advertisements or in other printed material where a provider has to inform about both the prefix and the short code number, the advertisement will typically follow this format:

Example 1 - Long version: Text Football to 72404 for latest football news.

Example 2 - Short version: football@72404

Crime rate is on the increase in Nigeria which is causing threat to lives and properties which in turn might affect the economy of the Nation, this research seeks a preventive measure to curb this ridiculous act. This paper therefore proposes the design of a system with an automated GSM-based vehicle inspection that will sending a short message to a particular designated shortcode, which in turns bring back all necessary information needed by the vehicle inspection team to verify the authenticity of the vehicle ownership, using an application PHP and MYSQL server in storing data being captured. Table 1 below gives the detailed comparison between the existing system and the proposed system.

Table 1: Comparison of the Existing System and the Proposed System

Existing System	Proposed System
Vehicle Inspection System is done manually	automates the vehicle inspection process
Characterize with fake/counterfeit vehicle particulars	Guarantee original vehicle particulars
Confrontation between law enforcement agent and citizen during vehicle inspection exercise	Reduce confrontation between law enforcement agent and citizen during vehicle inspection exercise
Authentications of vehicles registration particulars is difficult	Authentications of vehicles registration particulars is guarantee
Retrieval of information very difficult because of the large volume of file	Easy retrieval of vehicle information and control data concurrency
Lot of time is devoted to the filing	It is real time process during inspection
Lack of distributed database for the storage of files	Window workflow foundation will keep track of the movement of information online because it make use of distributed database

GSM Based System Related work

(Ramamurthy & Shashi, 2010) proposed the development of a Low-Cost GSM SMS-Based Humidity Remote Monitoring and Control system for Industrial Applications. They proposed a wireless solution, based on GSM networks for the monitoring and control of humidity in industries. This system provides ideal solution for monitoring critical plant on unmanned sites. The system is Wireless therefore more adaptable and cost-

effective. Utilizing Humidity sensor HSM-20G, ARM Controller LPC2148 and GSM technology, this system offers a cost effective solution to wide range of remote monitoring and control applications. Historical and real time data can be accessed worldwide using the GSM network. The system can also be configured to transmit data on alarm or at preset intervals to a mobile phone using SMS text messaging. The proposed system monitors and controls the humidity from the remote location and whenever it crosses the set limit, the LPC2148 processor will sends an SMS to a concerned plant authority(s) mobile phone via GSM network. The concerned authority can control the system through his mobile phone by Attention Command (AT) Commands to GSM MODEM and in turn to processor. Also the system provides password security against operator misuse/abuse. The system uses GSM technology thus providing ubiquitous access to the system for security and automated monitoring and control of Humidity.

(Ghose et al, 2011) presented the design and development of microcontroller based SMS gateway for GSM Mobile. In their work, a microcontroller based SMS gateway for GSM mobile has been designed and developed. Most of the SMS gateway system was controlled by PC based software where microcontroller only used for controlling and sending status of devices or any appliances connected with the system. An Ericsson T68i, one of the cheapest GSM mobile phone sets available with most of the advanced features, has been interfaced with a PC via RS232 serial port. The SMS packet has been analyzed and its different fields have been identified for the Grameen Phone, the largest GSM operator in Bangladesh. Then the PC has been removed from the system and the transmission and reception technique of SMS have been implemented into the PIC microcontroller. Successful completion of the design and testing of the SMS Gateway indicates that the PC as an SMS gateway can easily be replaced by a PIC microcontroller. Beside this, the additional IC, MAX232, used for voltage adjustment between the mobile and PC is no longer needed in the proposed micro-controller based system. It also reduces the complexity and the overall development cost of such a system. Therefore the system becomes smarter, efficient and portable. In addition, since the microcontroller can also be configured as a web server, this system can be accessed for controlling various devices in the remote place through the Internet. The developed system has been tested successfully. The system is also simple, smarter, portable, cost effective (as the PC has been removed) and low power consuming.

The development and implementation of a Global System for Mobile Communication (GSM) based control system for electrical appliances that enables the complete control of the interface on which it is based (Oke et al, 2013). GSM module was used for receiving short message service (SMS) from user's mobile phone that automatically enable the controller to take further action like switching ON and OFF electrical appliances such as fan, air- conditioner, light etc. The system was integrated with microcontroller and GSM network interface using C language. MPLAB software was utilized to accomplish the integration. The system is activated when user sends the SMS to the controller at home (regarded as Smart Home). Upon receiving the SMS command, the microcontroller unit then automatically controls the electrical appliances by switching ON or OFF the device according to the user's order. In other word, it reads message from the mobile phone and respond to control the devices according to the received message. The project which is development of a GSM based control system for electrical appliances was designed considering some factors such as economic application, design economy, availability of components and research materials, efficiency, compatibility portability and durability. The performance of the project after test met design specifications. However, the general operation of the project and performance is dependent on the user who is prone to human error such as entering wrong timing. Also the operation is dependent on how well the soldering is done, and the positioning of the components on the Vero-board. If poor soldering lead is used, the circuit might form dry joint early and in that case the project might fail. Furthermore, if logic elements are soldered near components that radiate heat, overheating might occur and affect the performance of the entire system. Other factors that might affect performance include transportation, packaging, ventilation, quality of components, handling and usage.

(Jain, Kumar, & Kedia, 2012) presented the Design and Development of GSM based Energy Meter. Traditional metering method for retrieving the energy data is not convenient and the cost of the data logging systems is high. So this paper

presents design and development of Automatic Meter Reading (AMR) system. AMR system is a boom for remote monitoring and control domestic energy meter. AMR system give the information of meter reading, power cut, total load used, power disconnect and tempering on request or regularly in particular interval through SMS. This information is being sent and received by concerned energy Provider Company with the help of Global system for mobile communication (GSM) network. Energy provider receives the meter reading within a second without visiting person. AMR minimize the number of traditional visits required by employs of energy Provider Company. This system not only reduces the labor cost but also increase meter reading accuracy and save hugs amount of time.

GSM based energy meter is easy to install and beneficial for both energy provider and consumer. AMR not only solves the problem of manual meter reading but also provide additional feature such as power disconnect due to outstanding dues, power reconnect after pay dues, power cut alert, tempering alert. AMR also gives the information of total load used in a house on request at any time. It sends a SMS alert to energy provider company whether a person using more than specify limit of load. The statistical load used and profile can help customer manage their energy consumption. This system is secure and reliable because it can be accessed only by an authorized person. If any un-authorized person tries to access the system this system send an alert to energy provider and also give warning of that unauthorized person. This device has the capability to revolutionize the energy meter market and will become help to country revenue by stopping the current theft and punishing the dishonest customers.

The study of Embedded Automobile Engine Locking System, Using GSM Technology, deals with the design & development of an embedded system, which is being used to prevent /control the theft of a vehicle (Pany & Choudhury, 2011). The developed instrument is an embedded system based on GSM technology. The instrument is installed in the engine of the vehicle. An interfacing GSM modem is also connected to the microcontroller to send the message to the owner's mobile phone. The main objective of this instrument is to protect the vehicle from any unauthorized access, through entering a protected password and intimate the status of the same vehicle to the authorize person (owner) using Global System for Mobile (GSM) communication technology. This system deals with the concept of network security. The main concept in this design is introducing the mobile communications into the embedded system. The entire designed unit is on a single board. This is a unique method of designing and assembling a low-cost, compact theft control system for an automobile. This instrument is an ultimate threat to vehicle thieves. By installing this instrument in the automobile engine it is very difficult to access by an unknown person, since it is based on GSM Technology. In future, there is no doubt, that all of the vehicles will be embedded with this unique kit. In addition to the above features we can also add extra features like thumb/face recognition to ascertain more security of the vehicle.

GSM based referral system for primary Health care centres in Nigeria. Referral system which is one of the strategies put in place for ensuring the best use of hospital resources and health care services. The design is an attempt to reduce paper work, forestall the problem of case notes getting lost in transit, shield the contents of case note away from patients which could also lead to a psychological breakdown by the patient and as a result of all these reduce mortality rate in the primary health care centres in Nigeria, an electronic referral system is design to alleviate all these aforementioned problems (Idowu & Ajayi, 2008).

Spiral model was the methodology used because it allows going back to earlier stages a number of times as the project progress. Interviewed of medical personnel and visit some primary health care centres (PHC) and it was discovered that some of the PHC have no internet or Landline facilities but they all almost have GSM network which mobile phone uses. They also gathered data that are related to referral of patients from PHC to tertiary hospitals and formulate a diagrammatic representation of the system in order to have a good database.

The designed known as Mobile Phone based Electronic Referral System (MERS) consists of three main menus namely: control panel, Hospital Records Management and Expertise. The package will be initializes by clicking on configure under the control panel. The forms designed for their system depict three different phases which are the input phase, processing phase and output phase. The results showed that patients' case notes (which encompass Patients' symptom, diagnosis, medication with the clinical number) were

transmitted using mobile phone on a Global System for Mobile Communication carries from the referral package within few seconds (Idowu & Ajayi, 2008).

In conclusion of their work, the MERS is designed such that there is one computer in each primary health care to access the computer at the teaching hospital. The system was developed with ability to transmit or send patients' case note which include the symptoms, diagnosis, medication with the appended patient' clinical number using mobile phone on a GSM carrier. The system has the potential to increase medical personnel productivity, reduce prenatal and neonatal mortality rates, improve medical care and minimize the cost of referral since GSM facilities are already on ground.

(Awodele et al, 2007) deals with An Improved SMS User Interface Result Checking System where they reviews the use of mobile phones for delivering examination results via Short Messaging Service (SMS) in a university where student who have written examinations and are anxious to see their results need to get their grades in a convenient and accessible way, whether in the comfort of their homes, on the road, or while at work.

At this time, a mobile phone is accessible to most students, and they take it almost everywhere with them. This technology can, therefore, be highly effective in bringing information to them quickly, easily, and while they are on the move. An SMS result checking system not only enables students to request their grades, the system can also deliver the grades to their phones as soon as the grades become available. This means that they can access their grades even in the remotest locations where internet service might be unavailable, preventing them from accessing the school's website.

This version is an improved version of the system presented previously (Adagunodo et al, 2009). The former system requires the user to submit an ID with a password, which is common to many SMS systems. The present system uses a social interaction with the password to reduce the incidence of guessing access codes occurring in the checking system.

The flow of the systems starts when a student sends a SMS (in the prescribed format) to the given number. The system then uses the content of the SMS to process the student's request, after which the student's result is sent back to the student via the same number.

In conclusion of their work the SMS result checking system is a cost effective and widely available means of communication for most students. The use of this medium enhances easy access to their result. The previous paper (Adagunodo et al, 2009) described the use of the student's ID with a password for obtaining the result, while the improved system uses social interaction based on surname of the individual involved in the checking system to request for the scores in addition to the password system.

This system uses a 2-tier level verification system involving the student's surname. The SMS result checking system tries to approach examination result checking from the point of social interaction between the students to improve the security to a certain level. The former system needs the students to produce the ID and a password, while in the new system the password is self-generated. This system tries to develop an open system which is based on trust so that students can access their result without being limited to a particular phone number. The extent of response to messages will depend on the network in use; therefore, the guaranteed delivery of grades within the shortest period is networked based. Another limitation is that the phone may have a low battery at the point of delivery to a student whereby the phone is switched off on its own. Further work is still to be done on the security protocols and to really ascertain the delivery of the various messages to the recipients.

3.0 RESEARCH METHODOLOGY

This paper designed a system where vehicle owners will be checked by vehicle inspection officers by sending a short message to a short-code which in turn responds by bringing all the details captured in the database during the vehicle registration. The system involves the design of an application that seats on the server and this house all information and data being collected from the vehicle owner during registration. The system registers all vehicles and captures all information about the vehicle and stores them in the database for future reference. The application is designed using Unified Modelling language (UML) Language and HTML, CSS, PHP and MYSQL server for the implementation of the system. The data can be retrieved during authentication, that is, verification by the vehicle inspection officer to ascertain whether the vehicle owner is truly the

person that he or she actually claims to be, if this system is actually implemented by the vehicle inspection team it will be a prominent way of reducing the rate at which vehicles are being stolen or vehicle particulars that are normally forged will be drastically reduced if not totally eradicated.

THE SYSTEM ARCHITECTURE

Mobile phone: The proposed system uses the user's identification number as primary identification (authentication) and all the staff phone number will be registered in the database. This offers a form of security by ensuring that only the registered staff with their mobile phone can request for the vehicle details. To request for a vehicle detail, the user sends vehicle registration number, along with a staff identification number (for security and secrecy) to the shortcode. The user sends the request data from his/her mobile phone by using shortcode to the designated phone/modem, the software in the server receives the data from the modem. The server (computer) which houses the details of vehicle registration. The computer then sends the results back to the user mobile phone, and this is read as a typescript message. The medium to act as an interface between the mobile phone and the computer is the GSM modem. In achieving this, the result is sent to the number of the SIM card that is use by the GSM modem to identify the owners. The mobile phone serves as an input and output device, sending and receiving messages (i.e. request data/detail vehicle registration results).

Private Phone/Modem Receives Data: The modem is connected to the computer's system using universal serial ports. The received request data is temporarily stored in the modem before being extracted by the system. The connection between the mobile phone and the GSM modem is wireless, and the connection between the computer and the GSM modem is physical, which is through the USB (universal serial bus) port.

Software in Server Accesses Modem for Data: This is used in creation and manipulation of information in the server, for maintaining the vehicle registration. This system uses time series forecasting method to predict the future using past and present data. This is also called the Work Space. This plays the role of "Read Only Memory" for all communications that need to be exchanged between the applications and mobile phones in use.

Server send Answers to Request Data: This module facilitates the input of acknowledged request data through program on the computer (Server).

Software Recalls Responses to Request from Database Table: This validates whether the request data adhered to the shortcode during the input process. If the user did not use the correct shortcode or use incorrect short-code, the "USE THE CORRECT SHORT-CODE" notification is displayed else the send request data to server invoked.

Display Result on Mobile Phone: This module displays the result of query with the software in the server by sending the owner vehicle registration details to the phone of the inspector agent.

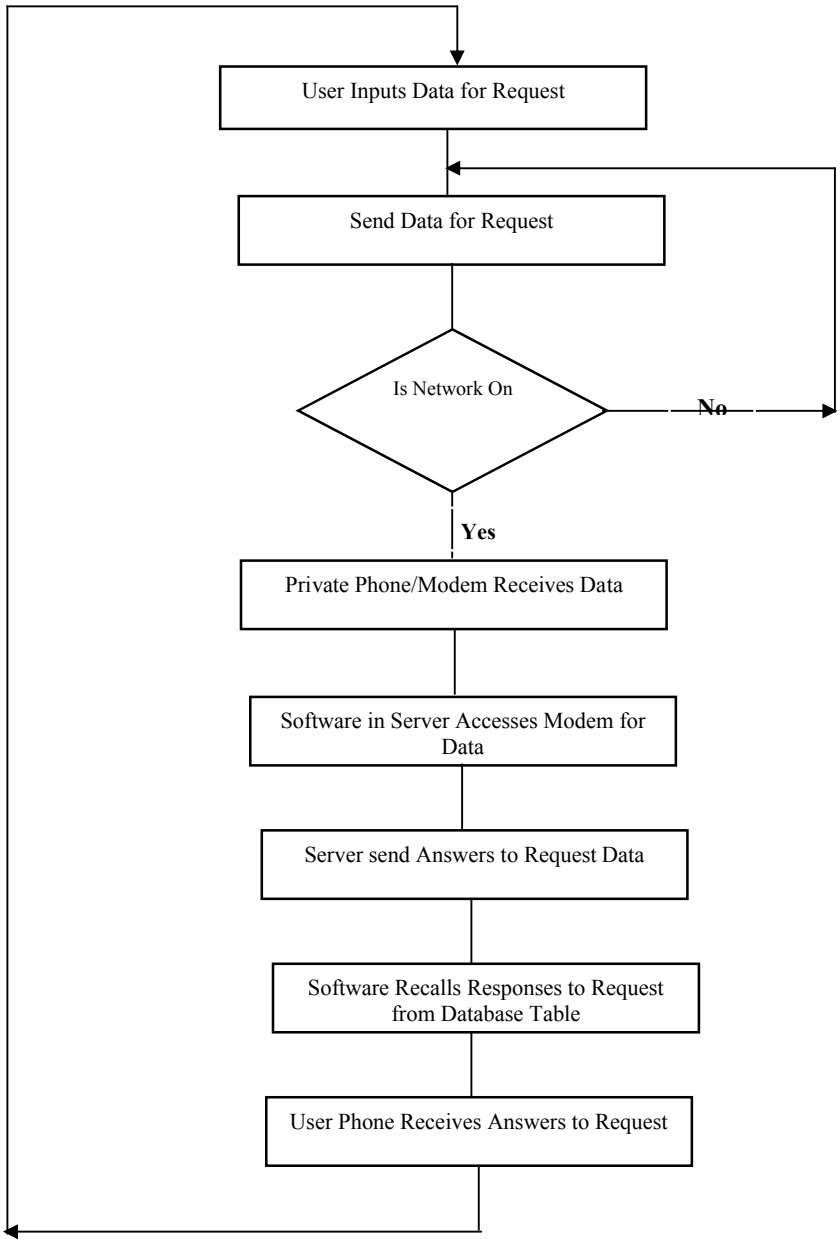


Figure 1: The Logical Design of the Proposed System

4.0 SYSTEM DESIGN

System design is the part where each element within the new system is structured in order to create and integrate the processes that meet the user needs and requirements. At the same time, system design must conform to the specifications and scope or boundaries established at the data analysis phase. All input requirements and expected output are identified and established.

Design is both a process and a product. The creative process of system design is the transformation of a process into a solution, otherwise known as "System Design"

Data Flow Diagram

This is a graphical representation of the flow of data through an information system. It mainly reveals relationships between entities by showing what data comes from where, where it is going to, as well as where it will be stored. The processes timing is not usually included in data flow diagrams.

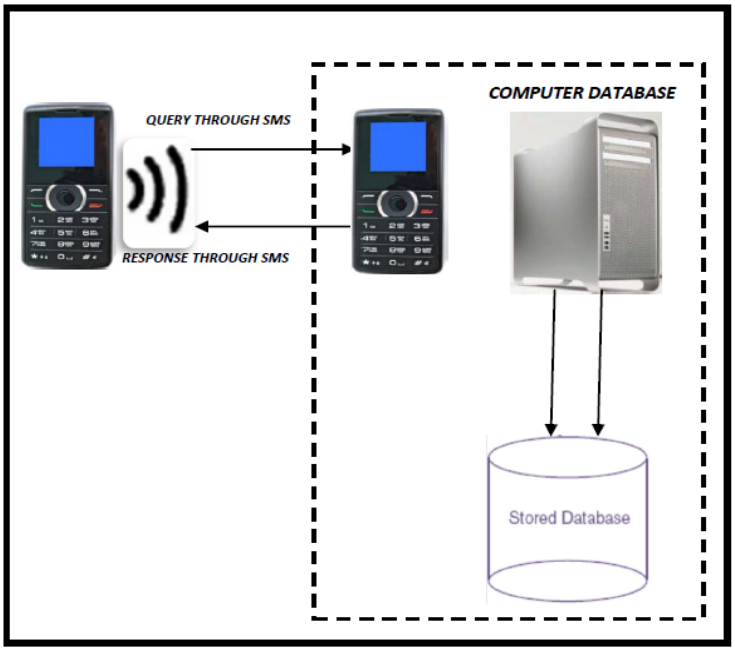


Figure 2: System Dataflow Diagram

Database Design and Specification

MySQL is a RDBMS (which has a free community edition) that can be used to develop database application both locally and on the internet. MySQL is popular and used by many webhost online for storing data on the internet. The current version that is used in this dissertation is version 5.5.24.MySQL storage engines provide more flexibility and offer more performance, it can host millions of database simultaneously without much commotion.

Pseudo code for the Proposed System

Pseudo-code is an informal way to express the design of a computer program or an algorithm. The aim is to get the idea quickly and also easy to read without details. It is like a young child putting sentences together without any grammar. The following are the pseudo code for the new system:

Pseudo code for the Proposed System

```
Begin
        Compose SMS
                SD 11391 Plate Number
        Send 35811
                Case Invalid
                        Return Back Message
                End Invalid
                Case Valid
                        Look-up Vehicle Info for
        Plate Number
                        If Found then
                                Return Vehicle
Info/Details;
                        Else
                                Return
"Information Not Found"
                        Endif
                End Valid
        End
        Stop
```

The pseudo code above is for the front end of the application system. This will allow an officer to compose an SMS with SD 11391 Plate Number then send it to a shortcode 35811.The application will look up the information and send the necessary data back to the user. If the information is not found, it will send a response indication that the information was not found

Back-End Pseudo code

```
Begin
    Display Welcome Screen
    Main Menu Display
    Select Menu Option
        Case Admin
            Enter Admin Username and Password
            If Username and Password are correct then
            Display Admin Main Menu;
            Select Option;
            Else
                MsgBox "Check your Username and Password"
            Endif
        End Select
        Case Home
            Display Statistics of Activities (glance view) going on in the Application
        End Select
        Case Add-Vehicle
            Enter a New Vehicle Details
            Compare Reg no (plate number)/Chassis number/Engine number
            IfAlready existthen
            MsgBox"Check Reg no (plate number)/Chassis number/Engine number"
            Endif
        End Select
        Case Vehicle Lists
            Display Vehicle Information Currently available
        Case Logs
            Display Users Accessing the Application through Shortcode Information
        End Select
        Case Logout
        End
End
Stop
```

The pseudo code above is for the welcome page to the organization website. This will allow the Admin to select the available menus options which any administrator can choose. It also include the pass/ authorization for the admin or the user. It is the validation and access for all users. The user is prompted here to enter a USERNAME and a PASSWORD for the administrator, the system verifies the correctness. It is from which the administrator manages all transactions. He/She can register a new vehicle details, modify, views registered vehicles, view details of numbers aceess the application through the short code and the access code at will.

Flowchart for the proposed system

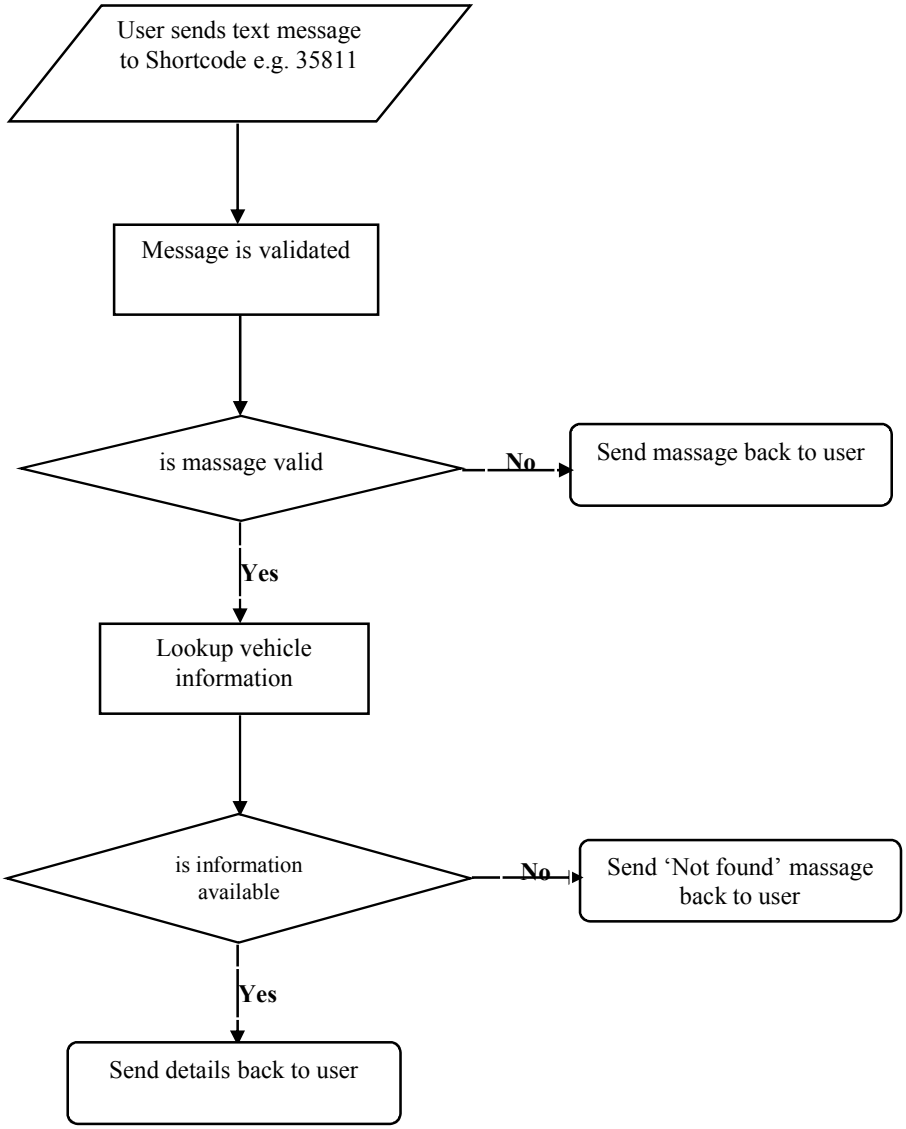


Figure 3: Flowchart for User Interaction with System

A GSM-Based Vehicle Inspection Management System is user friendly and has a Graphical User Interface. It is also interactive and will assist any law enforcement agency to collect, store, process and mange vehicle information and investigation for present and future needs of the Motor Vehicle Inspection Unit and other law enforcement agencies. System requirements include hardware and software requirements. Due to the rapid growth in technology, new systems are developed and presented in the market every day.

5.0 CONCLUSION

It is a well-known fact that insecurity is what most of the developing nation spend their affluence on. Many researchers have tried to have a lasting solution to these problems by trying to profound solution in different ways in like manner of this paper. GSM-Based Vehicle Inspection Management System is a widely available means of vehicle inspection for most law enforcement agencies. The use of this medium enhances easy access to vehicle owner's information via mobile phone. The system design which make use of SMS

and short-codes technology automated program that handle the response and typically requires the sender to start the message with a command word or prefix.

This system uses a 2-tier level verification system involving the law enforcement agency's mobile phone number. The GSM-Based Vehicle Inspection Management System tries to approach vehicle inspection from the point of social interaction between the vehicle users and inspection officers to improve the security to a certain level. The existing system needs the vehicle user to produce the registration documents, while in the new system the plate number is only required. With the use of this system the issues of vehicle stolen or collected at gunpoint will be reduced if not totally eradicated. The proposed system was developed using HTML, CSS, PHP and MYSQL server.

6.0 FURTHER WORKS

Further work should investigate mobile phone functionalities and its usage for more practicable platform for managing the process of inspection of vehicles. Future work should also consider how the system can be further secured from authorized users, and further research work should be carried out on the use of Multimedia Messaging Service (MMS) and Short Message Service (SMS) and its application to accommodate the picture of the vehicle owners.

REFERENCE

Nwobodo L.O., and Inyiama H.C., (2013). GSM Based Vehicle Inspection and Verification System, *International Journal of Engineering Research & Technology (IJERT)*, Vol. 2, Pg2393-2399.

Vandana .P. and Deepali .S., (2012). GSM Modem Based Data Acquisition System, *International Journal of Computational Engineering Research (ijceronline.com) Vol. 2, Pg1662-1667*.

Ramamurthy B., Bhargavi S., and ShashiKumar R., (2010). "Development of a Low-Cost GSM SMS-Based Humidity Remote Monitoring and Control system for Industrial Applications", *International Journal of Advanced Computer Science and Applications*, Vol. 1, Pg20- 26.

Ghose S.,Rahman Md. S., Sharmin D.,Hussain I.,Yousufzai T.K., (2011). "Design and Development of Microcontroller Based SMS Gateway for GSM Mobile", *International Journal of Advanced Engineering Sciences and Technologies*, Vol.2, Pg90-98.

Oke A. O., Emuoyibofarhe J. O., Adetunji A. B., (2013). "Development of a GSM based Control System for Electrical Appliances", *International Journal of Engineering and Technology* Vol.3, Pg443-448.

Pany J. K. and Das Choudhury R. N., (2011). "Embedded Automobile Engine Locking System, Using GSM Technology", *International Journal of Instrumentation, Control and Automation* Vol.1, Pg49-53.

Rappaport T. S., (2000). Wireless Communications, second edition, PHI New Delhi

Phanerus Technology, (2011). "Short Code Service API Documentation". http:/www.smsdam.com.

Oke A. O., Emuoyibofarhe J. O., Adetunji A. B., (2013). "Development of a GSM based Control System for Electrical Appliances", *International Journal of Engineering and Technology* Vol.3, Pg443-448.

Jain A., Kumar D., Media J., (2012). "Design and Development of GSM based Energy Meter", *International Journal of Computer Applications*, Vol.47, Pg41 – 45.

Idowu P. A. and Ajayi A. S., (2008). "GSM based referral system for primary Health care centres in Nigeria", *International Journal of Soft Computing* Vol.3, Pg421-427.

Awodele O., Adagunodo E. R., Akinwale A. T., Idowu S. and Agbaje M.(2007) "An Improved SMS User Interface Result Checking System", *Interdisciplinary Journal of Information, Knowledge, and Management* Vol.4, Pg51-62.

Adagunodo, E. R., Awodele, O., and Idowu, O., (2009). "SMS user interface result checking system" *Issues in Informing Science and Information Technology*, vol.6, Pg163-177.

Awotunde J.B., Adewunmi-Owolabi, F.T., Owolabi A.A., & Akanbi, M.B. (2014). Automated Global System for Mobile-Based Vehicle Inspection Using Short-Code: Case Study of Nigeria, *Computing, Information Systems, Development Informatics & Allied Research Journal*, Vol.5(3), Pg45-50

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

INTEGRITY ASSURANCE FOR SMALL SCALE DIGITAL DEVICES BASED EVIDENCE FOR CYBER CRIME INVESTIGATION

M. K. Muhammad
Academic Planning Unit, Federal University of Technology, Minna
muhammad_kudu@futminna.edu.ng

I. Idris
Department of Cyber Security, Federal University of Technology, Minna ismi.idris@futminna.edu.ng

I. Lukman
CODEL Unit, Federal University of Technology, Minna
lukman.ibr@futminna,edu,ng

ABSTRACT

The recent rapid development in the field of information and communication technology industry have made the concept of acquisition and analysis of digital evidence an increasingly important tool for uncovering digitally related crimes and preparing them as a reliable evidence for legal acceptability. In this paper, a new generalized framework for the acquisition of digital evidence was applied on multiple forensics tools. The forensic tools used were EndCase, AccessData FTK Imager, Mount Image Pro and Autopsy 4.0 and their individual features was compared in order to provide reasonable level of assurance to compare various level of integrity assurance to make them admissible as viable digital evidence in the law court during cyber related litigation.

Keywords: Cyber Crimes, Device, Digital, Digital Forensics, Encase., Message Digest, Scale, Small

4. INTRODUCTION

In recent years, digital forensics has changed the general approach used by the law enforcement domain to an invaluable tool for detecting and solving corporate cyber related crimes. Since digital forensic evidence play a vital role in solving cyber related crimes, it worth to be investigated in a forensically sound manner.

Digital forensics evidence can be termed as a set of binary digit numbers stored as files on an electronic storage media either mobile or otherwise. There are a number of characteristics that are to be considered; for example, the said evidence can be copied and modified, this alteration to the original information may not be identified or noticed when it is compared with the original source. It can also be integrated into other data format verification. Often, digital evidence may not be understood directly without technical process and knowhow, even interpreting it from public perception may requires the efforts of an experts, otherwise the entire presentation may on its own becomes abstract in nature. Digital evidence according to Baggili (2015), is a fragile piece of information that can easily be destroyed or become inadmissible for legal credibility after its collection as a result of modification either intentionally or otherwise.

Originality of digital evidence is surrounded with the challenges of how its integrity is preserved, and this is a fundamental requirement because of trust as the end point is human dependent. It is necessary to preserve the integrity of digital evidence during its entire life cycle in order to have a forensics value thereby making the assurance of such evidence an umbrella principle. Digital evidences (Hagy, 2007) are usually an extracted piece of information obtained from the crime suspects and taken to the forensics laboratory for examination. Only the conclusions which is usually in the form of reports are usually shared with the parties concerned, so the digital forensics process can be liken to a black box for cyber crime investigation (Saleem, 2015).

With these new methods of perpetuating digital crimes, there has to be emergence of new technologies and measuring devices that can be used to track the cyber criminals, they are called electronic evidences. This is an instrument that is fast becoming part of our daily life and is acquiring increasing importance in lawsuits. It is no longer understatement that traditional evidence is shifting from paper supporting documents towards a digital and virtual domain and its management processes are proportionally changing in this world of dynamic technology even in the court of law.

Solid state digital devices (SSDD) are majorly the most popular non-volatile solid-state technology in the world of information and communication technology today and it can be accessible by anybody for conveying information from one medium to the other, either for legal or illegal purposes. According to a study conducted by ITU, it was revealed that 86.7% of individual using one computing devices or the other are using a mobile device (Thing et al, 2010). Since small scale digital device (SSDD) have literally become a sort of digital behavioral archieves both at collective levels and individual. They are omnipresent recording of all users activities at the moment. It obvious that, during cyber crime investigation, these category of storage devices can be a reliable source of evidence in furthering and resolving a related legal case with more assurance (Saleem, 2015).

5. RELATED LITERATURE

Here some works that has been previously carried out on this subject matter was critically reviewed with the aim of knowing why digital evidence are not globally acceptable as viable evidence during cyber related crime investigation.

There is no doubt that digital forensic according to (Harrill & Mislan, 2007) is a viable research area today because of the innovations in the digital technology industry coupled with exponential growth in cyber-crimes especially in the information superhighways. Digital forensic is becoming more attractive to the academicians except that some scholars are claiming that there are some characteristics that are affecting the investigation processes. Some of these characteristics include the physical shape of the Ahuja *et al*, (2005) devices with respect to most of the recent reported crimes. For example, the tiny and adaptable nature of small scale digital devices makes digital forensics investigation more complex for the investigators. As a result of this, cyber criminals use flash memory technologies to perpetuate their illegal activities (Casey, 2014).

Over the years, digital forensics have transformed into discipline that requires a comprehensive forensics investigation process model. Different researchers have proposed several investigative process model (Brison, *et al*, 2006). However, these proposed model over the years lack practical evaluation especially on mobile storage devices. The role of testing and evaluating a harmonized investigative process model lies in ensuring that the model adhere to certain forensics standard (Reith, 2012). It is because of these inadequacy that has made the growth of digital forensics investigations on small scale digital devices very unpopular and cyber criminals explored this weakness to perpetuates several undiscovered crimes and in a few cases where they are discovered, the integrity of data presented before the court lacks expected merits (Casey, 2004).

6. PROPOSED METHODOLOGY

This section provides detailed information on the proposed methodology. In order to validate the reliability of evidence collection for the various experiments to be conducted for this study, similar works done by Arasteh *et al*, (2013) and Saleem (2015) were used as benchmarks. Data extraction algorithm of Saleem (2015) was expanded. Though the algorithm was limited to android devices but the human right and other legal privileges of cyber-crime suspects was the focus of their work. For the purpose of this work, the evidence extraction part was reviewed and expanded in order to make use of multiple forensics software tools. This modification introduced a new algorithm that was implemented for both the collection of evidence from any category of solid state digital devices (SSDD) in order to establish how the integrity of digital evidence can be preserved.

In this study, it is believed that applying multiple digital forensic tools will assist to determine if the contents of the solid state digital devices (SSDD) has been altered while on transit between the point of arrest or collection of the device and the point of examining the contents.

3.1 Case File Extraction and Modification

The evidence extraction tools were also evaluated for their ability to preserve the integrity of digital evidence. The following experiment was conducted.

3.2 Procedures

i. Digital evidence from SSDD was obtained.
ii. The evidence image file was opened using the hex editor in each of the tools and its contents were modified.
iii. The same case file was reopened with each of the tools.

3.2.1 Results

90% of the entire solid state digital device (SSDD) was in good working condition hence, evidence was obtained from all the storage devices presented during the experiments except some few files whose contents were damaged. It was noted that message digest (MD5) and digital hashes were used to preserve the integrity of digital evidence.

3.3 Extracting Evidence and Preservation of Integrity

Prior to the commencement of evidence acquisition process, it is obligatory to safeguard the device with Farady cage to avoid unnecessary alteration in case the system in use is on a network which could trigger events resulting in modification of contents of the SSDD object. This may affect the integrity of the expected result if the SSD is an Android devices that have option plug and play during operations. This is really helpful since for collecting data which otherwise could have been altered if the device is turned off when it was seized or collected from the crime suspect.

It therefore become pertinent to check if the SSDD is already connected, and replace it with the target SSDD or where the SSDD contents is to be transfer then, there may be need to look for an add-on application with some the popular forensics tool such as Efficient Generalized Forensics Framework Acquisition Application. There is need to then navigate through File Explorer in order to launch the add-on application if it does not come with the forensics tools. The application will automatically close all firmware processes running on the system being used for the experiment in order to avoid the issue of locking. In order to ensure integrity of the acquired evidence, the application comes will various tools various tasks such as hashing of each file before and after copy. The purpose of this is to keep tracks of activities on images/data before they were extracted and after the actual extraction.

3.4 Evidence Acquisition Process Model

Figure 3.6 present the new model for the acquisition of digital evidence from both Android and Non-Android storage device otherwise referred to in this study as small scale digital devices (SSDD). In the model, when an SSDD is mounted, the forensics tool used already have some enhanced functionality for computing the Message Digest algorithm and the SHA1 in order to avoid unnecessary human interaction with the entire process. The model automatically accesses the SSDD physical volume and other file structure including the FAT file of NTFS part. All the details of the images contained in the SSDD is accessed including date and time when the image was created, modified and other task that any user may have carried out on such data are noted and reported during analysis. All other required activities are included in the algorithm systems in figure 1.

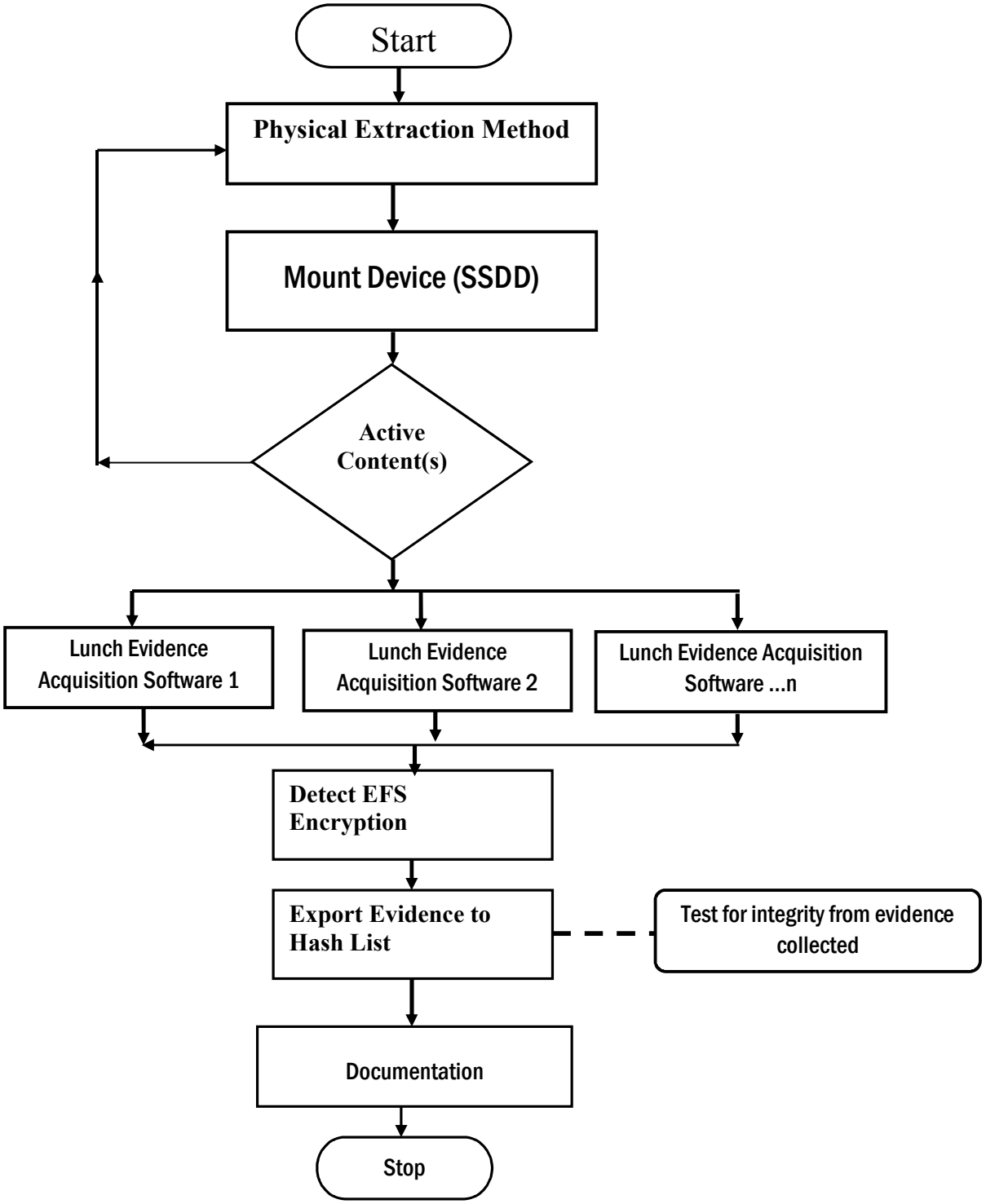


Figure 1: Generalized Evidence Acquisition and Integrity Check Model

3.5 Case File Extraction and Modification

The evidence extraction tools were also evaluated for their ability to preserve the integrity of digital evidence. The following experiment was conducted.

3.5.1 Procedures

i. Digital evidence from SSDD was obtained.
ii. The evidence image file was opened using the hex editor in each of the tools and its contents were modified.
iii. The same case file was reopened with each of the tools.

3.5.2 Results

90% of the entire solid state digital device (SSDD) were in good working condition hence, evidence was obtained from all the storage devices presented during the experiments except some few files whose contents were damaged. It was noted that message digest (MD5) and digital hashes were used to preserve the integrity of digital evidence.

3.6 Extracting Evidence and Preservation of Integrity

Before Acquisition process starts, it is necessary to shield the device with Farady cage to avoid network communication which could trigger events resulting in modification of file system's object. Mostly all the Android devices have option to plug-in a SD card while the device is powered-on (hot-plug) without removing battery. This is really helpful since for collecting data which otherwise could been altered if the device is turned off before the seizure process.

