
PROCEEDINGS OF THE INTERNATIONAL CONFERENCE IN INFORMATION AND COMMUNICATIONS TECHNOLOGIES

Cresta Golfview Hotel, Lusaka – Zambia

20TH – 21ST NOVEMBER 2019



Organized by The University of Zambia, Copperbelt University and Mulungushi University



Key Sponsors



ISBN NUMBER: 978-9982-70-985-9

PREFACE

The International Conference in ICT is an annual conference organised by the three public universities namely The University of Zambia, The Copperbelt University and Mulungushi University. This year's conference was sponsored by ICT Association of Zambia (ICTAZ), IEEE Zambia Section, Zambia Research Network (ZAMREN), Zambia National Airports and Huawei Zambia.

The conference received a total of 45 research papers and abstracts. Each submission was exposed to blind peer reviewing and was reviewed by a minimum of two experts and some articles up to four experts. EasyChair Conference Management Systems was used to manage the peer reviewing process. Experts were drawn from the Southern and East African Region Universities. A total of 37 abstracts were accepted presented and published in the proceedings. This includes a total number of 36 full papers that were accepted, presented and published in the proceedings. This translates to 80% acceptance rate.



Prof. Jamson Mbale
ICICT2019 Conference Chair



Prof. Douglas Kunda
ICICT2019 Conference Co-Chair

TABLE OF CONTENTS

Contents

PREFACE.....	i
TABLE OF CONTENTS	ii
LOCAL ORGANISING COMMITTEE	v
REVIEW COMMITTEE.....	vi
ABSTRACTS	vii
1.1 William Gomera and George Oreku	vii
Defect of mobile devices usage to Village Community Banks in supporting members' participation: The Case of Tanzania.....	vii
1.2 Valerianus Hashiyana.....	vii
A Smart Remotely Automated Geyser System for Energy Saving	vii
1.3 Victor Kasonde and Christopher Chembe.....	vii
An Analysis of the Best Practices for the Implementation of Data Centers in Zambia.....	vii
1.4 Davies Kintu and Christopher Chembe.....	vii
An Assessment of the Challenges and Benefits of Digital Broadcasting Migration in Zambia.....	vii
1.5 Darius Bwalya, Christopher Chembe and Victor Kasonde	viii
An Insight of Smart Health Care as a Growing Application Need in Zambia: A Systematic Literature Investigation.....	viii
1.6 Nchimunya Chaamwe	viii
A Survey on the State, Adoption and use of ICT in Self-Study by Students: A Case of the Copperbelt University, Zambia	viii
1.7 Hilaire Nkuzimana	viii
Multi-Objective Resource Scheduling for QoS-Based Cloud Computing	viii
1.8 Emmanuel Nyirenda and Christopher Chembe.....	ix
Meter Reading Android Application for Water Utility Companies	ix
1.9 Michael Bwalya and Christopher Chembe.....	ix
Using Cryptography to Secure Android based Mobile Application Systems.....	ix
1.10 Tuesday Bwalya, Akakandelwa Akakandelwa and Edward Mwalimu	x
The Adoption Footprints of Koha as A Library Management System (LMS) in Higher Education Institutions (HEIs) in Zambia.....	x
1.11 Ruth Mulabita and Monde Kabemba	x
Development of a Mobile Based Geographical Information System (GIS) for the Management of Boreholes in Chalala and Chilenje Townships	x
1.12 Victor Kasonde, Darius Bwalya and Josephat Kalezhi	xi

A Comparative Study of Databases for Web Applications Using a Simple but Comprehensive Framework.....xi	
1.13 Michele Mukeshimana, Hilaire Nkuzimana, Xiaojuan Ban, Abraham Niyongere and Jeremie Ndikumagenge	xi
Emotion Recognition using a Multi-view learning method	xi
1.14 Darius Bwalya, Kasonde Victor and Nchimunya Chaamwe.....xi	
Exploring the Need for Context-Awareness Mobile Applications-A literature Analysis Perspective	xi
1.15 Matengele Kaira and Jameson Mbale	xii
Vehicle Tracking System with Passive security using Pythagoras theorem.....	xii
1.16 Kennedy Kanja and Chimanga Kashale	xii
Exploring the Potential of Sentinel-2 Imagery for Land Use and Land Cover Mapping - A Case of Kopa area in Kanchibiya District, Zambia	xii
1.17 Macmillan Simfukwe, Kondwani Chanda and Christopher Chembe.....	xiii
An Expert System for the Diagnosis of Poultry Diseases	xiii
1.18 Hastings Maboshe Libati and Paul Chima	xiii
An Inquiry into the Public Sector Employee Resistance to Information And Communications Technology Based Changes: The Case of IFMIS, in the Ministry of Finance and General Education	xiii
1.19 Kuliwa Sibajene Syakwasia and Christopher Chembe.....	xiii
A Comparative Study of the Performance between Boys and Girls in Computer Studies in Three Selected Secondary Schools in Lusaka District, Zambia.....	xiii
1.20 Innocent Mwila and Jackson Phiri	xiv
Modelling to Help the Prevention of the Spread of TB through Improved Monitoring and Tracking Using Cloud Computing, Geospatial Data and Web Technologies.....	xiv
1.21 Raphael Banda, Jackson Phiri and Mayumbo Nyirenda	xiv
Model Development of a Recommender System for Cognitive Domain Assessment Based on Bloom's Taxonomies	xiv
1.22 Twataizya Minango, Jackson Phiri and Faustin Banda	xv
GIS Management Tool for Managing of Underground Electrical and Telecommunication Networks at the University of Zambia	xv
1.23 Felix Kabwe and Jackson Phiri	xv
A Framework for Digital Identity Management	xv
1.24 George Kasanga and Jackson Phiri	xv
Passenger and Luggage Tracking System Using Sensor Networks: For Livingstone and Lusaka Intercity Bus Terminal	xv
1.25 Aaron Zimba, Mumbi Chishimba, Christabel Ngongola and Tozgani Fainess Mbale.....	xvi
Demystifying Cryptocurrency Mining Attacks: A Semi-supervised Learning Approach Based on Digital Forensics and Dynamic Network Characteristics.....	xvi
1.26 Fickson Mvula, Jackson Phiri and Simon Tembo	xvi

A Conceptual Secure Blockchain Based Settlement And Clearinghouse For Mobile Financial Services In Zambia	xvi
1.27 Jonathan Phiri and Jackson Phiri	xvi
Crime Mapping Model Based on Cloud and Spatial Data - A Case Study of Zambia Police Service	xvi
1.28 Gabriel Chibuye, Jackson Phiri and Faustin Banda	xvii
A Spatial Framework for Managing Sewer and Water Networks Using Sensor Networks: A case of the University of Zambia	xvii
1.29 Banji Milumbe, Jackson Phiri and Mayumbo Nyirenda	xvii
Enhancing security of examination question papers through a tracking system based on spatial and cloud technologies.....	xvii
1.30 Bruce Mwiya, Mukwemba Mangangu and Beenzu Siachinji	xviii
Examining Antecedents of Social Media Adoption as a Business Platform among Student Entrepreneurs: Evidence from Zambia	xviii
1.31 Knox Kamusweke, Mayumbo Nyirenda and Monde Kabemba.....	xviii
A Data Mining Model for Predicting and Forecasting Fraud in Banks.....	xviii
1.32 Jimmy Katambo, Mayumbo Nyirenda and David Zulu	xviii
Distributed Spatial Search Using Paillier Cryptosystem and a Distributed Ring Algorithm	xviii
1.33 Zaima Banda and Mayumbo Nyirenda	xix
Understanding the Transactional Challenges Encountered in Village Banking While Exploring a Blockchain Enabled Solution: A Case on Zambia	xix
1.34 Tobias Makai and Mayumbo Nyirenda.....	xix
Categorisation of Sexual Reproductive Health Short Messages Texts into Thematic Areas Using Text Mining ..	xix
1.35 Makani Mweembe, Monica Kabemba and Jackson Phiri	xx
The use of Coding Environments to Teach Computer Programming to Children in Upper Primary	xx
1.36 Joe Phiri and Mayumbo Nyirenda.....	xx
Medical Equipment and Laboratory Services Support System	xx
1.37 Andrew Biemba and Mayumbo Nyirenda.....	xx
Augmented Reality Aided Navigation System Model for Support of Student Registration Process	xx
ICICT2019 PROCEEDINGS ARTICLES	xxii

LOCAL ORGANISING COMMITTEE

- | | |
|-------------------------------------|--------------------------------------------------------------------|
| 1. Prof. Jameson Mbale | Copperbelt University (Chair) |
| 2. Prof. Douglas Kunda | Mulungushi University (Co-Chair) |
| 3. Dr. Jackson Phiri | The University of Zambia (Editorial) |
| 4. Dr. Josephat Kalezhi | Copperbelt University (Editorial) |
| 5. Dr. Mayumbo Nyirenda | The University of Zambia (Secretariat) |
| 6. Dr. Christopher Chembe | Mulungushi University (Secretariat) |
| 7. Mrs. Monica M. Kalumbilo-Kabemba | The University of Zambia (Secretariat) |
| 8. Mr. Bonny Khunga | ZAMREN ((Secretariat)) |
| 9. Mr. Ariel Phiri | Zambia Airports Corporation Limited and
IEEE Zambia (Technical) |
| 10. Mrs Mercy Fandamu | Copperbelt University (Secretariat) |
| 11. Ms Mulenga Mwenya | ICT Association of Zambia (Treasury) |