Therefore, we have to check first if a SD card is already plugged, and replace it with a SD card containing updated version of Efficient Generalized Forensics Framework Acquisition App. We need to then navigate through File Explorer to launch the Acquisition application. The application will automatically short down all firmware processes running on the system in order to avoid locking problems. In order to ensure integrity of the acquired evidence, the application comes will various tools to perform other tasks such as hashing of each file before and after copy. The purpose of this is to keep tracks of images/data before they were extracted and after the actual extraction.

3.7 Acquisition Algorithm

The implementation details are provided in the following Figure 6 which shows the pseudo-code for the Acquisition Process:

The acquisition algorithm performs the follow tasks:

i. Copy Evidence from SSDD mounted on the system
ii. In this task, all the contents of the SSDD are copied into a Case file
iii. Hashing
iv. The task of Hashing is to ensure integrity of the extracted evidence and allows discovering if there is an alteration in the contents between when the evidence was extracted and when it was actually analyzed.

The acquisition algorithm uses the various features in the forensic evidence acquisition tool for performing needed tasks during the above processes.

This algorithm preserves the main directory structure, by duplicating the existing images/folders, files and other contents of the SSDD according to their original position on the storage device recursively. The hashing ensures integrity check before and after duplicating the device contents. The hashes are also written in the appropriate log file called case1and case 2.

3.7.1 Algorithm Acquisition

```
Input: A path P
Out: none
for all objects obj (folders, files and directories) in p do
        if obj is a directory then
                create a directory names p in SSDD
                                Recursively call
Acquisition(p/obj)
                                        else
                if obj is a file then
                                        compute
MD5/SHA1 hash of obj
                                        copy obj in path p
on the SSDD
                                                if obj has
not been copied then
                                                access to obj with
                        evidence acquisition  software
                end if
                        end if
                                        if obj is access then
                                        recreate database in
path f
                                on SSDD
                                end if
```

```
                end if
        end if
            compute MD5/SHA1 hash of evidence
            extracted obj on the SSDD
```

3.8 Returning the SSDD to its former state

If the device is not booted using the CRMI, the device can be return to the former state after completing the evidence acquisition. This process continues until all the evidences contained in all the seized mobile devices are fully acquired. If a file for the boot partition exists, a check will be conducted to determine if it is the correct original boot partition. When checking if it is correct original boot partition, the firmware version that is used in the targeted device is essentially important to note.

When the targeted mobile device is completely returned to its former state, the device should be unplugged until the device is used again in order to prevent data modification. If the device is an all-in-one type with battery, then cut off the power by using the power button and if the battery can be removed, it should be remove.

In case of turning the device off by using the power button, then it is recommended that the researcher should not use the menu functions of the recovery mode. The menu can be different for each vendor/device manufacturer or firmware version, but the reboot system now is usually included in the menu recovery mode. Before removing the battery, the USB cable must be separated first. Some mobile devices mount the user data partition by using the power provided by the USB cable if the battery is separated when the USB cable is still connected.

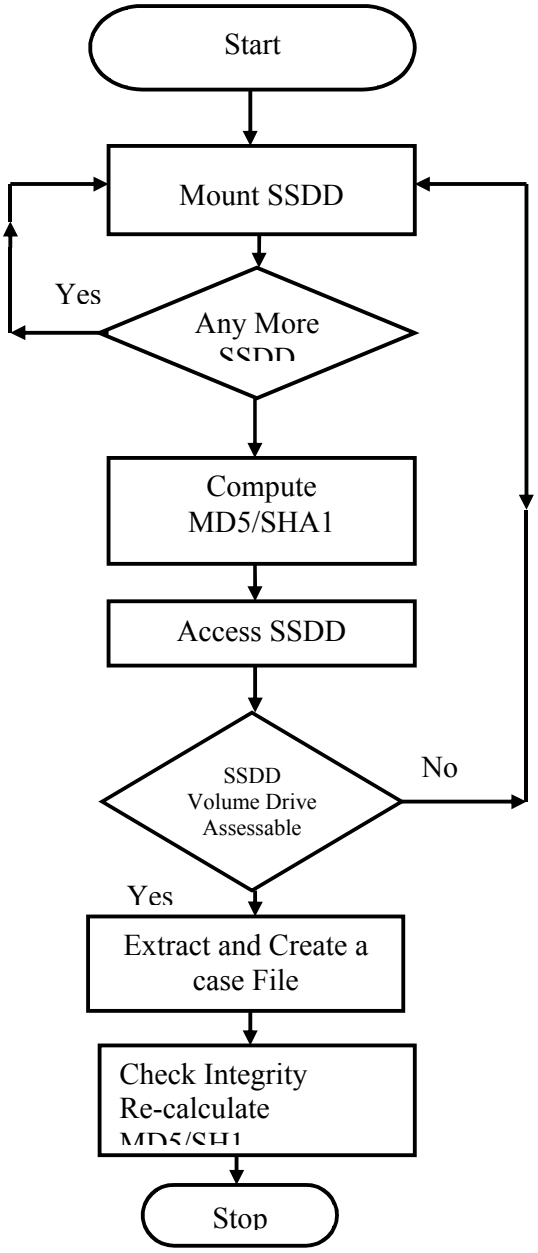


Figure 2: Evidence Acquisition Process Model

4. PRESENTATION OF DATA AND DISCUSSION

In this chapter, discussions on various approaches used during the extraction of digital evidence and how the adopted forensics tools was used are fully documented. Basic features of each of the forensics tools used were discussed with a comparative analysis of their attributes for providing adequate integrity on the extracted digital evidence.

4.1 Evaluation Criteria for Integrity Assurance

Digital evidence is ubiquitous, so digital evidence can come from various category of SSDD, regardless of whatever implication any individual may passive it. Digital evidence is by any mean crucial to the development of forensics industry hence, preserving the integrity of such extracted information from the devices used in perpetuating the crime in question thus important. There are many methods used in preserving the integrity of digital evidence, when attention is focus on the various approach that some of these tools handles accuracy, performances, vulnerabilities and complexity, they are differs in nature. In order to know how suitable they are in preserving evidence, table 6 provides 3 classes for evaluating integrity of digital evidence. The following preservation scheme was adopted from Saleem, (2015) to confirm the result of some of the experiment performed with the 4 forensics tools as indicated in Table 1, Table 2, Table 3, Table 4 and Table 5 respectively. Some of the criteria used includes:

i. Digital Hashes (MD5 and SHA1
ii. Digital Signature which rely on public key cryptography and require PKI at it backend.
iii. Cyclic Redundancy Checks (CRCs)

4.3 Comparative Analysis of Digital Forensic Tools

In providing reliable computer analysis and collection of digital evidence to meet the variety of needs in the field of forensics, digital forensics tools play a vital role. Most of these tools are used to conduct investigations of computer crimes by identifying evidence that can be useful in the court of law during cyber related crimes investigation. In addition to cyber related crimes investigation, these tools are used for the purpose of evidence extraction, debugging, data recovery among other in a secured environment which are usually refers as being forensically sound.

Table 1 shows comparative details of four evidence extraction tools with five parameters. From the table, the speed of acquiring evidence from solid state digital device (SSDD) is very slow on EnCase 7 and AccessData FTK Imager, although both of them are highly rated with respect to integrity assurance. But when large numbers of solid state digital device (SSDD) are to be consider for investigation, it will take longer time to extract evidence using EnCase 7 and AccessData. Unlike the Mount Imago pro and Autopsy, the speed of evidence acquisition was quite high. This gives an indication that Mount Image Pro and Autopsy 4.0.0 are good tools when speed of acquiring evidence is of high priority.

Criteria	EnCase 7	AccessData FTK Imager	Mount Image Pro	Autopsy 4.0.0
Speed Acquisition from SSDD	Slow	Slow	Very High	Very High
Scalability of Extracted Evidence	Selective	Selective	Any Image Format	Any Image format
Message Digest (MD)	Produce MD 5, MD3.	Produce MD 5, MD3 and MD1.	Produce MD 3	Produce MD 3
Hash File Level of Secured Evidence	Comprehensive High Level	Comprehensive High Level	Not Comprehensive High Level	Not Comprehensive High Level

Table 1 : Evidence Formats and Evidence Acquisition Tools

4.4 Analysis of time spend on each tool during evidence acquisition

Since the study focus on various category of SSDD, there would be need to consider time spend on different SSDD with respect to their storage capacity. Every digital forensic tools will spend different amount of time on each category of SSDD to access, extract and analyse the contents of each storage device.

To determine the integrity of acquired evidence, the need to know the quality of the digital forensics tools used with respect to whether the software tool is a free license, open source, the operating system platform with which the tool will run such as Microsoft windows and the need to also know the performance and cost of acquisition is very crucial. As shown in the table 2, for the purpose of study, open source of digital forensics tools was obtained and exclusively used.

	EnCase 7				AccessData FTK Imager				Mount Image Pro				Autopsy 3.0			
	Physical SSDD	Logical Volume	File	Folders	Physical SSDD	Logical Volume	File	Folders	Physical SSDD	Logical Volume	File	Folders	Physical SSDD	Logical Volume	File	Folders
USB Flash Drive	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mobile Phone	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SIM Card	✓	✗	✗	✓	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗
Memory Card	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 2: Behavioural Analysis of Evidence Acquisition Tools on set of criteria on SSDD.

Also in order to accomplish one of the set objectives of using more than one forensics tool to test for integrity of evidence, other evidence acquisition tools was also acquired from the open source platform. Therefore, Table 3 also show the analysis of four categories of tools used and their functionality was also compared using Cost, Performance, Platform Support and License criteria's. For example, table 3 shows that EnCase 7 is a commercial version and apart from the fact that its performance was very high, it supports both 32 and 64 bit windows operating system.

If EnCase 7 is compared with Mount Image Pro, it was shown in the table that Mount Image pro was Open Source, its run effectively on only windows 32 bit but it attract no cost in terms of acquisition.

	EnCase 7	AccessData FTK Imager	Mount Image Pro	Autopsy 4.0.0
Software License	Commercial	Commercial	Trial Version	Trial Version
Platform Support	Windows 32 Bit and 64 Bit	Windows 32 Bit and 64 Bit	Windows 32 Bit	Windows 32 Bit
Performance	High	High	Low	High
Cost	High	High	Free	Free

Table 3: Comparison of considered tools on the basis of features

In table 4, digital forensic investigation process was examined and four forensic tools was compared. As indicated in the table, Mount Image Pro does not have adequate features for keeping track of date and time when an evidence is acquired so it could not provide a good valid information on Preservation of evidence and is also not able to analyse evidence even though, it has a good reporting features. From the table, it obvious that Autopsy 3.0.0 do not have feature for examining extracted evidence

Tool Used	Preservation	Collection	Examination	Analysis	Reporting
EnCase 7	Yes	Yes	Yes	Yes	Yes
AccessData FTK Imager	Yes	Yes	Yes	Yes	Yes
Mount Image Pro	No	Yes	Yes	No	Yes
Autopsy 3.0.0	Yes	Yes	No	Yes	Yes

Table 4: Comparison of considered tools on the basis of Digital Forensic Investigation Process

In table 5, set of scalable criteria was used to examine each of the digital forensic tools used. The objective of this was to know further apart from table 4.6, how each of the tool handles other basic integrity criteria between the Fully, Partly or Nil. For example, in trying to know how variable like Automated MD5 Algorithm was treated on each of the tools, from table 5, the performance remark was for all the four forensic tools used.

This also show that EnCase 7 can be a more preferred digital forensics tools when knowledge of the details of deleted files from an solid state digital device (SSDD) is a critical factor to maintain assurance over a digital evidence.

Scalable Criterias	EnCase 7	AccessData FTK Imager	Mount Image Pro	Autopsy
Supported Image File Format	Fully	Fully	Partly	Partly
Show Deleted Files	Fully	Fully	Partly	Partly
Show Unallocated Clusters	Fully	Partly	Partly	Partly
Remove Hidden Attributes	Fully	Partly	Partly	Partly
Physical Drive Mounting	Fully	Fully	Partly	Partly
Extended Partition Support	Partly	Partly	Partly	Partly
Plug and Play Mount Option	Fully	Fully	Fully	Fully
File Activities Log Details	Partly	Fully	Partly	Partly
Automated MD5 Algorithm	Fully	Fully	Partly	Nil
VMWare Activities Log	Partly	Partly	Partly	Fully
Extensible Keyword Search	Partly	Fully	Partly	Partly
Artifact Analysis	Partly	Partly	Fully	28
Registry Analysis	Fully	Partly	Partly	Partly

Table 5: Comparison of considered tools on the basis of Digital Forensic Investigation Process

On any flash or memory card that has no folder, but has partition(s) information, the size of each partition was checked and the partition(s) are imaged and checked in line with the steps in the Algorithm. For the acquisition of the file allocation table (FAT), the partition table is automatically mounted in read only mode to guarantee data integrity. The consciousness here is that, if the time and date of the content of the partition table changes, the integrity of the content is loss and the set objective will not be met.

5.1 CONCLUSION

Having concluded an in-depth literature research into integrity of digital evidence and explored diverse reason why most cyber related extracted evidence from the various storage devices are not usually considered as legitimate evidence for consideration by the court during investigation, it can be concluded that the prosecutor of some of those court cases with respect to cyber-crimes lost out because, evidence are either extracted manually with already compromised human intervention. However, with some of the exercise and results of this work, it is obvious that using an automated forensic tools goes a long way to reduce the existing challenge of non-admissibility of digital evidence in the law court. It was also noted that digital evidence can be relied upon especially when the evidence are extracted in a forensics manners. The same way mobile telephone call logs are recognized and admitted in the law court, digital evidence that are extracted in a forensically manner should be recognize and admitted in the law court during cyber related investigations.

REFERENCE

Ahuja, M. K., and Thatcher, J. B. (2005). Moving beyond intentions and toward the theory of trying: effects of work environment and gender on post-adoption information technology use. *Management Information System quarterly*, 29(3), 427-459.

Arasteh, A. R., Debbabi, M., Sakha, A., and Saleh, M. (2013). Analyzing multiple logs for forensic evidence. Digital Investigation, 4, 82-91.

Baggili and Huebner, E. (2015). Computer forensic analysis in a virtual environment. International journal of digital evidence, 6(2), 1-13.

Brinson, A., Robinson, A., and Rogers, M. (2006). A cyber forensics ontology: Creating a new approach to studying cyber forensics. digital investigation, 3, 37-43.

Casey, E. (2004). Network traffic as a source of evidence: tool strengths, weaknesses, and future needs. Digital Investigation, 1(1), 28-43.

Casey, E. (2014). Digital evidence and computer crime: Forensic science, computers, and the internet. Academic press.

Hagy, D. W. (2007). Digital evidence in the courtroom: a guide for law enforcement and prosecutors. *National Institute of Justice*.

Harrill, D. C., & Mislan, R. P. (2007). A small scale digital device forensics ontology. *Small Scale Digital Device Forensics Journal*, 1(1), 242.

Reith, M. (2012). An examination of digital forensic models. *International Journal of Digital Evidence*, 1(3), 1-12.

Saleem, S. (2015). Protecting the Integrity of Digital Evidence and Basic Human Rights During the Process of Digital Forensics.

Thing, V. L., Ng, K. Y., and Chang, E. C. (2010). Live memory forensics of mobile phones. digital investigation, 7,S74-S82.

SESSION A:

Educational Technologies and E-Learning

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

COMPARATIVE ANALYSIS OF KNN AND SVM CLASSIFIERS FOR STUDENTS' ACADEMIC PERFORMANCE PREDICTION

M.G. Samuel
Department of Computer Science,
Federal University of Technology Akure, Akure, Nigeria
mojiasho@yahoo.com

M.O. Omisore
Institute of Biomedical & Health Engineering,
Shenzhen Institutes of Advanced Technology,
Chinese Academy of Sciences, Shenzhen, China
ootsorewilly@gmail.com

O.W. Samuel
Institute of Biomedical & Health Engineering,
Shenzhen Institutes of Advanced Technology,
Chinese Academy of Sciences, Shenzhen, China
timitex92@gmail.com

B.A. Ojokoh
Department of Computer Science,
Federal University of Technology Akure, Akure, Nigeria
bolanleojokoh@yahoo.com

O. Dahunsi
Centre for Information Technology and Systems,
University of Lagos, Yaba, Lagos State, Nigeria
dansylobss@gmail.com

ABSTRACT

The performance of students in tertiary institutions has been a major indicator for assessing the quality of students and their respective institutions over the years. Placement of students into appropriate department/faculty based on their individual competence would no doubt lead to the production of quality graduates who can contribute substantially to societal development. Recently, a number of studies have attempted to provide a means of predicting the performance of students with the aim of placing them into appropriate departments/faculty programmes. However, among several earlier proposed methods for students' performance, k-nearest neighbor (kNN) and support vector machine (SVM) seem promising based on their ease of use and high level of reliability. Therefore, a kNN and SVM predictive models were developed to evaluate and predict the performance of students in institution of higher learning. A comparative analysis of the performances of these two models were observed using 5-fold, 7-fold, and 10-fold cross validation methods. From the experiment, the best performance was achieved when 7-fold cross validation was used of kNN (69.00%) and SVM (72.24%), which indicates that SVM is a better predictor over kNN. Lastly, our findings show that SVM can help provide a more accurateand robust decision making tool with respect to student performance prediction.

KEYWORDS: University, Student performance, Predictive model, k-nearest neighbor, Support vector machine.

1. INTRODUCTION

A major setback in developing countries is low quality student graduates produced by their educational systems (Ekundayo and Ekundayo, 2009). Despite the annual increment in number of entrants into tertiary institutions in Nigeria, some local and foreign industries still find it difficult to make many of these graduates verily fit into their establishments, while others would rather send them on quick (few months) training for effective production. This in no way has helped both parties as huge amount of money is spent on training or longer times are wasted before such students become well equipped with the required skills and knowledge. To minimize these expenses, Nigerian educational sector introduced Industrial Training programme while the students are about completing their education, yet little or no improvements have been realized. This could be due to the fact that such students are always wrongly utilized while serving the industry, or they have weak theoretical background hence finding it difficult to merge with the real life applications in the industry. In a recent study (Ojerinde and Kolo, 2007), it is clear that the academic entry requirements into Nigerian Universities influence the quality of graduates from the institutions. A general believe is that students who had excellent scores during admission are likely to complete their education with excellence, and have greater edge in their future career.

Conventional methods for assessing student performances generally relies on subjective evaluations (Omisore and Azeez,2015) rather stakeholders such as, teachers, the West African Examination Council, the Joint Admission Matriculation Board, and University managements should take a deeper look into how examinations are conducted in terms of focus and emphasis on actual skills and competences. Strict attention paid to subjective methods does not assist the institutions because it rarely provides an efficient means of filtering inferior students; thereby students produce woeful results (Alexander, 1996). Importantly, counseling and administration department of tertiary institutions should pay more attention to determining academic status of student during their years of studies. If such institutions know in advance the weak students who are likely to have insubstantial performance in their present and future career, necessary actions can be taken to assist such students. Gaining insight into students' future career is very difficult except by intervention into the nitty-gritty of present academic progression of students. Furthermore, this has to be properly taken care of just as soon as students are given admission into institution of higher learning. However, a question that needs to be addressed is how institutions' management can presume the capabilities of such newfangled students. Successfully predicting the academic performances of students in institutions of higher learning would help provide a means of admitting and placing students into the right faculty programmes where they can develop substantial skills and knowledge.

Data mining, an approach that has been used in several domains to provide useful information for effective decision making has been rarely applied to educational data of students for the purpose of performance prediction.This technique often provides effective means of discovering hidden patterns in large dataset thereby leading to discovery of knowledge which aids decision making. Previous studies reported data mining as a veritable technique that can be applied to study available data in educational field in order to extract hidden knowledge (Galit, 2007). Also, it has been used to predict the performance of students at different levels in institutions of higher learning (Omisore and Azeez, 2015 and Asogbon et al., 2016).

Recently, two popular prediction methods among others that have gained wide range of applications across several fields are k-nearest neighbour (kNN) and support vector machine (SVM). SVM, a robust and accurate technique for pattern classification and knowledge mining, has been reported to have sound theoretical foundation and highly reliable. Unlike other algorithms, the application of SVM to educational data is yet to be fully explored. kNN, a labor intensive algorithm best adopted in situations of small training dataset, conforms with Euclidean distance measure in terms of distance metrics. Its application in the field of students' performance prediction is relatively low compared to other classical prediction algorithms. Although, few studies had used this two predictive tools to evaluate and predict the performance of students but a comparative study on the performances of these two methods has not been investigated. Therefore, this study is aimed at comparing the performances of kNN and SVM methods for student performance prediction.

2. LITERATURE REVIEW

The context of mining in educational data varies with the areas of application. For instance in this context, it is the extraction of meaningful knowledge from large repositories of data generated from learning activities (Asogbon et al., 2016 and Althaf et al, 2012) towards improving results obtained from conventional system. To achieve this, some techniques have been proposed recently in order to devise means of extracting useful knowledge from large amount of data. Finding major criteria for selecting mining technique, taxonomy of mining techniques, and identifying type of data to be mined, kind of knowledge to be discovered, or type of application to be adapted have been detailed in (Muqasqas, 2014, Han, 2012 and Kovačić, 2010).

The success of a student depends on patterns that exist in their physiological and psychological features (Porchea, 2010). However by observation, the relationship between psyche-physiological data is not linear hence very difficult to classify. Some studies with the aim of predicting student's performance are reviewed herein. Objectively, these studies focus on how to improve the quality of decisions necessary to impart delivery of quality education. The exponential growth of educational data in higher institutions of learning has given rise to different approaches of performance prediction. Lori et al, (2004)attemptedto justify some variables that may be related to predicting success of students who enrolled for a distant educational programme. The study concludes that learners' characteristics have a major impact on the manner in which online courses were designed and the pedagogy employed to deliver them. Oladokun et al, (2008) traced the poor quality of graduates from Nigerian Universities to inadequacies ofadmission systems in the country. The study employed artificial neural network to predict the performance of candidates considered for admission into University of Ibadan, Nigeria. In the study, 10 variables from demographic and educational backgrounds information of prospective students were transformed into a format suitable for the network processing. An output variable representing the students' final Grade Point Average on graduation was then predicted. The model operates with a promising accuracy rate of 74% but no psychometric factors were considered in the study.

Also, Stamos and Andreas (2008) utilized the promising behavior of neural network to predict students' final results. The network feature a three-layered perceptron trained with back-propagation. Experimentation was conducted with a case study of 1,407 profiles of students at Waubonsee Community College, Chicago, USA. The study concluded an average predictability rate of 68%. Survey in Romero C and Ventura(2007) describes educational data mining as a leading way to determine academic performance in learning institutions. This can be backed by an expert system developed in Hatzilygeroudis et al,(2004) to predict students' success in gaining admission into higher education institution through Greece national exams. In the study, prediction is made at three points with different variables considered at each point. An initial prediction is observed after the second year of studentship with specialization, sex, age, grades of students as input variables. This prediction gives a first indication of student's possibility to pass his/her exams otherwise specifying the effort necessary for student to pass. The prediction system demonstrates accuracy and sensitivity potencies of 75% and 88% respectively. Also in Gibson et al, (2012), the effects of students' characteristics, such as gender, ethnicity, and age on their studies were observed. The study considered 113,000 cases from a large database of national universities to determine the effects of these physiological factors. Analyses from multiple sessions across curricular areas shows students' performance has some form of non-linear relationship with various factors that make up their socio-demographic data.

Language is another impeding factor that can inhibit students' success especially in places where official language varies from local dialects or if the different ethnic groups in the nation do not converge on a central language. Effect of this factor was observed in Pandey and Pal (2011) by predicting the performance of students from different colleges of Awadh University in India. Although, the study considered variables from educational background information of the students, but also laid emphasis on language of instruction. Conclusions threw light to the fact that new-comer students are likely to always perform low. Furthermore, Sajadin et al, (2011) applied kernel method as mining techniques to analyze relationships between students' behavior and their success. The model wasdeveloped using smooth support vector machine and kernel k-means techniques to cluster final year students at University

Malaysia Pahang, Malaysia. Clustering results expressed a strong correlation between mental conditions and academic performance of students. Hence, psychometric information can be taken as an important factor while predicting students' academic performance.

In another recent study, Ahmed and Elaraby (2014) used ID3 classification technique to predict final grade of students in department of Management Information System, American University, Cairo, Egypt. Student information from data mart of 1547 records was used to train the model. Decision rules were used to find interestingness in the training data to predict and improve students' performance. Also, the model can help to identify students that might need special attention to reduce failure rate. However, none of the above review studies have investigated the performance of kNN and SVM with respect to students' performance prediction.

3. PROPOSED METHODOLOGY

3.1 DATA PRE-PROCESSING

Datasets of students relating to psychometric and physiological factors were collected by means of questionnaire from the department of Computer Science, University of Lagos, Nigeria. The dataset comprises of information of 300 students from all levels in 2013/2014 academic session. The data collected contains demographic, current and previous academic standings, departmental structure, and family backgrounds of students. 17 significant attributes were extracted for experimentation in this study. Table 1 presents the dataset attributes, description, and grading of attributes into categories.

Table 1: Attributes of student performance measure

Group Code	Group Name	S/N	Attributes	Attributes grading (From Class 1- 6)					
				1	2	3	4	5	6
A	Demography Information	1	Gender	Female	Male				
		2	Age (years)	14-20	21-25	>26			
		3	Entry Type	DE	UTME				
		4	Campus Location	Off-	Shared-	Own-			
B	Personal Information	1	Department	Poor	Average	Good	Best		
		2	Academic Advisor	Poor	Average	Good	Best		
		3	Curriculum	Poor	Average	Good	Best		
		4	Courses	Poor	Average	Good	Best		
C	Academics Information	1	Secondary School						
		2	Entrance Score	<60	60-70	>70			
		3	Grade Point Average	0–1.49	1.50–2.39	2.40-4.49	4.5-5.0		
D	Family Information	1	Family Size	<=2	3-4	>=5			
		2	Annual Income						
		3	Fathers Qualification	N/A	Elementary	Secondary	Bachel	Master	Ph.
		4	Mother's	N/A	Elementary	Secondary	Bachel	Master	Ph.
		5	Father's Occupation	N/A	Personal	Private	Civil	Public	
		6	Mother's Occupation	N\A	Personal	Private	Civil	Public	

3.2 KNN CLASSIFICATION

K-NN classification is a labor intensive algorithm best adopted in situations of small training dataset. The algorithm is found to conform to Euclidean distance measure in terms of distance metrics. The algorithm assumes a student, entity of our discourse herein, can be described as a set of values representing their attributes. Suppose a student (Si) is represented by the attributes described in Table 1, then relationship between the students and a set of other students is defined by their Euclidean distance. This describes closeness that exists between the set of known students ($S_1, S_2, S_3, \ldots, S_n$) and an unknown student(S_u) in terms of distance metrics. For instance, if (S_k) represents a known student with attributes given as ($S_{k1}, S_{k2}, S_{k3}, \ldots, S_{km}$), then the Euclidean distance between unknown student (S_u) and a known student (S_k) is given as:

$$E(S_u, S_k) = \sqrt{\sum_{i=1,j=1}^{n,m} (S_{uj} - S_{ij})^2} \quad (1)$$

If the tuples have numeric values on attributes $i = 1,2,3,\ldots,m$; then V-1 the cumulative aggregation of squares of each attribute differences is normalized before equation (1) is applied. To normalize the square of the differences, min-max normalization is applied. This function, given as equation (2), prevents attributes with large initial ranges from outweighing attributes with smaller ranges. It transforms numeric value V of attribute A to V^{-1} in range of $[0, 1]$.

$$V^{-1} = \frac{V - min_A}{max_A - min_A} \quad (2)$$

where min_A and max_A are the minimum and maximum values of attribute A.

In the case of nominal attributes where values are non-numeric, the distance metric can be deduced by correlating the corresponding values of attributes in tuples S_k and S_u. If such values are similar, the difference is taken to be zero (0); otherwise the difference is one (1) irrespective of their ordering. If the value of A is missing in tuple S_k and/or in tuple S_u, the maximum possible difference is assumed. For instance, in nominal categorical variables, a value of 0 is assigned if the corresponding values of the attributes are similar otherwise 1 is assigned. However, if either one or both of the corresponding values of A are missing, then the maximum possible value is assigned. For instance, if A is non-numeric and missing from both tuples S_k and S_u, the difference is taken to be 1.

3.3 SVM CLASSIFICATION

SVM algorithm based onGaussian Radial Basis Function (RBF)kernel is used in this study to predict students' performance in institutions of higher learning. The algorithm works by solving a single optimization problem through maximizingthe margins between all designed classes simultaneously. For instance, Figure 1 shows an illustration of linearly separable Crammer and Singer (CS) SVM based classifier that uses hyper-plane to separate the various classes considered.

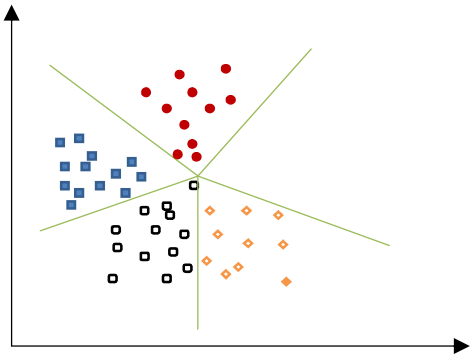


Figure 1: A five class problem based on CS classifier

CS inCrammer and singer (2002) proposed an approach that requires the solution of a single Quadratic Programming (QP) problem of size (k − 1) n, which uses less slack variables in the constraints of the optimization problem. This approach considers all available training dataset at once, and constructs *k* class categorization rules where *k* is the number of classes (Symeon et al, 2010).

Given a set of *n* training dataset $X = \{(x_1, y_1), \ldots\ldots, (x_n, y_n)\}$ where $x_i \in R^d, i = 1, \ldots, l$ are the input feature vectors, $y_i \in \{1, \ldots, k\}$ is the class output associated with the training dataset x_i, and d is the dimensionality of the input feature vectors. Solving this single optimization problem leads to construction of *k* decision functions. The m^{th} decision surface $w_m^T \phi(x)$ is determined by its normal vector $w_m \in R^d$ hence, separates training vectors of *m* class from others by minimizing the primal problem expressed in equation 3.

$$\min_{\{w_m\},\{\xi_i\}} = \frac{1}{2}\sum_{m=1}^{k} w_m^T w_m + C\sum_{i=1}^{l} \xi_i \quad (3)$$

$$subject\ to \quad w_{y_i}^T \phi(x_i) - w_m^T \phi(x_i) \geq e_i^m - \xi_i \quad i = 1,2,\ldots\ldots,l$$

where $\phi(.)$ denotes a function that maps the input feature vector x_i to an arbitrary-dimensional space $\mathcal{F}$where the dataset are to be linearly separable, C denotes the parameter that penalizes the training errors, $\xi = [\xi_1, \ldots, \xi_l]^T$is the slack variable vector, w_m is the weight vector associated with class m, $e_i^m = 1 - \delta_{y_i, m}$ and

$$e_i^m = 1 - \left(\delta_{y_i, m} \equiv \begin{cases} 1 & y_i = m \\ 0 & y_i \neq m \end{cases}\right) \quad (4)$$

where$\delta_{y_i, m}$ denotes the Kronecker delta function.

In equation 1, the constraint $m = y_i$ corresponds to the non-negativity constraint, $\xi_i \geq 0$, hence the decision function of the primal optimization problem is represented as equation (5).

$$\arg \max_{m = 1, \ldots, k} (w_m^T \phi(x_i)) \quad (5)$$

The dual problem of equation (3) involves a vector α having dual variables $\alpha_i^m \forall_m, i$. the w_m get defined via α as shown in equation (6).

$$w_m(\alpha) = \sum_i \alpha_i^m x_i \forall_m \quad (6)$$

for:

$$\begin{cases} C_i^m \leq 0 & y_i \neq m \\ C_i^m \leq 0 & y_i \neq m \end{cases}$$

The dual problem is expressed in equation (7)

$$\min_{\alpha} \quad f(\alpha) = \frac{1}{2}\sum_m \|w_m(\alpha)\|^2 + \sum_i \sum_m e_i^m \alpha_i^m \quad (7)$$

$$subject\ to \quad \left(\alpha_i^m \leq C_i^m \forall_m, \quad \sum_m \alpha_i^m = 0\right) \forall_i$$

The gradient of f is given in equation (8)

$$g_i^m = \frac{\partial f(\alpha)}{\partial \alpha_i^m} = w_m(\alpha)^T x_i + e_i^m \quad \forall i, m \quad (8)$$

Optimality of α can be checked using the quantity expressed in equation (9)

$$v_i = \max_m g_i^m - \min_{m:\alpha_i^m < C_i^m} g_i^m \forall i \quad (9)$$

where dual optimality holds when $v_i = 0 \forall i$

For a given i, the values of m that attain maximum and minimum values in equation (9) are expressed in equation (10).

$$M_i = \arg\max_m g_i^m \quad and \quad m_i = \arg\max_{m:\alpha_i^m < C_i^m} g_i^m \forall i \quad (10)$$

The Gaussian BRF kernel was also employed alongside with CS algorithm in the training process to implicitly transform input space (training dataset) into a linear separable feature space where linear classification are applicable, known as kernel trick. A kernel function K effectively computes the dot products in a higher-dimensional space represented by $\mathbb{R}^M$ while remaining in $\mathbb{R}^N$. For $\vec{x_i}, \vec{x_j} \in \mathbb{R}^N, K(\vec{x_i}, \vec{x_j}) = \langle \phi(\vec{x_i}), \phi(\vec{x_j}) \rangle M$, where $\langle .,. \rangle M$ is an inner product of $\mathbb{R}^M, M > N$ and $\phi(\vec{x})$ transforms $\vec{x}$ to $\mathbb{R}^M (\phi: \mathbb{R}^N \rightarrow \mathbb{R}^M)$. Hence, the Gaussian RBF kernel is mathematically expressed as equation (11).

$$K(\vec{x_i}, \vec{x_j}) = exp^{-\frac{1}{2\alpha^2}\|\vec{x_i} - \vec{x_j}\|^2} \quad (11)$$

where $\alpha = \sqrt{5 * dim(x_i)}$ and *dim* denotes dimension of the training dataset.

4. EXPERIMENTAL RESULTS

The performances of kNN and SVM methods in terms of student performance prediction are shown in Figures 2-4. Meanwhile, values in the diagonal of each confusion matrix in figure 2-4 represent the prediction of student performance across various classes as defined in Table 2.

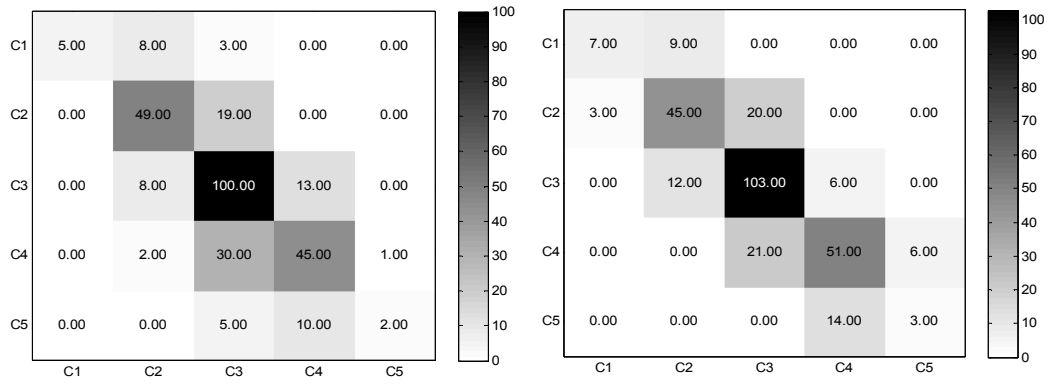
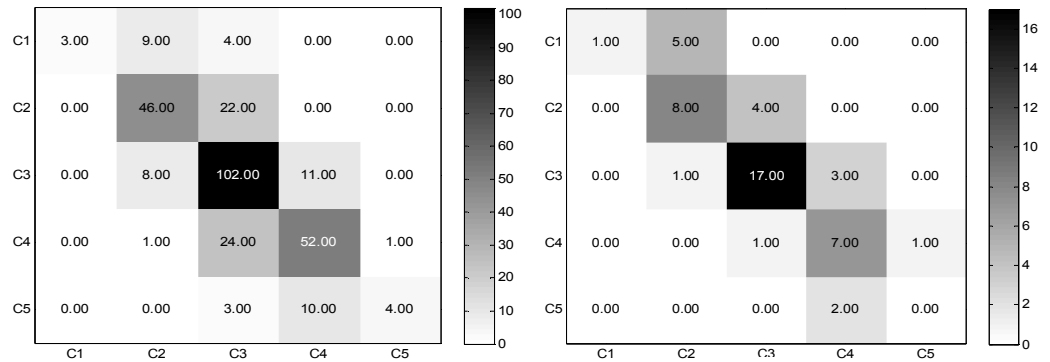
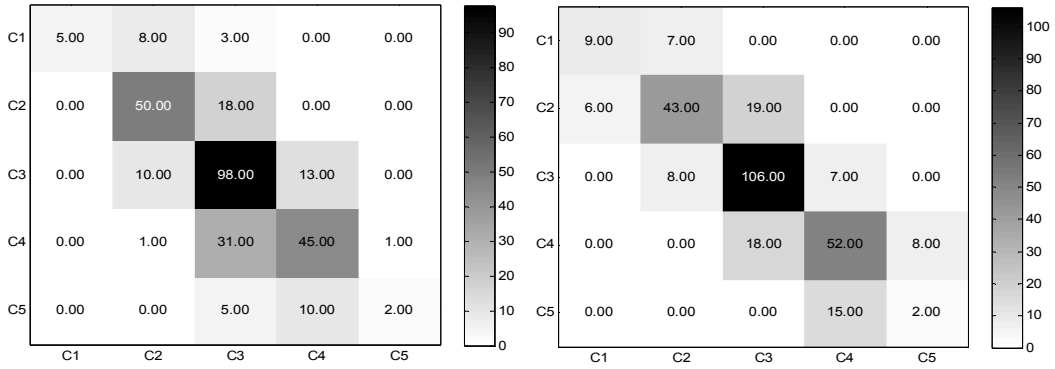


Figure 2: Prediction accuracy of students' performance; a) kNN 5-fold validation, b) SVM 5-fold validation



a) b)

Figure 3: Prediction accuracy of students' performance; a) kNN 7-fold validation, b) SVM 7-fold validation



a) b)

Figure 4: Prediction accuracy of students' performance; a) kNN 10-fold validation, b) SVM 10-fold validation

Table 2: Classification of student's output performance

S/N	GPA Range	Class Categories	Code
1	4.50 -	First class	C1
2	3.50 -	Second class	C2
3	2.40 -	Second class lower	C3
4	1.50 -	Third class	C4
5	0.00 -	Pass	C5

The average prediction accuracy for 5-fold, 7-fold, and 10-fold cross validation methods across all the class categories for kNN and SVM predictors is shown in Table 3.

Table 3: Average prediction accuracy of kNN and SVM models across all classes

S/N	Prediction	Accuracy
1	kNN-5-fold	67.00%
	SVM-5-fold	70.69%
2	kNN-7-fold	**69.00%**
	SVM-7-fold	**72.20%**
3	kNN-10-fold	67.00%
	SVM-10-fold	69.67%

Going by the results shown in Table 3, for the SVM models, it can be observed that the best performance prediction (72.20%) was archived with 7-fold cross validation SVM predictor while 70.69% accuracy was recorded for the 5-fold SVM model which is the next highest. Also, considering the kNN models across the three configurations, it could be seen that the best prediction accuracy (69.00%) was achieved when 5-fold cross validation technique was used while a lower accuracy (67.00%) was recorded for both 7 and 10-fold cross validation techniques. In general, the best prediction accuracy for both methods (kNN and SVM) was achieved at 7-fold cross (67.00% and 72.20 respectively). Meanwhile, SVM seem to perform better that kNN in terms of student performance prediction.

5. CONCLUSION

As found applicable in different areas, application of data mining can be very useful in improving the quality of education in the developing nations. A number of data mining techniques have been proposed for predicting likelihood and success rate of students in institutions of higher learning. In this study, we compared the prediction result of SVM and kNN classifiers since they have both received low attention compared to other classical prediction algorithms. These algorithms are better choices from linear and non-linear classification techniques respectively.

Experimental results obtained show that predictions based on SVM classification are more accurate compared to kNN-based predictions. This might be due to the fact that SVM makes use of hyper-plane that with margin wide enough to separate data points unlike kNN that approximates underlying distribution of the data in a non-parametric fashion. Also prediction obtained at 7-fold cross validation has highest accuracies for both classification techniques when compared with 5-fold and 10-fold cross validation. Inaccuracies in kNN can be due to outliers as the technique considers outlying data during classification. In contrast, SVM deals directly with data points around its hyper plane by using only the most relevant points to find a linear separation.

Despite the glories of both techniques, a major challenge with kNN is selection of value for k. Several experiments were carried based on trial and error before an optimal result was gotten at k=7. Also, deciding values for kernel function parameters sigma and epsilonin SVM took some time. An important practical question is proposing methods to solve these.

6. ACKNOWLEDGEMENTS

The authors would like to thank the management of University of Lagos, Nigeria, for providing us with the dataset as well as other materials that led to the realization of the project.

7. REFERENCES

Ahmed A and Elaraby I. (2014). Data Mining: A prediction for Student's Performance

Using Classification Method. *World Journal of Computer Application and Technology*, Vol. 2, No. 2, pp. 43-47.

Alexander, H., (1996). Physiotherapy Student Clinical Education: The Influence of Subjective Judgments on Observational Assessment.*Assessment and Evaluation in Higher Education*, Vol. 21 No. 4, pp. 357-366.

Althaf H. B. et al, (2012). Predicting Student Academic Performance Using Temporal Association Mining. *International Journal of Information Science and Education*, Vol. 2, No. 1, pp. 21-41.

Asogbon M.G. et al., (2016). A Multi-class Support Vector Machine Approach for Students Academic Performance Prediction. *International Journal of Multidisciplinary and Current Research*, Vol. 4, pp. 210 – 215.

Crammer K. & Singer Y. (2002). On the learnability and design of output codes for multiclass problems. *Machine Learning*, Vol. 47, No. 2, pp. 201–233.

Ekundayo M. S. and Ekundayo J. M., (2009).Capacity constraints in developing countries: A need for more e-learning space? The case of Nigeria.*Proceedings of Australasian Society for Computers in Learning in Tertiary Education Auckland.*Auckland, New Zealand December 6-9, pp. 243-255.