REVIEW COMMITTEE

1. Prof. Douglas Kunda - Mulungushi University, dkunda@mu.edu.zm
2. Prof. Jameson Mbale - Copperbelt University, jameson.mbale@cbu.ac.zm
3. Prof. Lisa Seymour - University of Cape Town, Lisa.Seymour@uct.ac.za
4. Prof. Hastings Libati - The Copperbelt University, libati@cbu.ac.zm
5. Dr. Josephat Kalezhi - Copperbelt University, kalezhi@cbu.ac.zm
6. Dr. Jackson Phiri - The University of Zambia, jackson.phiri@cs.unza.zm
7. Mr. Bonny Khunga - ZAMREN, khungab@zamren.zm
8. Mr. Ariel Phiri - Zambia Airports Corporation Limited and IEEE Zambia, ariel.phiri@zacl.aero, ahphiri@ieee.org
9. Dr. Christopher Chembe - Mulungushi University, cchembe@mu.edu.zm
10. Dr. Edmore Chikohora - Namibia University of Science and Technology, chikohora@nust.na
11. Dr. Derrick Ntalasha - Copperbelt University, dbntalasha@gmail.com
12. Dr. Mayumbo Nyirenda - University of Zambia, mnyirenda@unza.zm
13. Dr. Vincent Omwenga - Strathmore University, vomwenga@strathmore.edu
14. Mr. David Zulu - University of Zambia, makadanizulu@gmail.com
15. Dr. Shemi Alice - Copperbelt University, shemiap@gmail.com
16. Mrs. Monica Kalumbilo-Kabemba - The University of Zambia, monica.kalumbilo@cs.unza.zm
17. Dr. Okuthe Kogeda - Tshwane University of Technology, kogeda@gmail.com
18. Dr. Manoj Lall - Tshwane University of Technology, LallM@tut.ac.za
19. Dr. George Mufungulwa - The Copperbelt University, mufungulwac@gmail.com
20. Alinani Simukanga - The University of Zambia, alinani.simukanga@cs.unza.zm
21. Mukelabai Mukelabai - The University of Zambia, mukemuke2008@gmail.com

ABSTRACTS

1.1 William Gomera and George Oreku

Defect of mobile devices usage to Village Community Banks in supporting members' participation: The Case of Tanzania

Abstract: Village Community Banks (VICOBA) and mobile phones have become assistance to many rural societies in encountering financial problems. However, mobile phone usages to VICOBA have isolated many resulting to unsubstantial benefits that do not foster member's participation. The weak participation of members to Village Community Banks activities has been one of the major impediments to their success despite of owning of mobile devices to some members. This paper is exploring the situation through identifying the deficiency available in mobile devices to VICOBA microcredit programmes in enhancing member's participation and suggest the effective usage pattern. The study captured 7 groups of village community banks at Ndungu village with total of 70 members whereas 48 female and 22 male participants. The study employed interview, observation and focus group discussion for data collection. Content analysis and coding approaches used to analyse data and interpreting findings. The findings of the study showed that there is big potential in mobile phone use to village community banks activities. It could potentially enhance members' participation in microcredit programmes even though the current usage does not foster members.

1.2 Valerianus Hashiyana

A Smart Remotely Automated Geyser System for Energy Saving

Abstract: A Geyser system is one of the house hold appliances which consumes a lot of Electrical Energy specially since the system rely on the manual turning the system on and off. However most of the time we find or have a problem with the Geyser been on for too long in the house and there is no one to turn it off or on when needed. This project proposes a remotely controlled Geyser system for Energy Saving that monitors the use of the geyser system and helps with automatic control of on and off state of the Geyser unit. This form of automation can be done by implementing Internet of Things (IoT) by developing a Mobile Application that will be connected to a unit that enable users to control the On and Off state and schedule of the Geyser unit remotely using the Internet. This research suggests a better way of saving Electrical Energy through the use of a Mobile Application which can remotely control the Geyser unit. This approach can help resolve the way in which people manage the Geyser units in houses as it enables them to control these units remotely using a Mobile Application.

1.3 Victor Kasonde and Christopher Chembe

An Analysis of the Best Practices for the Implementation of Data Centers in Zambia

Abstract: the purpose of this study was to analyse the Best Practices for the Implementation of Data Centers in Zambia. Various literatures were reviewed to consider what other researchers have done. To this effect, the study has established that the advent of new technologies, such as blade servers, that require substantial incremental power and cooling capacity; the pressure to consolidate multiple data centres into fewer locations; the need for incremental space; changes in operational procedures; and potential changes in safety and security regulations converge to impose constant facilities changes on the data center implementation process in Zambia.

1.4 Davies Kintu and Christopher Chembe

An Assessment of the Challenges and Benefits of Digital Broadcasting Migration in Zambia

Abstract: Digitization is the current trend in Television broadcasting, both in Zambia and the world over. The International Telecommunication Union (ITU), set 2015 as the year for the entire broadcasting stations

in the world to go digital. Despite the advantages that come with digital migration, Zambia is yet to realize the full benefits of digital migration implementation. It's against this background that the study is to be conducted to assess the challenges of digital migration process and its benefits in Zambia. To this effect, a quantitative research design will be used on a sample of 300 stakeholders who have direct influence on the process to determine their views. The data collected shall be analysed using Statistical Package for Social Sciences (SPSS) to categorise the challenges of the migration process and its benefits. The study is aimed at assessing the digital terrestrial television migration implementation process; examine the challenges of switching from analogue to digital technology and its benefits.

1.5 **Darius Bwalya, Christopher Chembe and Victor Kasonde**

An Insight of Smart Health Care as a Growing Application Need in Zambia: A Systematic Literature Investigation

***Abstract:** With the current trend of world population exploding, there is a rising constraint on the provision of quality healthcare. This exactly describes Zambia in which according to the United Nations Department of Economic and Social Affairs: Population Division, the current population is now well over 18 Million against a limited number of medical staff, creating a situation where access to a doctor is characterized by long waits which escalate in hours in most cases. Add to that management of chronic illness. It is undeniably clear that there is a great need to explore other viable healthcare alternatives. Thanks to ICT technology, there is an emerging form of healthcare provision called Smart Health Care in which Cloud Computing and Internet of Things (IoTs) is expected to complement each other, creating a great possibility to automate health care. Before Zambia adopts Smart health Care, it must be understood and appreciated by the stakeholders. This paper is therefore a systematic literature investigation of an insight of Smart health Care provision. The paper highlights what the components of Smart Health Care (Cloud Computing and IoTs) are and how the two technologies can collaborate as well as the challenges arising from this. Finally the paper shows how this technology can be adopted in Zambia in its endeavor to improve provision of healthcare.*

1.6 **Nchimunya Chaamwe**

A Survey on the State, Adoption and use of ICT in Self-Study by Students: A Case of the Copperbelt University, Zambia

***Abstract:** The past decade has seen a proliferation of ICT tools in Universities. Universities all over the world including Zambia have taken advantage of this increase by integrating ICTs in the teaching and learning of students. The internet is one such tool that has revolutionized the way learning and teaching is being conducted in Universities. Little research has been done though to verify the process of how university students adopt and use ICTs for their learning in Zambian Universities. The research looked at Student experience in using ICTs in their learning, especially during their self-study at The Copperbelt University (CBU) in Zambia. The results showed that most students have adopted the use of ICTs in their learning. Of the surveyed students, 94.7% of them indicated that they consulted the internet sometimes or most of the times when studying. This is despite the obstacles they faced when accessing the internet. The paper recommended the improvement in internet connectivity and the increase in wireless access points*

1.7 **Hilaire Nkunzimana**

Multi-Objective Resource Scheduling for QoS-Based Cloud Computing

***Abstract:** Cloud computing represents the new computing paradigm. Due to its essence and nature it has brought some challenges such as security and quality of service. In order to ensure good quality of service, one should carefully manage computing resources available in the cloud. In this paper, we propose a cloud computational resource scheduling algorithm based on a genetic algorithm. First, the genetic algorithm is improved and applied to solve the problem of virtual machine allocation in cloud computing. We describe the*

modelling process for chromosomes and genes that correspond to the virtual machine scheduling problems. This scheduling algorithm aims at optimal placement of virtual machines in order to make full usage of the CPU and hence curb the waste time and energy of idle machines.

1.8 Emmanuel Nyirenda and Christopher Chembe

Meter Reading Android Application for Water Utility Companies

Abstract: *Water utility companies have a billing process which involves reading the meter as well as knowing the consumption of its customers. The major challenge faced by water utilities in Zambia is inaccurate customer billing. Affecting hugely this inaccurate customer billing is the manual meter reading process. For most water utilities in Zambia meter reading is a very complex task. The current method for meter reading uses manual work of reading the meter and updating details of the customers. The readings are maintained in a record book and due to manual processing of data there are possibilities of generating an incorrect bill. This process cause extra work load on person who is capturing meter readings. The manual process is associated with the following challenges; transfer errors, poor handwriting, manipulation of figures, lack of analysis of readings, low productivity, time pressure, increase in non-revenue water and meter readers not visiting the property. The other problem with manual meter reading is that both the customer and the water utility company have challenges in monitoring and analyzing of consumption trends for customers and it is very difficult to monitor meter readers in the field. Although technologically advanced solutions exist, most water utilities in Zambia have no budget for them. The proposed model comprised of the android mobile application for capturing readings in the field. A usability test was undertaken consisting of 36 respondents which comprised of 75% meter readers and 25% customers. The results obtained from the mobile application, showed that the capturing, uploading, monitoring and analysis of readings were very good for the users and customers who participated in the study. Three user experience areas were used to undertake the survey namely, perceived ease-of-use, perceived flexibility and perceived information accuracy. With the android application customers are also able to take their own readings and view graphical usage at any time anywhere.*

1.9 Michael Bwalya and Christopher Chembe

Using Cryptography to Secure Android based Mobile Application Systems

Abstract: *Currently mobile applications are playing major role in many areas such as banking, social networking, financial apps, entertainment and many more. With this increasing number of applications, security is an important issue. The growth of android market has increased security risk and thus focus should be given to the security. Security is the biggest issue in the field of mobile technology. Therefore, mobile applications need to be assessed and ensure that secure coding practices have been followed during development. Data that is shared on an unsecured channel is vulnerable to attacks and to stop unauthorized access to this data, there is need to encrypt the data before it is sent to the server. In this research work, different cryptographic algorithms for encrypting data and secure data sharing in mobile applications across communications channels were examined. Simulation methodology was used to investigate a suitable cryptographic algorithm and to design a security framework for mobile applications to solve mobile application security problems. The proposed framework employs the use of Advanced Encryption Standard (AES) algorithm for encrypting meter readings data being exchanged between a smart phone and the server. The results obtained from the simulation of the security framework, showed that the four fields namely: Account number, image path, meter number and phone number on which AES encryption was applied were in an unreadable format (ciphertext), implying that the fields have been successfully encrypted. This solution allows application users to transfer (upload readings) data between a smart phone and database server in a secure manner without facing the problem of data attack. Data being uploaded to the server is encrypted before it is transferred and decrypted once it reaches the server side. This solution addresses android application security in the application and network communications layers and data transmission. The research paper ensures information security is guaranteed between an organisation and its customers.*

1.10 Tuesday Bwalya, Akakandelwa Akakandelwa and Edward Mwalimu

The Adoption Footprints of Koha as A Library Management System (LMS) in Higher Education Institutions (HEIs) in Zambia