GalitB. Z. et al, (2007). Examining online learning processes based on logfiles analysis: a case study.*Research, Reflection and Innovations in Integrating ICT in Education*, pp. 55-59.

Gibson A. et al, (2012). An Inquiry into Relationships between Demographic Factors and Teaching, Social, and Cognitive Presence.*Internet Learning*, Vol. 1, No. 1, pp.7-17.

Han J. et al, (2012). *Data mining concept and Techniques*. Third Edition, Morgan Kaufmann Publishers, Elsevier inc., 225 Wyman Street, Waltham, MA 02451, USA, pp. 285-350.

Hatzilygeroudis I. et al, (2004). PASS: An Expert System with Certainty Factors for Predicting Student Success.M.Gh. Negoita et al. (Eds.): KES 2004, LNAI 3213, pp. 292–298.

Kovačić Z., (2010). Early Prediction of Student Success: Mining Students Enrolment Data.*Proceedings of Informing Science and IT Education Conference*.Cassino, Italy, June 21-24, pp. 647-665.

Lori B. et al, (2004). Student Traits and Attributes Contributing to Success in Online Courses: Evaluation of University Online Courses.*Journal of Interactive Online Learning*, Vol. 2, No. 3, pp. 1-17.

Muqasqas S. A. et al, (2014). A Hybrid Classification Approach Based on Decision Tree and Naïve Bays Methods, *International Journal of Information Retrieval Research*, Vol. 4, No. 4, pp. 61-72.

Ojerinde D. and KoloT. N., (2007). Influence of some Variables on the Degree of Prediction of First Year Grade Point Average (FGPA) by Universities Matriculation Examination (UME) Scores. *Journal of the Association for Educational Assessment in Africa*, Vol. 4, pp. 191-206.

Oladokun V. et al, (2008). Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course.*The Pacific Journal of Science and Technology*,Vol. 9, No. 1, pp. 72-79.

Omisore O.M. and Azeez N.A., (2015). Predicting Academic Performance of Students in Higher Institutions with k-NN Classifier. *International Conference on Computer Science Research and Innovations*. University of Ibadan, Oyo State, Nigeria, August 20-22.

Pandey U. and Pal S. (2011). Data Mining: A Prediction of Performer or Underperformer using Classification, *International Journal of Computer Science and Information Technology*, Vol. 2, No. 2, pp. 686-690.

Porchea, S. F. et al, (2010). Predictors of long-term enrolment and degree outcomes for community college students: Integrating academic, psychosocial, socio-demographic, and situational factors. *The Journal of Higher Education*, Vol. 81, pp. 750-778.

Romero C. and Ventura S., (2007). Educational data mining: A survey from

1995 to 2005.*Expert Systems with Applications*, Vol 33, pp 135-146.

Sajadin S. et al, (2011). Prediction of Student Academic Performance By An Application Of Data Mining Techniques. *International Conference on Management and Artificial Intelligence*, IACSIT Press, Bali, Indonesia, pp. 110-114.

Stamos T. and Andreas V.,(2008). Artificial Neural Network for Predicting Student Graduation Outcome.*Proceedings of the World Congress on Engineering and Computer Science*. San Francisco, USA.

Symeon N. et al, (2010). Incremental Training of Multiclass Support Vector Machines. *International Conference on Pattern Recognition*. pp 4267-4270.

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

DEVELOPMENT OF AN IMPROVED CAMPUS WIDE DATACENTER NETWORK FOR NIGERIAN UNIVERSITIES

E. O Nonum
Dept. of Computer Science/Mathematics, Novena University Ogume, Delta State, Nigeria
oyibow@yahoo.com

P. O Otasowie
Dept. of Electrical Electronic Engineering, University of Benin, Edo State, Nigeria.
potasowie@uniben.edu.ng

K.C. Okafor
Dept. of Electrical Electronic Engineering, Federal University of Technology, Owerri, Nigeria
kennedy.okafor@futo.edu.ng

ABSTRACT

The use of Information Technology for educational processes in the 4th industrial revolution currently places demand on security, speed performance, and effective service delivery. Most Campus Wide Data Center Networks (CW-DCNs) suffer from the above perspectives. Cases of traffic congestions caused by unbalanced traffic distributions and security vulnerabilities are obvious. Nevertheless, it is difficult to address these challenges with the existent Ethernet-based (IEEE 802.11) DCN architecture. This paper proposes services convergence with CW-DCN as a viable alternative with a high potential to tackle the above problem dimensions due to its flexibility, scalability and reliability. The work presents a WiMax based Data Center Network (WiMax-DCN) that offers secure and optimal service delivery in Nigerian Universities. More specifically, the design focused on a comprehensive secured server scheduling mechanism using IBM Flex manager and Cisco Stateful Packet Inspection Firewall (CSPIF). In the design, two different optimization objectives were considered, with one targeting the unbalanced traffic distribution in legacy WLANs and one maximizing the total network Quality of Service (QoS) under the constraints of limited wireless resources. Discrete event multithreading Riverbed Modeller 17.5 was used as a heuristic tool for the two problems. This was used to carry out extensive simulation study for validating the system architecture. The results demonstrate that WiMax-DCN offers improved performance over WLAN based DCN. Consequently, this work therefore recommends WiMax-DCN (with supports for backward compatibility) to Nigerian tertiary institutions for enhanced security, cost effectiveness, and optimal service delivery.

Keywords: Campus Networks, Services Convergence, Security Vulnerabilities, QoS, WiMax-DCN, Riverbed Modeller

7. INTRODUCTION

1.1 Background Study

In Nigeria today, about 97% Campus Networks (CNs) uses hotspot solution based on IEEE 802.11 Ethernet technologies, (i.e. Wifi802.11a/b/g/n), White Paper- Security, (2011). This technology is a very popular. Unfortunately, it has generated a kind of distrust regarding security and QoS in mission critical environments. When either confidentiality, authentication (open system and shared key authentication), and integrity is compromised, this could create a terrible havoc in the campus network architecture.

Campus Wide Network (CWN) is a cloud computing based network that uses WiMax Infrastructure, firewall gateway and virtualized server compute clusters to provision services to end users or university community. This is quiet similar to Smart Green Energy Management System Network (SGEMSN) discussed by K.C Okafor, et al. (2016a). Security and QoS metrics are essentially important in these types of networks. Infact, these are particularly critical in large scale wireless computing.

In the conventional WLANs, the Service Set Identifiers (SSIDs) is the network name used by the mobile users to gain access to Access Point (AP) which is attached to the network. Hence, authentication is accomplished using SSID. In this case, the mobile user sends out a probe frame (active scanning) with desired SSID. The AP send back a probe response frame and client node finally accept the SSID, Timing Sync Function (TSF), timer and PHY setup values from the AP frame. This is basically a weak security framework because the SSID may be sent by the AP in its broadcasted beacon frames.

Therefore, IEEE 802.11 defines authentication and encryption services based on the Wired Equivalent Privacy (WEP) algorithm. i.e. 40-bit encryption key. Temporary Key integrity protocol (TKIP) and WAP2 are inclusive in IEEE 802.11 security framework. Another technique of authentication in campus WLAN is via user device MAC address Look-Up Table in the AP. Besides, an Access Control (AC) list is kept in each AP with centralized database records of users in RADIUS server. However, MAC address based authentication offers a similar weak security because the equipment can be stolen.

Despite the above security mechanisms, the following security issues exist with IEEE 802.11 campus networks (Young, 2015):

i. No per-packet authentication
ii. Vulnerability to disassociation attacks
iii. No user identification and authentication
iv. No central authentication, authorization, and accounting support
v. RC4 stream cipher is vulnerable to known plaintext attacks
vi. Some implementations derive WEP keys from passwords
vii. No support for extended authentication; for example: token cards; certificates/smartcards;
One-time passwords; biometrics; etc.
viii. Other key management issues are re-keying of global keys, absence of dynamic, per-STA unicast session.

Future of campus WLAN will be based on Advanced Encryption Standard (AES), Extensible Authentication Protocol (EAP), authentication frameworks to support multiple authentication types, and support for inter-network roaming.
However, these does not still satisfy the requirement for a comprehensive enterprise campus network. Also, their datacenter integration lacks resilience, adequate security and QoS performance Rajiv, and Surya, (2013), Udeze, et al. (2014), Ugwoke et al. (2015).

Figures 1, 2, 3a,b shows the Network Operating Center (NOC) of the existing campus networks which have the above issues. These campus networks have no support for on-demand services, robust security as well as tighter QoS integration.

Figure1: NAU DCN Switching System, 2015

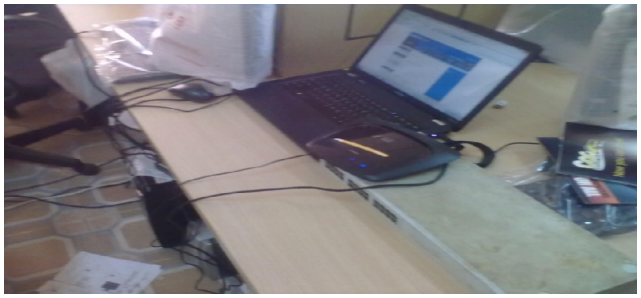


Figure 2: ELDI Hotspot NOC, 2015

Figure 3a: UNN Hotspot DCN (Source: UNN DCN, 2015)

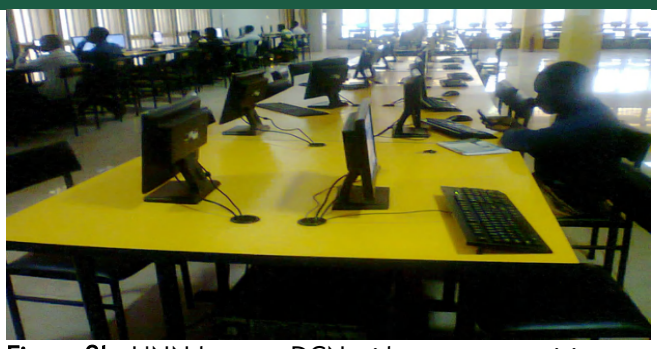


Figure 3b: UNN hotspot DCN with user connectivity, (Source: UNN DCN, 2015)

However, the use of WiMax technology defines a security sub-layer specifically dedicated to provide privacy, confidentiality and authentication to the final users on its protocol-stack. WiMax security system is based on the principles of authentication and encryption, which makes it more secure than the legacy WLAN networks.

In the proposed WiMax based CWN, after the base station (BS) or Radio Network Controller (RNC) authorizes the network user; additional encryption mechanisms are used to maintain data confidentiality and integrity.
For this purpose, the user device sends to the BS a request for encryption keys called Traffic Encryption Keys (TEKs) in a response message. These messages are immediately encrypted with a key that is only known by both parts. The algorithm used for encrypting the TEKs could be Triple Data Encryption Standard (3DES), AES, and WPA-2, White Paper Security, (2011). Once the TEKs are known, AES is then used for data encryption.
Some of the advantages of the encryption mechanism implemented by WiMax which is not found in WLAN IEEE 802.11are:

- Use of very robust algorithms for scalability.
- Support of dynamic keys generation with a variable Time to Live (TTL) session.
- Independent encryption for each service flow allowed.

Interestingly, educational institutions as well as mission critical organizations can use the WiMax based converged infrastructure to centralize the management of their IT resources, so as to consolidate their systems, increase resource-utilization rates at lower costs. This converged

infrastructure can foster these objectives by implementing pools of computers, storage and networking resources that can be shared by multiple applications and managed in a collective manner using policy-driven processes.

1.2 Research Objectives And Contributions

The objective of this work is to establish a well secured and scalable network, that will guarantee confidentially and QoS in WiMax campus networks. In designing the WiMax CWN, the main goals are to maintain a robust security, throughput, low latency, fast convergence and scalable network that are flexible with less administrative overhead. The network must be capable of supporting high performance workloads such as user video, voice and data traffic besides conventional Ethernet traffic like web applications. In this research, the following contributions were made:

i. Development of a secured re-engineered model for CWN based on WiMax RNC, and virtualized server cluster backend.
ii. Development of discrete event process algorithms for traffic flow paradigms in the CWN cloud datacenter while making a QoS metrics comparison with Non-virtualized WLAN network. This comparative evaluation shows the relevance of the CWN architecture.
iii. Validation analysis using Riverbed Modeller for QoS metrics such as throughput and latency for optimal quality of service performances (for end user web transactions).

The rest of the paper is organized as follows. Section 2 discussed related research efforts with limitations of traditional WLAN networks. Section 3 presents the CWN system description as well as the functional operation. Section 4 presents the system deployment and the performance evaluations. Section 5 concludes the work with recommendations and future directions.

8. LITERATURE REVIEW

Several research efforts have been made in the domain of datacenter networks used for critical business services. These will be highlighted in this section. Also, security, and QoS contributions in WLAN and WiMax networks will be briefly reviewed.

C.C.Udeze, et.al (2014) proposed Reengineered DataCenter (R-DCN) architecture for efficient web application integration. In the work, an attempt was made to address network scalability, efficient and distributed routing, packets traffic control, and network security using analytical formulations and behavioural algorithms in some selected architectures. Though their work was not specific on the type network used for the R-DCN, critical evaluations were made on the basis of QoS metrics to justify their design.

K.C.Okafor et.al. (2016a) developed cloud network management architecture for grid performance using procedural benchmarking on BCube and DCell architectures. The network sought to house a grid application using task scheduling and resource allocation algorithms. The tradeoffs of these schemes were not discussed.

Yong, et al. (2011) presented a Datacenter network that uses Wireless Link (WLS) to provide a feasible wireless DCN. The work considered two wireless scheduling optimization objectives in the DCN viz: one targeting the unbalanced traffic distribution and one maximizing the total network utility, under the constraints of limited wireless resources and co-channel interference. The optimization problem was resolve and simulated demonstrating the effectiveness of their proposal.

Udeze, et.al. (2014) provided a comprehensive discussion on various DCN architectures, their merits and demerits but not in the context of WiMax DCN.

Similarly, C.Guo et al. (2009) proposed BCube, as a new network architecture specifically designed for shipping-container based, modular datacenters. At the core of the BCube architecture is its server-centric network structure, where servers with multiple network ports connect to multiple layers of COTS (commodity of-the-shelf) mini- switches. Servers act as not only end hosts, but also relay

nodes for each other. The work developed path adaptation algorithm using its path selection procedure.

Faisal A. et al (2010) introduced Campus as a framework that supports the development of multi-agent low cost security system driven by wireless sensor network. In the paper, it was opined that the challenge for pervasive computing is the seamless integration of computer support with users' activities in a very dynamic setting, with deep human and resource mobility. In their work, campus is designed to provide the necessary infrastructure for ambience intelligence applications.

Arinze, et al. (2014), presented cognitive mini-pop Access Point (AP) architecture for high density wireless environments. The architecture was derived from the Tranzio-Wavion radio infrastructure model is based on quality of service optimization. MATLAB Simulink 2009b was used to configure parameter settings and values for the physical layer of the cognitive minipop_AP communication network based on 802.11 IEEE standards. The results of model physical layer provide an economical and flexible solution that encourages efficient network resource utilizations.

Okafor, et al. (2013) presented a cost effective wireless hotspot model called iWLAN for Electronics Development Institute (ELDI), Awka, Nigeria, that supports multiple user sessions simultaneously. Their research focus was on SMART intranet traffic engineering and QoS guarantees in the iWLAN model. With the former, the work aimed at distributing the bandwidth in the iWLAN according to available throughput allocation criterion. With the latter, the objective was to ensure that the performance metrics (throughput and delay) experienced by user allows for flexible data communication within the high density zones.

Shuang, and Issac (2014) discussed and made comparison between WiMax and WLAN technologies while carrying out a wireless network coexistence deployment with OPNET 9.1.

From security perspective, Sanjay and Nicole (2010) presented an assessment of WiMax security while enumerating the threats to such networks, viz: Rogue base stations, DoS attacks, Man-in-the-middle attacks, and Network manipulation with spoofed management frames. The work discussed the protocol stack of WiMax in a comprehensive manner. Several differences were highlighted between Wifi and WiMax instance, the WiMax MAC layer uses a scheduling algorithm while contention access used in the Wifi MAC layer. Other areas of marked differences include: authentication, authorization, encryption and availability.

Similarly, K.C.Okafor, et al. (2016b) presented Vulnerability Bandwidth Depletion DDoS Attack (VBDDA) model with Cisco Nexus 9000 firewall as an embedded network device with support for Virtual DDoS protection as a threat mitigation procedure. Also, the work investigated on security QoS profiling for DDoS traffic flows using advanced Cisco technologies.

Also, M.Alzaabi et al. (2013) carried out an investigation into a) the weakness and threats on WiMax security algorithms and b) the best method that could prevent DoS attacks prior to the authentication algorithm. The work also, presented the architecture of WiMax while identifying the main and sub layers that these security algorithms are performing their functions from within. A new method was incorporated with the authentication algorithm to improve the efficiency of the security of WiMax. Obviously, some of the efforts made in the context of IEEE 802.11 have deficiencies which are clearly highlighted below.

2.2 Limitation of legacy IEEE 802.11

The following are the observed research gaps from the previous works in literature, viz:

1. Robust security integration in a CWN from the perspective of the RNC and Stateful Packet firewalls has not been explored in WiMax-CWN.
2. A coverage constraint across geographically dispersed access points for data traffic is usually present. This now requires a leverage on extended service set infrastructure resulting in enormous cost investment in generic hotspot WLANs.
3. In generic 802.11 WLAN hotspot, traffic does not flow across the core unless the switching system must control traffic between localized access points and flows within the same RF neighbourhood, making users to experience packet drops at busty traffic periods.

4. Policy enforcement for network performance is dependent on the hotspot infrastructure set ups which lacks comprehensive evaluations on QoS metrics and security vulnerabilities.
5. Most of the works have issues of poor QoS owing to high bit error rate at peak traffic periods with high network density.

9. SYSTEM DESCRIPTION

This work will adopt the WiMax IEEE 802.16e which operates on a carrier frequency below 11GHz, having frequency bands of 2.5GHz -5.7GHz for access connectivity as shown in Figure 4. The bandwidth range is 1.5MHz -2MHz, while the radio frequency is Orthogonal Frequency Division Multiple Access (OFDMA) with Quadrature Amplitude Modulation (QAM).

The system consists of the node transmitter (base station), channel, and RNC, firewall and cluster servers. The transmitter and receiver components consist of channel coding and modulation subsystems.

In the physical setup, channel coding transforms the baseband signals in order to improve communication performance by increasing the robustness against channel impairments such as noise, interference and fading.

The process involved in the CWN channel coding are data randomization, Forward Error Correction (FEC)/convolutional encoding and Interleaving.

At the physical layer, interleaving is done by spreading the coded signals in time before transmission. Incoming data into the RNC interleaver is randomized into permutations in order to avoid error bits. During the modulation phase, coded bits are mapped to the In-phase, and Quadrature (IQ) constellations. All the radio base stations communicate with the RNC. The duration of the simulation for the proposed CWN is 200secs. The transport layer forwards the traffic to the backend servers which run on virtualization.

3.1. CWN Environment for Series Convergence

Following the issues from the existing CWN, this research seek to develop a more robust, scalable and flexible CWN that will support services integration and application convergence while offering the best quality of service to end users. Figure 4 shows the contextual CWN based on WiMax infrastructure that could supports various CWN applications and services. As shown in Figure 4, the CWN as a converged infrastructure operates by grouping multiple information technology components into a single, optimized computing package.

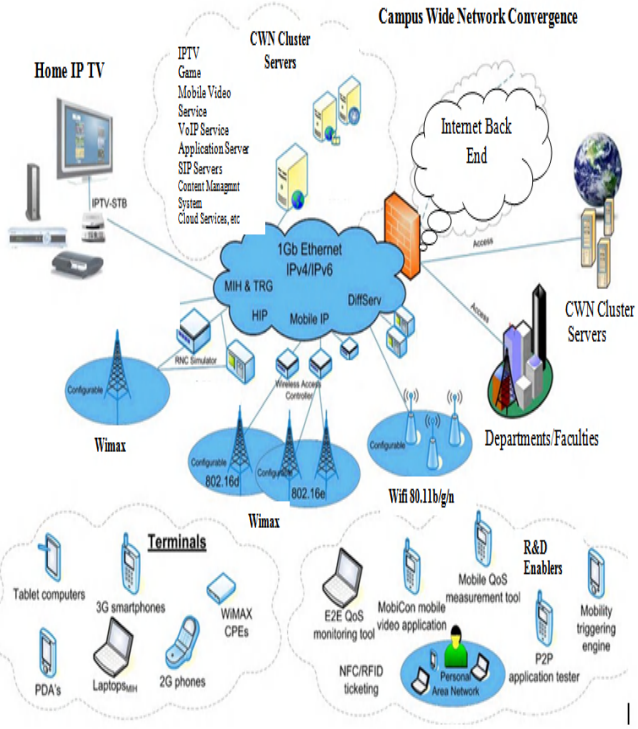


Figure 4: Proposed CWN Scenario for Services Convergence

The components of the converged CWN infrastructure in Figure 4 include: servers, data storage devices, networking equipment and software for IT infrastructure management, and automation. The core network supports IPTV Set Top Box (STB), diffserv, mobile IP, Host Internet Protocol (HIP), remote network connector, backend storage and servers, terminals and other converged services. These have demand for QoS performance and security at large.

For instance, the Internet Protocol television (IPTV) traffic could be delivered using the Internet protocol suite over a secured QoS based packet-switched

network instead of being delivered through traditional terrestrial, satellite signal, and cable television formats.

Unlike downloaded media, the CWN IPTV offers the ability to stream the media in smaller batches, directly from the source. In the system, the HIP runs on Internet Protocol (IP). i.e. cloud Internet. This has two main name spaces, IP addresses and the Domain Name System. HIP separates the end-point identifier and locator roles of IP addresses.

The HIP provides secure methods for IP multihoming and mobile computing for CWN. The effect of eliminating IP addresses in application and transport layers is a decoupling of the transport layer from the internetworking layer.

3.2. Proposed Reengineered CWN

The CWN from WiMax perspective consists of three main sections, User Terminals, Access Service Network and Connectivity Service Network discussed below.The block diagram model of the converged CWN is shown in Figure 5. This shows the following block components viz: User Terminal Units (UTU), Access Virtualization layer (AVL), Core speed redundancy layer (CSRL). They are described below.

i. User Terminal Units (UTU)

This is the termination point of the CWN where users with their client machines (workstations, PDAs, PCs, Laptops) can gain access to the network. Owing to virtual machine logic instantiation in the enterprise server, upon authorisation users can make connection and access resources in the network. Beside the security configurations done at the other layers such as hybrid speed redundancy layer and access layer, high level security is implemented at the user domain since it supports extensive platform security.

ii. Access Virtualization Layer Block

This layer runs on an enterprise server architecture having two layers of caching services, one for virtual machines (VM) outer loop and the other for the servers attached to the VMs and its related applications. This layer is designed to have resilience, scalability, robustness owing to its hardware configurations. Also, beside the caching services, bandwidth optimisation is realised in this layer. Media Access Control (MAC) controllers and servers (database, application and web servers) are the devices located in this layer. This layer allows for terminal connectivity with data center network (CWN) switch.

The user domain devices (PCs, PDAs, Laptops, iPAD and mobile phones) only need to have a compatible remote desktop protocol (RDP), Media Access Control (MAC) which is implemented by MAC controller for connectivity to a terminal CWN switch. The MAC controller also implements Carrier Sense Multiple Access and Traffic Arbitration Messaging Protocol (CSMA-TAMA) which makes for an efficient and flexible data throughput while optimizing bandwidth in the proposed reengineered CWN architecture.

By connecting to a terminal RNC, Virtual Local Area Network (VLAN) sessions are created which runs and controls user access.

Programs like anti-spyware, anti-viruses could run both on the client machines and high-powered computing virtual machine engine server (VM-server) to protect user data. User data is always stored in a centralized location server other than the user work stations thereby simplifying data and recovery processes. The anti-spyware and the anti-viruses are updated from the virtual server ensuring that all the definitions are up to date. Applications are also updated centrally from the virtual server.

At the server backend, the VM-server is a high computing scheme deployed to provide remote accessibility and efficient service provisioning to all the physical machines in the network. Through the concept of virtualization, various server instances are created, running different services. This virtualized server provides cheap client hardware and having the users connected to it to run more powerful applications such as word processors, internet browsers, e-mail clients. In this work, various server instances were created in the VM box.

iii. Core Speed Redundancy Layer Block

Hybrid speed redundancy layer block is modelled to completely replace the core and distribution layer of the traditional data center network. In other words, a two-layer data center model was adopted owing to its advantages over the three-layer model.

In this work, the high speed switching will be achieved by the use of RNC multi-protocol label

switch (MLS) with VLAN segmentation for efficient packet delivery from sources to destinations and vice versa in this layer. Apart from its dual routing and switching functions at a very high speed; the RNC has in-built VLAN capability for VLAN segmentation of data center networks.

At the core, the server keeps specific user account of the client machines and gives them the ability to access any virtual application or server for the organisation. Audit logs, traces and transaction times are kept by the server. The server has security layer setting used to configure it for authentication which is how it verifies client identity. This allows a secured socket layer (SSL) certificate which enables the client machines to connect. Using this method reduces the risk associated with authentication credential from suspicious users and networks. The encryption on the server has different levels used to protect the content of the server from interception from suspicious parties. On the server, group policies are configured by default.

Some of these policies include connection policy, device and resource redirection policy, licensing policy, profiles policy, remote session environment policy, security policy, session time limit policy. On the virtualized server, the application services runs on the browsers which is modelled as a Transmission Control Protocol (TCP) http service from the application supported services of the server. Also, the Gigabit Ethernet (GE) is the link interface that connects the access/virtualization layer and the hybrid speed redundancy layer. This maintains the Fibre Channel protocol over 40Gb CWN server cluster.

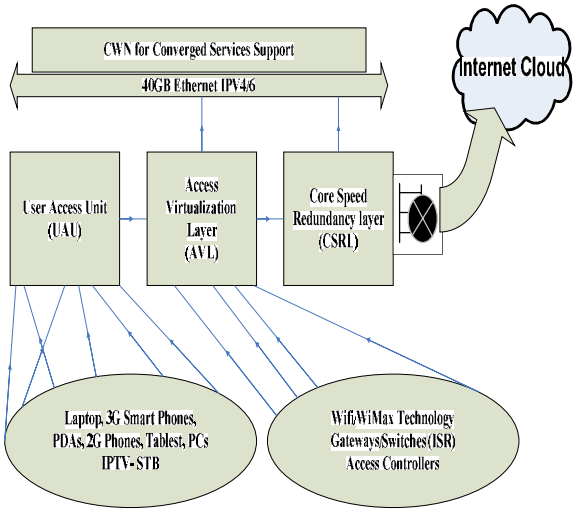


Figure 5: Block Diagram of CWN for Converged Services

10. SYSTEM IMPLEMENTATION

The CWN validation was carried out using multithreaded Riverbed Modeller 17.5. The work adapted similar design philosophy from existing CN designs (such as UNN, UNIZIK, and ELDI). The various objects were generated from the Object palette library of Riverbed Modeller. The entities and their attributes were fully configured as discussed in section 4.1. The developed CWN was analysed with a simple topology that included a mixture of WiMax LAN components, the RNC, the firewall (CSPIF) and the IBM server clusters.

Figure 6 shows the design implementation with WiMax technology as the backbone of the CWN hotspot. From the subnet site, the WiMax nodes (clients) are connected to the network servers via the WiMax RNC connections. The hardware multiplexing with virtualization scheme was included in the server cluster interconnection layouts. The CWN algorithms for efficiency were deployed in a validation scenario in Figure 6. The settings for evaluating the algorithm were enabled. In the validation, the WiMax nodes emulated the MAC layer protocol while the traffic to the RNC servers was properly harnessed.

4.1. Generation of Datasets/Validation

In the design, trace-driven simulation with genuine traffic trace files was collected from the WiMax DCN to evaluate the performance of the system. The procedure on how the validation was carried out and the outlined processes used in generating the graphical plots are detailed below

i. The heuristic Riverbed tool was launched with scenarios created for event script trace files.

ii. The validation trace file is created and the trace events (discrete events) are configured to remain compatible with C based environments like MATLAB, MIDE, Catastelia, OMNET++, NS2-PT++ and stored in the Opnet_model logs.

iii. Next, the object palette tool was used to generate the WiMax block sets as well as the RNC, firewalls (Cisco Stateful Packet Inspection Firewall (CSPIF), IBM flex routine (IBM Flex manager). This was imported into the work area environment.

iv. The design involving the use of application configuration tabs, objects and tools are used to configure all the parameters.

v. The workflow is then enabled i.e. characterizing traffic mix in the trace file event (importing the algorithms).

vi. The traffic mix is configured which involves the background traffic flows based on the configured scenario algorithms of Proposed CWN.

vii. The discrete event parameters for WLAN and WiMax architectures were configured while enabling the event statistics, animation and results window. Finally, the discrete event simulation was run and results for analysis collected comprehensively.

In the design, traffic was measured on the client and server sides. As shown in Figure 6, the RNC is a governing element in the CWN (radio access network) and is responsible for controlling the Node *Bs* that is connected to it.

The RNC carries out radio resource management, some of the mobility management functions and is the point where encryption is done before user data is sent to and from the mobile. The RNC connects to the Circuit Switched Core Network through its internal logic and then connect to Media firewall Gateway (MFGW). This also connects the hosts servers in the network.

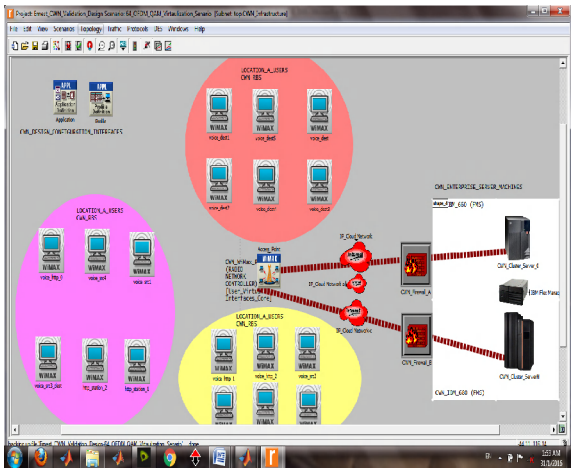


Figure 6: CWN validation system with Cisco Stateful Packet Inspection Firewall (CSPIF)

Figure 7 shows the CWN RNC with virtual interfaces. In this case, the VLAN procedure, and traffic policy configurations were enforced. The RNC makes the CWN a virtual cell wireless infrastructure in which all devices are on a single channel and are subject to any single controller. So instead of having APs on different channels, all APs are on a single channel.

It is opined that what is critical for CWN is not only the achievable performance, but consistency of performance. With virtualized server clusters, all of the network zones are inter-connected using the dedicated, low latency, high throughput global backbone network that is owned and operated by the institution as shown in Figure 6.`

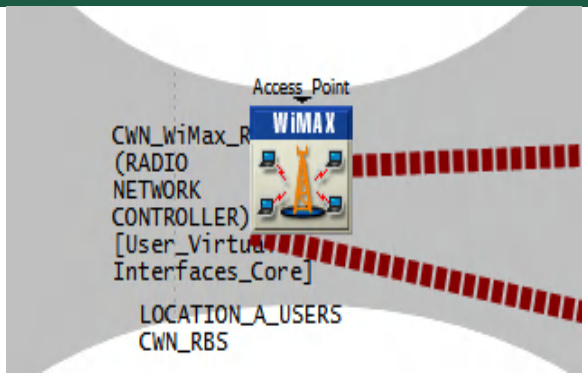


Figure 7: CWN RNC with virtual interfaces

Figure 8 shows the CWN WiMax users with voice and http traffic. The logical connections between the network elements are known as interfaces. The interface between the RNC and the Circuit Switched Core Network (CS-CN) is called *Iu-CS*. The interface between the RNC and the Packet Switched Core Network is called *Iu-PS*. Other interfaces include *Iub* (between the RNC and the Node B) and *Iur* (between RNCs in the same network). *Iu* interfaces carry user traffic (such as voice or data) as well as control information and *Iur* interface is mainly needed for soft handovers involving more than one RNC. The adopted CWN server cluster in Figure 8 has support for the following, viz:

- Remote control of hardware and operating systems
- Web-based management with standard Web browsers (no other software is required)
- Scriptable command-line interface and text-based serial console redirect
- System-independent graphical console redirection.
- Remote back support for continuity and data recovery.

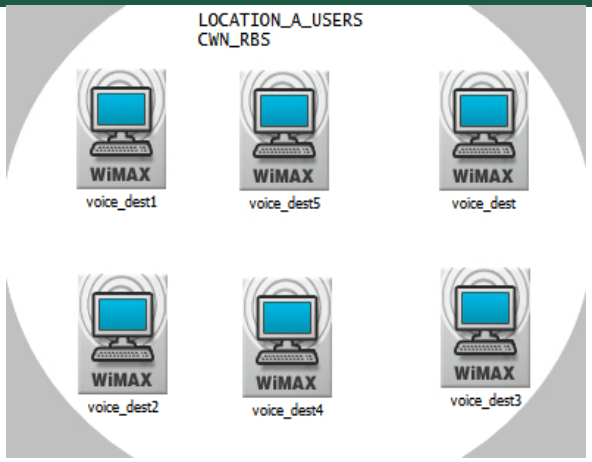


Figure 8: CWN WiMax users with Voice and Http traffic

Recall that the *Iub*, *Iu* and *Iur* protocols are used for carrying both user data and signalling (that is, control plane).

In general, the following formulations are valid from Figure 6, considering Node B.

- The signaling protocol responsible for the control of the Node B by the RNC ie. Node-B Application Part (NBAP). It is subdivided into Common and Dedicated NBAP (C-NBAP and D-NBAP), where Common NBAP controls overall Bode B functionality and Dedicated NBAP controls separate cells or sectors of the Node B. NBAP is carried over *Iub*. In order for NBAP to handle common and dedicated procedures, it is divided into: Node B Control Port (NCP) which handles common NBAP procedures and Communication Control Port (CCP) which handles dedicated NBAP procedures.
- Control plane protocol for the transport layer is called Access Link Control Application Protocol (ALCAP). Basic functionality of ALCAP is multiplexing of different users onto one RNC transmission path using channel IDs (CIDs). This ALCAP is carried over *Iub* and Iu-CS interfaces.
- Signaling protocol responsible for communication between RNC and the core network is called Radio Access Network

Application Part (RANAP), and is carried over *Iu* interface.
- Signalling protocol responsible for communications between RNCs is called (Radio Network Subsystem Application Part (RNSAP) and is carried on the *Iur* interface.

Figure 9 demonstrate the firewall with CWN Server Clusters running virtualization services. This is represents the CSPIF (firewall) whose internal specifications are as follows:

- 10 blade slots in 9U (racks)
- Shared Media tray with Optical Drive and USB 2.0 port.
- One (upgradeable to two) Advanced Management Modules
- Two (upgradeable to four) Power supplies
- Two redundant High-speed blowers
- Two slots for Gigabit Ethernet switches which can also have optical or copper pass-through)
- Two slots for optional switch or pass-through modules which can have additional Ethernet, Fibre Channel, InfiniBand or Myrinet 2000 functions.
- Four slots for optional high-speed switches or pass-through modules can have 40Gbit Ethernet or InfiniBand 4X.
- Optional Hard-wired serial port capability

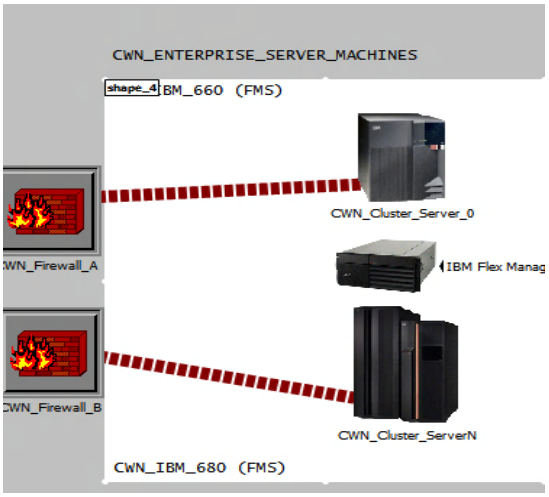


Figure 9: Firewall (CSPIF) with CWN Server Clusters running Virtualization services.

5.1. Performance Evaluations

The security integration of Figure 6 was successfully configured for deployment. In the analysis, server scheduling mechanism using IBM Flex Manager and Cisco Stateful Packet Inspection Firewall (CSPIF) were considered. From figure 6, two different optimization objectives were considered. First, unbalanced traffic distribution in legacy campus WLANs was observed. The QoS under the constraints of limited wireless resources in WiMax design was investigated. The performance both networks were considered using throughput and latency metrics

Now, throughput or bandwidth is an important measure of CWN performance considered in this work. It is the measure of the average amount of data that can be transferred through the network per amount of time. It is common for cloud providers to offer throughputs of around 300Mbps, and this may exceed the rate of data transfer required by the software application. But there are obviously applications where throughput will be a critical factor; anything involving video data, scientific data, data being streamed by Internet of Things (IOT) devices, or 'real time' big data systems such as the proposed CWN. Figure 10a shows the plot of virtualization throughput obtained from the secured backend. In the CWN design, the workload was run several times and the maximum throughput value observed and reported. the figure shows file workloads approaching a maximum rated throughput of 2500000bytes/sec (65.80%) initially. throughput for the bulk workload in CWN is much higher, exceeding throughput on the default workload of the traditional WLAN (Unizik network i.e. 34.21%) by 31.59 %. this observation underscores the fact that the pattern of the network traffic represents one of the most significant factor of network performance. in the bulk workload, the direction of traffic results in better and higher throughput performance after a period of 50secs.

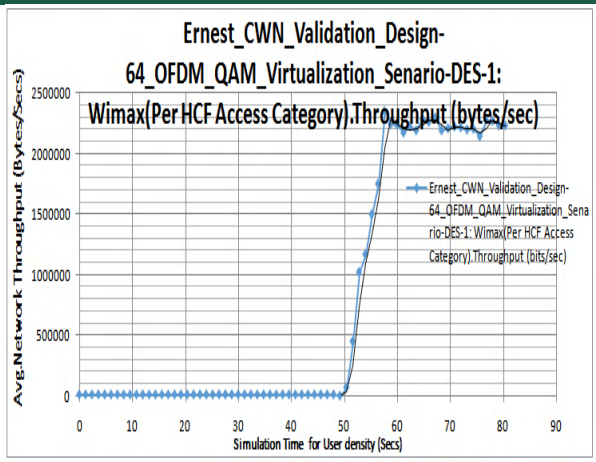


Figure 10a: CWN Virtualization Throughput Scenario

It was shown that the consolidation of multiple workloads onto fewer physical servers through the use of server virtualization concept can change the dynamics of traffic flow within the CWN generally. As observed in Unizik network, without server virtualization, their DCN administrators use less expensive, fixed-configuration switches close to the physical servers. These rack switches provide a couple of uplinks to the core network. This arrangement works because the majority of the servers connected to the rack switches have vastly underutilized network links, making it possible to aggregate all that traffic onto only a couple of uplinks to the core network while keeping the oversubscription ratio within generally accepted limits. In this case, traffic aggregation is occurs within the rack switch itself. This will lead to a potential disaster as the network grows in traffic size.

With server virtualization, traffic aggregation occurs at the physical server level. Besides, multiple workloads run on the same physical server and sharing the same network links. As a result of workload consolidation through the use of server virtualization, the network links to the physical servers are more heavily utilized. Using a rack switch to uplink to the core in an effort to aggregate traffic that has already been aggregated can result in bottlenecks and hindrances to network performance and throughput.

In the CWN server virtualization, the placement of workloads can change dynamically depending upon server utilization. Technologies such as live migration allow server administrators to move workloads to different physical servers easily and quickly. Some virtualization solutions even offer the ability to automate this process; VMware's Distributed Resource Scheduling (DRS). With workloads now moving freely across the server farm, CWN administrators no longer have the option of using locality. Where locality had been used to help minimize the number of uplinks to the core network, the number of uplinks may now need to be increased. Figure 10b illustrates the behavior of CWN with optimal virtualization throughput scenario. This is obtained when the workload dynamics changes.

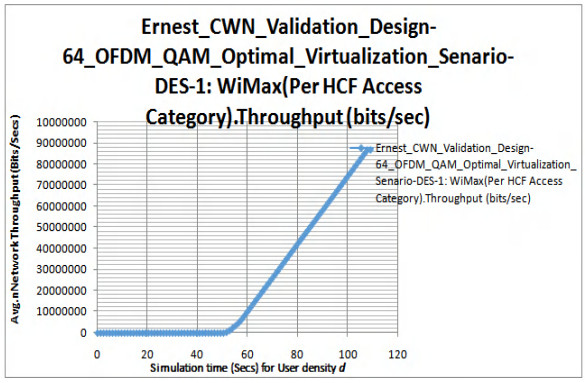


Figure 10b: CWN Optimal virtualization throughput Scenario

As part of the design requirement, the DCN layer performance was studied for both the proposed WLAN and WiMax CWN scenarios in Figure 11. In a production environment, CPU utilization plays a significant role in reaching acceptable network throughput. To process higher levels of throughput, more CPU resources are needed in a typical CWN. The effect of CPU resource availability on network throughput of virtualized applications is even more significant.

As shown in Figure 6, running IBM flex server series requires a certain amount of fixed CPU resources. This depends on the configuration of the server. In addition, because all the elements of the networking from physical to the application layer are virtualized, processing network transactions will be somewhat

more expensive for virtualized applications than for applications running directly on the physical platform in traditional WLAN.

Figure 11 illustrates corresponding CPU utilization for the throughput experiments of WLAN and WiMax shown. From the plot, it was observed that the proposed CWN offered a throughput of 54.05% while that of WLAN offered 45.95%. One of the great advantages of secured virtualization in CWN is its flexibility and the ability to consolidate multiple servers in virtual machines on a single physical platform.

In some cases, such virtual machines hosted on the same physical platform need to be networked together. Virtual infrastructure allows the transparent implementation of such a network configuration. When operating inside a virtual machine, the workloads use the same networking mechanism as they would use when running in physical machine using physical network cards (NICs).

In many cases, the maximum throughput that can be reached between virtual machines is comparable to throughput between a virtual machine and a physical server.