Abstract: Koha has been one of the widely used free and open source library management systems (FOSLMS) in libraries. Libraries in higher education institutions (HEIs) in Zambia like any other libraries in other countries have not stood aloof in the use of Koha to automate their operations. This study therefore thought to investigate the use of Koha among HEIs in Zambia. The study was quantitative in design and across-sectional survey was used. A complete census method was used in which all the 54 HEIs with libraries that have automated their operations were surveyed. All the 54 respondents answered the questionnaire; resulting in 100% response rate. The research findings show that 41 (76%) of HEIs have automated their libraries using Koha. The study further shows that, 14 (34.14%) Colleges of Education and (31.70%) Universities in Zambia have adopted Koha. Furthermore, 10 (24.39%) TEVET colleges and 5 (12.19%) were found to be using Koha. This implies that nursing schools have the lowest adoption footprints of Koha. It also been established that public owned HEIs have adopted Koha more than the privately owned HEIs. The study has also shown that there is no association between the adoption of Koha in HEIs and demographic variables as the hypothesis tests conducted using Chi-square at a significance level of 0.05 (95%) show no association between gender and adoption of Koha ($\chi^2= 1.612$; $df=1$; $p=0.204$), between categorized age and Koha adoption ($\chi^2=2.614$; $df=1$; $p=0.181$), academic qualification of librarian and Koha adoption ($\chi^2= 5.61$; $df=1$; $p=0.517$ and between work experience and adoption of Koha ($\chi^2=4.160$; $df=2$; $p=0.124$).

1.11 Ruth Mulabita and Monde Kabemba

Development of a Mobile Based Geographical Information System (GIS) for the Management of Boreholes in Chalala and Chilenje Townships

Abstract: Domestic and commercial groundwater usage is inevitable in the absence of the right amount of water supply in terms of availability, quality, cost and pressure. This leads to people accessing ground water via the use of boreholes. They find boreholes to be self-sufficient save money and at the same time make money. In some areas boreholes are drilled by those who can afford for business purposes such as hand pumps, wells or solar pump. There are different mechanisms for drilling these boreholes. The massive borehole drilling has a great effect on ground water. The ground water might be contaminated as these boreholes are sunk right next to septic tanks. An improperly designed, located, constructed, or maintained septic system can leak bacteria, viruses, household chemicals, and other contaminants into the groundwater causing serious problems. In order to control and manage the usage of ground water responsible authorities require the necessary type of information to support decision making and access to the resource. The objective of this study was to develop a Mobile based geographical information system for the management of boreholes in Chalala and Chilenje Townships. A mobile phone GPS facility was used to map the boreholes. This connects to the geographical information systems (GIS) database that stores all spatial data for the operational areas. A Mobile phone was bought with GEO-ODK application installed on it for data collection. The GPS location is the key for data analysis. 100 households in Chilenje and Chalala were visited for this survey. Another 20 from different occupations were also interviewed, this included employees from WARMA, Borehole Drillers and Farmers. The involvement of the residents in the registration process ensures ownership and responsibility of the all process. The study proved that there were a number of challenges faced in the collecting of data on borehole information. These challenges can be overcome by developing the Web-based system and using data collection tools that send the information to the server. The Geo-ODK is open source software. The system is based on open source application and postgresql is used for database management. The system has to be able to be connected with other GIS application for future enhancements. There is need to enhance this system in future to enable it provide real time data. The same Android phone will be used by the Hydrologists employed by the ministry to do spot checks on areas where people have sunk the boreholes. A Web-based Geographical information system if implemented will make it easier for all stakeholders to monitor the boreholes that are already drilled and the yet to be drilled boreholes.

1.12 Victor Kasonde, Darius Bwalya and Josephat Kalezhi

A Comparative Study of Databases for Web Applications Using a Simple but Comprehensive Framework

Abstract: The purpose of this study was to compare database technologies for Web Applications with regards to their performance, security, cost, usability, support and scalability. The database technologies considered included MySQL, PostgreSQL, Microsoft SQL Server database, Oracle, MongoDB, SAP Hana, MariaDb. The results indicated that MySQL provided many performance- features that helped web-based applications perform at a rate that modern businesses demand. The study further revealed that MySQL is user-friendly and provides many security features for its implementation than other database technologies. Other features that makes MySQL more usable than others included the fact that it is readily available and can be customized to suit a particular user need besides the high level of security capabilities that it provides. The study recommended that web developers have to consider the benefits of MySQL database technologies for possible implementation and it is hoped that this research will be of great value to the web developers in determining the features that their web application need and ultimately the right database to implement.

1.13 Michele Mukeshimana, Hilaire Nkunzimana, Xiaojuan Ban, Abraham Niyongere and Jeremie Ndikumagenge

Emotion Recognition using a Multi-view learning method

Abstract: The increasing use of the computing devices and applications in human daily life, triggers the need of a natural human computer interaction. Emotion Recognition using multiple features using semi-serial fusion method is proposed. The study analyses the impact of the feature combinations in enhancement of the recognition enhancement. The paper presents the use of the multi-view learning principle to fusion multiple features in audio and facial expression-based for unimodal emotion recognition. The results show that proposed method is effective. In comparison to other methods using one type of features or combining them in a concatenated way, this new method outperforms others in recognition accuracy, execution reduction, and stability.

1.14 Darius Bwalya, Kasonde Victor and Nchimunya Chaamwe

Exploring the Need for Context-Awareness Mobile Applications-A literature Analysis Perspective

Abstract: The prevalence of smart mobile devices has brought about the need for applications to be location aware, content-aware, activity-aware, environmental and user-aware, a technological phenomenal now being referred to as Context-awareness mobile application. Great strides have been made in contextualization by means of Global Positioning System (GPS) enabled applications and Pervasive Computing. In this paper, we aim to understand mobile application context-awareness and why its application scenario should be significant in the Zambian context. We begin by examining literature on context-awareness to fully appreciate its meaning, identify factors or components of context and group into location-aware, Activity-aware, Environment-aware and User-aware and we further show why Zambian mobile app developers, business people including medical experts should think of implementing and utilize this form of smart dynamic technology to present more relevant information, functionalities to their clients depending on circumstances.

1.15 Matengele Kaira and Jameson Mbale

Vehicle Tracking System with Passive security using Pythagoras theorem

Abstract: *The REST based Vehicle Tracking System with Passive Security using Pythagoras theorem is a low cost vehicle tracking system that is adept for tracking in regions that experience poor network availability (this is the case in Zambia). Most tracking systems require active input from a user to repeatedly send commands to the vehicle unit usually through SMS to get back coordinates which are displayed on a map. The difference with the proposed approach is that the coordinates which will be polled at a specified interval will be cached on local storage and then a RESTful service will be invoked by a C program running on an Arduino microcontroller at a specified interval, this will then log all previously collected coordinates into a database on a server allowing for replay of the movements of the vehicle. A web application that has google maps integrated into it will then query these coordinates and display the latest coordinates on the map. The proposed system will also employ a concept of passive security by allowing the user to pre define an allowable radius on a server which the vehicle tracking system will test against every time it polls coordinates. The moment the vehicle is outside the allowable radius, the vehicle unit will utilize Pythagoras theorem to determine the car has left the allowable radius without a need for server side computation. This can then result in the following actions; the engine will cut and the doors will lock. And an alert will be sent to the authorities and the owner that the vehicle was driven outside the allowable radius. This feature as you can imagine will be very useful for protecting vehicles that are stolen in a stealthy manner or when the owner is not within reasonable proximity of the vehicle to know it has been stolen. Applicable examples could be when the owner is asleep or has travelled outside the country and left the car at the airport. Pythagoras theorem has been selected as the passive security features initiation mechanism because it is a simple formulae that almost every student is familiar with thus very easy to explain. However when applied to GPS coordinates directly the resulting hypotenuse is in Cartesian coordinates and not a distance in meters or kilometers that can be easily interpreted by a human being in relation to his or her environment, hence there is need to modify the implementation of the algorithm slightly so it converts the result into meters and kilometers. Another reason Pythagoras theorem has been selected is because most vehicle units have little resources available for complex computation thereby forcing most system to push computation onto the server side which requires more server resources than the distributed approach suggested in this research. Furthermore if connectivity between the unit and the server is unavailable the passive security features residing on a server become ineffective.*

1.16 Kennedy Kanja and Chimanga Kashale

Exploring the Potential of Sentinel-2 Imagery for Land Use and Land Cover Mapping - A Case of Kopa area in Kanchibiya District, Zambia

Abstract: *Accurate land cover mapping is an important component in land use land cover change analysis. Land cover mapping is accomplished using Remote Sensing by way of image classification. With recent developments in Remote Sensing, freely available high resolution images are now available with a resolution of as high as 10m. This study therefore explored the potential of the freely available Sentinel-2 imagery for land cover classification. Sentinel-2 image was downloaded from ESA website and Landsat-8 OLI image was downloaded from USGS website. Random Forest (RF) algorithm was used to classify the two images. The overall classification accuracy was 60% and 72% for landsat-8 OLI and Sentinel-2 imagery respectively. The Kappa indexes were 52% for Landsat-8 OLI and 61% for Sentinel-2 imagery. These results proves that a higher classification accuracy can be achieved by using 10m resolution bands from Sentinel-2 imagery. This implies that land use land cover change analysis can be achieved more accurately and at a finer scale using Sentinel-2 imagery. Land use land cover change analysis is an important component in climate change analysis.*

1.17 **Macmillan Simfukwe, Kondwani Chanda and Christopher Chembe**

An Expert System for the Diagnosis of Poultry Diseases

***Abstract:** An expert system is a computer program that can perform tasks that are normally performed by human beings, using intelligence. Expert systems keep knowledge in their knowledge bases and use this knowledge to make inferences and perform reasoning. Expert systems are widely used in various fields such as diagnosis of diseases, diagnosis and repair of machinery, creation of investment portfolios, analysis of the chemical composition of substances, etc. Poultry farming has become a very popular economic activity in Zambia. Thousands of Zambians are engaged in poultry farming as a business, because it is easy to run a poultry business and it also does not require a lot of capital to start. However, the birds, especially the broilers, are susceptible to a variety of diseases, most of which require special veterinary knowledge to correctly diagnose. Unfortunately, there are very few veterinary clinics in Zambia, especially in rural areas. Hence, in times of epidemics, the poultry farmers have nowhere to run to for professional help and end up incurring losses when their birds die. In order to assist the poultry farmers with disease diagnosis and treatment advice, we have developed an expert system that is capable of diagnosing poultry diseases. The aim is to detect the diseases as early as possible in order to treat them early and avoid epidemics. In this paper, we present the expert system for the diagnosis of poultry diseases that we built.*

1.18 **Hastings Maboshe Libati and Paul Chima**

An Inquiry into the Public Sector Employee Resistance to Information And Communications Technology Based Changes: The Case of IFMIS, in the Ministry of Finance and General Education

***Abstract:** Change is inevitable in organizations. It happens continuously and often at such a rapid rate to the extent that it influences the efficiency and effectiveness with which an organization carries out its business mandate. Because change has become an everyday part of organizational dynamics, employees who resist change can actually cripple an organization. One of the major changes taking place today in businesses is technological change. Change in technology can have both positive and negative effects on the performance and attitude of employees. The general objective of this study was to establish the reasons why employees resist technological change in this case the Integrated Financial Management Information Systems (IFMIS) based change in public organizations. The study also sought to identify change management strategies and approaches that can be used to manage or prevent resistance to change. The study employed a descriptive approach to establish factors that influence employees' resistance to Information and Communication Technology (ICT) change in the Ministries of Finance and General Education of the Government of Zambia. Major reasons for employee resistance to technological change were found to be uncertainty of the outcome of the technological change, lack of capacity building and skills training, poor and untimely communication about change, lack of management support, lack of employee involvement in the change process and fear of loss of their control. The study also suggests other avenues for further studies on the same topic*

1.19 **Kuliwa Sibajene Syakwasia and Christopher Chembe**

A Comparative Study of the Performance between Boys and Girls in Computer Studies in Three Selected Secondary Schools in Lusaka District, Zambia

***Abstract:** This study investigated the performance between secondary school boys and girls and their attitudes towards computer studies in Lusaka district, Zambia. The study adopted a descriptive survey method. The instruments used were a questionnaire and past academic performance results in computer studies. Firstly, to determine difference in performance, a sample of 387 grade 9 results for pupils who sat for their computer studies examination in 2018 from three selected schools were collected and analyzed. Secondly, a structured questionnaire on pupils' attitude towards computer studies with a battery of 19 questions on a five point Likert scale was administered on a sample of 131 grade 9 computer studies pupils preparing for their 2019 Junior Secondary examination. The respondents' results and responses were scored*

and analyzed using t-test and descriptive statistics. The results of the study revealed a significant difference between secondary school boys' and girls' performance in computer studies. However, there was no evidence to show that all boys perform better than girls. Findings on boys' and girls' attitude towards computer studies showed no significant difference. However, the study found boys to have a slightly higher positive attitude than girls. Evidence suggests that agents of socialization play a major part in the observed performance difference. Recommendations were made based on findings of this study. Pupils and parents should be encouraged to cultivate the right attitude towards computer studies. Teachers and schools should provide selective help to improve the attitude of girls towards computer studies and guard against stereotyped views in the way they interact and organize their classes. The Ministry of Education should provide schools with adequate computers, electricity, qualified teachers and all the required teaching and learning materials.

1.20 **Innocent Mwila and Jackson Phiri**

Modelling to Help the Prevention of the Spread of TB through Improved Monitoring and Tracking Using Cloud Computing, Geospatial Data and Web Technologies

***Abstract:** Today, stakeholders in the fight against Tuberculosis (TB) face a dilemma of the quick spreading disease that must be fought with minimal resources. The scarcity of the resources available for the fight against TB calls for creative and strategic ways if this fight is to be won. Information and Communication Technologies (ICTs), cloud computing, geospatial data analysis and web technologies presents an opportunity towards the creative fight against TB. The ability of these technologies to capture and present spatial data in real-time on mapping applications makes them a good candidate for exploration in the fight against TB. The aim of this paper is to develop a decision support model that would help in the prevention of the spread of TB through improved monitoring and tracking using cloud computing, geospatial data and web technologies. This system will display on a mapping application the geographical distribution of TB cases and other reports which can be used to make decisions in the prevention of TB. Using the information provided by this system, TB program managers will be able to put the scarce resources where it matters more in the fight against the disease - where the cases are.*

1.21 **Raphael Banda, Jackson Phiri and Mayumbo Nyirenda**

Model Development of a Recommender System for Cognitive Domain Assessment Based on Bloom's Taxonomies

***Abstract:** Assessment is used a lot in the schools and universities and is synonymous to evaluation. It is used to measure learners' competences and proficiency after undertaking some course or a set of instructions. Online or hard copies Tests used to assess learners are considered as assessment tools. These tools especially when using outmoded tools can be a big challenge in assessing the learners. Some of the challenges range from flawed data entry owing to manual entries and failure to interpret and systematically award marks according to the questions asked. We introduce a system that uses the question ranking system based on action verbs of Blooms Taxonomies (BT). This technology automatically assigns each question a term weight according to the level of the verb on BT. The traditional way of assessing learners does not take care of verb weights as a result the output or interpretation of such results is exceedingly flawed and its credibility questionable. We shall develop an advanced model to decode the Bloom's Taxonomy of the Cognitive Domain in awarding marks to questions. The system shall also assist educators to find appropriate action verbs to use to assess their candidates.*

1.22 Twataizya Minango, Jackson Phiri and Faustin Banda

GIS Management Tool for Managing of Underground Electrical and Telecommunication Networks at the University of Zambia

***Abstract:** The University of Zambia's electricity and telecommunication utilities are supplied by underground cables. However, the current management system lacks spatial information about these cables. This is due to the fact that the current system is based on out dated maps and drawings that lack locational data. This problem has resulted in the cutting of these utility cables during day to day construction of new infrastructure within campus. To efficiently manage the underground electrical and telecommunication cables, an improved system is required. In order to develop this system, the cables have to be mapped and digitally stored. Digital mapping of these cables will play a vital role towards their management. This would provide for accurate location of these cables as well as easily make better decisions when installing new network connections. Therefore, this paper focuses on the development of the management system of the existing underground electrical and telecommunication cables at the University of Zambia. To achieve this, the use of free and open source tools such as GPS, PostgreSQL/PostGIS and QGIS desktop software were used to establish a digital management system to easily update, store and analyze the utility network lines. A desktop and web based GIS prototypes were developed in this study.*

1.23 Felix Kabwe and Jackson Phiri

A Framework for Digital Identity Management

***Abstract:** The growth in the use of services on the World Wide Web has resulted in the proliferation of online cyber mischief, misrepresentation, and cybercrime. Cyber fraudsters and criminals hide their online identities to steal services and other valuables or harm innocent online services' users. We will reflect on the major sources of identity attributes currently being used in the application and registration forms for the various services offered both in the cyber and real space. The study shall also consider an information systems model that would explain factors that influence users' decision of who uses these tokens, when the tokens are used, and why identity tokens are used. This study ruminates identity tokens paradigms and their levels of importance. The study further explores the extraction of key identity attributes from identity tokens like identity documents, application and registration forms for the various services offered both in the cyber and real space. The techniques that have been used include data collection techniques, data mining techniques and statistical analysis. The primary data obtained from questionnaires was analysed using Statistical Package for Social Science (SPSS) and Excel. The respondents were drawn from Banks (14%), Churches (12%), Government of the Republic of Zambia (6%), Hospitals (16%), Insurance (10.7%), Mobile Phone companies (2%), and less than 1% from Pensions. Others were Schools (21%), Universities (16%), and Utility companies (1.3%). The paradigms under study include Usefulness, Trust, Ease of use, Image, and User satisfaction. The results helped in establishing a framework that was guiding the research.*

1.24 George Kasanga and Jackson Phiri

Passenger and Luggage Tracking System Using Sensor Networks: For Livingstone and Lusaka Intercity Bus Terminal

***Abstract:** Technology has revolutionized communication and how services can be delivered to the intended end users .today many areas such as supply chain, logistics, health, aviation and education have been greatly influenced. Technology adoption Model (TAM) has been widely used in studies to ascertain acceptance of new technology in various sectors. Three variables are considered perceived usefulness (PU), perceived ease of use (PE), perceived risk (PR) Adopting and implementing appropriate technology has become a source of competitive advantage for organizations. This study aims at understanding how technology can be used to streamline operations and services of bus operators at intercity bust terminus in Lusaka, Zambia The proposed solution makes use of wireless sensor networks (WSN).*

1.25 **Aaron Zimba, Mumbi Chishimba, Christabel Ngongola and Tozгани Fainess Mbale**

Demystifying Cryptocurrency Mining Attacks: A Semi-supervised Learning Approach Based on Digital Forensics and Dynamic Network Characteristics

***Abstract:** Crypto currencies have emerged as a new form of digital money that has not escaped the eyes of cyber-attackers. Traditionally, they have been maliciously used as a medium of exchange for proceeds of crime in the cyber dark-market by cyber-criminals. However, cyber-criminals have devised an exploitative technique of directly acquiring crypto currencies from benign users' CPUs without their knowledge through a process called crypto mining. The presence of crypto mining activities in a network is often an indicator of compromise of illegal usage of network resources for crypto mining purposes. Crypto mining has had a financial toll on victims such as corporate networks and individual home users. This paper addresses detection of crypto mining attacks in a generic network environment using dynamic network characteristics. It tackles an in-depth overview of crypto mining operational details and proposes a semi-supervised machine learning approach to detection using various crypto mining features derived from complex network characteristics. Such an approach is helpful during security mitigation by network security administrators and law enforcement agencies.*

1.26 **Fickson Mvula, Jackson Phiri and Simon Tembo**

A Conceptual Secure Blockchain Based Settlement And Clearinghouse For Mobile Financial Services In Zambia

***Abstract:** Developing Countries in Africa in general and Zambia in particular, have seen a rapid rise in use of mobile payment platforms. This has not only revolutionized access to finance for the poor but also allowed them access to other financial products such as savings or insurance. Mobile financial wallets are being used by different mobile network operators to extend their product offering beyond the traditional voice services. Equally other non-mobile network providers have joined the race in providing mobile money wallets. As a result of these different mobile wallet providers, subscribers are presented with a problem of interoperability between them where transfer of value from one provider wallet to another on a different network is not possible. In this study we first review the technical landscape and features of mobile payment systems in Zambia and then assess the feasibility of using blockchain technology in proposing a settlement and clearing system that will allow mobile money interoperability. A decision model is used to test what form the proposed prototype system design would take. A prototype is then designed in which amounts being interchanged between providers are managed as assets on a permissioned blockchain. The system runs a distributed shared ledger which provides non-repudiation, data privacy and data origin authentication, by leveraging the consistency features of the blockchain. Development of the prototype is being undertaken as the third objective of the study.*