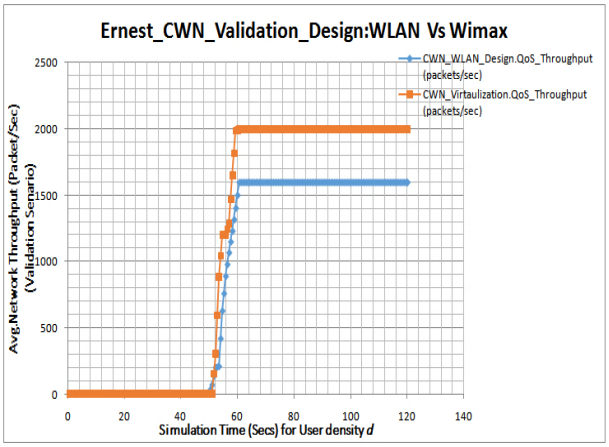


Figure 11: Throughput Validation for WLAN and WiMax (DCN layer)

Recall that the Radio Network Controller (RNC) is the governing element in the CWN radio access network (CRAAN) as shown in Figure 6. This is responsible for controlling the node *Bs* that is connected to it. The RNC carries out radio resource management, some of the mobility management functions and is the point where encryption is done before user data is sent to and from the mobile.

The RNC connects to its output Circuit Switched Core Network and which now links the Firewall Media Gateway (FMG) in the packet switched core DCN. Figure 12 shows the throughput validation for WLAN and WiMax (RNC layer).

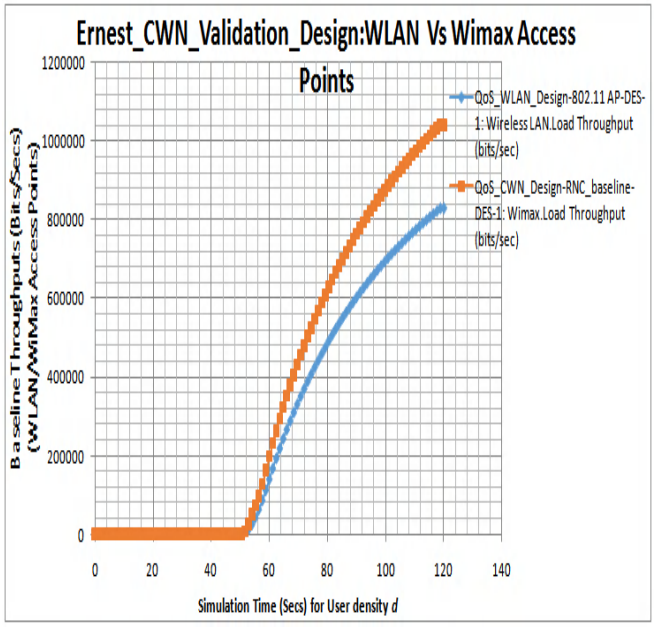


Figure 11: Throughput Validation for WLAN and WiMax (RNC layer)

Another interesting metric in the developed CWN DCN is latency. Generally, response time always has two parts viz: the time taken for the server to receive information, compute the response, and send out the response, and the time taken for data to pass through the network between the client computer and the server, and back again. The latter is the network latency, also known as Round Trip Time (RTT) latency.

While there are usually ways to reduce the latencies and improve response time inside the compute/data centre (such as use more CPUs, faster processors, faster RAM memory, or a faster storage medium, or improve the algorithm and/or software implementation), the network represents fixed performance of the cloud service which cannot be modified in a production scenario. Latency measured in context is the round trip time. Delays introduced by network hardware, delays due to error correction on data packets, all still contribute to this latency.

From Figure 12, the stable state network latency is about 0.0016secs which is comparably better performance conventional Unizik network in an ongoing study.

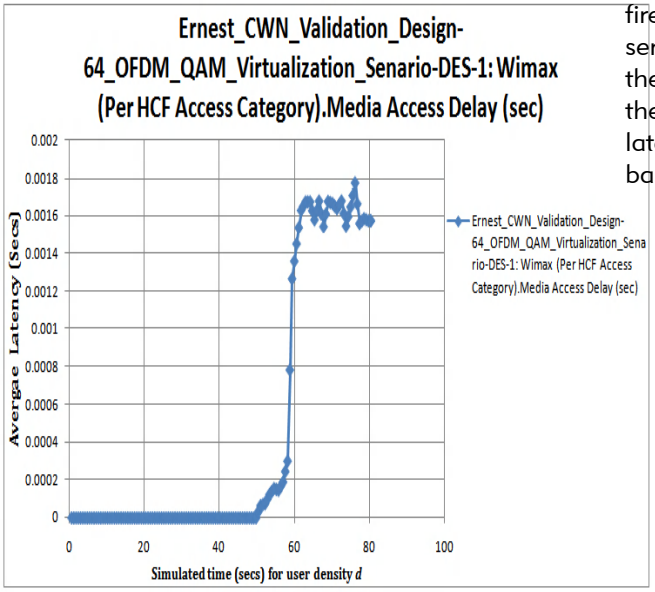


Figure 12: Throughput latency behavior

6. CONCLUSIONS

This research focused on developing a Campus Wide Network based on WiMax infrastructure which polls services from a virtualized cloud backend server clusters through a backend firewall. This pervasive network service seeks to execute innovative applications that can meet campus user needs. Based on the IEEE 802.16 standard, the WiMax wireless transmission of data offered a better platform for service provisioning. With WiMax system, it was observed that this can expand the wireless access coverage, improve the quality of service and the security deployments. The network could provide enough bandwidth and a wireless alternative for last mile broadband access. Considered as one of the next generation technologies, this ubiquitous wireless network of broadband access can allow more convenient and secure roaming services. As basis for this work, a description of the implementation of CWN for a production university network services was carried out. In context, Nnamdi Azikiwe University campus Awka, Electronic Development Institute, (ELDI) and University of Nigeria Nsukka (UNN) were used as a study testbed to understudy the exiting network performances.

From the limitations of the existing infrastructure, this work integrated WiMax with existing Wifi to establish a three-layer configuration for CWN. Hence, the system capacity and performance of the wireless network were expanded via virtualization, security and QoS metrics. Beside the WiMax RNC security integration, Cisco Stateful packet Inspection firewall was introduced as a gateway to secure backend servers. The results from functional experiments shows that the proposed CWN offers better performance compared with the legacy WLAN in terms of security, throughput and latency. The following recommendations were outlined based on the findings of this research, viz:

1. A good collocation strategy within a third-party data center can be used to house the CWN cloud applications for security, latency monitoring, troubleshooting, support, etc. Collocation facilities for CWN cloud services can offer excellent connectivity and cross-connections. This will support collaboration of resources, improve performance and reduce latency to end users.
2. Service Level Agreement (SLAs) could be used to prioritize particular applications based on security, performance and availability.
3. There is need for collaboration between various institutions in Nigeria and other critical stakeholders to develop a well articulated policy on integration standards and procedures for CWN.
4. Additionally the following delivery platforms could be adopted for the CWN services, viz:

- Managed hosting by adopting a hosting provider to host or manage CWN infrastructure.
- Cloud computing by using an on-demand cloud-based infrastructure e.g. Amazon Elastic Compute Cloud.
- Application led platforms (such as SaaS) should be used to develop and deploy campus wide application services.

REFERENCES

Arinze C.O, Idigo V.E, Ohaneme C.O, Agu V.N, Ezechukwu A.O, "Simulation and Evaluation Of High Density Cognitive Hotspot Network Based On IEEE 802.11 Minipop Access Point Series", International Journal of Computers and Distributed Systems, Vol.4, Issue 2, 2014, Pp.1-10.

C.Guo, G.Lu, D.Li, H. Wu, X.Zhang, "BCube: A High Performance, Server-centric Network Architecture for Modular Data Centers", SIGCOMM'09, August 17th–21st, 2009, Barcelona, Spain.

C.C.Udeze, K.C. Okafor, C.C.Okezie, I.O.Okeke, G.C.Ezekwe, "Performance Analysis of R-DCN Architecture for Next Generation Web Application Integration", In IEEEXplore Digital Library, 6th IEEE International Conference on Adaptive Science &Techonology (ICAST 2014), Covenant University Otta, 19th-31st, Oct,2014.Pp.1-12.

Faisal Alkhateeb,Zain Al-Abdeen Al-Fakhry,Eslam Al Maghayreh, ShadiAljawarneh& Ahmad T. Al-Taani, "A multiagent based based system for securing university campus",IJRRAS, Vol2.No.3, 2010, Pp.223-231.

F.N.Ugwoke, K.C.Okafor, V.C.Chijindu, "Security QoS Profiling Against Cyber Terrorism in Airport Network Systems", In IEEE International Conference on Cyberspace Governance – Cyberabuja 2015 November 4-7, 2015. Pp.241-253. IEEEXplore Digital Library,4th-7th, Oct. 2015,Abuja,DOI:10.1109/CYBER-Abuja.2015.7360516.

K.C. Okafor, CC.Udeze, C.M Nwafor, A.C Abarikwu,"IWLAN: An Implementation Model For High Density Smart Intranet Communications, (A Case for ELDI)", In International Journal Of Engineering And Computer Science ISSN:2319-7242 Vol. 2 Issue 6, 2013, Pp.1766-1776.

K.C.Okafor, F.N. Ugwoke, A.A.Obayi, V.C.Chijindu, O.U.Oparaku, "Analysis of Cloud Network Management Using Resource Allocation and Task Scheduling Services", In International Journal of Advanced Computer Science and Applications (IJACSA), England, U.K Vol. 7, No. 1, (2016a), Pp.375-386.

K.C.Okafor, Anulika Okoye Joy, G.Ononiwu "Vulnerability Bandwidth Depletion Attack in Distributed Cloud Computing Network-a QoS Perspective", In International Journal of Computer Applications (IJCA), USA.Vol.138. No.7, March (2016b).

M.Alzaabi, K.D Ranjeeth, T.Alukaidey and K.Salman,"Security Algorithms For Wimax", International Journal Of Network Security & Its Applications (IJNSA), Vol.5, No.3, 2013, Pp.31-44.

Rajiv Ranjan, Rajkumar Buyya, Surya Nepal, "Model-driven provisioning of application services in hybrid computing environments", Future Generation Computer Systems 29 (2013) 1211–1215.

White Paper- Security in WiMax 802.16-2009 networks, Jan, 2011.

www.riverbed.com-Online: https://supportstaging.riverbed.com/bin/support/static//doc/opnet/17.5.A/online/itguru_17.5.PL5.

Shuang Song and Biju Issac, "Analysis ofWifi and WiMax and Wireless Network Coexistence", International Journal of Computer Networks & Communications (IJCNC) Vol.6, No.6, November 2014.Pp.63- 78.

Sanjay P. Ahuja, Nicole Collier, "An Assessment of WiMax Security,"Communications and Network, 2010, 2,134-137.

Young Kim, ELEN 6951- 802.11b Wireless LAN Authentication, Encryption, and Security, Available Online: http//repo.hackerzvoice.net/depot_madchat/reseau/wireless/wifi_report5.pd.Retrieved on 13th Jan, 2015.

Yong Cui, Hongyi Wang, Xiuzhen Cheng, "Wireless Link Scheduling for Data Center Networks", ICUIMC '2011 Seoul, Korea

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

DEVELOPMENT OF E-SLATE SYSTEM FOR ENHANCING READING AND LEARNING

M. K. Muhammad
Department of Computer Science,
Federal University of Technology, Minna
muhammad_kudu@futminna.edu.ng

A. M. Aibinu
Department of Mechatronic,
Federal University of Technology, Minna
maibinu@gmail.com

M. B. Abdullahi
Department of Computer Science,
Federal University of Technology, Minna
el.bashir02@futminna.edu.ng

ABSTRACT

The need for increasing reading and learning ability in the developing countries as necessitated the introduction of methods such as Universal Basic Education (UBE) Scheme. Despite the success of the introduced scheme shortage of manpower seems to hinder the progress of the scheme. This shortage has greatly affected the reading and learning ability of many. Thus, this work presents the development of a device for improving the basic reading ability of the rural communities in Nigeria. Review of note taking techniques in Nigeria higher institutions have been presented in this work. Also, the design and fabrication of an E-slate for aiding learning and reading has been developed in this work. The design of the handheld version incorporate the use of microcontroller based system and logic circuitry with memory system in determining correct answers to a particular question. The developed system has been tested and was found appropriate for rural communities in the developing nations taking Nigeria communities as case studies.

KEYWORDS: Multimedia, Multimedia application, Learning, Teaching.

1. INTRODUCTION

Recently, the Federal Government of Nigeria introduced the Universal Basic Education Scheme as a way of improving the basic educational level of its citizen (Chukwunke & Chikwenze, 2012). One of the objectives of this scheme is to provide free and compulsory universal basic education for every Nigerian child of school age (Edho, 2009). Despite the fact that the scheme has proved effective in most of the Nigeria cities, it still faces lots of challenges in the rural areas, hence, making it difficult in actualizing the primary goal of the scheme. These challenges among others include the unwillingness of qualified teachers to work in the rural areas and lack of modern teaching facilities (Chukwunke & Chikwenze, 2012).

Schools in the rural areas are the worst hit by lack of teaching and learning materials (Edho, 2009). Also, most teachers posted to rural areas do reject their posting for lack of conducive environment for teaching, inadequate provision of basic amenities that may facilitate effective teaching and productivity of teachers among other things (Edho, 2009). In tackling this challenge, it was suggested in (Edho, 2009) that rural teachers should be employed to work in their vicinity. However, the lack of capable and qualified teachers in most rural areas makes it impracticable.

Reading and writing are the basic essential needs to become an effective and productive member of any literate society. In a bid to improve these basic skills, a study was carried out on the use of ICT in developing reading and writing skills in children with hearing impairment (Bano & Abdulhameed, 2007).

This paper is targeted towards improving the reading ability of rural communities by the use of embedded device called E-Slate (Electronic Slate). It particularly focuses on providing a learning aid for the Nigeria citizens especially those in the rural areas. The device apart from being a learning aid, can also be regarded as part of effort to achieve better writing and reading ability for rural dwellers in Nigeria.

The remainder of this paper is organized as follow: Section II provides review of learning devices used in aiding reading and teaching in Nigeria while section III provides a detailed explanation of design and development of the E-slate. Conclusion ends this paper in section IV.

2.0 REVIEW OF SPECIAL LEARNING DEVICES IN NIGERIA

In (Bano & Abdulhameed, 2007), the objective of the study was to see how effective the use of video clipping will be in developing reading and writing skills of children with hearing impairment. Their findings show that video-clipping proves effective in teaching writing skills among children with hearing impairment. However, a major drawback in the approach was its ineffectiveness in teaching reading skills.

Information and communication technologies (ICT) are often associated with high-technological devices such as computers and software, but ICT encompasses more "conventional" technologies such as radio television and telephone technology (UNESCO, 2006). ICT has been used as a way to transmit, store, create and share or exchange information. The five key ways in which ICT can be used to support literacy as highlighted by UNESCO includes: development of professional teachers, creating local content, broadening access to literacy education (UNESCO, 2006). This could be achieved through the use of radio for radio lessons, learning via television, reading tutor software, multimedia software for enhancing reading skills, audio books, electronic books and online texts (UNESCO, 2006).

As part of effort towards improving literacy education, a simple inexpensive, multilingual people's computer (SIMPUTER) was developed in (Duveskog, et al, 2004). It is a small hand-held computer designed by the Simputer Trust, a non-profiting organization. It was aimed at enabling the widespread use of computers in India and other developing countries. A programming tutorial was developed on the Simputer platform and the learning environment was tested in Tanzanian context. An analysis of the students' feedback shows the feasibility of Simputer as a learning platform. However, the need to have a basic knowledge of computer usage is a major drawback.

The inadequate access to updated information in rural areas is a disturbing challenge which has contributed greatly in the poor effective teaching and learning in such environments (Nwaji, 2011).

An important aspect of learning process that is often overlooked by researchers is the skills or technique for effective note taking among students particularly in Nigeria, despite the fact that it makes or mar the success of the students. There are numerous techniques that are used in taking notes with the goals of aiding student understanding of topic or subject, enhancing their capability of recalling essential information and assisting the students in revising for test and exam. Some of the technique of note taking among students of higher institution in Nigeria includes the Cornell method. This method entails dividing the notebook page into different part, with each part performing different function. Keywords are recorded on the left merging, notes or content are recorded on the right column while summary is recorded at the bottom of page. A similar method is the Two Column method where the page of the notebook is divided into two columns with keywords recorded on the left column with the corresponding description presented on the right column. Other methods include the mapping, outlining, charting, tape recording and the sentence

methods with each of the methods having their strength and weaknesses as discussed in Section III.

2.1 REVIEW OF NOTE TAKING TECHNIQUES IN HIGHER INSTITUTION IN NIGERIA

This section examines the specific techniques of notes taking in higher institutions in Nigeria including their strengths and weaknesses.

2.1.1 CORNELL METHOD

This is one of the most popular notes taking method in the world though rarely used among Nigeria higher institution students. This approach involve dividing the page of the note into two, with a margin of about 2.5 inches drawn to the left, which is used to record key points that will aid in recalling the actual point made during the lecture (Pauk & Owens, 2013). The notes or content was written down at the other right margin of the page, which is the largest space as shown in Figure 1. A summary space is usually at the bottom of the page where the students summarize the note taking after lecture. This method helps in organizing, understanding, recalling of key points using some few sentences and serves as a quick study guide for exams. However, it requires good listening and writing skills by the student.

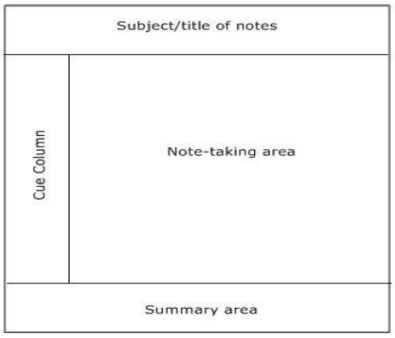


Figure 1: Cornell System of Note Taking

2.1.2 TWO COLUMN METHOD

This is another popular method of note taking in Nigeria which is similar to the Cornell method. In this method the page of the notebook is divided into two columns with keywords or ideas recorded on the left column while the corresponding explanation or discussion of the keyword is presented in the right-hand column (Chukwunke & Chikwenze, 2012). This approach enables easy scanning of note in locating specific information. Nevertheless, the student must be attentive throughout the lecture to enable capture essential information.

2.1.3 OUTLINING METHOD

This method is popular among students in Nigeria. It entails taking note with short sentences comprising of many heading and sub-headings as may deem fit by the students (Unit, 2009). These short sentences help to organize the students understanding of the lectures and ideas in an indent format structure in such a way that the general idea on a topic is captured in the first sentence or indent, followed by supporting information in the second level indent. Other indents may also be included to further provide supporting information that will aid the student understanding after the lecture. It has the merit of easing review process by converting main points into questions as well as minimizing editing of notes. However it requires more thought in class for accurate organization and difficult to used when the lecturer talk fast.

2.1.4 CHARTING METHOD

This is another popular notes taking method among students though rarely used by students in Nigeria higher institution. this method entail the student dividing the page of the note into rows and columns that is taking note in a tabular form with each column carrying different headings pertaining to the topic that will be covered by the lecturer (Bano & Abdulhameed, 2007). This approach has the merit of assisting the students to extract the most relevant information from the lecture, minimize the amount of writing necessary by the student during lecture, providing easy way of examining comparison and relationships that exist between certain topics or ideas as well as aiding review and facts memorization by students. However, it can be a hard system of note taking to learn and used. More so, there is need for the students to know the content that will be covered by the course lecturer before the class begins.

2.1.5 MAPPING METHOD

This is a popular note taking approach that is commonly used among students of higher learning in Nigeria. It entails the student using various minds mapping approach or spider diagram with the intent of recalling all important points in the lecture delivered. Usually, the main theme is at the centre while all other ideas are linked as branches around the theme (Meleisea, 2009). It has the advantages of enabling the students to visually keep track of the lecture by maximizing active participation, it also minimize the level of thinking needed by the students to establish relationships between ideas and also ease the process of note editing and probably adding colours if necessary. Nevertheless, it is difficult for the students to be able to track changes in the content of the lecture from major points to facts.

2.1.6 SENTENCE METHOD

This is the most popular method of note taking in Nigeria higher institution. It involves noting every new thought, topic or fact on separate lines numbering each point as you progress (Bano & Abdulhameed, 2007). It provides a more organized way of recording most of the information in the lecture, and also helps when there is a lot of information to be recorded by the students without knowing how each idea fit together. However, it is difficult to edit note and differentiate between major and minor points.

2.1.7 TAPE RECORDING METHOD

This approach of note taking is gradually becoming popular among Nigerian higher institution students as it enables students to record the lecture directly as it is being delivered using phones, Laptops and other tape recorder (Unit, 2009). It enables the student to concentrate on listening during classes as they can form their note afterwards by listening to the recorded lecture, it also helps record a more complete and accurate notes and facilitate effective reviewing of the materials. However, student may be bored and lose concentration during lecture.

The Cornell method helps organized notes, aids in identifying key word, concepts and serves as a quick study guide for exams. However, it requires good listening and writing skills by the student. The two column method enables easy scanning of note in locating specific information. It has a disadvantage of requiring the user to have good listening and writing skills by the student. In the case of the outlining method, Ease review process by converting main points into questions as well as minimizing editing of notes. This method requires more thought in class for accurate organization and difficult to used when the lecturer talk fast. Charting method minimize the amount of writing necessary, provide easy way of examining comparison and relationships that exist between certain topics or ideas as well as aiding review and facts memorization by students. However, it can be a hard system of note taking to learn and use. Also, there is need for the students to know the content that will be covered by the course lecturer before the class begins. The mapping method enables the students to visually keep track of the lecture by maximizing active participation, minimize the level of thinking needed by the students to establish relationships between ideas and also ease the process of note editing. It is difficult for students to be able to track changes in the content of the lecture from major points to facts. Sentence method provides a more organized way of recording most of the information in the lecture, and also helps when there is a lot of information to be recorded by the students without knowing how each idea fit together. In this method, it is difficult to edit note and differentiate between major and minor points. The tape recording method assists the student to concentrate on listening during classes as they can form their note afterwards by listening to the recorded lecture. It also helps record a more complete and accurate notes and facilitate effective reviewing of the materials. However, student may be bored and lose concentration during lecture.

3.0 METHODOLOGY

The methodology adopted in the design and implementation of the e-slate is presented herewith.

3.1 DESIGN OF THE E-SLATE DATABASE SYSTEM

The overall system consists of a remotely located database system connected via GSM modem. The database was developed to contain all the training materials including e-class in different languages, e-resources that enhance reading ability. The database was linked to the internet for access in the targeted communities. The block diagram is as shown in Figure 2.

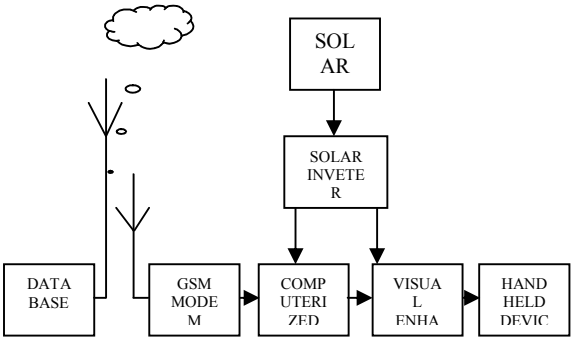


Figure 2: Block diagram of the overall system.

3.2 IMPLEMENTATION

The handheld version of electronic circuits was designed and programmed to have the capability of connecting to the internet.

In essence, the E-slate has books that contain some set of questions that needs correct answers. When the book is placed on the logic sensors, the logic sensors represent the options to the questions in the slate book.

The circuit diagram for the electronic part of the system is shown in Figure 3.

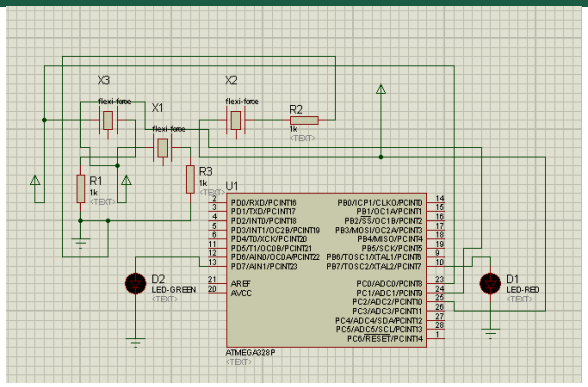


Figure 3: Electronics circuit diagram for E-Slate

By placing the questions in the slate book on the sensors and E-slate shown in Figure 3, the students could made their answer choices as either A, B or C, and hence the corresponding sensors is been pressed. The micro-controller however, has a SD – card on, it where there correct answers are been stored. So when a correct answer is chosen, the E-slate displays the green LED, which signifies the right answer choice, but when the wrong answer choice is been chosen, the Red LED is been displayed to signifies that the choice of answer was wrong. So the process could be repeated for any subject of multiple choices. The whole system is shown in Figure 4.

Figure 4a: E-Slate system

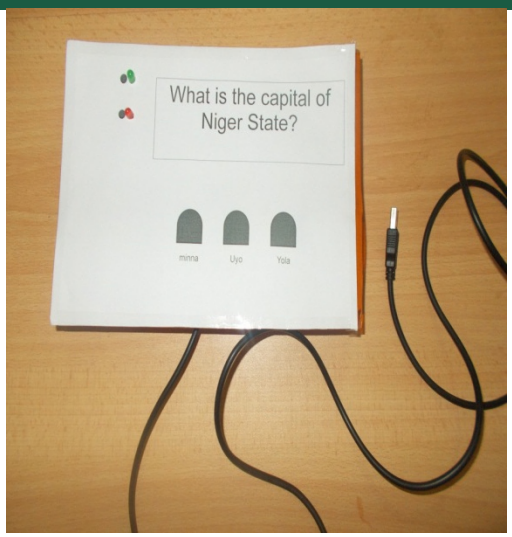


Figure 4b: Question placed on E-Slate

4.0 CONCLUSION

Multimedia is very vital tool to teach in the society. There is need for Nigerian education system to incorporate teaching with aid of multimedia and curriculum should have such element In this work, the development of E-Slate for learning and reading in Nigeria has been presented. The device is a low cost device that can be adopted around the world for teaching in remote locations.

5.0 REFERENCES

Bano H. and Abdul Hameed, 2007. The use of ICT in Developing Reading and Writing Skills in Children with Hearing Impairment. *Proceedings on the World Congress on Engineering and ComputerScience*, San Francisco, USA. Pp 65-66.

Chukwuneke B. U. and Chikwenze, A. R, 2012. The Extent of Implementation of Universal Basic Education (UBE) Program in Nigeria: Focus on Basic Science Curriculum. *Journal of Research and Development*, Vol. 4, No. 1, pp 45-46.

Duveskog M., et al, 2004. Simputer as a Platform for ICT Education in Tanzania. *Proceedings on the International Conference on Advanced Learning Technologies*, pp 1018-1023.

Edho O. G, 2009. The Challenges Affecting the Implementation of the Universal Basic Education (UBE) in Delta State, Nigeria. *Journal of Social Science*, Vol. 20, No. 3, pp 183-187.

Meleisea, E., 2006. *Using ICT to Develop Literacy*. Available from http://files.eric.ed.gov/fulltext/ED494262.pdf.

Nwaji O. J, 2011. Implementing the Universal Basic Education (UBE) Program: Issues and Problems. *bong .T.B, Educational Mentally Retarded children in Nigeria*, Ibadan : Claveriancon Press.

Pauk, W. and Owens, R. J., 2013. How to study in college, Cengage Learning. Unit, E., 2009. *The challenges affecting the implementation of the universal basic education (UBE) in Delta State, Nigeria*, J Soc Sci, Vol. 20, No. 3, pp 183-187.

UNESCO, 2006. *Using ICT to develop Literacy*, UNESCO, Bangkok, Thailand.

SESSION D:

Management of National Database and other Digital Assets

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

A FRAMEWORK FOR INTEGRATED WEB BASED SIM REGISTRATION SYSTEM (IWSRS)

Ibrahim S. Shehu
Department of Computer Science, Federal University of Technology, Minna
ibrahim.shehu@ futminna.edu.ng

Solomon A. Adepoju
Department of Computer Science, Federal University of Technology, Minna
solo.adepoju@ futminna.edu.ng

Garba Suleiman
Department of Computer Science, Federal Capital Territory College of Education, Zuba
sulgarba@gmail.com

Enesi F. Aminu
Department of Computer Science, Federal University of Technology, Minna
enesifa@ futminna.edu.ng

Agada A. Emmanuel
Department of Computer Science, Federal University of Technology, Minna

Hussaini I. Aliyu
Department of Information System, Stratford University, USA (India Campus)

ABSTRACT

The relevance of Information and Communication Technology in our routine activities cannot be over emphasized; hence there is need for daily, weekly or monthly communication among us. The introduction of Global Satellite for Communication (GSM) and the internet is a welcomed development that has eased our schedules which can be done within a predefined period. In addition the issue of security has to be taken into considerations while using some of these technological innovations. As a result of usage of mobile for communications and to ensure security of lives and properties, there is a need for a secured database of mobile phone subscribers in order to help the government most especially in securing a nation as most recently the issue of insurgencies, militancy, kidnapping and other vices that are causing disharmony, anarchy in our society today. In view of all these challenges, this paper proposes an implementable framework for unifying or integrating the nations Subscriber Identification Module (SIM) registration activities in view of solving the issue of multiple registrations by mobile phone subscribers, in-adequate database and constant collections of subscriber's details while registering another SIM. More so benefits of the proposal herein are enormous to stakeholders in the Nigerian Telecommunication Industry which include reduction in time, cost and resources expended during and after SIM registration.

Keywords: Subscriber, SIM Registration, Centralized Database, Mobile Network Operators, Telecommunication Industry, Nigeria

11. INTRODUCTION

According to Ali (2012) Information Systems (IS) has become a crucial point for organizations to survive in technology-focused environment. The amount of resources needed for information system infrastructures in organizations are large, hence there is need to engage or use a robust, well secured systems that will give out the best services required to handle organizational demands. Collaborating the extent of technology, Abdullahi and Hassan (2011) asserted that technology advances, particularly in the area of information and communication keep growing on daily basis, taking advantage and keeping abreast of these technologies is a paramount concern to organizations.

The advent of mobile communication systems has revolutionized the way information are sent or communicated (Adeyegbe n.d.) (Mallikharjuna, Naidu & Seetharam n.d.). The development has brought about ideas, innovation in the field of telecommunication. This is evident from revelation in the work of Diana (2008) stating around two third of the world's populations are connected with mobile phone as of 2008, while statistic from The International Telecommunication Union shows an estimated value of 6.8 billion mobile subscriptions worldwide as of February 2013, which directly represents 96 percent of the world population (ITU 2013).

Nigeria is not left out from the rapid development the world has recorded in the area of mobile communication as the country strives to become a technology driven economy (ITU 2013). In 2011, the US-Embassy in Nigeria at its economic section presentation on Nigeria's telecommunication fact sheet placed telecoms sector as the most viable and fastest growing industry of the Nigerian economy, creating direct and indirect jobs. However, Mbendi (2014) maintained the telecommunication services which began in Nigeria in 1986 via a connection between London and Lagos remain under developed before the sector was deregulated fully in 2001. The deregulation exercise brought about lesser restriction in the sector leading to inauguration of a full GSM services in Nigeria (Mbendi 2014). Despite the aforementioned changes, there still remain some big challenges, such as service quality and hideous crimes committed using mobile phones (Agba 2001). Many believes the government is to be blamed because after the inauguration of the GSM services in 2001, Subscriber Identification Module (SIM) cards were offered to mobile subscribers without the proper requirement for registration, to aid identification of the subscribers.

In 2002, there was an attempt by the regulatory body, Nigeria Communication Commission (NCC) to enforce subscriber registration. The effort did not yield much result because emphases was not on the documentation of telephone lines that are in use against the name and full identity of those who purchase them for use. Consequently, security implications surfaced as a result of the negligence. According to NCC (2016a) more recently the quest for subscriber registration has risen. NCC and several other partners rolled out compulsory registration of mobile lines for existing and new subscribers to check some of these hideous crimes in the society and also to improve service delivery. Therefore the necessity to review the existing SIM registration system and provide unique general platform that is cost effective for subscriber authentication begins.

Finally, registration of identity information to activate a mobile SIM card, are fast becoming universal in Africa, with little to no public debate about the wider social or political effects. Also, SIM registration represents a form of communications surveillance that reduces the anonymity once afforded, perhaps unintentionally by prepaid airtime. These identification mandates may bring modest security benefits, although as noted, the evidence for such claims remains inconclusive (Kelvin & Aaron 2014).

2. LITERATURE REVIEW

Subscriber Registration is a documentation of the mobile lines that are actively in use against the names and full identity of those who have purchased them for utilization (Omo-ettu 2012). It implies getting good details of all the things that happen in the network such as the identity of all subscribers to network amenities. Such information is needed to manage today, and to strategize for the future. To utilize the

aspect of it that can help combating cases of kidnapping, terrorism, social harassment and to also to serve as a larger whole part of planning the national life, education, health, transport, and so on.

The SIM card has been one of the important technologies deployed or utilized by Mobile technology, whereby each phone uses a unique identifier to use the available networks. The SIM card is designed in various formats such as full-size SIMs, mini SIMs, micro-SIMs, embedded-SIMs and nano-SIMs but not limited to the mentioned as technology continues to expand. According to Elaheh (2013) the SIM is security element used in the authentication of the subscriber before granting him/her access to the mobile network. The ingenuity of the SIM lies on the fact that it is a separate tamper resistant module which can be installed or removed from the mobile phone. However, with the advances in wireless and storage technologies there is proposal to replace the current SIM by the so-called soft SIM, which consists of a tamper resistant module soldered on the mobile phone and a software SIM downloadable over-the-internet.

To use this SIM card, a subscriber need to register this SIM card as a documentation to ascertain the ownership of the SIM and to be able to use it on the available network which varies from different mobile network operators. According to Etisalat-Nigeria (n.d.) during registration a subscriber is required to make available for use full names, residential address, age, date of birth, state of origin, occupation, photograph, nationality, religion, Subscriber SIM MSISDN, SIM serial and bio-metrics information like thumb print. Outbound calls and SMS services will not be accessible by the subscriber until the SIM is fully registered. However, subscriber can receive calls and SMS that are inbound on an unregistered number for 30 days period starting from the when the first call from the SIM card was made. Deactivation of SIM card occurs after 30 days if the SIM remains unregistered at the end (Etisalat-Nigeria n.d.) (Bingham 1999).

According to *The mandatory registration of prepaid SIM card users* (2013) in many countries around the world, consumers can buy prepaid or 'Pay As You Go' mobile SIM cards from retail outlets usually with little or no paperwork involved. Unlike pay-monthly mobile SIM contracts, the activation and use of prepaid SIM cards does not always require the subscriber to register or present any identity documents at the point of sale. The researcher further stated that, in countries where prepaid SIM registration is not required, mobile subscribers can access mobile services more easily, but can also voluntarily register with their mobile network operator (MNO) in order to use additional services that require identification, such as mobile banking.

Mean while in Nigeria, there are issues of security of the present SIM card registration process whereby Mobile Network Operators allow their agents register mobile subscribers using several customized SIM registration applications, databases. A similar case to Bangladesh where cases of SIM cards registered with false information by dealers are widely used for committing crimes (Nazmul, Mohammmed, Raisul and Nazia, 2014). To lessen the rate of crime committed by using mobile phones in Bangladesh, Nazmul et. al, (2014) proposed a cloud based system model which shows an online method of SIM card purchasing and registration. Their proposed method resolve the conventional system of mobile purchasing and registration in Bangladesh and most importantly the issue of security in order to reduce the rate crime are committed using mobile phones and also encourage mobile user to register their SIM cards without any stress.

2.1 CURRENT STATUS OF SIM SUBSCRIBERS IN NIGERIA

The Federal Government through the Nigeria Communication Commission (NCC), the regulatory agency mandated the compulsory registration of subscribers SIM as a tool to combat insurgency, terrorism and most importantly security issue, sadly adequate and up-to--date subscribers details are yet to be perfected either due to sale of unregistered SIM card, double registration, provision of change/update of certain sensitive information of the subscribers. According to NCC (2016a) the objectives of the NCC when it mandated nationwide registration of SIM users in March 2011 were to:

- Assist security agencies in resolving crime and by extension to enhance the security of the state;
- Facilitate the collation of data by the Commission about phone usage in Nigeria;
- Enable operators to have a predictable profile about the users on their networks; and
- Enable the Commission to effectively implement other value added services like Number Portability among others.

Due to the importance attached to it and through supervision, by July 2013, Bio-key International (2013) asserted the NCC had reportedly uploaded more than 110 million entries to its database facility, including users' biometric details (thumbprints) and issues approximately 8 million new SIMs annually through its multiple mobile network operators. To further enumerate the extent of SIM subscribers in the country, NCC (2016b) reported that more than 188 million mobile phone subscribers are connected to one mobile network service to another. However, only 139,143,610 million of these lines were active by the end of 2014 as shown on table 1.

Table 1: Subscriber statistics

	Operator	2014
Connected Lines	*Mobile (GSM)*	184,782,512
	Mobile (CDMA)	3,743,811
	Fixed Wired/Wireless	365,871
	Total	188,892,194
Active Lines	*Mobile (GSM)*	136,772,475
	Mobile (CDMA)	2,187,845
	Fixed Wired/Wireless	183,290
	Total	139,143,610

According to NCC (2016c) at the end of third quarter 2015, NCC had reported an increased figure for the active lines to 150,660,631 million with the mobile network operators taking 98.52% of the total number of active lines as shown on table 2.

Table 2: Operators data

Latest Data: Quarter 3 - 2015	
Operator	Sept 2015
MTN Nigeria Communications	62,493,732
Globacom Limited	31,306,472
Airtel Nigeria	31,134,625
EMTS Limited (Etisalat)	23,492,214
Sub-Total (GSM)	***148,427,043***
Visafone Limited	2,031,802
Multilinks Telkom	10,213
Sub-Total (CDMA)	***2,105,981***
Visafone Limited	63,396
Multilinks Telkom	2,923
VGC/MTN	9,731
21st Century Technologies	100,986
IPNX	2,879
Globacom Limited	11,658
Sub-Total (Fixed/Fixed Wireless)	***191,573***
TOTAL	***150,660,631***
% of Mobile (GSM)	98.52
% of Mobile (CDMA)	1.36
% of Fixed/Fixed Wireless	0.13

Statistics presented in table 1 and 2 shows the Nigerian telecommunication sector has significantly grown since the three companies (M-TEL, MTN and ECONET) were awarded licenses to operate the GSM in January 2001, though operations started in August of the same year, mobile network operators attained about 500,000 subscribers in 2001 (Olayiwola 2010). According to Olayiwola (2010) the industry grew to over 7million subscribers in 2004; in December 2008 the subscribers in the market grew to 62.99million. An addition of 22.59 million subscribers in 2008 alone represented 56% annual growth rate. Recent figure as at January 2009 put the subscribers' base at 64.16million.

From 2009 to date, Nigeria had gradually become the most competitive markets in Africas telecommunication industry with four active GSM mobile network operators (i.e. MTN, GLOBACOM, AIRTEL and ETISALAT) and two active CDMA companies namely VISAFONE and MULTILINKS. The development has resulted in much lower tarrifs, a wide variety of innovative services, attractive offers and improvements in service quality in order to differentiate and set the brands aside.

Having said much on the growth of SIM subscription in Nigeria, the researchers are worried if the mobile network operators will keep to NCC's directive of continual SIM registration of subscribers considering the enormous challenges the current systems are faced with. However, if the directive is continued with the current SIM subscription statistics, it could be the largest and most comprehensive biometric database ever assembled on one platform in the Africa, through the SIM registration exercise.

3.0 METHODOLOGY

The researchers identified three (3) key activities for achieving the research aim of proposing an integrated framework for the SIM registration in the country in figure 1.

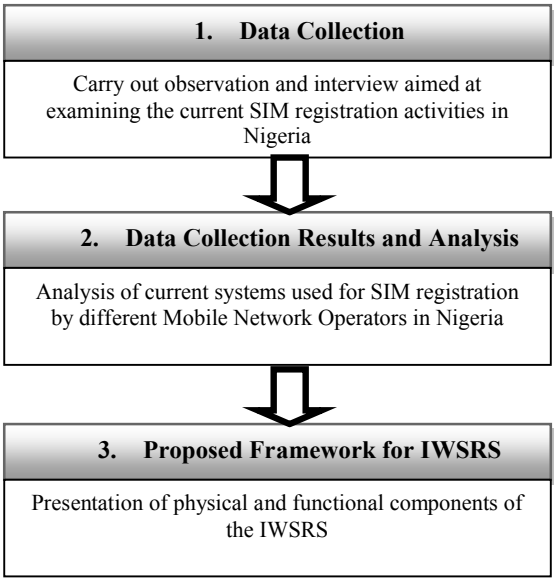


Figure 1: Research methodology activities

3.1 Data Collection

The data collection was carried out in Minna, the capital city of Niger State. Accessible data collection methods were used to examine the current state of SIM registration activities in Nigeria.

A. Observation Method

Due to the relevance attached to the collection of accurate data from the right and reliable source, the researchers set out on observing registration procedures at different location of the city i.e. MTN and Etisalat customer shops located at Tunga, Globacom customer shop located along city gate. Airtel shop located at stadium road. Various registration agents were also visited at Obasanjo complex along Mobil and Uche communication limited at Bosso. This method was utilized for the following rationalities:

- To have first hand information about the mobile network operators and their agents carrying out the registration at various points without embellishment
- To avail the researchers the chance of observing the whole system, its structures and requirements

B. Interview Method

It involves questioning and evaluation. The personal consultations seem to be the most compelling tool in the methods used for data collection. The method gave the researchers insight to certain operational activities that cannot be accessed by absolute observation.

A face-to-face interaction occurred between the research team and the following questions were asked to attain the level of acceptability, efficiency, and privacy from quotas interviewed;

- Staffs of mobile network operators and their agents
  1. Why do you ask for subscribers SIM registration?
  2. What is your opinion on the level of perception the subscribers have on the issue of SIM registration?
  3. Do you believe having a unified or integrated subscriber registration system online could be more cost effective to your business than the current customized standalone subscriber registration systems you use?

- Registered subscribers
  1. How many SIM do you have?
  2. How long does it take you to register your SIM?
  3. Can you register your SIM from different mobile network operators all at one point?
- Unregistered subscribers
  1. Why is your SIM not registered?
  2. Are you actively connected (e.g. can you make and receive calls)?
  3. Do you plan to register your SIM in the future?