1.27 **Jonathan Phiri and Jackson Phiri**

Crime Mapping Model Based on Cloud and Spatial Data - A Case Study of Zambia Police Service

***Abstract:** Abstract—Crime mapping is a strategy used to detect and prevent crime in the police service. The technique involves the use of geographical maps to help crime analysts identify and profile crimes committed in different residential areas, as well as crafting best methods of responding. The development of geographic information system (GIS) technologies and spatial analysis applications coupled with cloud computing have significantly improved the ability of crime analysts to perform this crime mapping function. The aim of this research is to automate the processes involved in crime mapping using spatial data. A baseline study was conducted to identify the challenges in the current crime mapping system used by the Zambia Police Service. The results show that 85.2% of the stations conduct crime mapping using physical geographical maps and pins placed on the map while 14.8% indicated that they don't use any form of crime mapping technique. In addition, the study revealed that all stations that participated in the study collect and process the crime*

reports and statistics manually and keep the results in books and papers. To address the second objective, the results of the baseline study were used to develop the business processes and a crime mapping model, this was implemented successfully. The proposed model includes a spatial data visualization of crime data based on Google map. The proposed model is based on the Cloud Architecture, Android Mobile Application, Web Application, Google Map API and Java programming language. A prototype has also been developed and is in the testing phase. Test results of the proposed system shows improved data visualization and reporting of crime data with reduced dependency on manual transactions.

1.28 **Gabriel Chibuye, Jackson Phiri and Faustin Banda**

A Spatial Framework for Managing Sewer and Water Networks Using Sensor Networks: A case of the University of Zambia

Abstract: *The University of Zambia has a lot of utility networks that run on the surface and underground. There's uncertainty of the utility lines locations and their depths. This causes destruction of water pipes and sewer lines when undertaking new constructions or any excavations. The lack of accurate location of underground and un-updated lines makes maintenance, management and planning of the utilities very difficult. Therefore this study attempts to build a framework for Managing Sewer and Water Networks Using Sensor Networks at the University of Zambia. The Resident Engineer's office only has some old scanned PDF sketches of both sewer and water networks. The field technicians locate most lines and other facilities using memory and experience. A Web-based GIS or WebGIS application was used to develop the spatial framework. WEBGIS is a powerful mapping and analytical functionality expressed within a web browser. The web application had some tools embedded in it including; query, measure, add point/vertex, zoom, view, search and identify. Using internet of things (IOT) technology, water level sensors were placed on some critical manholes to help with the monitoring of the Sewerage network. Utility coding and tagging was done using QR code technology. This hugely helped in the identification of all the infrastructure which were presented as either nodes or vertices each with a unique identification number. The QR codes could be scanned using any QR code reader and could provide attribute data including the location for easy identification for any field personnel.*

1.29 **Banji Milumbe, Jackson Phiri and Mayumbo Nyirenda**

Enhancing security of examination question papers through a tracking system based on spatial and cloud technologies

Abstract: *This study first investigated the transportation of examination materials in relation to examination malpractices and leakages. Based on the findings, the study explored the use of GPS, GPRS/GSM and cloud technologies to enhance the distribution process of examination question papers and answer scripts. We proposed a GPS tracking system which would ensure real-time tracking and monitoring of the examination distribution process. The main advantage of this system is that the transportation vehicles will always be tracked and remain monitored such that if there is any holdup, GPS tracking allows immediate action to be taken. Also since GPS allows constant updates in real-time, any sort of uneven event happening, the main station will be informed immediately. This helps timely action to be taken and also eliminates human interventions or rather depending on humans to report any uneven event. The baseline study indicated that the challenges faced in the distribution process which was mainly lack of reliable transportation vehicles which accounted for 67% of those who did not have any designated vehicles.*

1.30 **Bruce Mwiya, Mukwemba Mangangu and Beenzu Siachinji**

Examining Antecedents of Social Media Adoption as a Business Platform among Student Entrepreneurs: Evidence from Zambia

***Abstract:** The advent of technology has created phenomena never before imagined. Communication now takes place at lightning speeds and the dynamics of life itself are changing because of advancements in technology. The business environment has not been spared from the ramifications of this change. From the perspective of Schumpeter, the disequilibria nature of these technological advances forms the heart of entrepreneurship. Despite the rapidly growing phenomena of social media and entrepreneurship, there is scanty literature on the subject especially in developing countries and none in Zambia. This study, therefore, sought to assess the antecedents of adoption of social media as a business platform among student entrepreneurs in an under-researched Sub-Saharan country, Zambia. Data was collected from 110 student entrepreneurs from a Public University in Zambia. The antecedents of social media adoption as adapted from the UTAUT (Universal Theory on Adoption and Utilisation of Technology) model are examined by correlation and regression analyses. The study revealed that facilitating conditions, perceived risk, perceived trust, perceived enjoyment and perceived expectations are indeed significant to the adoption of Social media platforms among student entrepreneurs. The findings indicate that there is a multiple correlation coefficient with a large effect size (0.719) and coefficient of determination of 48.8%. By focussing on one public University, generalisability of the findings is limited. Future studies should endeavour to use a wider sample. That notwithstanding, for policymakers, the study shows that there is a need to come up with stronger ICT policies and programs that facilitate and make it cheaper for the adoption of social media for business purposes. For entrepreneurs, the evidence shows factors to take into account when considering the adoption of social media as a business platform. In terms of originality, this paper extends the scanty literature on the underexplored phenomenon of social media usage and student entrepreneurship to the under-researched African context.*

1.31 **Knox Kamusweke, Mayumbo Nyirenda and Monde Kabemba**

A Data Mining Model for Predicting and Forecasting Fraud in Banks

***Abstract:** Banks generate and handle millions of transactions across their platforms. These transactions contain significant patterns and trends which are hidden but needed for knowledge discovery and actionable insight. Uncovering these patterns and trends has always been a challenge for most financial institutions due to the large volumes of transactions and the ever changing patterns. This has made the patterns and trends more sophisticated to discover and has left many institutions without better knowledge and insight into the types of transactions taking place. This study aims to use data mining to construct a model that discovers hidden patterns and is able to predict and forecast fraud on the basis of those discovered patterns. The study was guided by two objectives; firstly, to find hidden patterns and trends in data, and secondly to construct a model that predicts and forecasts fraud on the basis of discovered patterns. Using the proposed model patterns of unusual transactions done with an upwards trend were discovered. Furthermore, the prediction score and forecasting indicated that these patterns of transactions were likely to continue presenting high chances of fraud happening hence the importance of more research into data mining for fraud detection.*

1.32 **Jimmy Katambo, Mayumbo Nyirenda and David Zulu**

Distributed Spatial Search Using Paillier Cryptosystem and a Distributed Ring Algorithm

***Abstract:** The problem of lack of anonymity and confidentiality can be experienced by both those who collect statistical data online and those who provide data. One end may be secure, for example, the one providing data and yet the other end, for example, the one collecting data, may not be secure. In another scenario both the data provider and collector may seek anonymity. Preventing decryption of data provided while providing aggregated results is the best solution for such scenarios. To achieve this, this paper proposes a protocol that puts into application Homomorphic Encryption and a Distributed Ring algorithm, to ensure anonymity of data of both parties involved in a spatial search, i.e. a data provider and a searcher. Firstly, we identify a*

Homomorphic Encryption technique that can work best for a spatial search by reviewing literature on Homomorphic Encryption techniques. Among the Homomorphic Encryption techniques reviewed were Rivest, Shamir and Adleman (RSA), El Gamal cryptosystem, Goldwasser-Micali cryptosystem, Benaloh cryptosystem, Paillier cryptosystem and Fully Homomorphic Encryption (FHE). After a comprehensive study, Paillier Homomorphic Encryption technique was identified as the best approach to be employed in securing a spatial search. Secondly, we propose a protocol for distributed spatial searching using Paillier cryptosystem and distributed ring algorithm principles. Finally, a proof of concept prototype using the proposed approach was implemented. From initial experiments conducted using the proposed approach it is evident that the bigger cost comes from the communication over the network and less from the encryption algorithm itself. As compared to the usefulness of the approach the time to execute the algorithm is negligible.

1.33 **Zaima Banda and Mayumbo Nyirenda**

Understanding the Transactional Challenges Encountered in Village Banking While Exploring a Blockchain Enabled Solution: A Case on Zambia

Abstract: — *Village Banking concept seems to be growing at a faster rate in the Zambian environment as most people are now joining simple groupings as a way of providing financial support to one another. They achieve this by making contributions on an agreed periodic basis towards the group's treasury. Individuals have a plethora of payment options to choose from, namely bank transfers or mobile wallet payments or cash. In today's environment we see mobile devices as a fundamental part of our daily lives especially when they are coupled with good internet service which allows Individuals to connect to vital applications and interact with social media platforms. Individuals usually fail to detach themselves from these mobile devices. It is this pervasive nature of these devices that we build our strategy on. We set up a payment ecosystem that will allow members to move funds into and out of the village banking platform with ease and in a manner where everyone in the group will be able to see and monitor the activities in the group clearly. We achieve this by designing a mobile application on a distributed blockchain ledger. The ledger provides the advantage of transparency and redundancy in the group. Our solution basically provides an alternative payment method that makes it easy for individuals to exchange and move funds across different platforms irrespective of the bank or mobile money organization they are affiliated to. Because of these different platforms people usually face challenges when they have to transfer funds to others in a village banking setup, this therefore necessitated for us to conduct a survey of the different payment modes people use and then solution an application that will address their issues. Our solution provided an environment where all members are able to track and monitor how funds are moving in the group.*

1.34 **Tobias Makai and Mayumbo Nyirenda**

Categorisation of Sexual Reproductive Health Short Messages Texts into Thematic Areas Using Text Mining

Abstract: *In a quest to enhance behavioral change among adolescents in Zambia, the National HIV/AIDS/STI/TB Council in collaboration with UNICEF developed the Zambia U-Report platform to grant young people increased access to information on various Sexual Reproductive Health topics via Short Messaging Service (SMS) messages. Over the years, the platform has accumulated millions of text messages which need to be categorized into key thematic areas for better tracking of sexual reproductive health knowledge gaps among young people. Categorization of messages is currently done manually – a process that has proven to be both tedious and time-consuming. In this paper, we present a proof of concept for the automation of the categorization process of text messages that over 198,150 U-Report subscribers have been sending to the platform, into key thematic areas for better analysis using existing data and text mining techniques. The paper discusses the model currently being used to categorize information on the platform and the proposed prototype that automates the categorization of the unstructured text messages into key thematic areas. Through an exploratory research method using semi-structured interviews and observation, we established how the categorization of messages is currently done and obtained a list of categories that U-*

Report counsellors have over the years adopted. We collected a copy of the U-Report database and used the list of categories as a primary guide to build and train a text categorization model using supervised machine learning algorithms for data and text mining. Preliminary results show that an accuracy of 75% can be reached using a Support Vector Machine. Though not as accurate as the counsellors the faster categorization gives huge benefits.

1.35 **Makani Mweembe, Monica Kabemba and Jackson Phiri**

The use of Coding Environments to Teach Computer Programming to Children in Upper Primary

Abstract— *In the recent past there have been several debates as to whether computer programming (computer programming) must be taught to children from as young as five years old. Some thinkers have argued that programming is too complex to be taught to children, on the other hand other thinkers have argued that just like reading and writing are important for children so is computer programming-computer programming is the new literacy. There is some evidence that teaching code to children offers many opportunities for learning and personal growth, exploration and creativity, mastery of new skills, and ways of thinking. Computer programming, like writing, is a medium of human expression. Through this expressive process, children learn to think, feel and communicate in new ways. Problem solving isn't the primary goal of teaching young children to code, the goal is to teach them to think systematically. This research focuses on assessing the use of block-based/environments environments to teach children computer programming and explores the addition of programming topics to the Primary school ICT syllabus in Zambia. The study uses observation and survey research with a closed ended researcher administered questionnaire. 60 pupils are being observed currently and 2 curriculum development specialists. The coding environments to be used are Scratch and Alice and 60 respondents will be exposed to these environments. The review of the ICT curriculum will be reviewed to understand the practical activities in the ICT syllabus that can be supported by the environments.*

1.36 **Joe Phiri and Mayumbo Nyirenda**

Medical Equipment and Laboratory Services Support System

Abstract: *Medical equipment are key to the successful implementation of the ART program and for uninterrupted services there is need for equipment to be monitored and managed through a structured program. The Ministry of Health and partner organizations has procured over the past 12 years approximately 700 chemistry, 900 hematology, and 1000 CD4 and over 2000 other medical equipment to support services at the different levels of health Care. This paper focuses on the development and implementation of a Medical Equipment Decision Support Information System for tracking the availability of consistently functional equipment throughout the 6 USAID supported provinces of Zambia. USAID through the PEPFAR program has also provided equipment use and training support for 500 facilities in the six provinces to optimize outcomes of equipment placements and ensure quality management. Equipment monitoring, and tracking is critical to determining the overall consistent functionality and care of analyzers and has a significant bearing on the quality of services.*

1.37 **Andrew Biemba and Mayumbo Nyirenda**

Augmented Reality Aided Navigation System Model for Support of Student Registration Process

Abstract: *Universities worldwide are expanding structurally year by year as the number of students gets bigger and bigger with every enrolment period. Due to the enormous size, there is a high chance that some students would not be able to find their way around campus with parents and newly enrolled students being the ones greatly affected. One of the key processes that universities undertake with every enrolment period is*

the student registration process. Student Registration is a systematic process that involves but is not limited to fees settlement with the bank, course selection and submission for approval by responsible department. In this study, the student registration process was modelled and integrated with a proof of concept Navigation system that makes use of the new technology, Augmented Reality. This project takes advantage of the exceptional data processing abilities that are possessed by smartphones in conjunction with their various services such as Global Positioning System (GPS), wireless networks, and compass to model a Navigation System that makes extensive use of Augmented Reality to support the student registration process. Augmented Reality (AR) is a type of interactive, reality-based display environment that takes the capabilities of computer-generated display, sound, text and effects to enhance the user's real-world experience. The Global Positioning System (GPS) is a space-based satellite navigation system that provides location and time information in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites. A navigation system is a system that aids in navigation. A compass is an instrument used for navigation and orientation that shows direction relative to the geographic cardinal directions (or points).

ICICT2019 PROCEEDINGS ARTICLES

THIS PAGE IS INTENTIONALLY LEFT BLANK

Defect of Mobile Devices Usage to Village Community Banks in Supporting Members' Participation: The Case of Tanzania

William Clifford Gomera
College of Business Education (CBE)
Dar Es Salaam, Tanzania
c.gomera@cbe.ac.tz

George S. Oreku
The Open University of Tanzania (OUT)
Dae Es Salaam, Tanzania
george.oreku@gmail.com

Abstract—Village Community Banks (VICOBA) and mobile phones have become assistance to many rural societies in encountering financial problems. However, mobile phone usages to VICOBA have isolated many resulting to unsubstantial benefits that do not foster member's participation. The weak participation of members to Village Community Banks activities has been one of the major impediments to their success despite of owning of mobile devices to some members. This paper is exploring the situation through identifying the deficiency available in mobile devices to VICOBA microcredit programmes in enhancing member's participation and suggest the effective usage pattern

The study captured 7 groups of village community banks at Ndungu village with total of 70 members whereas 48 female and 22 male participants. The study employed interview, observation and focus group discussion for data collection. Content analysis and coding approaches used to analyse data and interpreting findings.

The findings of the study showed that there is big potential in mobile phone use to village community banks activities. It could potentially enhance members' participation in microcredit programmes even though the current usage does not foster members.

Keywords— Village Community Banks, Microcredit Services, Mobile Devices

I. INTRODUCTION

Microcredit is a vital tool for poverty alleviation all over the world especially in developing countries. Microcredit Programs (MCPs) are considered as one of the effective tools to alleviate poverty in developing countries [1]. These programs introduced to targeted mostly resource limited societies that can conduct business and other productive activities [2]. Moreover, improving access of formal and informal credits by low income earners is a key point to boost development as it enables large base of population to involve in different activities like agriculture, small business and small scale manufacturing which improves living standards [3].

As for other Developing countries, Tanzanian Government is tremendously conscious about poverty alleviation, where Village Community Banks (VICOBA) found to be one of the partners on the move [4]. Since, VICOBA have tried to fill the financial gap to low income earners, active participation of members to these institutions is an important aspect to be addressed by practitioners and research world. Large part of low income earners communities in developing countries depends on informal credits where most of them are group based microcredits

institutions like VICOBA. Informal microcredit providers remain strong financial sources to poor whom do not have access to either formal or semi-formal (microfinance) sectors [5].

VICOBA are member – based microcredits institutions which owned and run by members for members [6, 7]. These groups are voluntary formed to target members after awareness campaign. [7]. Initiators and other NGOs are providing technical skills to group members through training in regular meetings. Training offered based on credit and savings, reporting and operating guidelines, leadership, conflict resolution, rules and regulations, business and basic financial literacy [6]. VICOBA operate by asking member to save money before taking loan at low interest rate. The loan taken starting with low amount and increase as borrower properly paying the previous loan.

Also members needed to attend regular meeting scheduled for specific interval of period mostly weekly basis. This affect the attendance and members' participation due to difficulties in attending the meeting [6, 7], hence affect members participation in VICOBA [8]. However, there is weak participation of members to VICOBA microcredit services due to time constraints and geographical distance factors. There are reasons which impede members' participation on VICOBA microcredit programmes especially of their regular meetings [9]. Some of the reasons are; low level of education to some of members, limited access to information, less mobility, busy schedules to members [7, 8]. Members' participation can be enhanced if and only the challenges affects their participation will be addressed.

On the other hand mobile technologies have been realized by researchers as one of the potential tool to address the identified challenges [9, 10, 11]. Mobile technology is a powerful tool to address social challenges and change a fate of poor people. Moreover, mobile device has being identified as a development tool that can easily address the geographical and time constraints [12].

Due to the penetration of mobile phone and its usage to almost people of all classes, the communication has increased its pace despite of the distance. From farmers to business personnel, mobile phone provides users with information when needed. Moreover mobile phone offers a greater scope in reaching people in town and rural areas. It has improved accessibility to information and enhances mobile services which in turn are instrument on bringing a chance in daily life [3].

There are number of literature focus on mobile phone usage pattern even to rural areas [4], ways members of VICOBA can benefit from mobile phone usage [13, 14, 15, 16, 17, 18] and the benefit that can be derived to enhance full participation on VICOBA activities [19, 20, 21, 22]. However, there is limited information that addresses poor participation despite of the usage of mobile phone in VICOBA activities.

Therefore, a care should be taken to conclude that usage of mobile phone is sufficient and can enhance activities and services of any group of people. The uses of mobile phones to practical activities at the grass root level need to be explored to find out if are sufficiently by itself. This calls our interest as researchers to find out deficiency existing in mobile device usage to enhance members' participation to VICOBA microcredits services. Thus we may also find out the effective use of mobile devices to enhance that member's participation.

It is the focus of this study to explore deficiency available in using mobile devices to VICOBA microcredit programmes in enhancing members' participation and suggest the efficiency usage to these microcredits institutions. Therefore, objective of our study is to investigate deficiency on usage pattern of mobile devices to VICOBA microcredit services in enhance members' participation in rural Tanzania context.

This study attributes to the recent broad circulation and usage of mobile technology as the technology open up door of opportunities to informal microcredit groups. The establishment of the current usage pattern and identifying challenges thereon will pave the way for future studies and practitioners to find out solutions the practical solution.

II. LITERATURE REVIEW

A. Microcredits to resources limited societies

Microcredit services in developing countries are still young, small and limited especially in resource limited societies [23]. Most of low income earners are not aware of the service structure offered by Commercial Banks and how should use the facilities to establish their business and conduct different projects [24]. In addition, services in commercial banks are still limited to those customers who can afford collateral based loans.

Therefore, group based microcredit services are playing a very important role to cover the financial services gap that left by formal financial institutions. Members of microcredits institutions at their grass root level identify and utilize opportunities and resources in the sustainable manner that could not be achieved if forced could come from above [6, 20]. Moreover, low income earners perceive informal and semi – formal microcredits institutions, the most accessible source of capital while loan from formal banks are ranked as the least accessible financing alternative [25]. In this social condition some Microcredit institutions especially VICOBA in Tanzania are of great support of savings, loans, investing and social assistances for low income earners.