3.2 Data Collection Results and Analysis

The two method of data collection used by the researchers gave a better understanding of the problems inherent in the various SIM registration systems used by mobile network operators and their agents presently in the country.

3.2.1 Observation

The findings gathered during the observation process shows the way SIM registration is done using the current systems as depicted in figure 2.

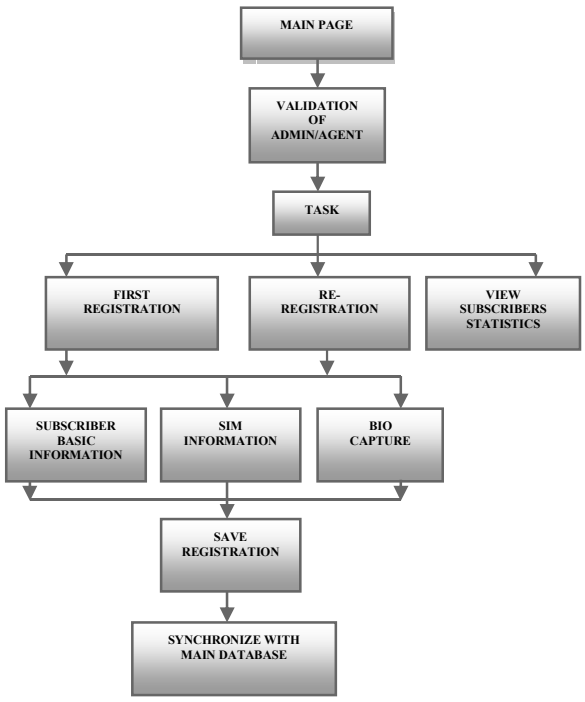


Figure 2: Top down model of the current SIM registration system

Figure 2 depicts the processes involved in the current SIM registration systems; the configuration of the current system varies from one mobile network operators system to another. With the specification given by the regulator (e.g NCC), each mobile network operator has individual Java/Visual Basic programmed software installed on customized mini-laptops distributed to their agents (registered and non-registered) to register subscribers. The system avail the operators' opportunity to register exiting and new customers to its database and likewise forward to NCC and other relevant agency if the need arise.

Functionalities peculiar to these current systems are a single interface which houses the subscriber basic information fields, SIM information fields, bio-data capture fields (photo finger biometric identification). Each subscriber fills a pre-registration form to quicken the process, after a successful process the system pops up message for subscriber registration synchronization with the mobile network operator's main database. The subscriber SIM gets activated within 4hours to 24hours of a successful registration.

The shortcoming of these systems includes; time wastages, duplication of data from different locations, financial waste due to multiple registration outlet on individual platform, fraudulent act due to unregistered agent. The demerits enumerated below can be overcome with the proposal made in this paper.

Multiple registration: The inability of the current system to verify the status of subscriber SIM registration at any registration point, subscribers with unknown SIM status tenders to re-register the same SIM which directly leads to multiple registration entry at the mobile network operators end.

Agent loop holes: The current system is operated by agents, most not registered. Since the authentication of most agents handling the current systems cannot be verified, loop holes are therefore created. These loop holes can lead to criminal and fraudulent activities like hoax calls, terrorism and kidnapping.

Cost implications: The absence of a unified or integrated registration system adversely results into individual platform for all mobile network operators to avail the subscriber chance of registration to enable SIM activation, it directly lead to high cost of buying individual platform to reach millions of subscribers.

Delayed registration: The current system uses an end-to-end retransmission strategy to make sure that data arrives correctly at the mobile network operator's end, which directly leads to delay activation and multiple registration by subscribers.

3.2.2 Interview

The results from the research interview are presented in a Simple Percentage Method (SPM) for easy analysis and interpretation.

In the first interview, a total number of twenty four (24) staffs and agents of mobile network operators were interview (i.e. for each mobile network operator three(3) staffs and three(3) agent was considered at different locations). Results presented in table 3.

Table 3: Interview results for staffs of mobile network operators and their agents

S/N	Questions asked	Response (in number and %) in relation to opinion
1	Why do you ask for subscribers SIM registration?	20(83%) maintained SIM registration is a policy from the Nigerian Communication Commission that should be obeyed while 4(17%) had no idea to why it is done 16(67%) revealed that subscribers do not see any importance of SIM registration, thus so many false information are provided during registration while 8(33%) believes they do. 21(88%) believes the prospect of the IWSRS while 3(12%) did not see the importance of proposing the IWSRS thus doubts its effectiveness against the existing standalone SIM registration systems.
2	What is your opinion on the level of perception the subscribers have on the issue of SIM registration?	
3	Do you believe having a unified or integrated subscriber registration system online could be more cost effective to your business than the current customized standalone subscriber registration systems you use?	

In the second interview, a total number of twenty four (24) registered subscribers for all the mobile network operators were interview (i.e. for each mobile network operator six (6) registered subscribers were considered at different locations). Results presented in table 4.

Table 4: Interview results for registered subscribers

S/N	Questions asked	Response (in number and %) in relation to opinion
1	How many SIM do you have?	19(79%) uses at least two (2) SIM from different mobile network operators while 5(21%) uses a single SIM from a single mobile network provider 22(92%) believes the time taken to register a SIM is not reasonable and should be less than 20minutes while 2(8%) believes the time taken is ok. This question was directed to the 19 registered subscribers that revealed they make use of at least two(2) different SIMs from different mobile network operators. 19(100%) revealed no, that they had to go to each mobile network operator to have their SIMs registered while 0(0%) none had no contrary view
2	How long does it take you to register your SIM?	
3	Can you register your SIM from different mobile network operators all at one point?	

In the third interview, a total number of twenty four (24) unregistered subscribers for all the mobile network operators were interview (i.e. for each mobile network operator six (6) unregistered subscribers were considered at different locations). Results presented in table 5.

Table 5: Interview results for unregistered subscribers

S/N	Questions asked	Response (in number and %) in relation to opinion
1	Why is your SIM not registered?	8(33%) did not see the importance of doing so while 16(64%) could not imagine going through the stress 4(17%) are actively connected even as unregistered subscribers while 20(83%) are not actively connected on any network because they have failed to register their SIM. 17(71%) intend to register soon when they have the time while 7(29%) revealed that they will only be interested to register their SIMs when the stress involved is minimized
2	Are you actively connected (e.g. can you make and receive calls)?	
3	Do you plan to register your SIM in the future?	

3.3 Proposed Framework for IWSRS

This section of the methodology presents an implementable framework for achieving the proposed IWSRS.

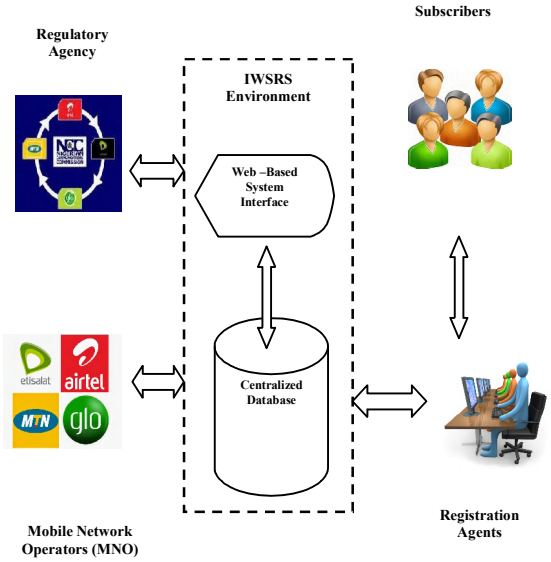


Figure 3: Physical components of the IWSRS

Figure 3 shows the interaction established between the different physical components and stakeholders of the IWSRS. A subscriber communicates their personal data for SIM registration to a registration agent. The registration agent interacts with the IWSRS to record the subscriber data. The mobile network operators and the regulatory agency interact with the IWSRS to perform functions peculiar to them as described on figure 4.

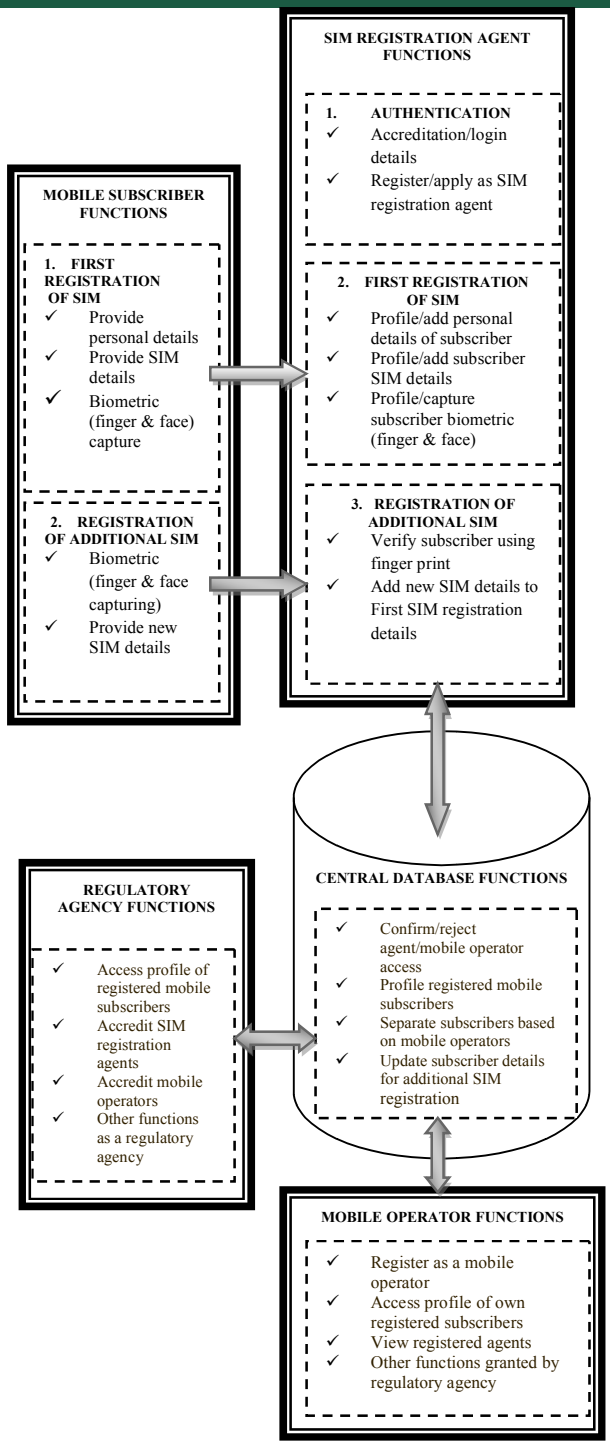


Figure 4: Functional framework of the IWSRS

Figure 4 represents a functional framework of the IWSRS. The framework consists of various functionalities to achieve the research aim. These functionalities are spread across the different physical components shown in figure 3.

A. Subscribers

A new subscriber has to register all his basic details through an approved and accredited agent by providing personal details such as fingerprint/face capturing, name, phone number, date of birth, place of origin, SIM details and other relevant information. This information are then stored through the web application into the database for future retrieval.

B. SIM Registration Agent

An agent can only perform the registration of SIM subscribers if given the license or authority to do so by the Government regulatory agency. Such agent will be given a login details so as to have access to the web application in other to register the subscribers.

Meanwhile, to register a new SIM, the agent has to enter the subscriber's details provided by the subscriber. To register an additional SIM by an already registered subscriber, there is no need of providing the personal details since such subscriber has registered a SIM earlier, what the agent need to ask or to be provided by the subscriber is the already registered SIM number and the fingerprint. On providing such details, the information about such subscriber is display on the screen for viewing in order to confirm authenticity of the subscriber and earlier registration. The Agent only have to now enter the new SIM number and the already saved details of such subscriber is updated in the database without re-registering subscriber's details again.

C. Central Database

A web application designed in a simple user friendly manner is linked to a central database (data repository). The central database has several functions as described in figure 4. The central database provides the unification of subscribers SIM registration as proposed by the research in such a way that every subscriber SIM registration is verified to see if there is an existing registration details of such subscriber in the database.

D. Mobile Network Operator (MNO)

The mobile network operators are companies licensed to provide GSM services by the Government. They perform certain functionalities as proposed by the researchers which include, registration as an MNO, accessing profile of own registered subscribers and other functions that may be granted by the regulatory agency.

E. Regulatory agency

The main function of a regulatory agency (such as NCC) is to regulate, coordinate and monitor the usage of SIM registered and other functions assigned to it by the law, hence in this case their functions is to give web application access to licensed MNOs, registered agents by providing them login details. It also perform the duties of managing the central database in order to ensure confidentiality, data integrity, access control, data management and concurrency handling.

4.0 CONCLUSION

SIM registration for mobile subscribers will be no more a rigorous and time waste issue as the paper proposes an implementable framework for a unified or integrated subscriber registration system where subscribers register their SIM with any accredited agents no matter the network they are using. All what is required is for the subscriber to register through the agents, who must have been accredited or licensed to do so. The web application central database developed will profile and separate details of subscribers of each mobile network operator. In addition the issue of registering another new SIM has been simplified, what is needed is the biometric fingerprint of the previous registration which the system will use to display the details of the subscriber so as to ascertain the authenticity of earlier registration. Afterwards, the new SIM to be registered will be entered at a column provided which will be updated automatically and linked with the previous details and biometric done earlier. In addition, each mobile network operator will be able to access their subscriber's details from the central database when needed through the Government regulatory agency. With this platform, administrative challenges that have bedeviled issue of SIM registration in Nigeria would have been resolved to a certain level.

Finally, though the researchers proposed a web based implementation of the IWSRS for effective usage of functionalities, it is appropriate to point out that web based systems are vulnerable to malicious attacks. The proposed IWSRS is not an exception due to the fact that some subscriber data have market value. So, data encryption at the highest level should be worked upon by interested researchers.

5.0 REFERENCES

Abdullahi, FB & Hassan, T 2011, 'Design And Implementation Of A Web-Based GIS For Patients Referral To Hospitals In Zaria Metropolis', *IJRRAS* vol.8(1), pp.109-114.

Adeyegbe, SO, n.d., *Years Of E-Registration: Keeping Pace With It Revolution In Testing: The WAEC Experience*, Viewed on the 6th December, 2015 < http://Www.Iaea.Info/Documents/Paper_1162a16530.Pdf >

Agba, PC 2001, *Electronic reporting: Heart of the new Communication Age*, Enugu, University of Nigeria press ltd.

Ali, KS 2012, 'Design Web-Based Of Hajj Registration System for Iraq', Master Of Science Of Information Technology, Universiti Utara Malaysia.

Bingham, J 1999, *Essence of SIM Registration worldwide*, London, McGrew.

Bio-key International, 2013, *Nigerian Communications Commission Deploys BIO-key Fingerprint Technology for National SIM Card Registration Program*. Viewed on the 6th January, 2016 < http://www.bio-key.com/press/nigerian-communications-commission-deploys-bio-key-fingerprint-technology-for-national-sim-card-registration-program >

Diana, A 2008, 'A Study on BiometricRecognition in Embedding Web Browers in 3G Mobile Phone Applications', *International Journal of Emerging Technology and Advanced Engineering*, Vol. 3 (1), pp. 1-5.

Elaheh, V 2013, 'Evolution Of The SIM To ESIM', Master of Science In Communication Technology, Norwegian University of Science and Technology.

Etisalat-Nigeria, n.d., *SIM registration*. Viewed on the 5th January, 2016 < http://etisalat.com.ng/sim-reg/ >

ITU 2013, *ICT statistics facts figures*, Viewed on the 8th November, 2014 < http://www.itu.int/en/ITUD/Statistics/Documents/facts/ICTFactsFigures2013.pdf >

Kevin, PD & Aaron KM 2014, 'The Rise of African SIM Registration: The Emerging Dynamics Of Regulatory Change', *First Monday*, vol.19, pp. 2-3

Mbendi 2014, *Telecommunications in Nigeria – Overview*, Viewed on the 2nd December, 2014 <http://www.mbendi.com/indy/cotl/tlcm/af/ng/p0005.htm >

Mallikharjuna, RN, Naidu, MM, & Seetharam, P, n.d., *An Intelligent Location Management Approaches In GSM Mobile Network*, Viewed on the 7th December, 2015 < http://Arxiv.Org/Ftp/Arxiv/Papers/1204/1204.1596.Pdf >

Nazmul, H, Mohammed, R. A., Raisul, I. & Nazia, M. (2014). Mobile Phone SIM Card: A Security concern in the Perspective of Bangladesh. *Proceeding of Computer and Information Technology (ICCIT), 2013 16th International Conference*. pp 213-217. Retrieved on the 7th May, 2016 from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6997302&url=http%3A%2F%2 Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6997302

NCC, 2016a, The aims and objectives, *Viewed on the 5th January, 2016 < http://www.ncc.gov.ng/index.php?option=com_content&view=article&id=122&Itemid=113 >*

NCC, 2016b, *Subscriber Statistics*, Viewed on the 10th January, 2016 < http://www.ncc.gov.ng/index.php?option=com_content&view=article&id=125:subscriber-statistics&catid=65:industry-information&Itemid=73 >

NCC, 2016c, *Operators data*, Viewed on the

10th January, 2016 < http://www.ncc.gov.ng/index.php?option=com_content&view=article&id=70&Itemid=67 >

Olayiwola, WB 2010, 'Mobile Telecommunication Customer Loyalty in Nigeria: Determining factors', Master Business Administration, Munich.

Omo-Ettu, T 2012, *NCC N6b: We need Telephone Subscriber Registration not SIM Card Registration*. Viewed on the 3rd December, 2014 < http://www.myondostate.com/myondostate/seeinterview.php?id=64 >

The Mandatory Registration Of Prepaid SIM Card Users 2013, Viewed on the 22nd December, 2015 < http://www.gsma.com/publicpolicy/wp-content/uploads/2013/11/GSMA_White-Paper___Mandatory-Registration-of-Prepaid-SIM-Users_32pgWEBv3.pdf >

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

A HASH-BASED SYSTEM FOR ENFORCING THE INTEGRITY OF DIGITAL ASSETS

A. E. Ibor
Cross River University of Technology,
Calabar, Nigeria.
ayei.ibor@gmail.com

W. A. Adesola
Cross River University of Technology,
Calabar, Nigeria.
shobisi@yahoo.com

ABSTRACT

The integrity of digital assets has become increasingly important owing to the high rate of cybercrimes and cyber-related offences in recent times. The pervasive nature of the Internet and the concomitant migration of most business functions to the cloud make it pertinent to deploy more veritable means of enforcing the integrity of data, information, databases, disk volumes and data warehouses. One of such veritable means of affirming the authenticity of the contents rendered both offline and online is the use of hashing. In this paper, the processes involved in creating a database of hash codes, which will contain hash values that are unique for some datasets was discussed. Emphasis was laid on the use of the hash codes to verify the non-alteration or otherwise of stored contents in a bid to enforce data integrity. It is assumed that the paper will help various categories of users to protect their digital assets without having to resort to computationally expensive mechanisms.

KEYWORDS: Hash Function, Data Integrity, Digital Assets, Cryptography, Forensic Methods

12. INTRODUCTION

Data stored by individuals, organisations, and the government requires enhanced security measures and policies geared towards maintaining the confidentiality, integrity and availability of its contents. The proliferation of networks and advancement in technology characterised by the wide appreciation and use of information technology principles and best practices have paved way for contents of devices, especially those connected to a network to be vulnerable to various degrees of security breaches and exposure Li (2012). Ahmad & Ahmad (2010) assert that data and information are valuable assets to the decision making process of an organisation. In their work, they highlighted that the availability and confidentiality of data is paramount to the continuity of business. To this end, there must be adequate layers of security put in place to allow the data held on storage devices or data in transit to maintain its originality at all times.

Many traditional methods of data security such as the use of passwords may not be sufficient in today's sophisticated information era. This is so because passwords can be stolen, forgotten, and are also transferable (Sun, 2011). In the event of such an incident occurring, the data being protected can fall into the hands of intruders, who may in turn distort, destroy, modify or expose the data. Such incidents can lead to huge financial loses as well as loss of business services and classified data. Imagine a situation where confidential product information falls into the hands of competitors in a business scenario. The aftermath of a security breach of this magnitude can destroy the life of a business and its stakeholders.

As discussed in Roussev (2009), hashing as a forensic method of securing data including data-fingerprinting provides for the integrity of the data

involved. This is connected to the fact that forensic methods such as hashing can be deployed in recreating the timeline of events, which culminated in the change of state of a computer system. Hashing and data-fingerprinting allow the integrity of stored or transmitted data to be protected such that it is possible to confirm when data has been modified or altered. A hash function takes a variable input and produces a fixed length output. In this way, when a bit string representing the value of a digital asset such as a file is altered, the hash of the original file and its image will not match. It is therefore not far-fetched to vividly identify that a data string has been altered.

Most critical services require that the data being processed and disseminated is accurate at all times. Storing or transferring erroneous data can be problematic to the furtherance of an organisation's goals and objectives. The worst hit may be financial data, which must be accurate for a financial transaction to be accepted between the parties involved. More and more devices as well as users are exploiting the power of the Internet to share data and information. Consequently, there must be a veritable means of confirming that the information transmitted at one end of a transmission point is the exact information received at the other end. There is therefore the need to apply forensic methods in establishing the security of data for business and personal use.

This paper will therefore present a hash-based method for enforcing data integrity protection, and will include a hash database that will be used to store the extracted hashes of digital assets, for which alterations on them can be easily confirmed by matching the hash of the data before and after the alteration.

2. HASH FUNCTIONS AND ALGORITHMS

Roussev (2009) opines that the investigation of large volumes of data should begin with hashing. Hashing is an effective tool for validating the integrity of data and also to identify known contents in the data being protected. Hash functions are used to effect hashing. The main function of a hash algorithm is to produce a fixed length digest from an arbitrary string of binary data used as input to the hash function.

Hash functions should be collision resistant since they are used to verify that the bit string in the data being considered is not altered. Being collision resistant implies that two different inputs cannot produce the same output at any point in time. Such computational infeasibility authenticates hash functions as effective tools for enhancing the integrity of data on which they are being used. There are a plethora of hash functions, which are designed to be collision resistant. Some of these hash functions, which have cryptographic properties include MD5 (Message Digest 5), SHA-1(Secure Hash Algorithm), SHA-256, and SHA-512.

Other hashing algorithms include checksums, polynomial hashes, universal hashes and so on.

However, deploying cryptographic hash functions is more efficient with modern day hardware and the speed of implementation of such hash functions is also higher than conventional hash algorithms. Using a hash function to verify the integrity of data is basically to apply the hash function on the target data, which can be an entire drive, partition or a set of files. As stated in Bui et al (2003), the malicious activities of hackers permit them to modify the state of certain files in a bid to create backdoors and rootkits as well as Trojans and viruses that can compromise the security of the files held on storage media. One significant solution to these is to have a comprehensive backup of critical data on which cryptographic checksums such as MD5 or SHA-1 are performed regularly.

It is common to find incidences in an organisation that can compromise the state of the data and configuration files including data modification, deletion, corruption, and distortion. In such instances, comparing the files with their checksums can reveal whether or not such files have been compromised. It is noteworthy to mention here that a change in the bit string of a data file can have significant effect on the contents of the data. Rowlingson (2004) as well as Preneel et al (1993) believe that taking the hash of data or files is sufficient to demonstrate the integrity of the contents of such data or files. This is also supported by Yannikos et al (2013), showing that the use of hash functions and algorithms enhance data integrity with emphasis on maintaining the originality of the data being used.

3. ACQUIRING THE HASH OF A DIGITAL ASSET

Stored and transmitted digital assets such as files and disk volumes can be hashed to protect the contents of these assets. Several methods can be deployed in the process of acquiring the hashes of digital assets. One of such is the use of forensic tools such as Prodiscover, Encase, SIFT, FTK and so on. The use of these tools is complemented by the DD command in most Linux distributions including BackTrack and Kali Linux. This paper adopted an algorithmic approach for the acquisition of the hashes of digital assets. The design of the hash process was implemented using the Java programming language and sample output was shown in this work. The backend was implemented using MySQL on XAMPP server, which is an open source relational database management system (RDBMS).

3.1 DESIGN METHODOLOGY

This section considered an object-oriented design approach that was based on the use of the use case, class, and activity diagrams for the design of the proposed system. Each of these design components was adequately discussed including the interactions among the various elements of the design. A process model of the system was also depicted in this work to give a vivid description of the architecture of the design. The process model will tend to provide an insight into the internal structure of the system and provide a concise picture of the interactions within it.

3.2 PROCESS MODEL OF THE PROPOSED SYSTEM

The process model of the proposed system is depicted in Figure 1. As shown in Figure 1, user authentication precedes the hashing of the digital assets. Each digital asset that is prone to integrity violation is hashed and the hash value stored in a database for reference purposes. In the event of a security breach including but not limited to unauthorised modification of contents, the hash of the altered digital asset is compared with the hash of the original digital asset to find a match or mismatch. A match shows that the integrity of the digital asset was not violated while a mismatch portrays a modified digital asset. When modification is an authorised update, the digital asset is hashed and the initial stored hash is updated in relation to the current update. However, an unauthorised modification of the digital asset produces a hash value that is at variance with the stored hash for that asset thereby divulging the possibility of a security breach on

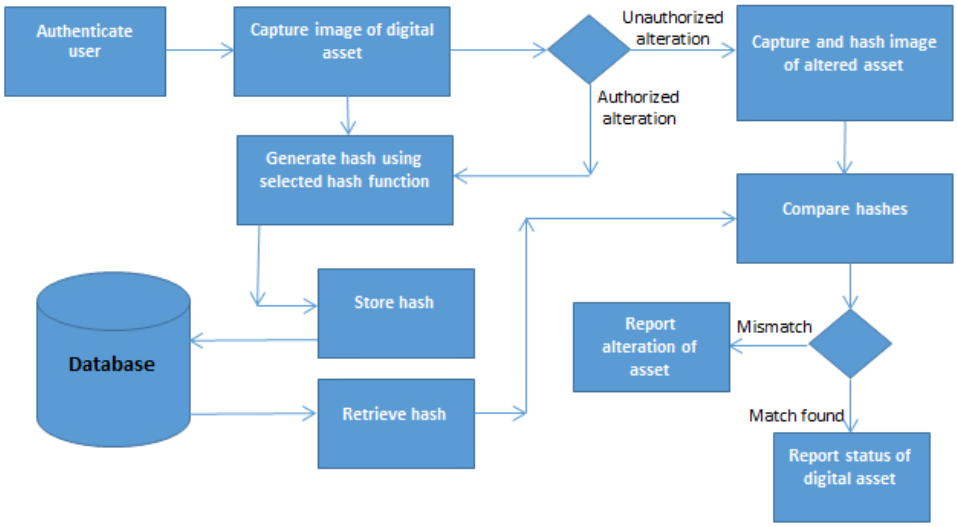


Figure 1: Process model of the proposed system

such an asset. This security is then reported and adequate measures taken to curtail it.

3.3 USE CASE DIAGRAM

The capturing of the dynamic behaviour of any system describes the most vital aspect of the modeling of such a system. A system's dynamic behaviour basically shows the reaction of the elements of the system at its point of operation (or when it is running). For every system, static behaviour is not sufficient for its modeling, hence the relevance of the use case diagram for this purpose.

As discussed in Rumbaugh et al (2004), in a use case diagram, there are internal and external agents known as actors, which depict the behaviour of the system. This follows that use case diagrams consist of actors, use cases and their relationships. With a use case diagram, the system as well as the subsystem of an application can be modeled such that a single use case diagram can capture a particular functionality of a system while several use cases can be used to model an entire system dynamically. The use case diagram for this implementation is depicted in Figure 2.

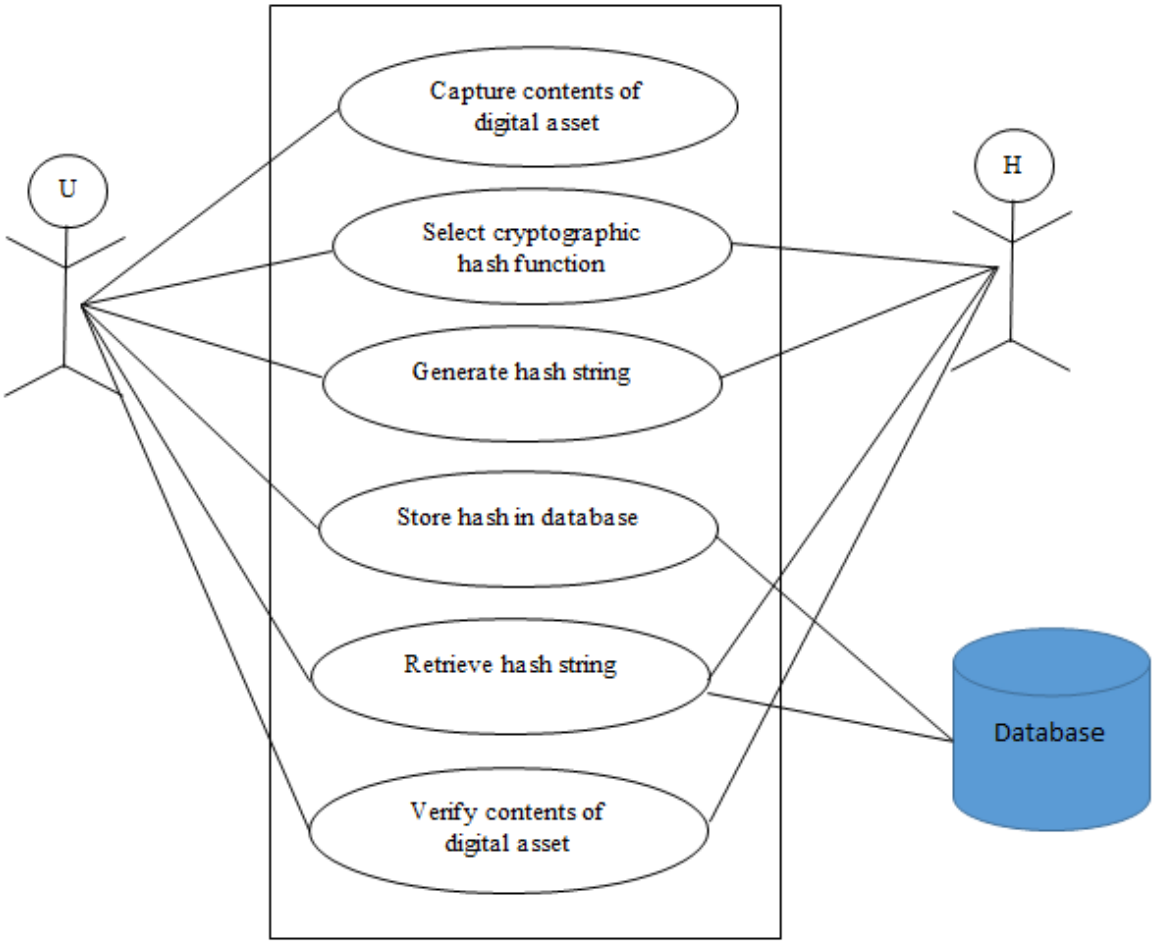


Figure 2: Use case diagram of the proposed system (U – User, H – Hash function)

The textual description of the use case diagram of Figure 2 is as follows:

U1: Capture contents of digital asset: this step is initiated by the user. The user can be a system administrator, a power user, data manager, and so on. At the stage, the user captures the relevant information that needs to be hashed either as a file or an entire disk volume.

U2: Select cryptographic hash function: the user in this step selects a hash function which is already publicly available. This paper will be based on three hash functions namely MD-5, SHA-1 and SHA-256.

U3: Generate hash string: the user invokes the hash function to generate a fixed length hash string from a variable length input. The generated hash string must be collision resistant

U4: Store hash in database: In this step, the generated hash strings for each digital asset is stored in a database for future retrieval and verification of the integrity of the digital asset involved.

U5: Retrieve hash string: the stored hash string can be retrieved when issues of data integrity arise.

U6: Verify contents of digital asset: In the event of an integrity issue on an identified digital asset, its retrieved hash string can be matched with a current version of the hash of such an asset to verify the alteration or otherwise of its contents. Where there is a mismatch between the two, it can be easily confirmed that an alteration has been effected on the digital asset and vice versa.

3.4 CLASS DIAGRAM

The class diagram as shown in Booch (2005), depicts the internal relationships among the different entities like the user, hash function, and so on. In other words, it can be stated that the class diagram shows the static structures of the system thereby displaying its logical platform to the user. The generic class diagram of the user and hash function in the proposed system is shown in Figure 3.

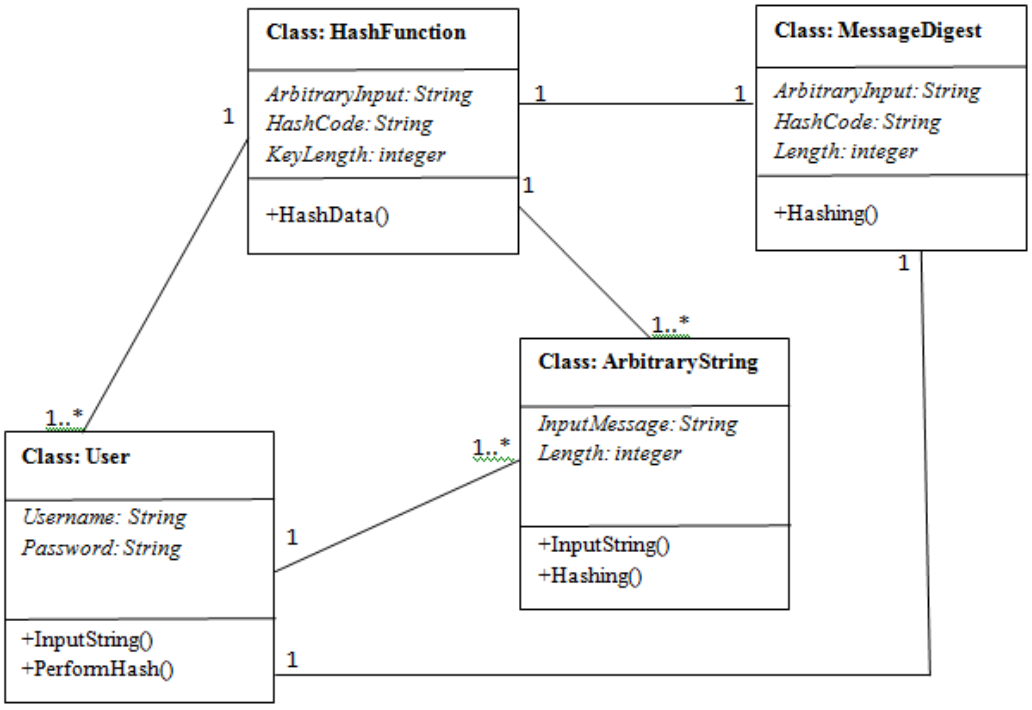


Figure 3: Class diagram of the proposed system (1:1 - one-to-one relationship, 1:1..* - one-to-many relationship, 1..*:1 - many-to-one relationship)

As shown in Figure 3, Class:MessageDigest represents the message digest of the hash algorithm used to verify the integrity of stored or transmitted data. Class:User represents the users of the system who are forensically protecting their data, Class:HashFunction represents the hashing algorithm that generates the message digest used for data integrity checking and Class:ArbitraryString represents the input to the hash function. This message digest serves the security token for the data being protected. The interactions and associations between the four classes are also shown in Figure 3.

The hash function as a class has such attributes as arbitrary length input, generated hash code and the length of the hash key. This class has a method called HashData() that hashes the arbitrary length string to produce a fixed length message digest. As shown in the diagram, there is a one-to-one mapping between the hash function and the message digest produced as a result of the hashing process. The message digest is produced from the arbitrary length input, which can be a character string, a file or an entire disk. Many arbitrary length strings can be hashed by one hash function, thus representing a many-to-one relationship)1..*:1). At the same time, each user of the system can supply one or more arbitrary length strings to the hash function. There can be one or more users that can deploy the hash function for the generation of the message digest. However, only one message digest can be returned per arbitrary string fed into the hash function. This implies that two arbitrary strings cannot produce the same message digest thereby enhancing the security of the protected data. Each user that deploys the hash function must be identified by his or her login credentials including the username and password. This adds another security layer to the hashing process by identifying the actual user whose identity is mapped to the login credentials.

3.5 ACTIVITY DIAGRAM

Activity diagrams display the procedural flow of control between two or more class objects during the execution of an activity Booch (2005). The respective activity diagram is shown in Figure 4.

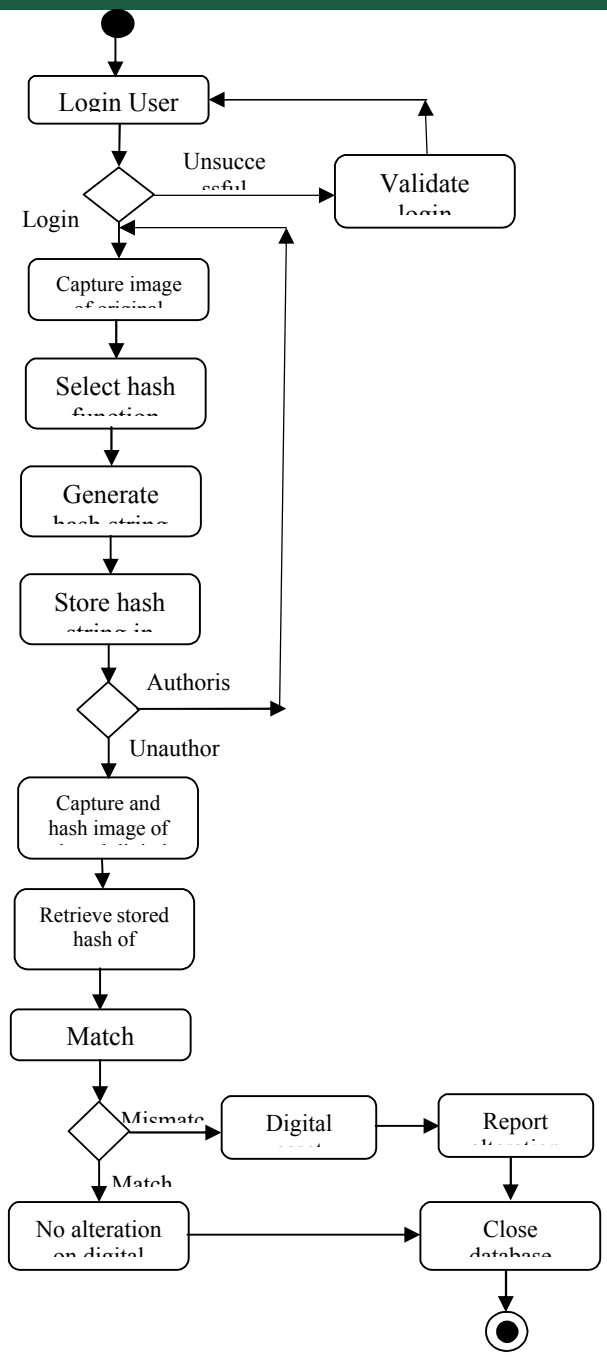


Figure 4: Activity diagram of the proposed system

The arbitrary input string captured into the hash function undergoes a hashing transformation process that sees it being hashed into a fixed length message digest that can be used to verify the integrity of digital assets. The integrity of any

digital asset can be verified through the hashing process – a situation that proves that the message has not been altered in any way by any process. When the hash of the original digital asset and its image match, it can be verified that the data, file or disk volume has not been tampered with. However, a mismatch shows a section of the contents of the digital asset has been modified arbitrarily and as such does not contain the exact data originally intended. This security check can help in protecting data from unauthorized modification or alteration.

4.0 EXPERIMENTAL RESULTS AND DISCUSSION

This section described the various experiments that were performed to authenticate the efficacy of the proposed system. As shown in the design of section 3, the various components of the proposed system will be implemented using Java (programming language) and MySQL (RDBMS). The experiments were run an Intel® Pentium ® CPU N3520 running Microsoft Windows 10 64-bit operating system, dual core 2.16GHz with 4.0GB RAM.

4.1 PSEUDOCODE

This section present a pseudocode for the acquisition of hashes of digital assets meant for the verification of their alterations or otherwise by the processes that use them. This pseudocode is presented as a step-wise process with control structures that also allow for selection and repetition of some processes. Three hash functions are considered in this work. These include MD5, SHA-1, and SHA-256 hash functions respectively. The user is given the options to choose from any of the hash functions used in this work to effect an integrity-check process for the digital assets or use all the available hash functions for more critical assets. The pseudocode for this implementation is shown below:

1. *Login module:*
    a. *if (new user) {*
        i. *create user profile*
    b. *} else {*
    c. *repeat*
        i. *authenticate user with username and password*
    d. *Until (login is successful) }*
2. *Capture image of original digital asset - odigasset*
3. *Select hash function; MD5, SHA-1,SHA-256*
4. *Generate hash string of odigasset - hashold*
    a. *If (MD5) {*
        i. *md5HashString = HashGeneratorUtils.generateMD5(argument);}*
    b. *if (SHA1) {*
        i. *sha1HashString = HashGeneratorUtils.generateSHA1(argument);}*
    c. *if (SHA256) {*
        i. *sha256HashString = HashGeneratorUtils.generateSHA256(argument);}*
5. *Open database connection*
    a. *insert hash records of odigassets into hashTable in database*
6. *if (authorised update) {*
    a. *update hash database with hash of updated asset }*
7. *if (unauthorised alteration) {*
    a. *capture image of altered digital asset - adigasset*
    b. *generate hash of adigasset - hashnew*
    c. *query database for hashold*
    d. *compare hashold with hashnew*
    e. *if (match found) {*
        i. *Odigasset not altered*
    f. *}else*
        i. *report alteration on odigasset*
8. *Close database*
9. *Stop*

The pseudocode depicts the structure of the proposed system. As shown in the pseudocode, the user logs into the system and is able to capture the hashes of the digital assets he intends to protect. The generated hash of the original digital asset is represented as hashold in the pseudocode. These hashes are stored in a database and can be retrieved in the event of a suspected unauthorised

modification of the target digital asset. When an alteration is suspected, the state of the digital asset at the time of the alteration is hashed – this hash is depicted as hashnew in the pseudocode. A comparison between hashold and hashnew is carried out to detect altered contents of a specific digital asset. This follows that at any point in time, the value of hashold must be equal to hashnew for a digital asset to maintain its integrity unless otherwise such an alteration is an authorised update.

.

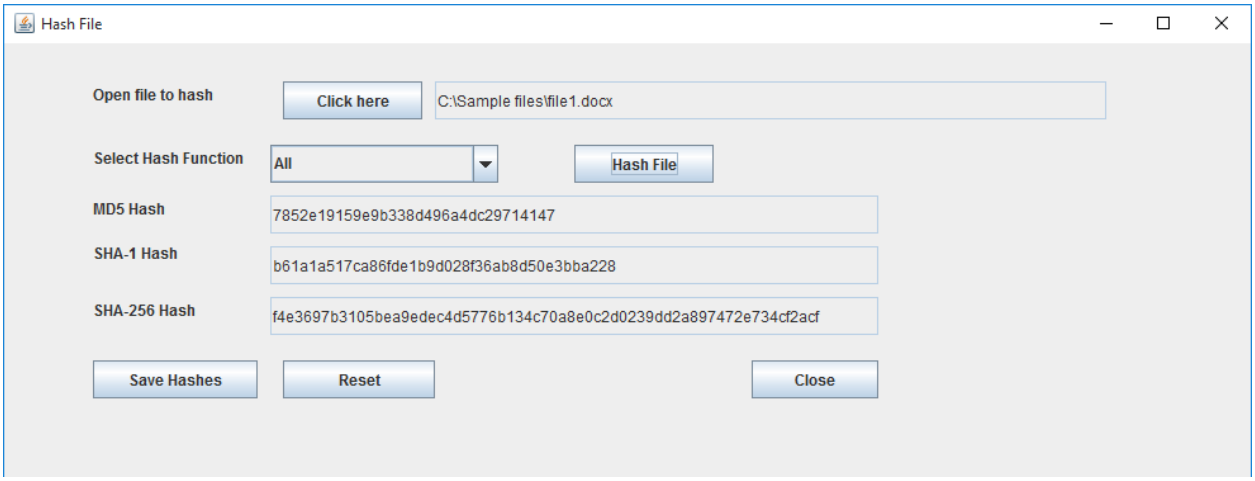


Figure 5: Interface showing file hashed with MD5, SHA-1 and SHA-256 respectively

.

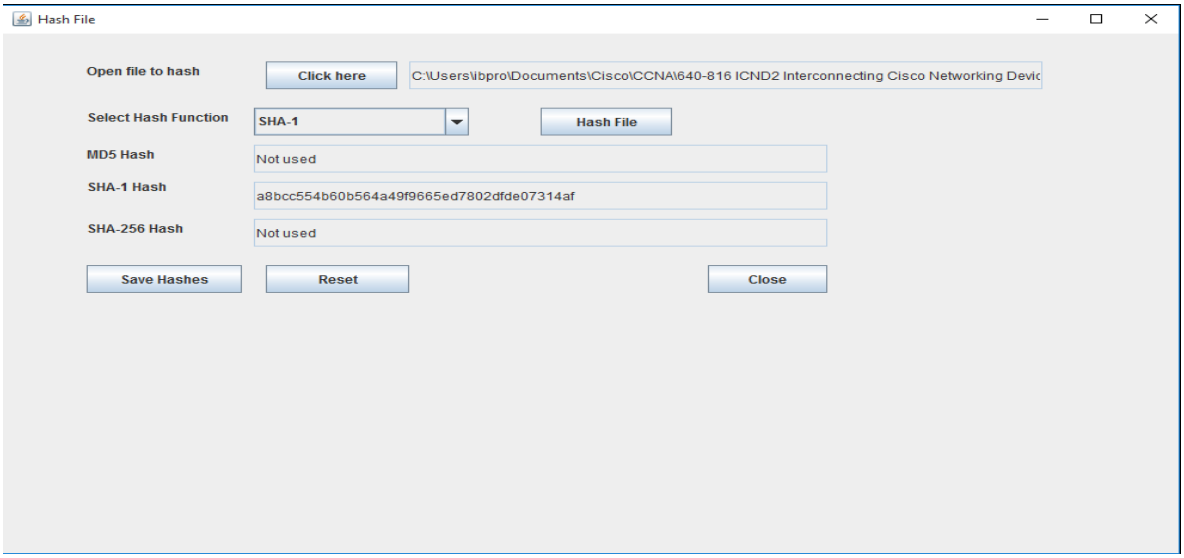


Figure 6: Interface showing file hashed using only SHA-1 hash function

4.2 SAMPLE RUN

The output from the execution of the Java code for this system is shown in Figures 5, 6, and 7 respectively. After the authentication module, where the user logs into the system with valid authentication credentials, the user can select the file to hash as shown in Figure 5.

The file selected is then hashed using all or one of the specified hash functions (MD5, SHA-1 and SHA-256). The entire file, irrespective of its size, is hashed into a fixed length string that is collision resistant. Figure 6 depicts the same file hashed using only SHA-1 hash function.

The hash(es) generated for a file are stored in a database (MySQL) when the ***Save Hashes*** button is clicked, and used to verify the integrity of the file on a later date. When an update is authorised, the hash of the file can be updated to reflect the authorised update. However, the hash will change. This change can be used to detect an unauthorised alteration of the file. All stored hashes can be

List of file hashes for MD5, SHA-1 and SHA-256 algorithms

filepath	md5hash	sha1hash	sha256hash
C:\Users\ibpro\Documents\PORTFOLIO 5.docx	f3a17714a27e8017e29b2b8fb352a998	0a88ed1ef42a7d8d01239e1bbc1dc36366687a31	5332c027e68f8a9af7fc32ce7e567a9df5c8a8fff9bdf4e607a3...
C:\Sample files\file1.docx	7852e19159e9b338d496a4dc29714147	b61a1a517ca86fde1b9d028f36ab8d50e3bba228	f4e3697b3105bea9edec4d5776b134c70a8e0c2d0239dd2...
C:\Users\ibpro\Documents\Books\Advanced-Persistent-Thr...	dbbe50972f16fd40dd5ee845b4cdc8b7	14b5a94b5de723de1927b8b0ac40df1df03d0fb0	cc3581fc53f982d8f7309e3599dc7afd4d613276b542f40ba1...
C:\Users\ibpro\Documents\Conferences\Conference reque...	344f2cbd7c81c52035578a424698c96a	Not used	Not used
C:\Users\ibpro\Documents\Cisco\CCNA\640-816 ICND2 Int...	Not used	a8bcc554b60b564a49f9665ed7802dfde07314af	Not used

Figure 7: A list of hashed files stored in the database.

All stored hashes can be verified for integrity whenever there is an identified breach on the files. Also, a periodic integrity check can be carried out to ensure that the contents of the files have not been altered. The integrity check module as shown in Figure 9 retrieves the stored hash of the file and compares it with the current hash of the file to detect a change in bit string in the stored contents of the file. If such a change is detected, and for an unauthorised alteration, the user or security administrator can easily report a breach on the file(s).

In Figure 9, it is clearly shown that the stored and current hashes of the file are equal. This is an indication that the file has not been altered. However, when the file is altered, it will generate a different hash from the stored hash as shown in Figure 10.

It is glaring that an update can be authorised and as such the user or security administrator can update the hash of the file in the database to reflect the update using the ***Update*** button on the interface (Figure 9). When such an update was not authorised, a security breach report can be generated to depict the state of the file at the time it was accessed.

Hashes of deleted files can also be erased from the database to prevent data redundancy. The ***Delete*** module as shown in Figure 11 can be used to achieve this task. The hash database is queried using the file path of the hashed file. All stored file

hash record module is invoked. When a file path is selected from the combo box, and the ***Delete File Path*** button is clicked, the hash record of the file is erased from the database. This allows the database to store only hashes of files and other digital assets currently residing on the disk volume. However, hashes of files deleted from the disk or other digital assets can be kept for future use where necessary

Figure 11: Interface showing the Delete module for erasing hashes of deleted files.

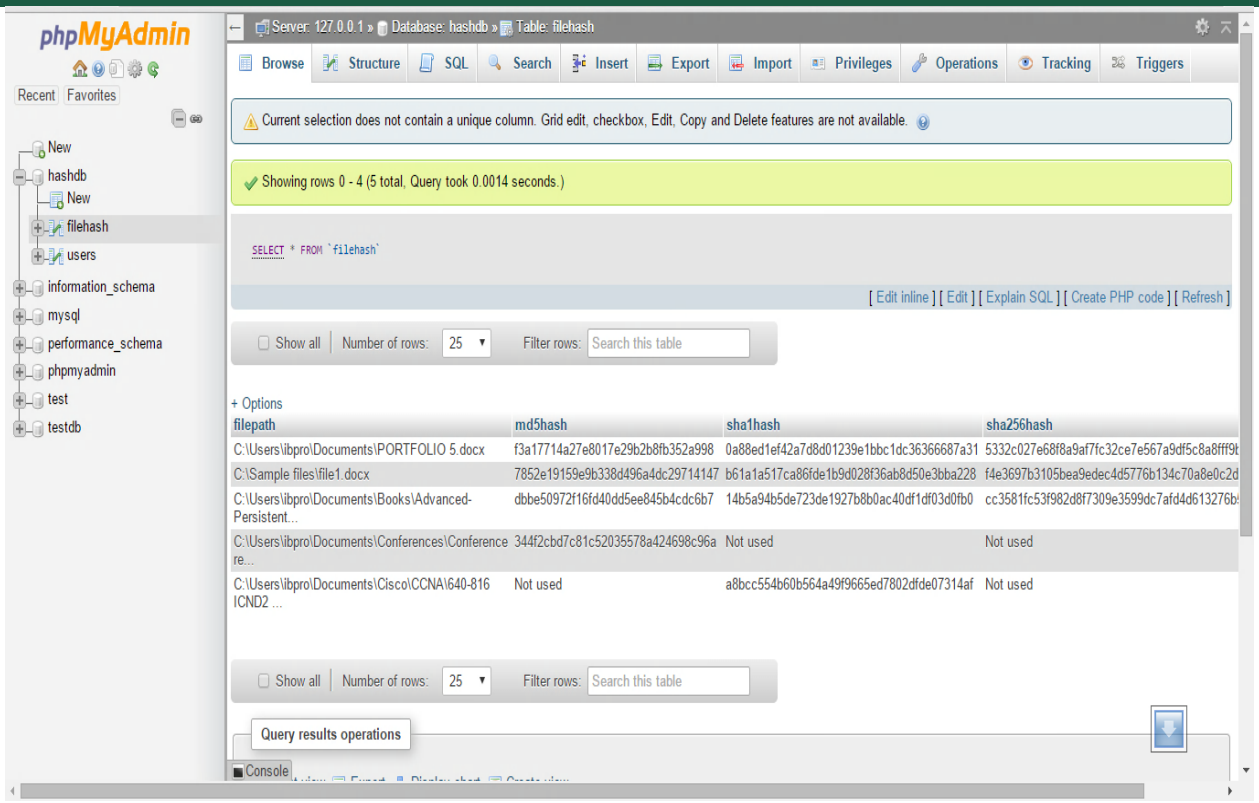


Figure 8: Interface showing hashes of selected files stored in MySQL database

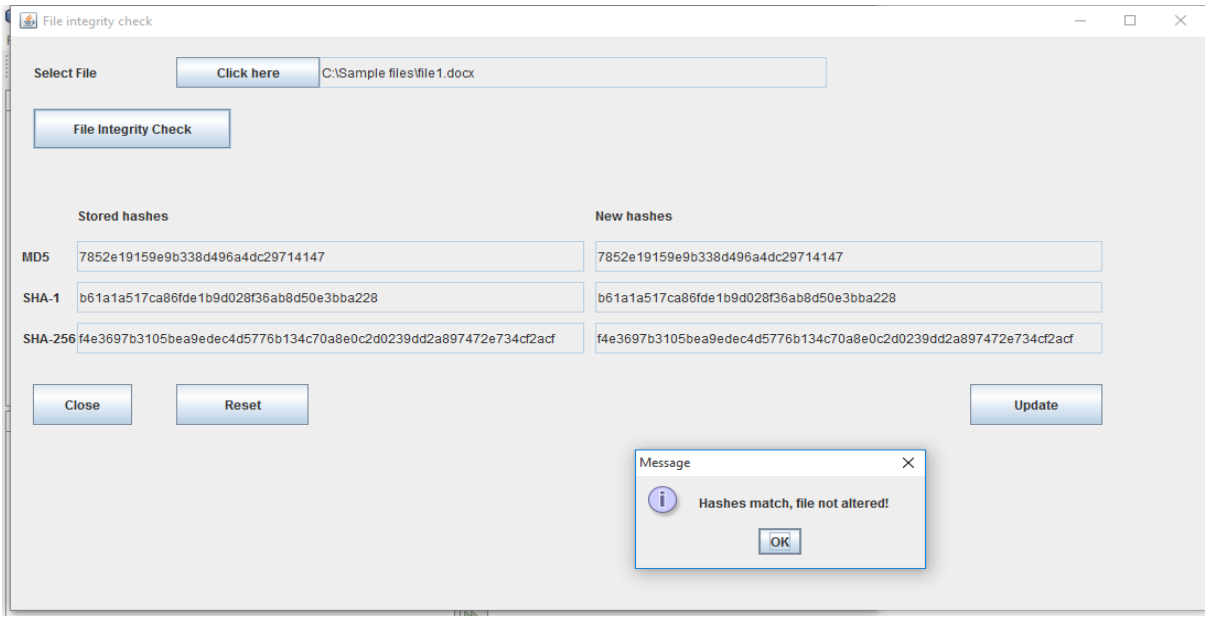


Figure 9: File integrity check module showing an unaltered file

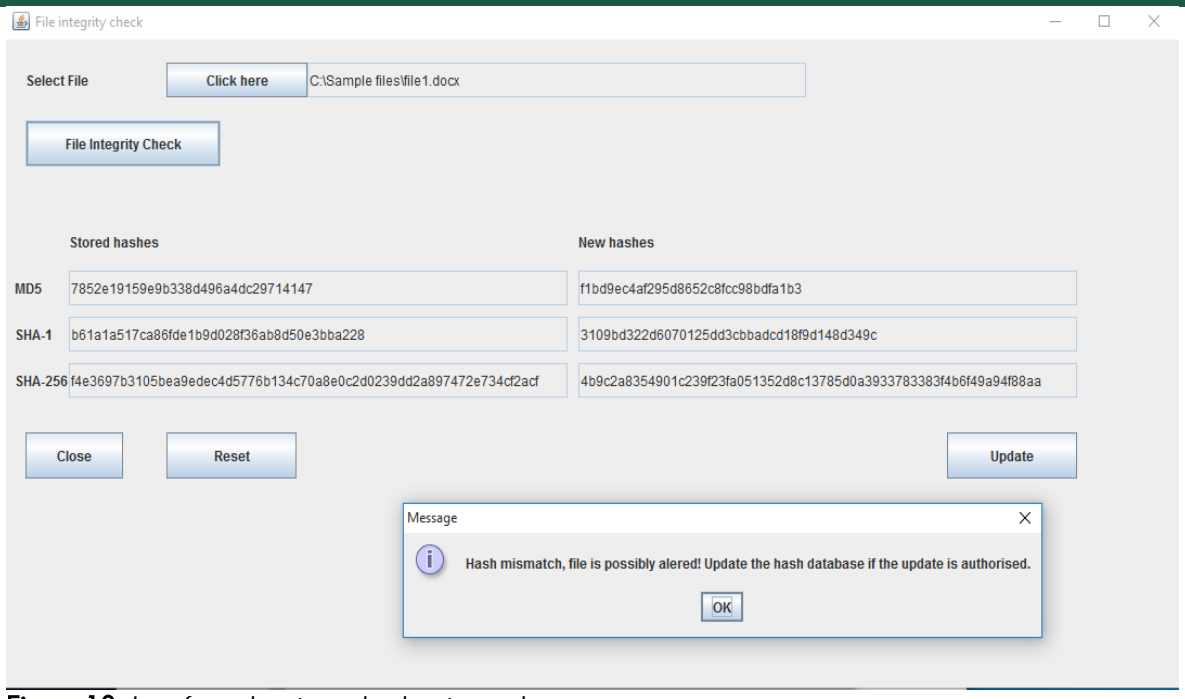


Figure 10: Interface showing a hash mismatch

5.0 CONCLUSION

The integrity of files, disk volumes and databases require periodic checks for security breaches. This is not unconnected with the fact that the Internet has made it possible for users to freely have access to digital assets even at remote locations. The aftermath of this possibility is not far – fetched.

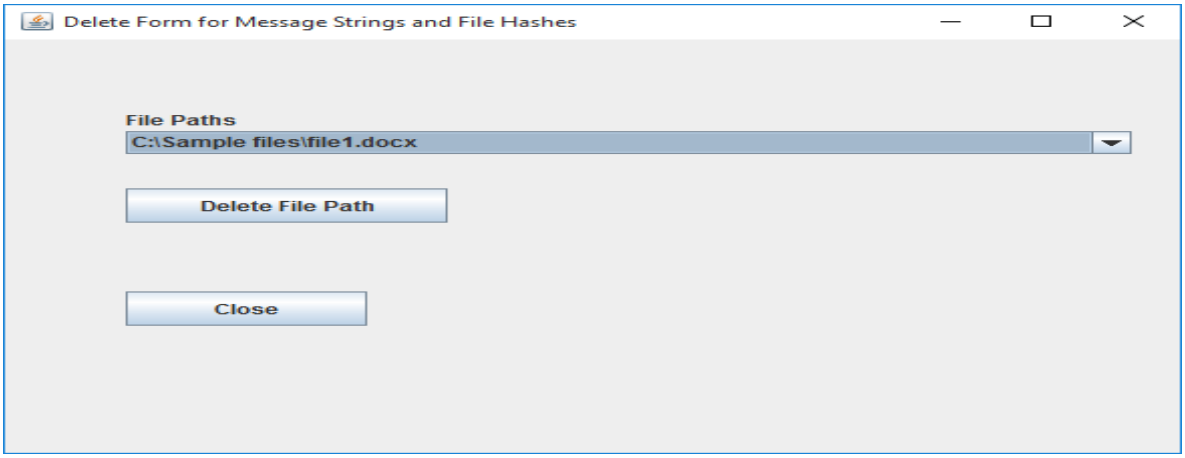


Figure 11: Interface showing the Delete module for erasing hashes of deleted files

Data theft, loss, distortion, and compromise are common consequences of the lack of integrity checks on digital assets. To this effect, it can be seen that hashing serves as an effective method of verifying the integrity of digital assets in whatever form they may appear. When a file, disk volume or database is hashed, a change in its contents will generate a different hash from the initially stored hash value of the same asset. This can be used to verify the alteration or otherwise of the asset. As shown in this work, the hash of a file can go a long way to identify when a file is accidentally, legally, or criminally modified at any point in time during its use. This work will be relevant to all categories of users and organisations in protecting the integrity of their digital assets.

6. REFERENCES

Ahmad, S. and Ahmad, R., 2010, June. An improved security framework for data warehouse: a hybrid approach. In *2010 International Symposium on Information Technology* (Vol. 3, pp. 1586-1590). IEEE.

Booch, G., 2005. *The unified modeling language user guide*. Pearson Education India.

Bui, S., Enyeart, M. and Luong, J., 2003. Issues in Computer Forensics. *Santa Clara University Computer Engineering, USA*.

Li, F., 2012, August. Study on security and prevention strategies of computer network. In *Computer Science and Information Processing (CSIP), 2012 International Conference on* (pp. 645-647). IEEE.

Preneel, B., Govaerts, R. and Vandewalle, J., 1993, August. Hash functions based on block ciphers: A synthetic approach. In *Annual International Cryptology Conference* (pp. 368-378). Springer Berlin Heidelberg.

Roussev, V., 2009. Hashing and data fingerprinting in digital forensics *Computing in Science and Engineering*, *7*(2), pp.49-55.

Rowlingson, R., 2004. A ten step process for forensic readiness. *International Journal of Digital Evidence*, *2*(3), pp.1-28.

Rumbaugh, J., Jacobson, I. and Booch, G., 2004. *Unified Modeling Language Reference Manual, The*. Pearson Higher Education.

Sun, X., 2011, December. The study on computer network security and precaution. In *Computer Science and Network Technology (ICCSNT), 2011 International Conference on* (Vol. 3, pp. 1695-1698). IEEE.

Yannikos, Y., Schluessler, J., Steinebach, M., Winter, C. and Graffi, K., 2013, January. Hash-based File Content Identification Using Distributed Systems. In *IFIP International Conference on Digital Forensics* (pp. 119-134). Springer Berlin Heidelberg.

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

CLUSTERING MIXED DATASETS WITH MULTI-SWARM OPTIMIZATION AND K-PROTOTYPE ALGORITHM

C. P. Oleji
Department of Computer Science
Federal University of Technology,
Owerri
olejichukwuemeka1@gmail.com,

E.C.Nwokorie
Department of Computer Science
Federal University of Technology,
Owerri
cenwokorie@gmail.com

F.E. Onuodu
Department of Computer Science
University of Port Harcourt, River State

O. D. Obinna
Business Administration,
National Open University Nigeria

ABSTRACT

Clustering is produced by grouping objects with high degree of relationship from object with low degree of relationship, such that object found in a group are highly similar and share common attributes that is distinct from the other groups. One major problem of data clustering is to know accurately the number of clusters that can be formed out of a set of data. Clustering mixed data set is also another problem that is faced with clustering models. This article proposed a hybrid clustering algorithm called Multi-swarmK-prototype clustering algorithm for clustering mixed dataset. Multi-swarmK-prototype algorithm consists of multi-swarm optimization and k-prototype clustering algorithm. The multi-swarm optimization algorithm was used to improve the convergence accuracy of the objective function of traditional k-prototype by random search for the initial global best value of k. Six datasets (yeast, soybean, Hepatitis, Australian Credit Approval, German Credit Data and Statlog Heart) obtained from University of California, Irvine (UCI) Machine Learning Repository was used to demonstrate the clustering performance of our algorithm. From the experimental results of soybean and yeast, the proposed Multi-swarmK-prototype had accuracy of (0.971 and 0.973) while MixK-meansKFon had accuracy of (0.86 and 0.84). From the experimental results of Hepatitis, Australian Credit Approval, German Credit Data and Statlog Heart datasets; Multi-swarmK-prototype algorithm had accuracy of (0.9169, 0.9789, 0.9495, 0.9208) while PSO based K-prototype algorithm had accuracy of (0.7521, 0.8229, 0.6261, and 0.8387). The proposed hybrid clustering algorithm is highly proficient for clustering mixed large datasets than MixK-meansKFon and PSO based K-prototype algorithm. It gave a very efficient clustering convergence and minimized the time complexity for clustering large dataset.

Keywords: Clustering, k-prototype algorithm, Mixed Dataset, Multi Swarm Optimization and Multi-model problems.

INTRODUCTION

Clustering is the process of partitioning a given *t* particles embedded in *g* dimension metric space in different groups or clusters, such that objects in the same group are highly similar and possesses distinct attributes from the objects in the other group. Clustering mixed large datasets is a fundamental problem in data science. Various researchers have tried to develop efficient clustering algorithms, but none was very efficient for clustering mixed datasets. Mixed dataset consist of numerical data (numbers like 1.2, 3.2, 0, 7 etc.) and categorical data (like gender, religion, rates, A, B, attitudes etc.). This paper proposed a hybrid algorithm called Multi-swarmK-prototype clustering algorithm for clustering mixed datasets. It consists of multi-swarm optimization and k-prototype clustering algorithm. Multi-swarm optimization is a variant of Particle swarm optimization (PSO) based on the use of multiple sub-swarms instead of one (standard) swarm. Particle swarm optimization is a population based stochastic optimization technique that can be used to find an optimal, or near optimal, solution to a numerical and qualitative problem (Eberhart and Kennedy, 1995). The general approach in multi-swarm optimization is that each sub-swarm focuses on a specific region while a specific diversification method decides where and when to launch the sub-swarm. The system imitates the social characters shown by swarms of animals. In this algorithm a point in the search space, which is a possible solution, is called a particle. The group of particles in a specific alteration is called 'swarm'. While looking out for food, the birds are either scattered or go collectively before they find out the place where they are able to locate the food. While the birds are on the search for food moving from one location to another, there is often a bird which is able to smell the food effectively, in other words, the bird discerns the location where the food is likely to be found, having superior food resource data. They convey the data, particularly the excellent data at any time while looking for the food from one location to another. Attracted by the excellent data, the birds will throng at the location where there is strong possibility for locating food. To really obtain the initial value for k, the multi-swarm optimization algorithm enables k-prototype to obtain global minimum, thus arriving at nearly closest objects efficiently. The multi-swarm framework is especially fitted for optimization of multi-model problems, where multi (local) optima exists (Backwell and Branke, 2004).

The proposed hybrid algorithm provides solution to the limitations of the initial selection of k-prototype algorithm. It provides intelligent search of the initial value of k for accurate convergence of the objective function of traditional K-prototypes clustering algorithm. In multi-model problems, it is important to achieve an effective balance instead of trying to achieve a compromise between exploration and exploitation which could weaken both mechanisms of the search process. Multi- swarm system separates them into distinct phases. Each phase is more focused on either exploitation (individual sub-swarms) or exploration (diversification method) (Hendtasll, 2005).Clustering in large data is required to identify the existing patterns which are not clear at first glance. The properties of large data pose some challenges against adopting traditional clustering methods:

(i) *Type of dataset*: Most clustering algorithms can efficiently cluster either pure numerical and categorical attributes but cannot cluster mixed datasets effectively. Some researchers had tried to solve this problem but ended up creating more problems to be solved. The collection of data in a real world contains both numerical and categorical data.

(ii) *Size of dataset*: The time complexity of the clustering process and clustering quality of different clustering algorithm depend on the size of dataset. Small datasets are clustered with k-medoid algorithm because it is very efficient to cluster small numerical datasets than the traditional k-means that is well known to cluster large numerical values.

(iii) *Handling outliers/noisy data*: Most times, high dimensional data used for real applications are wrongly computed from sensors, or during its tabulation. Noise which comprises of high or low values in a dataset, influence the results of most data mining techniques. In clustering algorithms, it affects the mechanism of clustering output most especially the ones that uses distance measure to

obtain objects' centroids. An efficient algorithm must be able to handle outliers/noisy data.

(iv) *Time complexity*: Most traditional clustering algorithms are used repeatedly to obtain better accuracy and quality clustering. Optimization of time complexity is precisely important for clustering high dimension mixed data.

(v) *Stability*: Stability corresponds to the ability of an algorithm to generate the same partition of the data irrespective of the order in which the data are presented to the algorithm. That is, the results of clustering should not depend on the order of data (Jain and Verma, 2014).

(vi)*High dimensionality (Curse of dimensionality)*: As the number of dimensions increases, the data become increasingly sparse, so the distance measurement between pairs of points becomes meaningless and the average density of points anywhere in the data is likely to be low. Therefore, algorithms which partition data based on the concept of proximity may not be fruitful in such situations.

(vii) *Cluster shape*: A good clustering algorithm should be able to handle real data and their wide variety of data types, which will produce clusters of arbitrary shape. Many algorithms are able to identify only convex shaped clusters.

Related Works

Prabha and Visalakshi (2015) proposed a new variant of binary particle swarm optimization and K-prototype algorithm to reach global optimization for clustering optimization problem. They used dataset from UC Irvine machine learning repository to evaluate the accuracy of their proposed system. Their results show that particle swarm based on k-prototype algorithm provided better clustering accuracy than the traditional k-mode and k-prototype clustering algorithm. Though their result produced accurate clustering, it is time consuming for a swarm to search for X particles moving around D-dimensional search space. Mohanavlli and Jaisakthi (2015) proposed a chi-square based statistical approach to determine the weight of the attributes which derive the distance matrix of mixed dataset. They demonstrated the performance of their algorithm using real life dataset (heart, credit and vote dataset) from UCI machine learning repository. Though they obtained better entropy than other existing clustering algorithms, their approach needs better startup weights for efficient distance computation to improve their results. They also have the problem of the initial selection of values for k. Jain and Verma (2014) proposed an approximate algorithm based on k-means. The algorithm provided mechanism to eliminate the drawback of k-means of uncertain number of iterations by fixing the number of iterations, without losing the precision. It is an innovation for clustering large dataset and it is very fast, scalable and has high clustering accuracy. But it could not handle categorical dataset well unless it is converted to equivalent numerical data. Their system inherited the problem of initial selection of k values which leads to premature convergence of the process. They stated that machine learning concept can be used to decide the priority of attributes instead of asking from the user. Manjinder, et al (2014) used adaptive K-means techniques for data clustering. It adapts itself according to the image based on color based clustering. The number of clusters using the color features is computed based on histogram analysis in gray format. The peak of the histogram is the main source of computation of number of colors in the image and based on the same, the image data are clustered. (Nielse et al, 2014) proposed extension of k-means clustering to mixed divergences. They extended the k-means++ seeding to mixed divergences and reported a guaranteed probabilistic bound. Then they described a soft clustering technique for mixed-divergences. Their work is still ongoing.

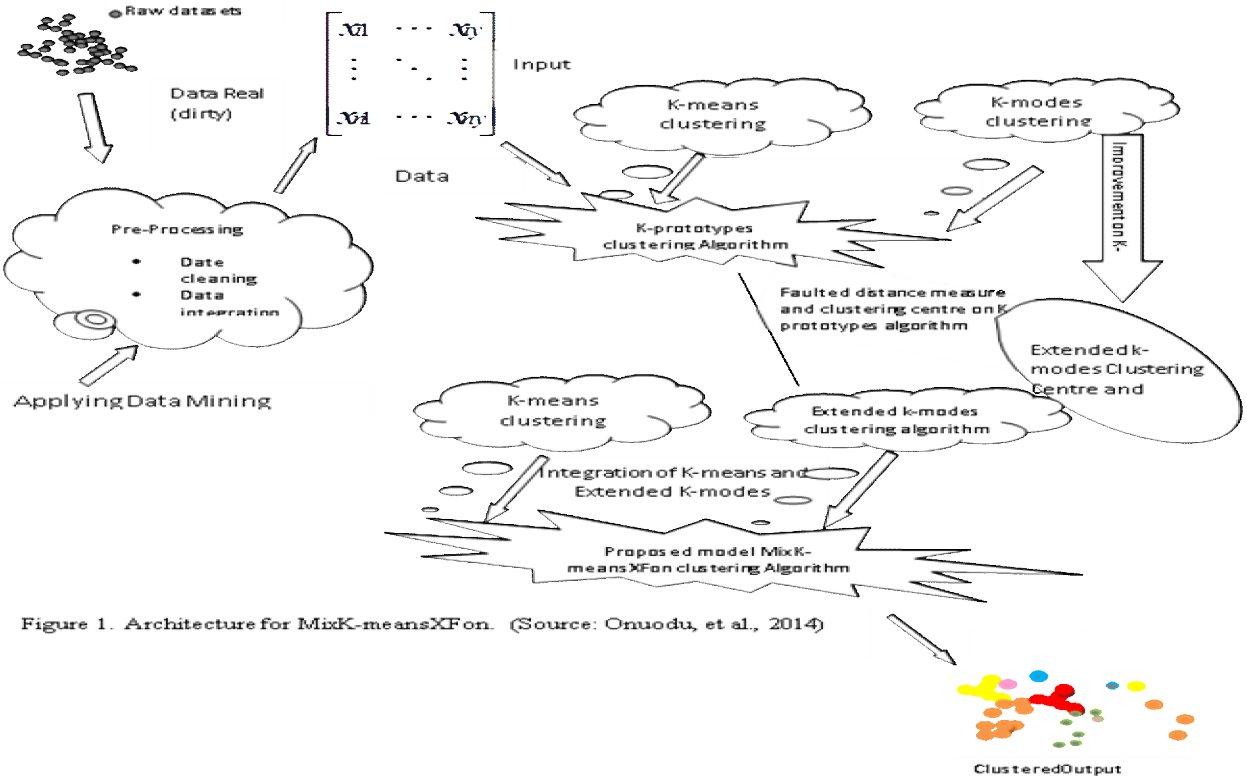


Figure 1. Architecture for MixK-meansXFon. (Source: Onuodu, et al., 2014)

Critical examination of the work of Prabha and Visalakshi (2015) gave a clear understanding of clustering as an optimization problem. Their work was a new variant of binary Particle Swarm Optimization and K-Prototype algorithms to reach global optimal solution for clustering optimization problem. Onuodu, et al (2014) proposed a new hybrid method called MixK-meansXFon clustering algorithm which extends K-means algorithm to categorical domain and mixed-type attributes. They also developed a new dissimilarity measure that uses relative cumulative frequency-based method in clustering objects with mixed values. The limitation of their new algorithm is that the value of k, which is the number of desired clusters, is still required to be given as input, regardless of the distribution of the data points in clustering mixed dataset. The architecture of their system is shown in Figure1. This paper adopts a hybrid clustering algorithm consisting of MixK-meansXFon (Onuodu, et al, 2014) and Multi-swarm intelligent optimization algorithm (Chen and Montgomery, 2011) that is a variant of Particle swarm optimization (PSO) based on the use of multiple sub-swarms instead of one (standard) swarm. This approach efficiently improves the clustering performance of the traditional k-prototype algorithm. The work of Onuodu, et al (2014) proposed a new hybrid method which extends K-means algorithm to categorical domain and mixed-type attributes. They developed new dissimilarity measure that uses relative cumulative frequency-based method in clustering objects with mixed values. The goal of this paper is to solve the problems of the initial values of k-means clustering algorithm.

Methodology

Object Oriented Analysis and Design Methodology were used for the design and implementation of the hybrid clustering algorithm with java programming language and MATLAB database. This is the process of defining the problem in terms of real world object with which the system must interact, and the software is used to explore various solution alternatives. Real world objects can be defined in terms of their classes, attributes and operations. This methodology features: the algorithm, high level model language, use case diagram, activity diagram and the architectural design of the system.

Proposed Hybrid Algorithm

The algorithm for Multi-swarmK-prototype clustering is shown in figure 2. After initialization, the multi-swarm algorithm iterates a main loop with four stages: Test for function change, particle update, attractor update and exclusion.
These stages are described in figure 2.

Step 1: Initialize. randomly identify each particle position and velocity in search space, set particles position are randomized by all attractors, then set swarm attractors to particles attractors and Set all store all function values to function floor.REPEAT WHILE swarm f (f is the number of swarms to be deployed)// Test for change

Step2: Compute function at swarm attractor of swarm n (n is the sub-swarm deployed at that region).

Step3: IF value of next result is distinct from previous iteration THEN
3.1: Re-compute function value at each particle attractor
3.2: Bring up-to-date swarm attractor and stock function values.

Step 4: WHILE particle k of swarm f, // Bring up-to-date Attractor.
4.1: Compute function at updated position and stock value.
4.2: If value of next result better than particle attractor value THEN
4.3: Particle attractor k: = position and value of particle k. (where k is the current position of the particle)
4.4 IF value of next result better than swarm attractor value THEN
4.5: Swarm attractor: = position and value of particle k.
FOR EACH swarm y ≠f (y is the sub-swarms) //Exclusion
Step 5: IF swarm attractor p_f is within r_{excl} of pyTHEN
5.1: Randomize the swarm with the worse swarm attractor value.
5.2: re-instruct particle attractors, compute f at each new position, stock these values, and place attractor of swarm to the position of its best particle). UNTIL number of function evaluates performed> max

Step 6: Apply K-prototype algorithm on each of the partition of the dataset to get the initial clusters of each partition by the initial value from the result of swarm to the position of its best particle.
Step7: Calculate the clustering centre of each clusters in all partitions separately.
Step 8: Calculate the objective function of the mixed dataset.
Step 9: Repeat step 4 for all partitions.

Step 10: Finally, based on the minimum distance criterion, assign each data points to the cluster to which it has minimum distance. The sum of squared error within groups is minimized.

Figure 2: Proposed hybrid algorithm

Use case diagram of the proposed system

The use case diagram depicts sets of activities the proposed hybrid algorithm performed to produce some output result. The use case diagram (figure 3) consist of the following: Data Mining model, Multi-swarm intelligent search algorithm, determination of distance measure for attributes in datasets, computation of similarity matrix for numerical attributes, computation of similarity/dissimilarity matrix for categorical attributes, computation of dissimilarity for mixed attributes to cluster datasets and the clustered results.

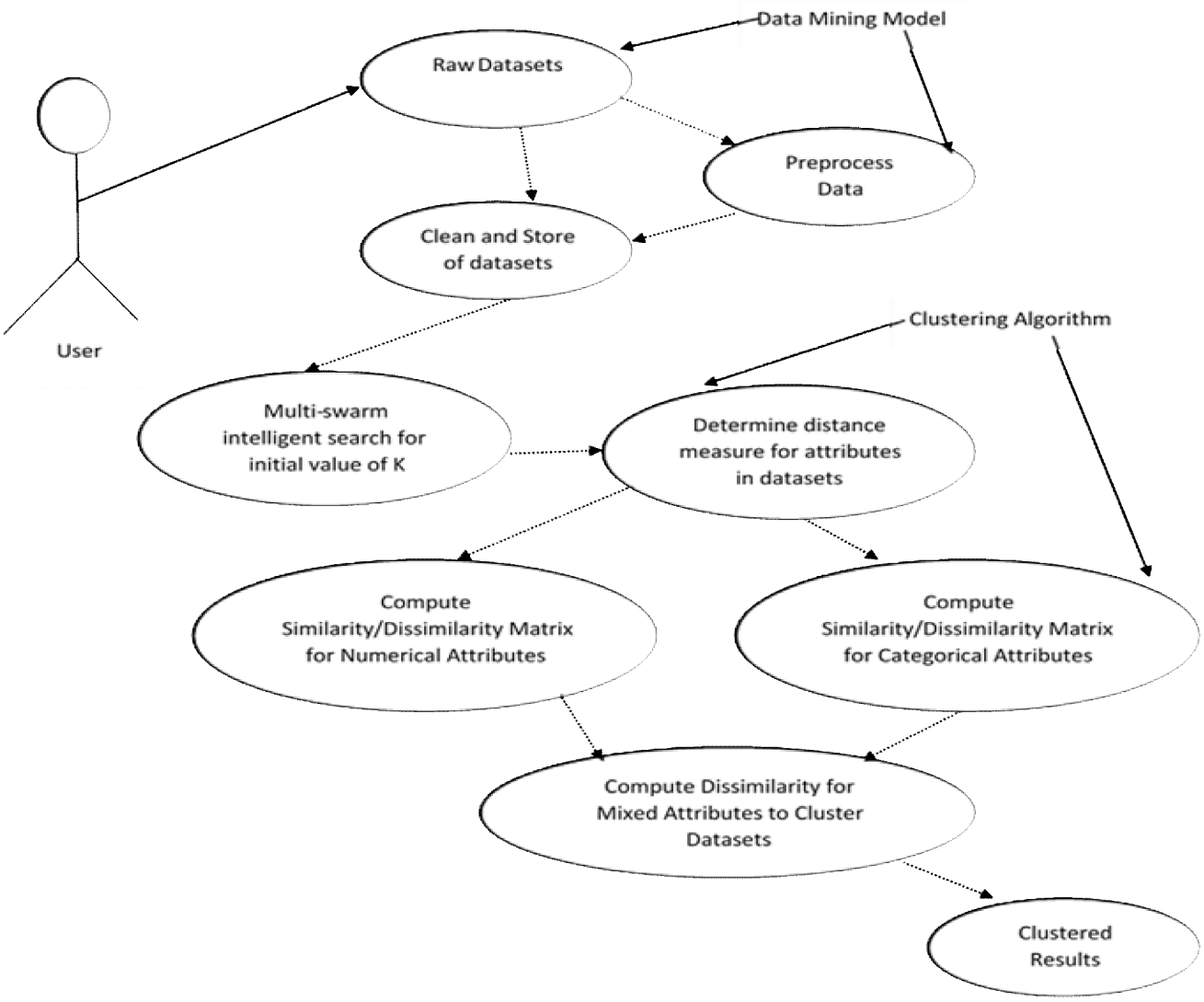


Figure 3: Use case diagram of the proposed system

.

Activity diagram of the proposed system

The proposed activity shows the dynamic behavior of the system which includes how data and information flows from one object to another. The activity diagram of the proposed system is shown in Figure 4.

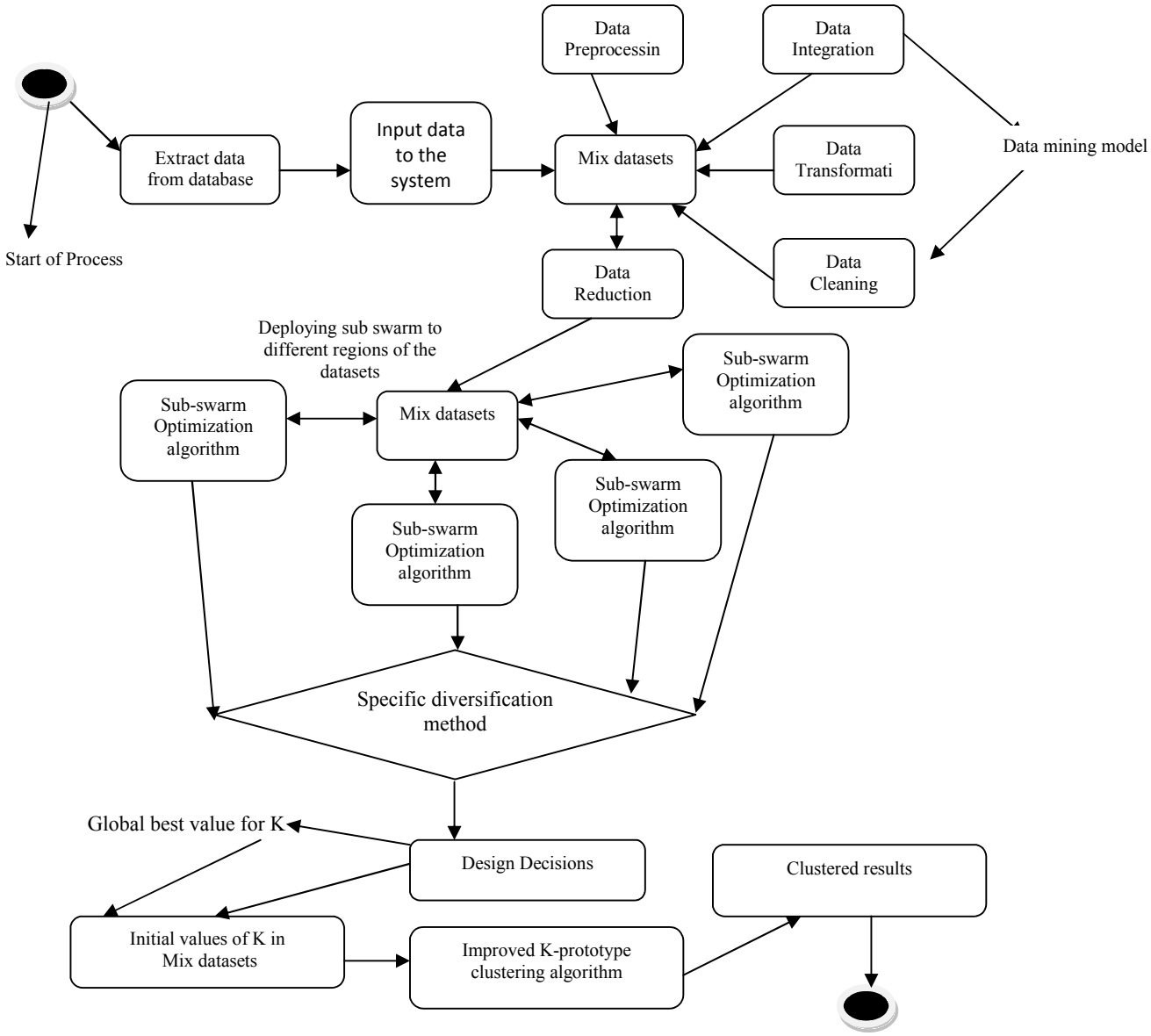


Figure 4: Activity diagram of the proposed system

The components of the activity diagram of the proposed system includes: data mining model, deployment of sub-swarms to different regions of the dataset and improved k-prototype clustering algorithm. The dataset extracted from UCI machine learning repository was processed (data processing, data integration, data transformation, data cleansing and data reduction) using data mining model. The result of the data mining model is used as input to the Multi-swarm intelligent search optimization algorithm to determine the landscape of the datasets. The mechanism of the particle attractor and swarm

attractor were used to deploy sufficient sub-swarm for the search of the initial global best value of k in the whole region of the dataset. The activities of these sub-swarms are controlled by specific diversification method via the design decision to choose the best global value of k (initial value of k) from the results of each of the sub-swarms. The global best value of k obtained by the swarm intelligent mechanism is used by the clustering algorithm to produce efficient clustering outputs.