A Number of researchers have raised their motive to study different aspects of community microcredit institutions including VICOBA. There are studies have cantered been on

limitation that hinder these institutions as pointed out by number of researchers as summarized in Table 1 below.

TABLE I. SUMMARY OF LITERATURE ON LIMITATIONS OF MICRO CREDITS INSTITUTIONS

I	More conscious about making profit [26].
II	Lack of financial resources to provide sufficient loan [27].
III	High interest rate [28, 29, 30].
IV	The administration section MCIs is not fit for purpose to communicate with clients after providing loans because of profit [31].
V	Rarely riches to extreme poor [3, 27].
VI	Rarely organize training and development programmes [32].
VII	Some institutions take long time to approve the loans [28].
VIII	Microcredit has limitation women accessing in their scheme [10, 24].

B. Microcredit Programme of the Village Community Banks

By 2011 was reported that 723 Micro-finance Institutions currently have 190,135,080 clients all over the world and 81.7% of them are extreme poor and 104,694,115 are women clients [3]. Tanzanian based research group reported that Micro-finance institutions playing an important role for poverty alleviation in Tanzania [4]. Some researchers also have pointed out some challenges facing these institutions on servicing their customers [6, 8, 9]. However, other researchers have pointed out different approaches on dealing with challenges affecting microfinance institution based on their types. As indicated that some microfinance institutions are business based and some are members oriented (community) based kind of financial institution, some are formal and others and informal [16, 33, 34, 35].

In Tanzania, VICOBA and other informal financial institution have reached 27.9% of resources limited societies who have no access to formal bank services (Verhan, Magesa, & Akidda, 2014). These institutes has almost 19 years of operation since their initiation in year 2000 by CARE Tanzania in Zanzibar and later on Orgut, SEDIT, World Vision and Women in Action (WiA) [34].

C. Participation of Members in VICOBA programme

It is important to know and move towards setting the position of low income earners in the society [3]. Low income earners play a key role in a private sector and micro, small and medium enterprises (MSMEs) in developing countries including Tanzania [36]. [19], suggests that low income earners directing up to 90% of their income to their family and communities. Moreover [22, 37], indicated their income are critical for food security as they cultivate up to 80% of food in many low and middle income societies. Also they play major roles in farm production and micro businesses; however, they remain constrained by financial challenges. Despite of forming large population of income generation, low income earners lack access to resources due to poor participation to microcredit activities, hence affect their rights to decision making [3, 38], this affect their participation on different financial activities [9].

Therefore, VICOBA's programme is promising to have future prospective, if challenges hindering members'

participation in VICOBA activities will be addressed [4]. There is a need of essential strategy to improve status and participation in VICOBA programmes [39]. Some researchers in Tanzania have tried to answer positively how VICOBA potentially contribute to poorer in social, economic and financial inclusion and shown the extent of acceptance of VICOBA activities in the society [20, 21]. Based on [15], findings that economic empowerment can be achieved through resource ownership and full participation in different projects, full enjoyment of different services, receiving relevant and continuity training and receiving reliable and affordable information sharing. However there is a limit on the way members' participation of VICOBA activities can be enhanced.

D. Mobile Phone to Rural Society

There is no doubt on how mobile phones have proven to be a tool for economic growth; hence, investing this technology to low income earners can improve the overall well-being of families and communities [11, 40]. Introducing activities of resource limited societies with the usage of mobile phones can accelerate social – economic development [40, 16] found that phone are the information related technology that has done the most to reduce costs, increase income and reduce uncertainty and risk through supporting the current reality of informal information, extend social and business networks and clearly substitute for journey. Mobile phones connect people and provide them with information that is useful in daily lives [42]. Their usage goes hand in hand with faster economic growth, job creation and profitable business.

There is strong link between increase in mobile phone and faster economic and business development [17, 43]. Moreover, Tanzania is not different from many other African countries in terms of mobile usage. During exploration of this research it was realized that there is an increase of mobile phone subscribers in each quarter of the Years 2019 as per Communication Regulatory Authority report. Subscription has increased by 168,726 during the first quarter of the year with the trend of total subscription of 43827857 in January, 43,946,080 in February and 43,996,583 in March[56]. This kind of increase in subscription fosters the economic growth of the country and contributes to the national income eventually.

Despite the benefit of mobile phone connectivity to most of developing word like Africa, Asia and Latin America. Low income earner in these region are not benefiting as much as expectation. This is because there number of daily activities not tailor-made accommodated by mobile technology to support the real need of the low income earners.

However, as suggested in [44] this is an agenda to focus on so as to increase benefit of mobile phone usage among low income earners. First is to position the phone as a life enhancing and income generating tool. The aim here is to create innovative programmes to increase the uptake of mobile phone amongst low income earners. Second is to set a mobile phone to be an operating tool at international and local assistance, with the focus to promote the mobile phone as an effective development tool which creates international and local opportunities. All these will help to identify

cultural relevant and acceptable ways of promoting mobile phone ownership [44].

III. METHODOLOGY

The study applied qualitative research design, to explore perception of members towards VICOBA microcredit services, deficiencies of mobile devices usage in VICOBE microcredits services that affect members' participation and effective ways can mobile devices used to enhance members' participation in VICOBA microcredits services. We used qualitative research approach due to our focusing on exploring the situation and requirement of detailed information that assist in establish a base on the perception and usage and deficiency thereon [45]. The explorative and interpretative approaches used due to the fact that the researcher wanted to make sense of words from human action [46].

Qualitative study in mobile technology and information system found to be very useful because it started on identify human behavior toward technology [47]. The interactive approach of practitioners and researchers has also been used in conceptualizing the practical solution of the challenges hinder members' participation on VICOBA activities. The explorative study made a base for design Science Research (DSR) aim on finding practical solution of the existing practical problem of the mobile devices usage fail to enhance members' participation on VICOBA microcredit services [12]. We collected qualitative data i.e. situation explained through words although, few numerical collected to show number of people events and percentage. These numerical also were presented in the study report. However, our explanation still base on qualitative kind of the study.

A. Area of Study

The case of Ndungu village in Same District of Kilimanjaro region, northern part of Tanzania was selected as area of study. The reasons for selecting this area was; familiarity of researchers with residents of the village this provide chance to get detailed information from participants. Second, its nature and characters of people in this area proposed to be good representative of other villages in developing countries concerning VICOBA programmes. Third, constraints faced by VICOBA members in this village resemble other developing countries in terms of formal financial institution support.

B. Participants

Participants of the study were 70 members selected from seven VICOBA in Ndungu village. These groups are; Yongoma Group, Mwamko Group, Amkeni Group, Jitegemee Group, Sasa Kazi Group, Juhudi Group and Changamkeni group. The reason for limiting number of groups was to limit the size of research being a detailed analysis. With 7 selected VICOBA groups seemed to be useful due to the complexity of the study that need in-depth study of three variables i.e. VICOBA activities, members participation and current pattern of using mobile phone in VICOBA activities and challenges thereon [12].

C. Data collection methods

Data collection methods included interview, observation and Focus Group Discussion (FGD).

a) *Interview: We used interview as it is usefully in the interpretive studies [47] that focus on the way members using mobile phone to receive microcredit services from VICOBA, deficiencies thereon and opinion on the way forward. Moreover, through interview individual perception and opinion on VICOBA operations, benefits and way forward were explained. We interviewed 30 members out of 70 of the entire study. Selection of 30 members out of 70 was done randomly by researcher. Specific and planned appointment made with interviewee was made to make sure that interviewee was comfortable with the schedule. Interview session lasted for 15 to 25 minutes, depending on the respondents' openness and awareness of VICOBA and technological opportunity available. Researchers structured the interview question in such a way that interviewee could give explain how the mobile devices used in VICOBA activities to enhance participation. This was done purposely based on the study conducted by (Joseph, 2011). During interview researchers took notes, however for more understanding the interview process conducted in Kiswahili.*

b) *Focus Group Discussion: The FGD was used as extension of interview that allowed participants to have wide explanation on the topic. Here participants shared experience on mobile phone usage, deficiencies of mobile phone usage that affect members' participation in VICOBA. Selection of study groups based on willingness of leaders and members participation to the study. There were 7 groups of 10 participants each, with average of 8 female and 2 male. Selections of participants were member nomination that made during weekly meeting. Nomination of participants made by members themselves based of representation of occupation criteria i.e. employees, business personnel and farmers. Researcher took notes for every member's contribution; moreover, Kiswahili language was used during discussion. This helped members to talk free and fluently since Kiswahili is a Tanzanian is a National language. The FGD used by researcher as team of researcher and practitioners to conceptualize on the practical problem and think on the potential practical solution. This formed a base of DSR that involve participatory approach.*

c) *Observation: We applied observation during weekly meetings in order to see the proportion of members and attendance on meeting. This method used due to the fact that researchers had time to attend those meetings as observers. Also researchers took notes on the way members behaved in meetings, their contribution, and ways they respond to or asked questions. Here researchers were complete observers of what was taking place. Researcher recorded data through handwriting of things taken place during meeting without audiotaping and by videotaping. The observation of each VICOBA group was also important to see if there was a different in approach from one group to another.*

D. Data Analysis

We used thematic analysis whereby respondents' ideas were grouped into different themes. This done initially by coding the questions into different groups and posing each explanation against each established code. All data obtained during interview, FGD and observation were recorded against pre-established themes that focused on usage pattern, deficiencies and opinion of the way forward in using mobile phone in VICOBA to enhance members' participation. After forming general categories, data coded and grouped into small related units, which were then categories into related themes for each group. The common concepts appeared in each discussion were noted to make group of idea, hence come up with items of the results and discussion of the study

IV. FINDINGS

The study involved three variables; members' participation, VICOBA microcredits services and usage of mobile devices. It was revealing that, the insight of members' participation on VICOBA microcredit services through mobile devices could make a significant change to people's life.

A. Description of Participants

Participants were 48 female and 22 male that involved in productive activities like; farming, micro-business and teachers of primary and secondary school. Some of micro-businesses owners dealt with food stuff, small shops and sell of other basic need items. About 60% of participants were of standard seven level, 22% form four, 14% certificate and diploma level, 4% degree level.