Analysis

Multi-swarm optimization algorithms are mainly used to handle multimodal environment problems, such as many peaks. Here, using multi-swarm to partition the entire space into many sub-spaces, each swarm optimizes the location of the best global particle in a local environment. It intelligently detects the most promising area using the swarm attractor or the parent swarm and then group of sub-swarms are deployed to search the local optimum within their own sub-space. Each child swarm has its own search region defined as a sphere of radius r and centers on its best particle $\vec{s}$. Hence, a particle $\vec{x}$ with its distance less than r from $\vec{s}$ belongs to the child swarm (Backwell and Branke, 2004). The distance d $(\vec{x};\vec{s})$ between two points $\vec{x}$ and $\vec{s}$ in the n-dimension space is defined as the Euclidean distance as follows:

$$d(\vec{x},\vec{s}) = \sqrt{\sum_{i=1}^{n}(x_i - s_i)^2} \quad \text{.....(1)}$$

According to our experimental experience, the more peaks in the landscape, the more child swarms are relatively needed. r is relative to the range of the landscape and the width of peaks. In general, we set r according to the following equation:

$$r = \sqrt{\sum_{i=1}^{n}\left(x_i^u - x_i^l\right)}/\left(W_{\min} + c\left(W_{\max} - W_{\min}\right)\right) \quad \text{....(2)}$$

where x_i^u and x_i^l are the lower and upper bound on the i-th dimension of the variable vector of n dimensions (Backwell and Branke, 2004). W_{min} and W_{max} are the minimum and maximum

The hybrid algorithm combines the attributes of k-means and attributes of the extended k-modes algorithms with the sub-swarms search to efficiently generate the initial value of k to cluster mixed-type objects. The new hybrid method is more useful because most frequently occurring objects in real world databases are mixed-type objects. The dissimilarity between two mixed-type objects X and Y, which are described by attributes Ar_1, A^r_2,.., A^r_p, $A^c_{p+1,...}A^c_m$, can be measured by:

$$d_2(X,Y) = \sum_{j=1}^{p}\left(x_j - y_j\right)^2 + \gamma\sum_{j=p+1}^{m} rcf\delta\left(x_j, y_j\right). \quad (3)$$

Where the first term is numeric attributes and the second term is the simple matching dissimilarity measure on the categorical attributes. The weight γ is used here to avoid favoring either the categorical or numerical attributes. The influence of γ in the clustering process is discussed in Huang (1998).

Using equation (1) for mixed-type objects, it is more convenient to modify the cost function Equation (3) as follows:

$$P(W, Q) = \sum_{l=1}^{k}\left(\sum_{I}^{n} w_{i,l}\sum_{j=1}^{p}\left(x_j - y_j\right)^2 + \gamma\sum_{I=1}^{n} w_{i,I}\sum_{j=p+1}^{m} rcf\ \delta\left(x_j, y_j\right)\right) \quad \text{....(4).}$$

Let

$$P_l^r = \sum_{j=1}^{n} w_i j \sum_{j=1}^{p}\left(x_{i,j} - q_{i,j}\right)^2 \quad \text{..... (5)}$$

And

$$P_1^c = \gamma\sum_{i-1}^{n} w_{i,l}\sum_{j=p+1}^{m} rcf\ \delta\left(x_{i,j} q_{i,j}\right)^2 \quad \text{... (6)}$$

Equation (4) can be re-written as:

$$P(W, Q) = \sum_{l=I}^{k}\left(P_l^r + P_l^c\right). \quad (7)$$

Since both P^r_l and P^c_l are nonnegative, minimizing P(W; Q) is equivalent to minimizing P^r_l and P^c_l for $1 \leq l \leq k$; and rcf is the relative cumulative-frequency for the mixed data. The hybrid method was simulated with Java programming and MATLAB database for data extraction respectively. The misclassification matrix of each clustering result is also used to compute the clustering accuracy. Huang (1998) proposed a measure of clustering results called the clustering accuracy, defined as follows: $r = \frac{1}{n}\sum_{l=1}^{k} a_l$, ... (8)

where a; is the number of data objects that occur in both cluster a_l and its corresponding labeled class,and n is the number of objects in the data set. Further, the clustering error is defined as e = 1- r.

High level model for the proposed system

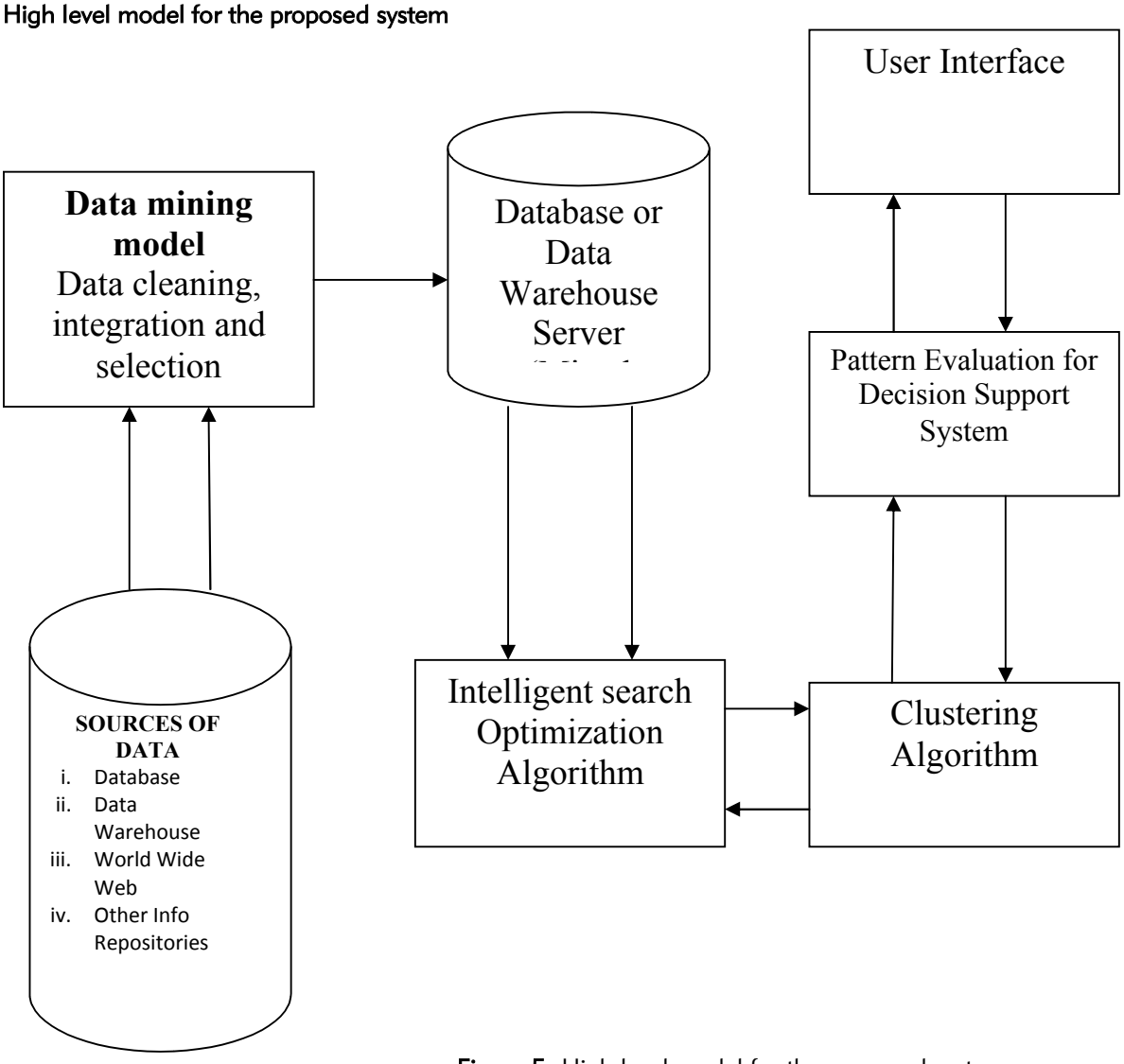


Figure 5: High level model for the proposed system

The High level model for the proposed system consist of source of data, data mining model, proposed database, intelligent optimization algorithm, clustering algorithm pattern evaluation and user interface .The high level model for the proposed system is shown Figure 5:

Proposed Architecture for Hybrid Clustering Algorithm

The Architecture of the Proposed Hybrid Clustering Algorithm is shown in Figure 6.

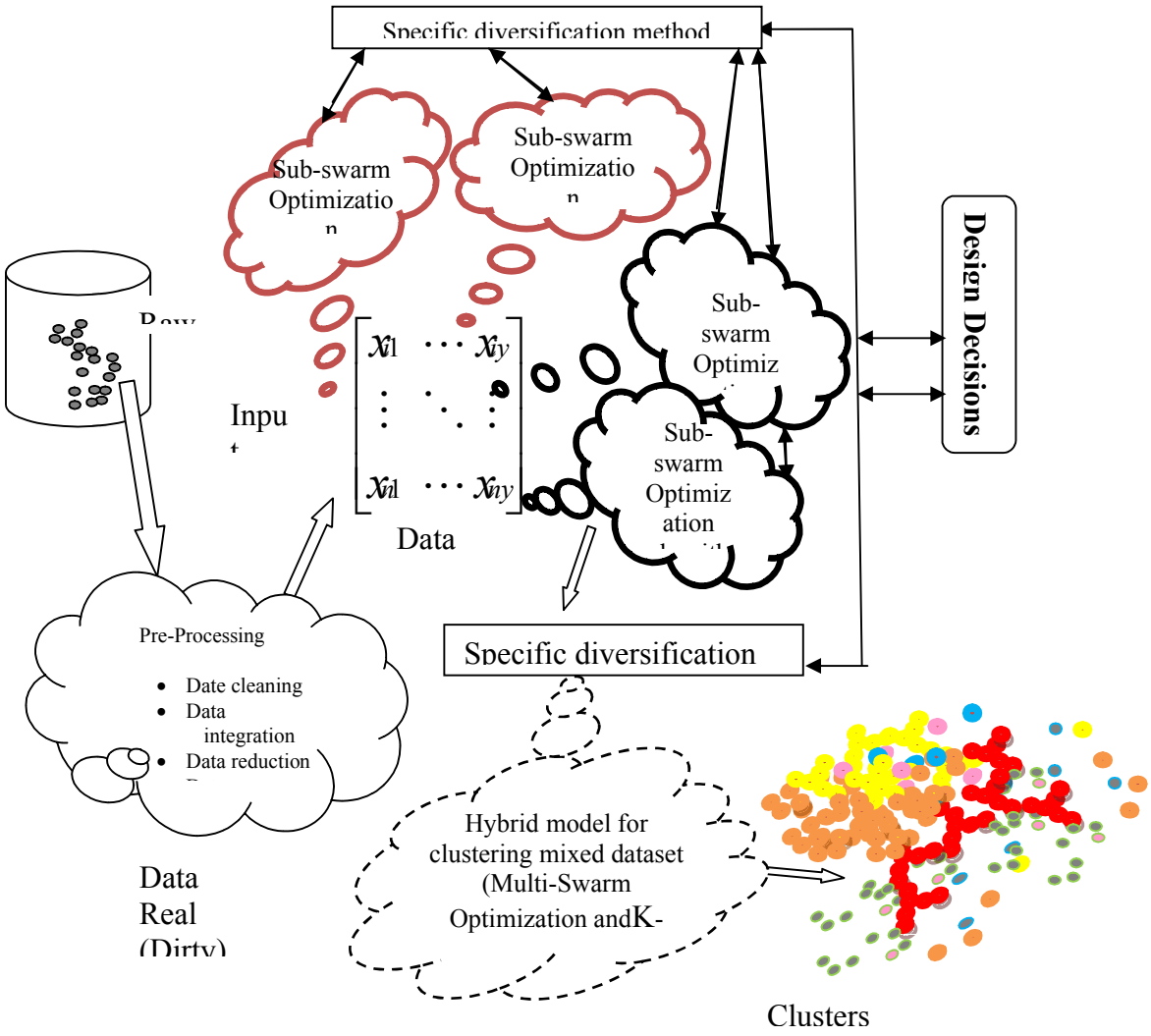


Figure 6: Architecture of Multi-Swarm Optimization and K-prototype

The hybrid algorithm improved the convergence accuracy of the objective function for the traditional k-prototype with the aid of multi-Swarm optimization. It enhances the clustering performance of k-prototype algorithm for clustering mixed dataset.

Results

The proposed system was developed with a very efficient user interactive interface. It enables the user to select different datasets and click on the cluster command button to produce the output. The result for soybeans clustered output is shown in Figure 7.

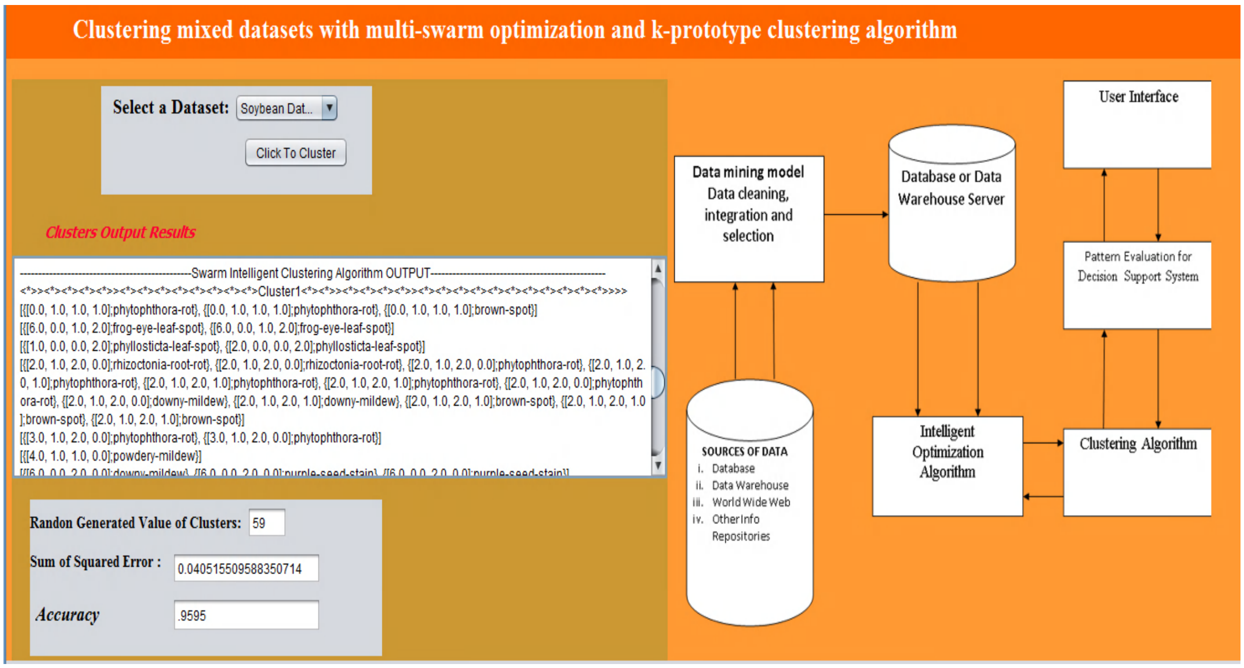


Figure 7: Soybean clustered output

The result for Yeast clustered output is shown in Figure 8.

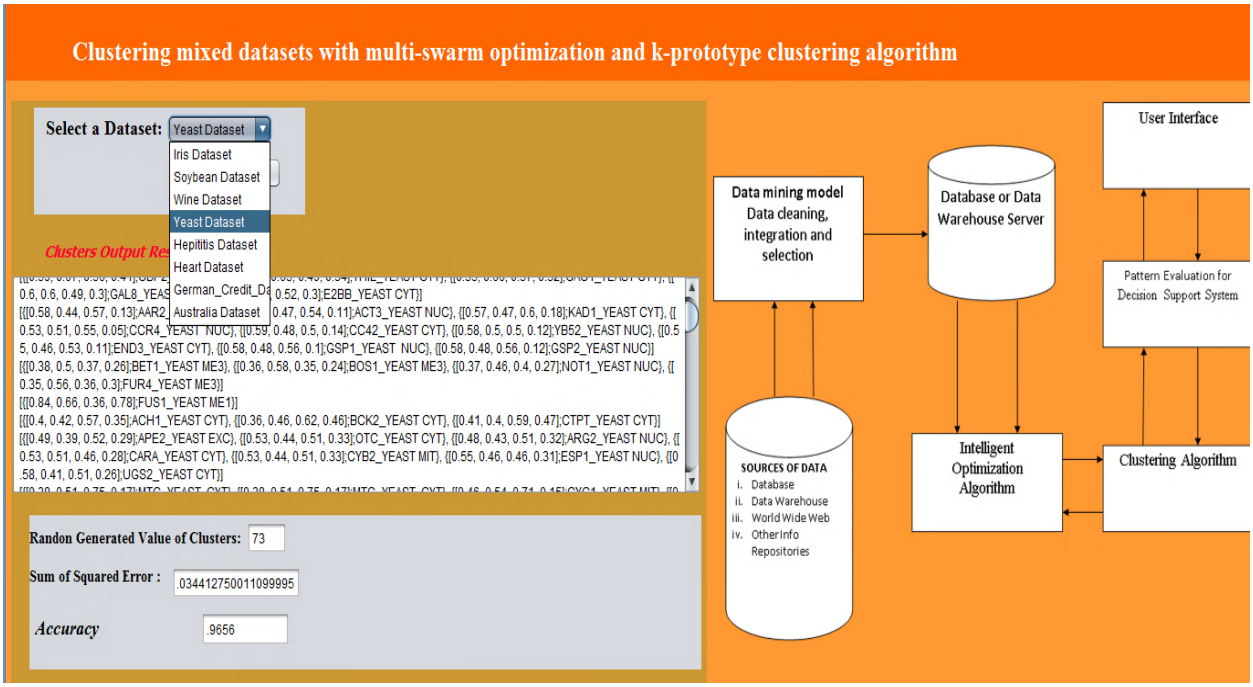


Figure 8: Yeast clustered output

Discussion of Results

To evaluate the accuracy of the proposed hybrid clustering algorithm, we used six datasets soybean, Hepatitis, Post-operative patient, Australian Credit Approval, German Credit Data, Statlog Heart and yeast datasets, obtained from UCI Machine Learning Repository to test the accuracy of the proposed hybrid algorithm. Two datasets Soybean and Yeast was used to compare the accuracy measures for K-modes, K-prototypes, Extended K-modes, MixK-meansXFon (Onuodu, et al, 2014), K-prototype (Huang, 1998) Extended K-modes (Arangnayagi and Thangavel, 2010) and the proposed hybrid clustering algorithm (Multi-swarmK-prototype clustering algorithm). The significant limitation of these algorithms was the selection of the initial value of K. The proposed Multi-swarmK-prototype clustering algorithm in this work solved this problem, by intelligent search for the best global particle for the initial value of k. This result strengthens the traditional K-prototype clustering algorithm to converge accurately and minimized the time constrain of the process. The rest of the dataset (Hepatitis, Post-operative patient, Australian Credit Approval, German Credit Data, and Statlog Heart) was used to compare the proposed Hybrid Clustering Algorithm and PSO based K-prototype algorithm (Prabha et al, 2015). The same Lambda values for the four benchmark datasets used for PSO based K-prototype algorithm experimental analysis (Prabha et al, 2015) was also used to analyze the performance of the proposed system, to ensure equal comparison of the performance of both hybridization of Multi-swarm and Binary particle Swarm Optimization Algorithms with k-prototype Algorithm. The lambda values are shown in Table 1.

Table 1: Details of Datasets and Lamda values (Prabha et al, 2015)

S/N	Hepatitis	Australian Credit Approval	German Credit Data	Statlog Heart
Lambda value	0.0533	0.0680	0.0995	0.0845

The comparison of the various clustering algorithms is shown in table 2.

Table 2: Comparison Analysis of the accuracy of proposed hybrid clustering algorithms and other algorithms for Soybean and yeast dataset

S/N	Datasets	MixK-meansXFon (Onuodu et al 2015)	K-prototype (Huang, 1998)	Extended K-modes (Arangnayagi, 2010)	K-mode	Proposed Hybrid algorithm
I	Soybean	0.86	0.69	0.83	0.37	0.971
Ii	Yeast	0.84	0.74	-	-	0.973

The results in table 2 are represented in Figure 9.

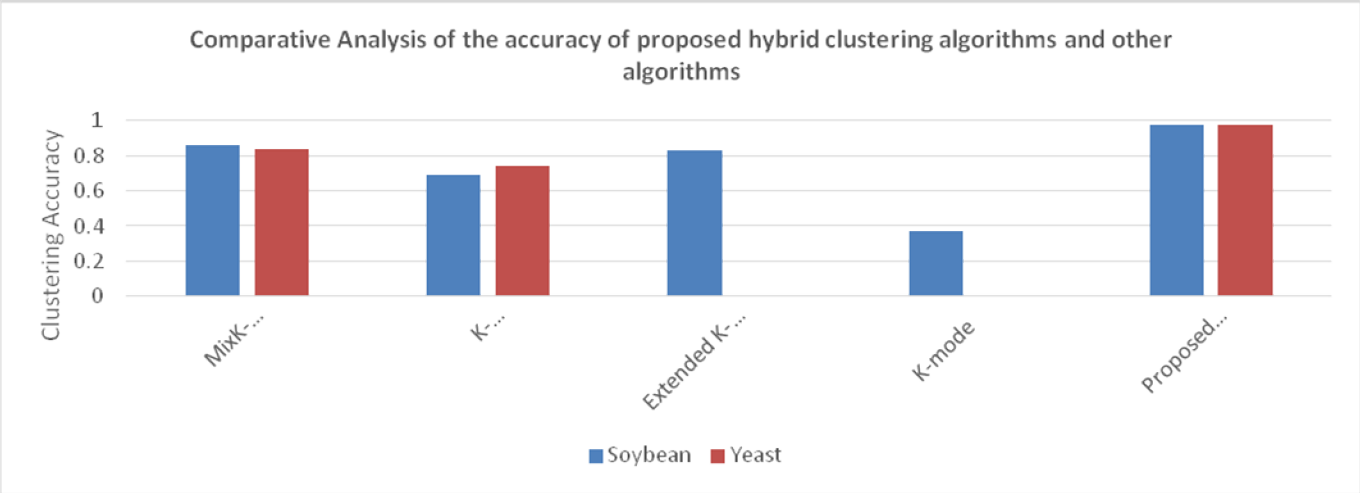


Figure 9: Graphical representation of Comparison of the clustering accuracy of proposed hybrid clustering algorithm and other algorithms

The comparison of the performance of Multi-swarmK-prototype and binary swarm based k-prototype clustering algorithm for clustering large dataset is shown in Table 3.

Table 3: Comparison of the accuracy values of Multi-swarmK-prototype and binary swarm based k-prototype clustering algorithm for clustering large datasets

S/N	Dataset	PSO based K-prototype algorithm(Prabha et al, 2015)	Proposed Multi-swarmK-prototype Clustering Algorithm
1.	Hepatitis	0.7521	0.9169
2.	Australian Credit Approval	0.8229	0.9789
3.	German Credit Data	0.6261	0.9495
4.	Statlog Heart	0.8387	0.9208

The results for the comparison of the performance of Multi-swarmK-prototype and binary swarm based k-prototype clustering algorithm for clustering large dataset is represented in Figure 10.

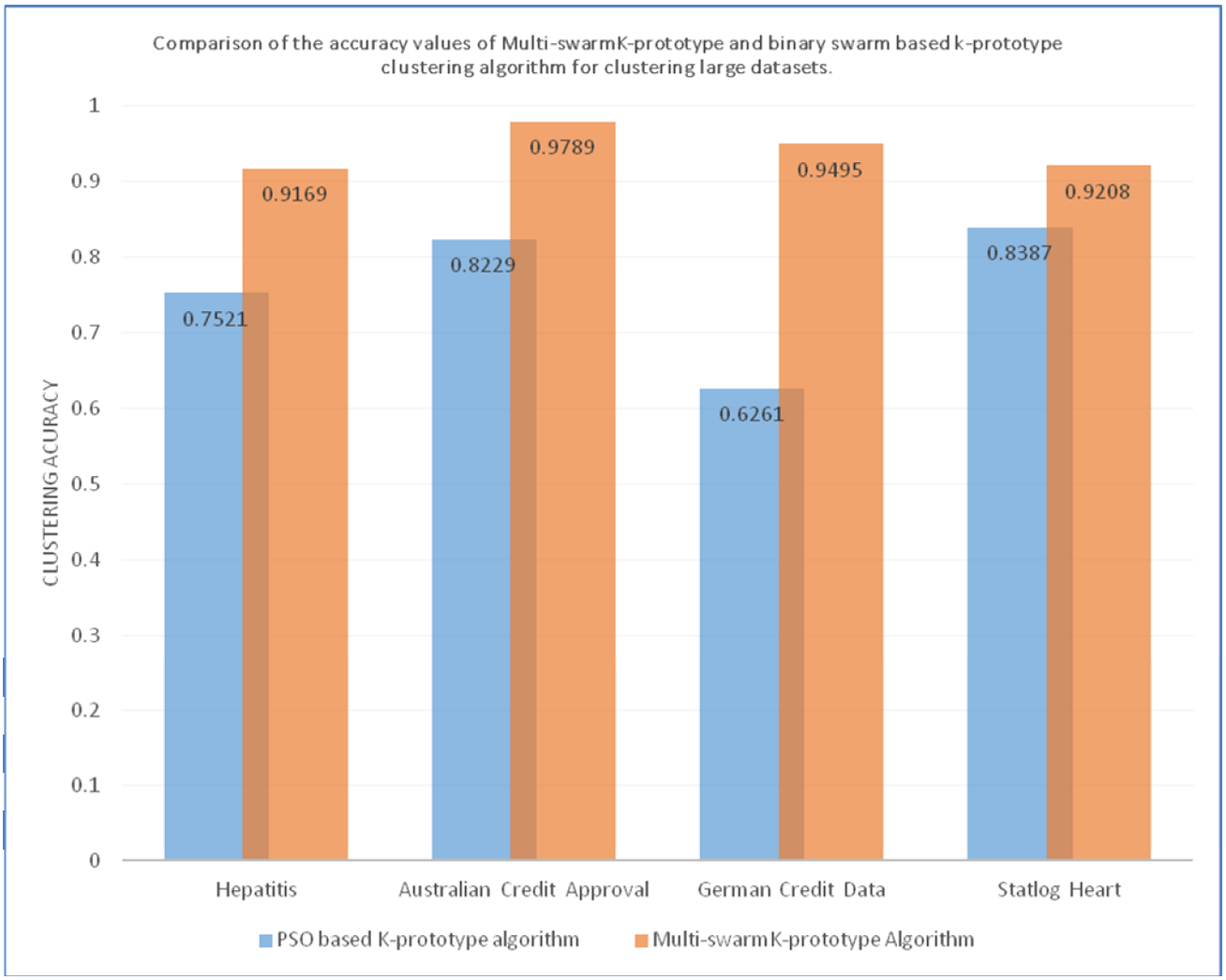


Figure 10: Graphical representation of comparison of the accuracy values of Multi-swarmK-prototype and binary swarm based k-prototype clustering algorithm.

The result from figure 9 shows that the proposed hybrid clustering algorithm is highly proficient for clustering mixed datasets. Also the results from Figure 10 shows that Multi-swarmK-prototype Algorithm had better clustering accuracy than the binary swarm based k-prototype clustering algorithm for the four datasets used.

Based on the comparative analysis, it is concluded that Multi-swarmK-prototype Algorithm has high clustering accuracy than the other clustering algorithms for mixed datasets. The comparative analysis of the time complexity of Multi-swarmK-prototype clustering algorithm and MixK-meansXFon algorithm is represented graphically in Figure 11:

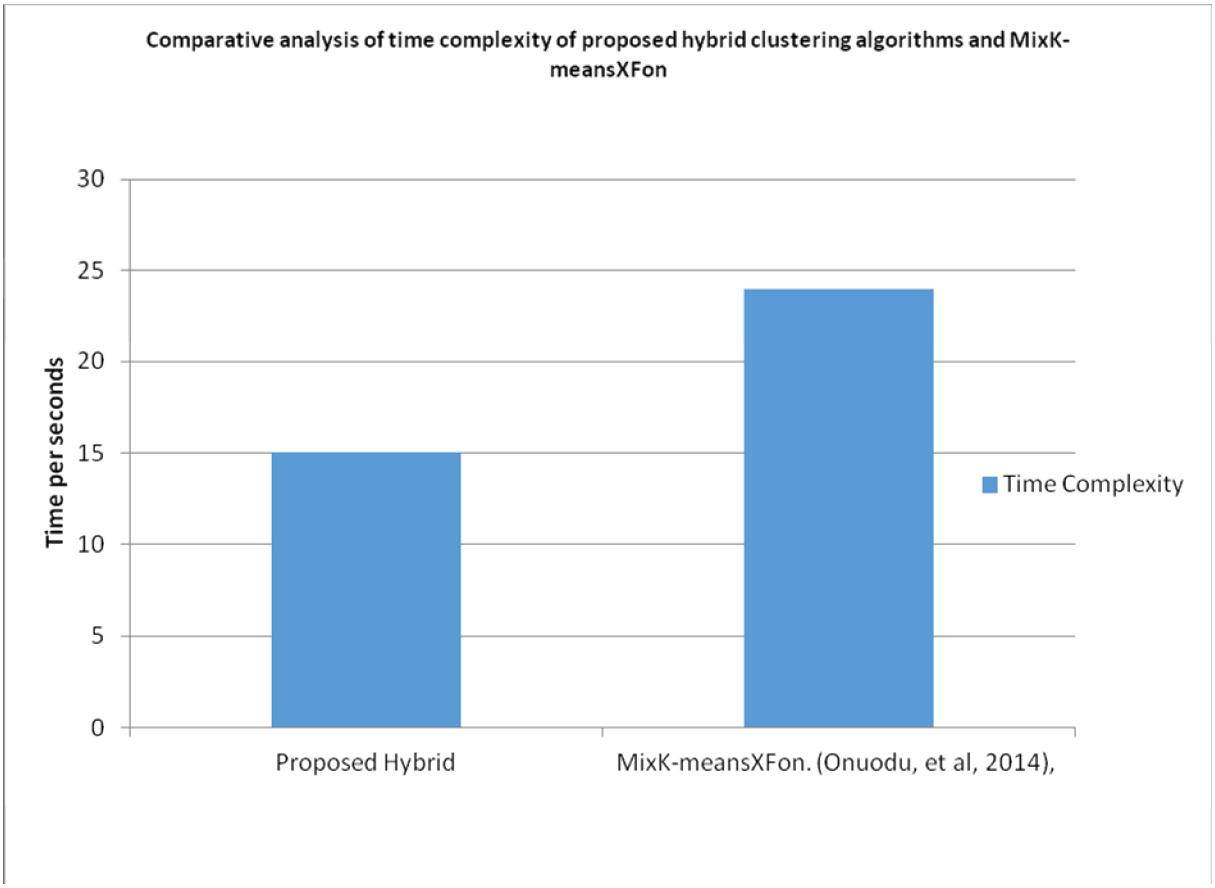


Figure 11: Graphical representation of Comparison of the time complexity of proposed hybrid clustering algorithms and MixK-meansXFon Clustering Algorithm.

The time complexity of the system was optimized by the mechanism of multi-swarm optimization search algorithm. It does not keep its initial optimal solution constant as in the case of other swarm optimization algorithms, it deploys sub-swarms to search for the best particle in the whole search space. The deployment of the sub swarm is controlled by the parent swarms, which detects the promising area in the whole space. And then deploys the child swarm or the sub swarms to explore the local optimum in a local area found by the parent swarm. It makes the sub-swarms spread out over the highest peaks and converge on the global optimum. The mechanism of multi-swarm optimization search algorithm for the initial value of k minimizes the time constrain of the clustering process.The time complexity is N^2, ignoring the constant terms and focusing on the dominant term in the expression. The order of growth for the time complexity is quadratic for the input size. That is, O (n) = N^2, where N is the total number of objects. This indicates that the new hybrid algorithm has O (n) running time.

Conclusion

A hybrid clustering algorithm which consists of multi-swarm and k-prototype for clustering mixed datasets is presented in this paper. It is efficient for random selection of the initial value of k. It also improved the clustering accuracy of the traditional k-prototype algorithm to obtain global optimal solution and high clustering accuracy. The mechanism of the multi swarm optimization algorithm using sub-warm to search through all the

region of the dataset for the initial value of k, minimized the time complexity of the process.

REFERENCES

Arangnayagi, S. and Thangavel, K. (2010), Extended K-modes with Probability Measure International Journal of Computer Theory and Engineering, 2(3), PP. 431-435.

Backwell, T. and Branke, J.,(2004), Multi-swarm Optimization in Dynamic Environments, Department of computing, Goldsmiths College, University of London New Cross, London SE14 6NW, U.K., pp. 1-12.

Chen, S. and Montgomery, J., (2011), Selection strategies for initial positions and initial velocities in multi-optima particle swarms, in proceedings of the genetic and evolutionary computational conference, pp. (53-99).

Eberhart, R., and Kennedy, J.(1995). A new optimizer using particle swarm theory. Paper presented at the Proceedings of the Sixth International Symposium on Machine and Human Science, Nagoya, Japan. Pp. 15-32.

Huang, Z., (1998), Extensions to the k-means algorithm for clustering large cybernetics 28C, 219-230.

Hendtasll, T.,(2005), A Multi-Optima particle swarm algorithm, in proceedings IEEE congress on evolutionary computation, pp 272-734.

Jain, M. and Verma, C, (2014), Adapting k-means for clustering in big data, International journal of computer applications (0975-8887) volume 101- No1.

Manjinder, K., Navjot, K. and Hsingh, H., (2014). Adaptive K-means clustering techniques for data clustering, International journal of innovative research in science, Engineering and Technology (An ISO 3297: Certified Organization) Vol. 3, pp.8-15.

Mohanavlli, S. and Jaisakthi, S.M.,(2015). Precise distance metric for mixed data clustering using Chi-square Statistics. Research journal for applied sciences, Maxwell scientific Organization, Engineering and Technology 10(12): 1441-1444.

Nielse, F. Nock, R and Dhun-ichi A.,(2014), ON Clustering Histograms with k-means by using mixed-divergences. www.mdpi.com/journal/entropy, 16, 3273-3301.

Onuodu, F. E., Nwachukwu, E. O. and Owolabi, O. (2014), Extension of K-means algorithm for Clustering mixed data, college of natural and applied science, University of Port Harcourt, Printed in Nigeria ISSN 1118 – 1931.Scientia Africana, Vol. 13 (No.2). pp. 2-19

Prabha, A. K. and Visalakshi, K. K., (2015), Particle swarm optimization based k-prototype clustering algorithm. IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 17, Issue 2, pp. 56-62.

nigeria computer society
www.ncs.org.ng
26th NATIONAL CONFERENCE & EXHIBITION

Full Paper

STPCLOUD: A SECURE TRUSTWORTHY PRIVACY-PRESERVING FRAMEWORK FOR CLOUD DATA

O. Agosu
Dept. of Computer Science,
Federal University of Agriculture,
Abeokuta
agosuoss@funaab.edu.ng

A. Onashoga
Dept. of Computer Science,
Federal University of Agriculture,
Abeokuta
bookyy2k@yahoo.com

O. Falana
Dept. of Physical Science,
Mountain Top University,
Magboro
ojfalana@mtu.edu.ng

O. Oyeleke
Dept. of Physical Science,
Mountain Top University,
Magboro
ooyeleke@mtu.edu.ng

ABSTRACT

In recent times, the growing need for data sharing in cloud computing environments is on the increase, with demands from the health informatics, national security and research institutions, in order to meet their organizational goals. However, the associated risk in security threats challenges cloud users' vision of computing as a utility and service. In attaining a sustainable level of trust in cloud services, Cloud data owners need to be guaranteed of a privacy preserving, highly secured, trustworthy and reliable computing environment playing host to their private or sensitive data, even in the course of a legitimate request by a data consumer. Most researches have investigated cloud insecurity with respect to cloud data outsourcing, with many proposed mechanisms to solve privacy, trust and security issues, but mostly in fragments. Therefore, this paper takes a survey of existing approaches to tackling this menace and proposes a better holistic framework known as the Secure Trustworthy Privacy-Preserving Framework for Cloud Data (STPcloud). The off-implementation evaluation and security analysis of this framework shows that it is reliable, implementable and as well meets the objectives of this research. In the near future, we would develop and deploy this framework and further evaluate its performance against standard threat models.

KEYWORDS: CLOUD COMPUTING, SECURITY, PRIVACY, TRUST, HOMOMORPHIC ENCRYPTION, PROXY ENCRYPTION, ANONYMIZATION, CERTIFICATE

1.0 INTRODUCTION

The recent advances in general computing has led to a combination of technological concepts such as virtualization, grid and distributed computing, which increases utilization of resources by driving physical resources virtually, and increases high-performance computing by providing computing resources on a more scalable, user-friendly, and pervasive model known as cloud computing (Robinson *et al.*,2010). It refers to the underlying infrastructure for an emerging model of service provisioning that has the advantage of reducing cost by sharing computing and storage resources, combined with an on-demand provisioning mechanism relying on a pay-per-use business model.

Cloud computing-based services are getting a great amount of success both in the market for individual users and small business organizations that do not want to spend money in buying and managing physical computing resources. Such services include online data storage facilities, and e-mail service provider facilities, web applications, and to mention but a few. The paradigm of Cloud computing can be described in simple terms as offering particular Information Technology Services hosted on the internet, the most common ones being Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Software as a Service (SaaS).

However, with Cloud computing gaining popularity, its resulting risks and challenges cannot be overemphasized. Its inherent features do not only have a direct impact on information technology (IT) budgeting, but also affects traditional security, trust and privacy mechanisms. Therefore, cloud users see this rising concept reliable only if they enjoy sufficient support from well-established security standards and reliable security mechanisms (Hogan *et al.*, 2011).

This aforementioned expectation in cloud computing cannot be said to have been met yet, as only functional issues such as data portability, interoperability and scalability, accessibility vulnerabilities, web application and virtualization vulnerabilities have received more attention from researchers and stakeholders (Badger *et al.*, 2011; Dawai Sun *et al*. 2011).

The remainder of this paper proceeds as follows. Section 2 presents security in cloud computing and background with STP Components. Section 3 describes STPCloud framework, Set-Up and Application. Section 4 highlights the security analysis and off-implementation evaluation of performance of the framework, while Section 5 concludes.

2.0 LITERATURE REVIEW

2.1 Security in Cloud Computing

Cloud-based data storage that requires security and privacy assurance are always deployed in private clouds. The Cloud Service Provider (CSP) needs to implement 'reasonable security' when handling personal information. Different companies may be involved in the cloud supply chain, and this can make it difficult to ensure that such security is provided all the way along the chain. At present, clients often know only the initial standard terms and conditions of cloud computing service providers. CSPs do not include any clause ensuring the level of security provided; they provide no guarantee as to the security of data and even deny liability for deletion, alteration or loss related to stored data.

The well-known approaches to users' data protection are generally based on concepts like authorization and access control, user authentication, intrusion detection and auditing, encryption and privacy policies to name a few (Kadhem *et al.*, 2009). This research work involves the extraction of useful notions and ideas from related approaches based on the concepts of anonymous access control (Jensen *et al.*, 2010), privacy preserving data publishing (Fung *et al.*, 2010), provable data possession (Ateniese *et al.*, 2007, 2008), data classification (Han *et al.*, 1991), and database encryption (Davida *et al.*, 1981).

2.2 Privacy in Cloud Computing

Westin (1960) defines privacy as the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. However, beyond developing secure mechanisms for ensuring the privacy of cloud users' data, there are stills some litigation and legal uncertainties in meeting trans-border data flow restrictions. According to International Association of Privacy Professionals (IAPP), privacy is the appropriate use of information under the circumstances. However, there is vagueness in the notion of what constitutes appropriate handling of data. This varies, as it is a function of some factors such as individual preferences, the context of the situation, law, collection, how the data would be used and what information would be disclosed.

Basically, privacy concerns in cloud computing are subject to lack of user control, potential unauthorized secondary usage, regulatory complexity (especially due to the global nature of cloud, complex service ecosystems, data proliferation and dynamic provisioning (Pearson and Yee, 2013).

Cloud users' data privacy is one of the various domain areas of cloud computing which lack the support of an accompanying standard of dependable maturity (Hogan et al., 2011; CSA, 2009).

Fully homomorphic encryption is identified as a promising solution for preservation of privacy in cloud environments but is constrained by its limited scope of utility arising from its capacity to serve only certain classes of cloud applications (Dijk and Juels, 2010). This technology allows manipulating encrypted data and performing logical operations on it without actually accessing the data in the clear. In particular, homomorphic cryptography is envisioned as the perfect technology for secure computation (or storage) delegation in the cloud. This work investigates how homomorphic cryptography can enable privacy.

Waqar *et al.*(2013) highlighted the possibility of exploiting the metadata stored in cloud's database in order to compromise the privacy of users' data items stored using a cloud provider's simple storage service. It, then, proposes a framework based on database schema redesign and dynamic reconstruction of metadata for the preservation of cloud users' data privacy.

2.3 Trust in Cloud Computing

Trust encompasses guarantee or assurance and confidence that people, data, entities, information or processes will function or behave in expected ways. Trust covers entities like devices, protocols, system users and etc. Extensively, trust can be regarded as a consequence of progress towards security or privacy objectives. Several researches have been conducted in the area of Cloud Trusted computing or the trustworthiness of cloud computing environment.

Li and Ping (2009) proposed a Multi-Tenancy Trusted Computing Environment Model (MTCEM) for IaaS delivery model. Its purpose is to guarantee a trusted cloud infrastructure to customers. The prototype showed low impact on system performance and the model is technically and practically feasible.

Huang *et al.* (2013) suggested a framework for integrating various trust mechanisms together. They suggested a policy-based approach of trust judgment by which the trust placed on a cloud service or a cloud entity is derived from a "formal" audit proving that the cloud entity conforms to some trusted policies.

Nagarajan *et al.*, (2014) proposed a trust enhanced distributed authorization architecture (TEDA) that provides a holistic framework for authorization taking into account the state of a user platform. The model encompasses the notions of 'hard' and 'soft' trust to determine whether a platform can be trusted for authorization. The result shows that such trust enhances better authorization in decision making, especially in a distributed environment where user platforms are subject to dynamic security threats.

2.4 Related Works with STP Components

Traditional security mechanisms predate evolving heterogeneous technology – dependent application domains with innovative service provisioning in dynamic environments. This consequentially calls for a collective interoperable security mechanism as seen in 'security, privacy and trust in cloud' suggested by AbManan *et al*. (2014), and 'security, privacy and trust in Internet of Things (IoT)' by Sicari *et al*. (2014), etc. However, narrowing down our research domain to data storage, outsourcing, protection and data share in the cloud, we reviewed research works closely related to our proposal.

Popa *et al*. (2011) proposed a secured storage system driven by attestation mechanism called CloudProof to guarantee confidentiality, integrity and write-serializability using verifiable proofs of violation by external third parties. Data owners encrypt their data with the private keys only known to them in achieving confidentiality. They use a block identifier to acquire the content of a block, making it possible to store data with block identifier and the contents of the block in the cloud. This system's attestation mechanism uses block hash as a signature verifying tool for proof for write-serializability using a forked sequence of the attestations and uses chain hash for a broken chain of attestations which are not sequenced correctly. Generally, the Attestations provide proof of sanity of data users, data owners and CSPs.

As a means of adding trust components to traditional security mechanisms in secure cloud data share, Zhou et al. (2013) proposed a trust-based cryptographic role-based access control (RBAC). The authors proposed that data owner can enforce authorization policies relating to roles trustworthiness using cryptographic RBAC. The trustworthiness of the system is evaluated using role inheritance concept of its authorization system.

Another interesting research solution for data processing and storage, specifically anonymized brain imaging data was ScaBIA (Gholami et al, 2013). This mechanism involves using PKI Authentication over HTTPS channel for PaaS middleware deployment by the system administrators and creation of researchers as users in the in Microsoft Azure cloud. Username/Password authentication is required from researchers to run statistical parametric mapping workflows within isolated generic worker containers. Access definitions are guided by RBAC model.

Kim and Timm (2014) also proposed a solution for data storage, processing and share called fermiCloud which uses PKI X.509 certificates for user identification and authentication. User identity is managed through a web interface called Open Nebula and an X.509 command-line authentication. Difficulty in cross-cloud data share resulting from the limitations of Access Control List (ACL) was resolved using existing local credential mapping service, which is now being adopted in cloud federations to authorize users across different cloud providers that have established trust relationships through trusted certification authorities.

Gholami and Laure (2015) proposed the BioBankCloud platform, which supports the secure storage and processing of genomic data in cloud computing environments. The cloud privacy threat modelling (PTM) approach was used as a building block for this framework. The PTM defines the model for processing next generation sequencing data according to the European Union Directive on Personal Data Processing i.e. DPD (EUD,1995). It is composed of security components such as a flexible two-factor authentication, RBAC and Auditing mechanisms in line with DPD.

However, this scheme does not deal with the scenario where a revoked user rejoins the group with different access privileges. The revoked user still has the decryption keys corresponding to ABE and hence in

theory can regain access to data he is not allowed to.

3.0 The Proposed STPcloud framework

Closest to this research proposal is the work by AbManan *et al.*(2014), in which a STP (Security, Trust and Privacy) framework for pervasive computing environment was proposed with a view to address how sensitive data could travel uninterrupted, unchanged and unseen by unscrupulous recipients. This proposed framework is yet to be implemented, and the implementation procedure in ensuring a secured privacy preserving trustworthy cloud was not made explicit.

However, the STPcloud Scheme proposed in this work, is different from all aforementioned approaches, in that, a hybrid trust evaluation scheme with the notions of Platform trust and user trust would be adopted as an extended abstraction of the proposed mechanism by Nagarajan *et al.*,(2014). This determines the trustworthiness of a cloud platform and its Authorization to requested services and cloud users. An additive homomorphic proxy re-encryption cryptographic scheme and Data anonymization by *K*-anonymity would be adopted to ensure privacy preservation of Cloud users' published data. For efficient authorization, an improved access control model for cloud computing would be adopted as an extension of AC3 (Younis *et al*., 2014).

In the proposed design, there are three (3) major players: *Cloud Data Owner (DO), Cloud Data User or Consumer (DC) and Cloud Service Provider (CSP),* while the core components of the framework comprises of the *Identity Management System (IDMS), Cloud Key Management System(CKMS), Access Control for Cloud Module (AC3), Policy Manager (PM),Trust Meter (TM), Data Manager (DM), Data Anonymizer (DA), and Data Share Subscription (DSS)* API as shown in *Figure 1.0*.

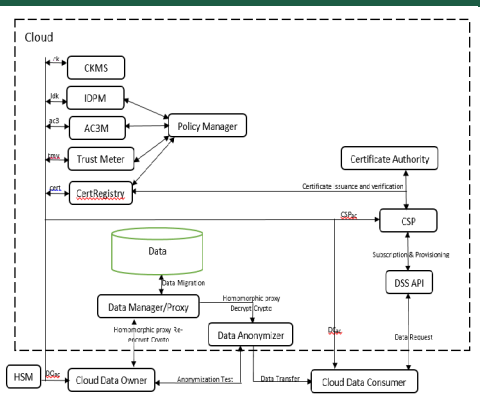


Figure 1.0: The proposed STPcloud Framework

In STPcloud, the proposed deployment environment is a public cloud providing a cloud infrastructure driven by robust computing resources in terms of bandwidth, CPU power, storage space and encryption services which are more reliable and powerful than personal servers, and as well affordable. The CSP is saddled with three responsibilities; data storage, data share request forwarding and re-encryption. After obtaining authentication and authorization permissions, the Data Owner stores encrypted data into the cloud using his HSM managed symmetric Secret Key. Upon the receipt of a data access or communication request, the CSP re-encrypt the already encrypted Secret Key. The DO manages the data share and access control policies over the cipher text.

By the STPcloud design in figure 1, the CSP lacks control over data since there is no provision for such as defined by the access control mechanism. Upon the data share request of a DC, the CSP stands as an arbiter relaying the data subscription request to the DO, as this can best be managed by CSP for a multi-owner data share subscription scenario. With adequate permission given and trust evaluation performed, the DC gets his data request under a trustworthy platform (Device-Based Trust assured environment) as a product of a homomorphically decrypted re-encrypted plaintext.

However the CA as an Attestation Party provides the reliability status of the cloud

Platform, components, properties and the entities such as certificates. The derived trust on the CA determined by the users of the public cloud is also measured by reputation, recommendation and direct trust, even though we assume our Third party is trustable.

The Data Manager, also known as the Proxy Data Server manages data operations for the cloud user while data is in transition from rest, transit, and to use. It serves the purpose of a proxy re-encryption server and data storage server.

The functions of the core components of STPcloud Framework are outlined as follows:

3.1 The Identity Provider Module (IDPM)

The Identity Provider Module is primarily responsible for the storage, maintenance and retrieval of cloud user credentials for either authentication, some other basic functions, or a path to authorization (Habiba *et al.*, 2014). It coordinates user identities certified by a trustable public key infrastructure (PKI) of a Certificate Authority offering identity- as- a-service (IDaaS). It is a provisioning framework linkable to the access control system and the Cloud Key Management System (CKMS). Every identity is mapped to an access control attribute for authorization to system actions, processes and resources via an Identity Provider API. After a first time identity account creation, the mechanism adopted here is the OpenID authentication-protocol-driven Single Sign-On (SSO). This mechanism helps mitigate the threat of password disclosure, since no OpenID-based Cloud service actually stores user's password information (Habiba *et al.*, 2014). The operations of the IDP module is detailed in the following Algorithms.

Algorithm 1: User creation

```
Procedure createUser( array userdetail)
begin
    if (!exist(userdetail['email'])) then
        if(validatePolicy(context createUser, userdetail['userId'] )==true) then
            do add(userdetail);
            keyset ← keyMapHsm(Pk_comm, Sk_comm, Pk_m, Sk_m);
            do certifyKey(userdetail['userId'], keyset);
            do wrapKey(keyset, userdetail['password']);
            do storeKey (userdetail['userId'], keyset);
            choose index i;
            assign UserType( getUserType[i], userId);
        endIf
    endIf
end
```

Algorithm 2: User Provisioning

```
Procedure string userProv(username, password)
begin
    If(IsValid(username, password)) then
        If(!exist(getOpenId(userId))) then
            openId ← generateOpenId(userId);
            If(!exist(getSessionId(userId))) then
                sessionId←generateSessionId(userId);
                If(!activateUser(userId) and (trustMeterStatus(userId)<>'Known')) then
                    do activateUser(userId);
                endIf
            endIf
        endIf
    endIf
end
```

Algorithm 3: User categorization

```
procedure createUserCategory
begin
array userClass
enum userCategory{Data Owner Do, Data Consumer Dc, Cloud Service Provider Csp};
    foreach usertype in userCategory do
        ini alize counter← 0;
        userClass[counter]← usertype;
        counter++;
    endForeach
end
```

Algorithm 4: Delete_User

```
procedure deleteUser(userId)
begin
        if(validatePolicy(context revokeUser, userId)==true)) then
                if(validatePolicy(context deleteUser, userId)==true)) then
                    If(exist(getUser(userId)) and (trustMeterStatus(userId)<>'Trusted')) then
                        do  deleteOpenId(userId) ;
                        do  revokePermission(userId);
                    endIf
                endIf
        endIf
end
```

Algorithm 5: Deactivate_User

```
procedure deactivateUser(userId)
begin
response ← 'false';
        if(validatePolicy(context deactivateUser, userId)==true)) then
          if(validatePolicy(context revokeKey, userId)==true)) then
                if(revokeCert(userId)) then
                   if(deleteKey(userId)) then
                       response ←'True';
                       return response;
                   endIf
                endIf
          endIf
        endIf
end
```

3.2 Key Management System (KMS)

According to Thilakanathan *et al. (*2013), when considering data sharing and collaboration, simple encryption techniques do not suffice, especially when considering key management. To enable secure and confidential data sharing and collaboration in the Cloud, there needs to first be proper key management in the Cloud. The mechanism proposed for key management in STPcloud is hybrid, in that the key management is both partly on premise and on-cloud. By the security principle of segregation of duties, the Cloud Key Management System (CKMS) is responsible for the generation, management and distribution, and revocation of key information for PKI communication, Data Share, Homomorphic Proxy Re-encryption (HPRE) and the Data Consumer's Encryption processes. However, the on – premise Hardware Security Module (HSM) based Key management System, which generates, stores and manages Data owner's encryption Key, remotely communicates with the CKMS over a secured Service Oriented Architecture (SOA). This provides a robust protection against both an external breach of the service provider as well as key related attack originating from a privileged user/employee of the provider. This model puts the customer in complete control of data encryption/decryption keys. The cloud service provider neither holds keys, has minimal knowledge of users, nor decrypt customer data, but can efficiently facilitates storage of the encrypted data. The key management procedures are premised on identity management (algorithm 1 and algorithm 5) and access control management operations (algorithm 10).

3.3 Access Control for Cloud Module (AC3M)

The Access Control for Cloud Computing (AC3) Model by Younis *et al.*,(2014) has been reviewed to be a better adoption for this framework but with some modifications to suit our trust and privacy objectives. Summarily put, in the reformed AC3 model, user classification is done on job basis, which translates to a mapping of the trustable users to the security domain that relates to their roles. Roles are assigned a set of the most relevant and required tasks for role actualization. Tasks don't only have their respective security classification for accessing the data or assets, but also the permissions needed for accomplishing any given task. A Policy Manager is utilized to deal with dynamic and random access behaviors of users by credit evaluation. Another major component of this module is the Security Tags Engine, used for issuing security tags in semi or untrusted environments and processes (e.g. the Proxy Server). Security Labels are attached to data or assets so as to secure access. Processes are executed based on the Task called and would require security tag to get data access. Classifications and security labels are used for controlling access to resources. Sensitivity labels are used to mark data internally according to their sensitivity and value. In essence, any task or process employed by a task needs a classification

to access resources, as there should be no access to any resource without a classification equal or superior to their source's sensitivity labels.

In Mathematical representation, the reformed AC3 components are defined as thus;

A set of Trust Labels *L* defined as

$L = \{$'*trusted*', '*untrusted*', '*uncertain*'$\}$;

A set of cloud users, $U = \{u_i: i \epsilon Z^+\}$;

A set of cloud trustable users, $U^l = \{u^l_i: i \epsilon Z^+\}$;

A set of Roles, $R = \{r_i: r_n \neq r_m \forall i, n, m \epsilon Z^+\}$;

A set Tasks, $T = \{t_i: t_n \neq t_m \forall i, n, m \epsilon Z^+\}$;

A set of Sessions, $S = \{s_i: s_n \neq s_m \forall i, n, m \epsilon Z^+\}$;

A set of Permission, *P* defined as;

$P = \{p_i: p_n \neq p_m \forall i, n, m \epsilon Z^+\}$;

A set of Data, *D* defined as

$D = \{d_{i\,=}\, (r_n c_m): \forall i, n, m \epsilon Z^+$, where *r* = row; c = column$\}$;

A set of mapping of U^l to *R*, User Assignment (*UA*) defined as;

$UA = \{u^l a_i = (u^l_i x\ r_i): (u^l_n x\ r_m) \neq (u^l_k x\ r_j) \forall i, j, k, n, m \epsilon Z^+\}$;

A set of mapping of *R* to *T*, Role Assignment (*RA*) defined as;

$RA = \{ra_i = (r_i x\ t_i): (r_n x\ t_m) \neq (r_k x\ t_j) \forall i, j, k, n, m \epsilon Z^+\}$;

A set of mapping of *P* to *T*, Permission Assignment (*PA*) defined as;

$PA = \{pa_i = (p_i x\ t_i): (p_n x\ t_m) \neq (p_k x\ t_j) \forall i, j, k, n, m \epsilon Z^+\}$;

A set of system access constraints K (for each access definition $ac \epsilon AC$,] a $k \epsilon K$ such that

$K = \{k_i: k_n \neq k_m \forall i, n, m \epsilon Z^+\}$;

A set of task classification C defined as

$C = \{c_i: c_n \neq c_m \forall i, n, m \epsilon Z^+\}$;

A set of Sensitivity labels from the Data Privacy Classification, *SL* defined as

$SL = \{$'*XP*', '*PP*', '*NP*'$\}$;

Definition 1

A data set/item is said to carry a sensitivity label *XP* (i.e. exclusively private) if it must be kept confidential.

Definition 2

A data set/item is said to carry a sensitivity label *PP* (i.e. partially private) if its integrity must be kept but may not be confidential.

Definition 3

A data set/item is said to carry a sensitivity label *NP* (i.e. Not private) if its integrity and confidentiality is insignificant to preserving cloud user's data privacy.

However, this technique satisfies the core objectives and constraints of Access Control which are Least Privilege Principle, Delegation of duties, and Separation of duties as contained in Algorithms 6, 7, 8, 9 and 10.

Algorithm 6: Function-Task mapping

```
procedure mapTaskFunct (array tasks, array functions)
begin
        foreach task t in tasks do
            foreach function f in functions do
                if((!exist(map(f,t)))) then
                    return map(f,t);
                endIf
            endForeach
        endForeach
end
```

Algorithm 7: Permission-Function mapping

```
procedure mapFunctPermission (array permissions, array functions)
begin
        foreach function f in functions do
            choose permission p ;
                if((!exist(map(f,p)))) then
                    return map(f,p);
                endIf
            endForeach
        endForeach
end
```

Algorithm 8: Role-Task mapping

```
procedure mapTaskRole (array tasks, array roles)
begin
        foreach role r in roles do
            foreach task t in tasks do
                if((!exist(map(r,t)))) then
                    return map(r,t);
                endIf
            endForeach
        endForeach
end
```

Algorithm 9: Data Sensitivity Parameterization

```
procedure setDataSecurParam(data)
begin
enum datasensitivity{'XP','PP','NP'};
    foreach relation r in relations do
        foreach attibute a in Attributes do
        choose index i;
        sensi   vity[r][a] ← map(data[r][a], datasensi   vity[i]);
        endForeach
    endForeach
return sensitivity;
end
```

Algorithm 10: Task-Permission

```
procedure mapPermission(task t, permission p, userId )
begin
    ini   alize accessStatus ← 'Denied';
    if(validatePolicy(context resourcePermission, userId)) then
        data ← resourceCoveredBy(t);
        sensLabel ← setDataSecurParam(data);
        taskLabel← correlate(map(r,t), map(f,t), map(f,p));
        trustMeterStatus← compute_UserTrust(userId)
        if((taskLabel >=sensLabel) AND (trustMeterStatus is 'Trusted' OR 'Indefinite') ) then
                accessStatus←'Granted';
        endIf
    endIf
    return accessStatus;
end
```

3.4 Policy Manager (PM)

This component helps in managing access control policies effectively, as it organizes relationship between the cloud stakeholders, utilities and third parties. It defines and enforces rules for certain actions such as auditing or proof of compliance to the IDPM, AC3, CA, CKMS and Trust Meter. These policies are defined around processes as *create*, *update*, *read*, *delete*, *import and export* etc.

3.5 The Trust Meter (TM)

Trust evaluation from the point of view of this architecture to be two, Device or Platform Trust and Operational Trust. The trust manager is meant to be coordinated by a trusted third party in estimating the trust value on the platform and identities. As for Platform trust, the collected evidences of the behavior of different applications installed on the platform is used to reason about its overall state, comparing it to known and acceptable reference state to decide whether or not it can be trusted. This engine measures the user trust of stakeholders in the cloud.

3.6 Data Manager (DM)

This is also known as the Proxy Data Server, which manages data operations for the cloud user while data is in transition from rest, to transit and in use. It serves the purpose of a proxy re-encryption server and data communication server.

3.6.1 Data at Rest

Data at rest is essentially the data that is stored persistently in some form e.g. as a file, in a database, etc. The goal of protecting data at rest is to prevent a third party from reading the data, should they gain access to the data in its persistent form. The security of data at rest is reliant on four key components, Key Management System, Access control Management system, Crypto-System adopted and the Security Policy for Key and Access. Having discussed extensively on others, the Crypto-System proposed for this framework is the Homomorphic proxy re-encryption. It is a combination of Gentry's (2009) Homomorphic cryptosystem and proxy re-encryption (Blaze et al., 1998). In a bid to allow operations to be performed on encrypted data without having to decrypt it, Gentry has shown that it is possible to analyze data without decrypting it. This would allow systems/applications to communicate securely with each other without ever having to exchange unencrypted data.

3.6.2 Data in Transit

The primary aim of protecting data in motion is to prevent a third party from eavesdropping on a conversation on the wire. Cryptographic protocols, such as Secure Sockets Layer (SSL) or Transport Layer Security (TLS), are typically used for protecting data in motion by establishing an encrypted and authenticated channel, but better achieved via a Public Key Infrastructure (PKI) to ensure an end to end secured communication (Figure 2.0). Here, the Data Owner subscribes to the client certification Service of a reputable Certificate Authority (CA) to coordinate the PKI mechanism. In STPcloud, it is best for such a CA to generate and protect the keys associated its Root CAs and subordinates with certified Hardware Security Modules (HSMs), capable of protecting against logical and physical attacks on the key store. However, the CA is also subjected to the scrutiny by the Trust Meter.

3.6.3 Data in Use

At this end, data is expected to be migrated to the data requestor i.e. the Cloud Data Consumer (DC). The DC, which could be the security agencies or the health data analyst could possibly have genuine reason to request for data. To sustain the confidentiality and identity preservation goal of this framework, we propose that the CSP provides a Data Share Subscription Service for all DCs. This is then forwarded to the DO, who in-turn transfers data in anonymized form (based on sensitivity parameterization) to the consumer under appropriate access and identity verification. Figure 2.0 shows the workings of the data export mechanism.

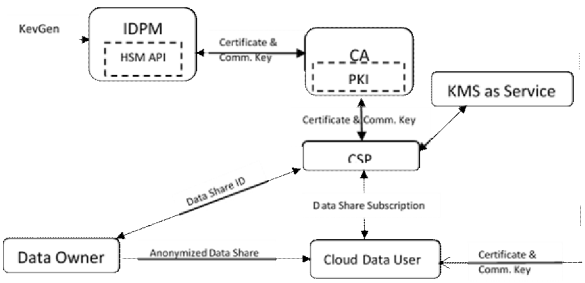


Figure 2.0: Data In Transit and Data In Use

3.7 Operational Flow of STPCloud Framework

The proposed STPCloud Protocol is constructed with the Public Key Infrastructure (for Communication), AES scheme (for encryption of plaintext data-in-transit), additive homomorphic proxy re-encryption for confidentiality and privacy (using the communication keys) of user identity and AES key, data share subscription identity, Identity provisioning and AC3 tokens for data subscription. The operational process of the STP framework comprises of the follows;

3.7.1 Data Upload and Outsourcing

The data owner, say Alice, having registered, verifies his identity with the IDPM by logging in with her username and password on the Data outsourcing application. She obtains a single sign-on token ss_{Alice} and a session token se_{Alice} to ensure her authentication and obtain a time-bound session to all available services defined for her identity type U. This attributes are wrapped and signed with signature sg by a certificate authority CA via the Certificate Registry CR. This forms the identity $idk_{Alice} = (ss_{Alice} \,||\, se_{Alice} \,||\, U_{Alice} \,||sg_{CA})$ of user Alice. However Alice generate a key pair $(Pk^{a}_{comm}, Sk^{a}_{comm})$ for a PKI - based end-to-end communication with the proxy server with $(Pk^{p}_{comm}, Sk^{p}_{comm})$. She generates a symmetric key Sk^{a}_{enc} from the Hardware Security Module (HSM) for encrypting data records with defined sensitive attributes. The encrypted data and password – wrapped key Sk^{a}_{enc}, which she sends to the proxy as message M for re-encryption homomorphically.

Suppose the optimal trust threshold δ_i is defined for cloud entities such as users, CA and platform, such that we can estimate the trustworthiness of these entities. Therefore, We proceed to authentication process, given a trust evaluated platform P with components C satisfying the required properties for a cloud service, with a trust level $x_0 \in X_0 \ni x_0 \geq \delta_0$ and user trust value $y_0 \in Y_0 \ni y_0 \geq \delta_1$, where $\delta_0, \delta_1 \in \delta$. Having met this conditions, the proxy maps Alice's anonymous identity $idk_{Alice}{}'$(generated by a transformation function f over idk_A i.e $idk_{Alice}{}' = f(idk_{Alice})$) to the access token $A^{u,d}$ appropriate for her user type $u \in U=\{DO,DC,CSP,CA\}$ with assigned

task t ϵ T={*read, write, execute, delete*} on a data set D of defined attributes a_i and row r_j, with varying sensitivity parameter $s \epsilon S$={*xp, pp, np*} defined as $D=\{d(a^s_1,r_1),\ldots.. d(a^s_m, r_n)\}$. However, $A^{u,d}$ is defined for user Alice as;

$A^{u,d}_{Alice}$ = [(*ac3*, U_{Alice})*verify(*U*, *f*'(*idk*$_{Alice}$')) * verify(*idk*$_{Alice}$', *sg*$_{CA}$) *verify(*U*, *ac3*,*D*)]

And returns *Null* if this permission assignment and verification fails.

3.7.1.1 Cryptographic Construction

The cryptographic protocols adopted in this proposed scheme are the homomorphic proxy re - encryption and AES schemes. All plaintexts are encrypted using symmetric AES algorithm, Data encryption keys are homomorphically encrypted before they are re-encrypted by the proxy server managed by the CSP. The end – to – end secured communication would be enhanced by the PKI technology. However Keys are stored and managed from the Hardware Security Module (HSM) devise and API.

Set_Up(ℓ): With the security parameter ℓ given, Let p be a ℓ−bit prime, and let *G*, *GT* be two cyclic groups of order *p*. Let *e* be a bilinear group, $e : G \times G \rightarrow GT$. Let g be the generator randomly selected from the group *G*, and $Z = e(g, g)$. The message space is *GT*. This algorithm reveals the global parameter param comprising of message space, plaintext space, and ciphertext space specifications. It serves as input in the following algorithm. Here all cloud users can generate the asymmetric Communication key pairs (Pk_{comm}, Sk_{comm}) certified by the CA, and publishes their public key Pk_{comm}. However the Cloud Data Owner (DO), having created an account via the IDP module is granted data upload permissions to generate keys for data upload and data release.

Upload Process: This involves the generation of encryption keys, proxy-rencryption keys, and encryption functions Enc, Dec, and transformation *f*, require to upload encrypted data on the Data Server. This entails a lightly encryption of data with 128-bit AES key in CBC mode and homomorphically re-encrypt its encryption key Sk_{enc}.

ENC.KeyGen: This algorithm randomly generates three separate symmetric keys k_1, k_2 and Sk_{enc} which are the support key, alternate key and content encryption key respectively from the HSM, defined such that $Sk_{enc} = k_1 + k_2$. The main AES symmetric key Sk_{enc} serves as the Message in this context. Given the asymmetrically generated key pairs (P_{do}, S_{do}) and (P_{du}, S_{du}) for the Data owner and Data Consumer respectively, the DO encryption key K_1 is such that;

K_1 = Dec(Enc(P_{do}, K_1), S_{do})

PRX.KeyGen: This algorithm allows Data Owner to generate proxy re-encryption key RX_{ou} by taking as inputs its public key P_{do} and the Data Consumer's public key P_{du}, via key wrapping over a transformation *f* such that the proxy can re-encrypt a ciphertext under P_{do} to a ciphertext under P_{du}. i.e $RX_{ou} = f(P_{do}, P_{du})$.

Enc: This algorithm encrypts a message say K_1 and encrypt with the public key of the data owner to generate a ciphertext C which is sent to the proxy-server.

$C = Enc(K_1, P_{do})$

Send C to Proxy and C_1= Enc(k_2, P_{do}) to HSM for storage.

PRX.Enc: This algorithm picks re-encryption key RX_{ou} to reencrypt the given ciphertext C under public key P_{do}, this algorithm will output ciphertext C′ under public key P_{du}. Proxy computes:

$Enc(K_1, P_{du}) = PRX.Enc(Enc(K_1, P_{do}), RX_{ou})$

$= PRX.Enc(C, RX_{ou})$

The content encryption Sk_{enc} computed from encrypted (k_1+k_2) representing our message is computed with DC's P_{du} and Sk_{enc}, this algorithm homomorphically encrypts a message C and outputs a reencryptable ciphertext C′.

Proxy computes with encrypted (k_1+k_2) by processing:

$C' = Enc(k_1+k_2, P_{du})$= Enc($k_1$, P_{du}) +H Enc(k_2, P_{du}) *mod* N^2

$= Enc(Sk_{enc}, P_{du})$

Where +H = additive homomorphism

N= Group size of encryption scheme.

3.7.2 Data Share Subscription

A data consumer, who is interested in the data outsourced by a DO (say Alice) to the CSP, must subscribe first on the outsourcing Application before such a data share request is forwarded to the DO. In order to gain access to the Data Share Service (DSS), a data consumer (DC) say Bob, acquires idk_{bob} for authentication. His data subscription description ds^{bob} is sent via the proxy server over the

secured PKI communication wire to the CSP. The CSP sends the data share Identity, *dsi* defined as $dsi_{bob} = [idk_{csp} \,||\, idk_{bob} \,||\, (ds^{bob} \,||sg_{CA})]$ to Alice. She receives an email, informing her of a data request, and the *ds* is added on his waiting list of request.

3.7.3 Data Access

The data owner, Alice, may grant $A^{u,d}{}_{Bob}$ to data consumer, Bob, after verifying dsi_{bob}. Alice validates the signed keys of the CSP, idk_{bob} and *ds* of Bob. This permits the generation and release of the key pairs (P_{du}, S_{du}) to Bob. He uses his private key S_{du} over the cryptographic construction as follows:

Dec: A decryption transformation that requires the DC's private key S_{du} on the ciphertext C' to fetch key Sk_{enc}, so as to decrypt *D'* to *D*.

$$Sk_{enc} = Dec(C', S_{du})$$
$$= Dec(Enc(P_{du}, Sk_{enc}), S_{du})$$

Hence,

$$D = Dec(D', Sk_{enc})$$

D represents the data record encrypted with the AES symmetric Master key Sk^{a}_{enc} used by Alice in encrypting the data. However, This plain data record *D* is subjected to anonymized before publishing using a privacy-preserving k-anonymity function μ over some defined quasi-identifiers such that:

$D'' = \mu(D)$

$A^{u,d}{}_{Bob}$ is granted to Bob for authorization to work on data with the corresponding C' using either the encrypted key Sk_{enc} or an API link D''_l to anonymized data. This link is added to his list of approved requests. However, the anonymized data is accessible by Bob only.

4.0 Security Analysis and Evaluation

Here, the security and performance of STPcloud is analyzed and evaluated. The security analysis entails encryption correctness, confidentiality and privacy preservation, Collusion Safety, Revoked User rejoin and confidentiality of stored data. The computational cost, end-to-end latency and throughput of the system would be evaluated during implementation. However, the system would also model some threat scenarios to Active Attack, Passive attack, and Insider attack.

4.1 Security Analysis

4.1.1 Correctness

Without any loss of generality, the data consumer, Bob, can successfully retrieve only the set of attributes values he obtains access token for, as specified on any data record d by the data owner, Alice. In the homomorphic proxy re-encryption process, for any given key pairs (S_a, P_a) and (S_p, P_p) generated by DO and DC respectively, combining the algorithms earlier stated should satisfy correctness conditions;

$Dec(S_a, E(P_a, M)) = M$, and

$Dec(S_p, F(Z(S_a, P_p), E(P_a, M))) = M$

4.1.2 Confidentiality and Privacy Preservation

The cryptographic approached used in this framework shows that only authenticated and authorized users of this system has the right to data access. The symmetric and asymmetric cryptosystems adopted for data and communication guarantees confidentiality. On the other hand, a privacy preserving framework is ensured by anonymizing identities and the use of *K*-anonymity approach for privacy preserving data publishing with defined class of sensitivity.

4.1.3 Collusion Safety

Suppose the data consumer, Bob, collude with the proxy server to cheat the data owner, Alice, if the proxy releases the support key k_1 and the Public key P_{do}, they would need the alternate key k_2 to form a key Sk_{enc} to record a successful cryptanalysis. Moreover, operations requiring the encryption key of the data records does not require decrypting Sk_{enc} first, as a result of the homomorphic computation.

4.1.4 Revoked User Rejoin

Whenever a revoked user attempts to rejoin the system, the system first revalidate his identity at the IDP Module, evaluate his trust value to meet the threshold and re-issue an appropriate access context to the user. If the set of attributes of the data record on which access is to be granted is same as old, the DO generates the old $A^{u,d}{}_{du}$ and invokes the AC3 Module to add it to $A^{u,d}$ with the corresponding $Enc(P_{du}, Sk_{enc})$. On the other hand if set of attributes is different, The DO generates a new $A^{u,d}{}_{du}$ and invokes the AC3 Module to adds it to $A^{u,d}$ with corresponding $Enc(P_{du}, Sk_{enc})$. Most importantly, any attempt by this user to access data has to ensure that the classification of the new task given to him dominates the security level of the data.

4.1.5 Trust Assurance

The trustworthiness of the entire framework is premised on the Platform – based trust evaluation

and the User – based evaluation of trust. For Platform – based trust we would employ the computation of the reliability of the cloud service properties, components and platform in form of *Hard Trust* and *Soft Trust*. For Hard Trust, we use A Simple Logic Language ALOPA policy (Nagarajan *et al.*, 2014) to determine if a platform satisfies a given property by extrapolating the available set of platform properties to the require set using the ALOPA rules. For Soft Trust, we believe that trust in the property certification authority (CA) is subjective and can vary depending on the context. This requires the formation of trust relationship among fellow data owners to evaluate the CA on trust operations based on past and present experiences. A Derived trust I_v is computed by combining the Recommendation trust I_r and Direct trust I_d, over a trust decay function β, is deduced. The Platform – based trust I_p and an instance n is given as;

$I_{p,n} = I_{h,n} \mathcal{O} \beta(I_{v,n})$

where $I_v = I_r \mathcal{O} I_d$, $\beta = (1 - e^{-(k,\Delta)})$, k is the decay rate and Δ is the number of years that has elapsed since the last trust opinion update.

However, the User Trust is evaluated by the Auditing component of the AC3 module, which evaluate each user's deviation from normal use of the system, in accordance with their access permission. The user trust I_u is computed as;

$I_{u,new} = \beta(I_{u,old})$

5.0 Conclusion and Future Work

Basically, minimal hands-on management and Cost efficiency are laudable attributes of cloud services that motivates data owners to outsource their data in the cloud. However, the security, privacy and trust issues identified with cloud models cannot be overemphasized. Data owners find it difficult to entrust their data with CSPs, who release their data without their knowledge, premised on some purported legitimate requests for data. In view of this, the paradigm shift in research direction would be to give the data owner full control over his data. Existing researches reveal that security, privacy and trust components of a cloud data outsourcing system lacks true interoperability and hence prompted our proposal of a holistic framework called the STPcloud.

In STPcloud, the data owner classifies data by sensitivity of the data records, and confidentially stores sensitive data by symmetric encryption. This data is then outsourced over a PKI – based communication channel for secured communication. The Key management system of framework manages encryption Keys generated and stored in Hardware Security Module (HSM), which guides against logical and physical attacks on the key store. However, the key to the encrypted data is transferred to the proxy server for its additive homomorphic proxy re-encryption, while anonymizing owner's identity. As for the trust component of the framework, the trustworthiness of the cloud platform in delivering services required by the system is estimated and users' direct and recommended trust is evaluated to support decision on authorization. A robust Access Control for Cloud Computing (AC3), was embedded to efficiently manage authorization requirement of the framework. The privacy of the shared data is preserved using k-anonymity. The off – implementation analysis and evaluation of STPcloud shows that the privacy, security and trust of the entire data outsourcing process can be guaranteed. In the near future, we would develop and deploy this framework and evaluate its performance against standard threat models.

REFERENCES

Abmanan Jamalul-Lail, Mohd Faizal Mubarak, Mohd Anuar Mat Isa,Zubair Ahmad Khattak (2014). Security, Trust and Privacy: A New Direction for Pervasive Computing.Recent Researches in Computer Science. ISBN: 978-1-61804-019-0.

Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., and Song, D. Provable data possession at untrusted stores. In ACM CCS (2007).

Badger, L., Grance, T., Patt-Corner, R., & Voas, J. (2011). Draft Cloud Computing Synopsis and Recommendations. National Institute of Standards and Technology (NIST) Special Publication 800-146. US Department of Commerce. May 2011. Available online at: http://csrc.nist.gov/publications/drafts/800-146/Draft-NIST-SP800-146.pdf (Accessed on: November 20, 2012).

Blaze M., Feigenbaum J., Keromytis A. D. (1998). Keynote: Trust management for public-key infrastructures. Infrastructures

(Position Paper). Lecture Notes in Computer Science 1550, 1998; 59–63.

CSA (2012) Trusted Cloud Initiative. http://www.cloudsecurityalliance.org/trustedcloud.html. Data Loss Prevention (2012) http://datalossprevention.com/ TCG, TPM main – part 1 design principles, version 1.2, revision 103, Trusted Computing Group, July 2007.

Davida G. I., Wells D. L., Kam J. B. (1981). A database encryption system with subkeys. In ACM Transactions on Database Systems (TODS), Vol. 6 n.2, p.312-328,June 1981 [doi>10.1145/319566.319580].

Dawei Sun, Guiran Changb , Lina Suna and Xingwei Wanga (2011). "Surveying and Analyzing Security, Privacy and Trust Issues in Cloud Computing Environments", Procedia Engineering, Vol. 15, 2011, pp. 2852-2856.

Dijk M. V., Juels A. (2010). On the impossibility of cryptography alone for privacy- preserving cloud computing. HotSec'10 Proceedings of the 5th USENIX conference on hot topics in security. Article No. 1-8. USENIX Association Berkeley, CA, USA.

EUD (1995), E. U. Directive, "95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data," Official Journal of the EC, vol. 23, 1995.

Fung B.C. M., Wang K., Chen R., Yu P. S. (2010). Privacy-preserving data publishing: A survey of recent developments, ACM Computing Surveys (CSUR), v.42 n.4, p.1-53, June 2010 [doi>10.1145/1749603.1749605].

Gholami A., Svensson G., Laure E., Eickhoff M., and Brasche G. (2013). "Scabia: Scalable Brain Image Analysis in the Cloud," in CLOSER 2013 Proceedings of the 3rd International Conference on Cloud Computing and Services Science, Aachen, Germany, 8-10 May, 2013, pp. 329–336,2013.

Gholami A. and Laure E. (2015), "Advanced cloud privacy threat modeling," The Fourth International Conference on Software Engineering and Applications (SEAS-2015), to be published in Computer Science Conference Proceedings in Computer Science and Information Technology (CS/IT) series.

Habiba U., Masood R., Shibli M.A. and Niazi M. A. (2014). Cloud identity management security issues & solutions: a taxonomy.Complex Adaptive Systems Modeling: a Springer Open Journal. DOI: 10.1186/s40294-014-0005-9.licensee Springer. Published: 11 November 2014. http://casmodeling.springeropen.com/articles/10.1186/s40294-014-0005-9.

Han J, Cai Y, Cercone N. Concept based data classification in relational databases. AAAI workshop on knowledge discovery in databases;1991. pp. 77-94.

Hogan, M., Liu, F., Sokol, A., Tong, J. (2011). 'NIST Cloud Computing Standards Roadmap' National Institute of Standards and Technology, Special Publication 500-291.

Huang, L.T., Deng, S.G., Li, Y., Wu, J., Yin, J.W. and Li, G.X. (2013), "A trust evaluation mechanism for collaboration of data-intensive service.

Jensen M, Schäge S., Schwenk, J. (2010). Towards an anonymous access control and accountability scheme for cloud computing. Cloud Computing (CLOUD). 2010 IEEE 3rd International Conference on. Miami: IEEE; 2010. 540- 1.

Kadhem, H., Amagasa, T. & Kitagawa, H. (2009), A novel framework for database security based on mixed cryptography, in 'Internet and Web Applications and Services, 2009. ICIW '09. Fourth International Conference on', IEEE Computer Society, Venice Mestre, Italy, pp. 163–170. URL: http://dx.doi.org/10.1109/ICIW.2009.31.

Kim H. and Timm S. (2014) "X.509 authentication and authorization in fermiCloud," in Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on, pp. 732–737, Dec 2014.

Li W., and Ping L. (2009) Trust Model to Enhance Security and Interoperability of Cloud Environment. In: Cloud Computing, Lecture Notes in Computer Science, Springer, 5931:69-79.

Nagarajan A., Varadharajan V., Tarr N. (2014). Trust enhanced distributed authorisation for web services. Journal of Computer and System Sciences 80 (2014) 916–934.

Pearson S. (2011). Toward Accountability in the Cloud. IEEE Internet Computing, IEEE Computer Society, July/August, 15(4):64-69.

Pearson S., Yee G. (2013). "Privacy, security and trust in cloud computing," in Privacy and Security for Cloud Computing (S. Pearson and G. Yee, eds.), Computer Communications and Networks, pp. 3–42, Springer London, 2013.

Popa R. A., Lorch J. R., Molnar D., Wang H. J., and Zhuang L. (2011). Enabling Security in Cloud Storage SLAs with CloudProof. Proceedings of the 2011 USENIX conference on USENIX annual technical conference, Pages 31- 31.

Robinson N., Valeri L., Cave J., Starkey T., Graux H., CreeseS.,and Hopkins P., (2010).The Cloud: Understanding the Security, Privacy and Trust Challenges. Final Report TR-933-EC.30 November 2010.Prepared for Unit F.5, Directorate-General Information Society and Media, European Commission.

Sicari S., Rizzardi A., Grieco L.A, Coen-Porisini A. (2015). "Security, Privacy and Trust in Internet of Things: The road ahead" , Science Direct Computer Networks 76, 2015, p 146- 164.

Thilakanathan D. , Chen S. , Nepal S. , Calvo R., and Alem L.(2014)."A platform for secure monitoring and sharing of generic health data in the cloud," Future Generation Computer System, 2014, pp. 102– 113.

USGLB (1999). U. States., "Gramm-leach-bliley act." http://www.gpo.gov/fdsys/pkg/PLAW-06publ102/pdf/PLAW-106publ102.pdf, November 1999.

privacy/generallyacceptedprivacyprinciples/downloadabledocuments/gapprac_%200909.pdf

Waqar A, Raza A et al. (2013). A framework for preservation of cloud users' data privacy using dynamic reconstruction of metadata, Journal of Network and Computer Applications, vol 36(1), 235– 248.

Westin A (1967) Privacy and Freedom. New York, USA, Atheneum.

Younis A. Y., Kashif K., Madjid M. (2014). An access control model for cloud computing. Journal of Information Security and Applications archive Volume 19 Issue 1, February, 2014. Pages 45-60.

Zhou L., Varadharajan V., and Hitchens M. (2013). "Integrating trust with cryptographic role-based access control for secure cloud data storage," in Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on, pp. 560–569, July 2013