B. Services offered in VICOBA Microcredits Programme

We found that most of VICOBA members opted for the microcredit programme that fit them; hence there is no constant programme for all VICOBA. However, in general the following microcredits programmes have almost available to all groups except minor deviation. We found that most of VICOBA microcredit programmes are conducted in their weekly meetings. Services offered to members include; savings, loan, training and business networking.

Researchers found that, the main programme of VICOBA is to assist member to increase savings, provide members with loan, provision of training and members networking for marketing and supplies.

There different kind of savings found in this study. These savings are received on weekly basis that collected from members in form of; contributions (wekeza), and social funds (huduma kwa jamii). Also entry contribution know as initial share (hisa anzia) which paid by member to join VICOBA. We also, found that funds contributed used in provision of loan at low interest rate, help VICOBA members who are in need and used in operation activities. Loan given without collateral, since, contribution, and community knowing each other is enough to provide loan to a person in need.

For the training, we observed that VICOBA invite different professionals to equip members with basic skills on loan, business management, food process and preservation and production of local products like soap. This help to equip VICOBA members with knowledge of different aspects on finance, business and productive activities as most of

characteristics of VICOBA members are of low education and people that doing micro business and small scale farming for the family welfare.

Business networking was observed to be very another important service offered by VICOBA as the members are allowed to come with different services and products during weekly meeting. Also it provide forum for suppliers and consumers to meet and exchange product. Moreover give chance for business opportunities after meeting. Member argued that VICOBA microcredit programme is less cost, managed in a voluntary bases with minor cost been taken care by members themselves. It can be assessed by all poor at the grass root, since share value of VICOBA is planned by group members themselves. Also the bottom up approach decision making, planning, implementation and ownership in the model of operation used.

We found that some of VICOBA have no any other services than financial service. Some respondents expressed their feeling in focused group that services provided by these institution should not be limited to financial services only, instead should include services like training, market information and business networking.

C. Importance of VICOBA to members

From data collected the study found that microcredits play a significant role to member. We found that 47% of members that participated in the study use funds from VIKOBA for business and farming, 22% of them uses school fees for their children, 20% uses to improve living standard and 11% uses for different purposes. Although, all respondents agreed to benefit from VICOBA, there are different uses of money received from VIKOBA. The statement can be backed up with the evidence from table 2 below which shows how they have being benefited in different ways.

Moreover loan provided to members after soliciting funds from savings are collateral free. The loan was given at very low interest of 3% to 10% depending on the decision of group members. The interest income received from loan advances benefiting all members of the group depend on the amount of shares contributed by a particular member. Participants of the study envisaged that, VICOBA microcredit programmes is one of the better tool for community emancipation socially and economically in Ndungu village.

This is evidently proven by members who claimed to benefit from resource mobilization process at the group level and benefited with the training programme and business networking at individual level during group meetings. Our study confirmed that VICOBA have large impact to their members. The information received from the study indicated that VICOBA members contribute little amount of money each week (most of them is the balance after house hold consumption) then after a certain period of time receive loan from VICOBA savings without collateral.

VICOBA microcredit programme have positively contributing to member development and economic growth by improved quality of life and social wellbeing. Also, this scheme has enable member to build up their awareness, knowledge, skills and capital generated from within themselves. Furthermore, members participated in this study

admitted that their financial situation improved as well as had better living condition. Some members admitted that *“through VIKOBA have being able to start their own small local shop to sell vegetables, fruits, and cooking needs. Hence assist to support family needs”*.

TABLE II. BENEFITS OF VICOBA TO MEMBERS

Uses of loan advances	Frequency	Percentage
Business	33	47
School Fees	15	22
Buying house items including furniture	14	20
Other Uses	8	11

D. Perception of VICOBA Services to members

We were interested to find out the extent participants perceive the programme/ services offered by VICOBA. During group discussion participants expresses their opinions basing on their perceptions towards VICOBA activities. Most of group participants admitted that VICOBA programme are well known to them. This backed up with the argument that before joining the group a well training on how generally VICOBA operating is conducted. Moreover, some of participants have being members of VICOBA for years and know exactly how services offered.

We found that some of VICOBA do not provide services on time. This noted in some FGD that VICOBA microcredit services are not offered on time. The commonly mentioned services that do not often offer to VICOBA are training. This explained that it take long interval between training sessions, some are taking up to six months since the last training. For that case has being proposed to see how this service can often available to members as it is potential.

To some extent participants were satisfying with VICOBA microcredits programmes; however some mentioned issues that need attention. Some participants mentioned time foe meetings consuming time hence affect participation members. We confirmed poor participants of member on VICOBA activities due to; most members occupied with family and community services, low level of education, lack of leadership ability, experience and lack of confidence. Also some of members have very small number of shared that they are not qualifying for large loan which can enable them to start business.

E. Deficiencies of mobile devices usage in VICOBA

Most participants explained that their interaction with VICOBA through mobile devices is not of high quality due to the fact that usage of mobile devices has not substituted the need of physical. Physical contact has being identified to be cha contact that consuming time of most low income earner. Also they argued that the usage of mobile devices foe normal calls, short messages, mobile money services and online banking have not help the transparency of their savings and loan.

Moreover, it has identified by participants that mobile phone do not help member with sophisticated information of other services like. *“We use my mobile phone for some of the microcredits services from VICOBA but there some important services that are not offered through these devices”*.

For already existing services like mobile money, we observed that the service help only sending money to VICOBA account or a person responsible for collections. However, the services do not provide balance of the savings or loan balance in an individual account. Therefore they argued that this service has no different with manual contribution of money because it can help the member to access his account without make a physical visitation to VICOBA secretariat.

For the case of short messages participants identified challenge that the method of information sharing is not consistence and sophisticated. This happened because it is not automated that a certain particular message with be sent to members for a specific purpose rather most of these messages based on reminding members either to attend meetings or to make payment required.

Therefore, researchers decided to find out views from participants on the challenges face by members in using mobile devices in VICOBA microcredit programme.

The study found that in some points all members are using mobile devices in VICOBA activities. However, members identified number of challenges that are associated with usage of mobile devices in VICOBA microcredit services and hence affect members' participation. During FGD the participants identified these deficiencies as listed in Table 3 below.

TABLE I. DEFICIENCIES OF MOBILE DEVICES USAGE IN VICOBA MICROCREDIT SERVICES

S/No	Deficiency on mobile devices usage in VICOBA microcredits programme
1	Failure of mobile devices to substitute weekly meeting of VICOBA
2	Usage of mobile devices has not supported provision of relevant information
3	Mobile devices usage has not enhance the frequently of training provision to members
4	Mobile devices usage has rarely riches to extreme poor especially women
5	The mobile devices usage have not enhanced the training services to members
6	Poor communication between VICOBA leaders and members
7	Lack of transparency of members' contributions, interest earned and loan provided
8	Lack of timely report on the performance of specific VICOBA

F. Proposed usage of Mobile Devices to VICOBA Microcredit services

Participant positively responded to the interests of having effective ways that can reduce physical contact and substitute for frequent movements. During interview and focus group discussion participants suggested that VICOBA has to find ways that can minimize time spent on service deliverance and meetings. As they identify time and physical meeting affect the members' participants in VICOBA microcredit programme.

Furthermore they admitted that receiving services at any time anywhere will help them to participate in VICOBA activities at the same time dealing with their daily activities. The usage of mobile devices to receive services from

VICOBA has attracted attention during FGD because most of participants claim to have bust schedule in their agricultural of business activates therefore they have no idle time to make follow up of microcredit programme.

Discussion on VICOBA mobile services took an interesting turn as participants admitted that they had no idea on the possibility of VICOBA to offer services to their members through mobile phone since they are not expert on how these services can be undertake. However, they argued that mobile devices has to target on minimizing time for weekly meeting, increase transparency, reduce time lead for services and being able to compensate for human movement and cost minimization.

Participant highlighted some of VICOBA microcredit service that can be offered through mobile devices hence to enhance members' participation. The identified services were financial services (money transfer, loan repayment and loan disbursement), training services, weekly report, members' status and business network service. Basing on these this requirement the researchers call an idea of researcher developing the Mobile application for VICOBA.

Participant went far to the point of identifying the basic characteristics of the mobile service tool. They expressed that technological tool to be built to have the following features: - friendly user interface, offered in Kiswahili, offered in all multimedia i.e. video, audio and document. They proposed that the tool should consider both rural and rural environment, nature and challenges of rural society and size of the mobile devices owned by most of rural people. Participants showed willingness to use mobile phone to receive services from VICOBA. Some commented that *"If mobile phone will give chance to see balance, other information and training at any time anywhere it will be a value addition tool for members it may improve trust of members on VICOBA"*

V. DISCUSSIONS

This study was guided by set of research questions which were used to set the base of discussions. The following were research questions:

A. RQ1: What are the perceptions of members in VICOBA microcredits services?

From finding we confirmed that VICOBA microcredit programme uses direct management with the poor, and look to the individuals and their immediately needs, also helps community to generate economic growth through market driven business initiatives. On top of that, micro-credit decreases individualisation and create a synergy well-built for group of people to domination resources get liberated to run some of their needs independently. Also members receive financial and other resource ownership freedom for income generation. The study observed that access to micro loan at low interest rate without collateral improved self-sustainable, family support and increase resource ownership. It also confirmed that before joining VIKOBA most of the participants owned very few resources and some owned nothing, however through credits programme members can own assets and money deposited as shares in VICOBA.

The main beneficial of VICOBA microcredit program are low income earners. VICOBA, described as a model of

micro-credit programs being; target to help extreme poor and their families to overcome poverty, provides collateral free loan. Loan proposals are sanctioned based on 'trust', mainly given for income gathering activities and small business development provide door-step banking services to the poor. That is why people do not need to go to the bank, bank will come to them, borrowers must join a credit union to get a loan, repayment is to be paid in weekly instalment, new loan being available to beneficiaries as soon as they repaid the previous loan, voluntary savings scheme whereby a certain amount deposited to the credit union for every specific period.

Research findings confirmed the findings from several studies made in Tanzania which showed that VICOBA microcredit programme has an exclusive impact on economic development of low income earners [4, 6, 8, 20, 21, 33, 39]. Moreover, VICOBA programme support business initiatives and working close with poor people as pointed out by [48, 49], that, "Micro credits institutions represents a sea change in the way financial capital used to stimulate economic growth in developing countries".

We confirmed that members of VICOBA contribute their shares and social protection mostly on weekly basis, these contributions are of three types; initial shares ("hisa anzia" in Kiswahili), social fund ("huduma kwa jamii" in Kiswahili) and savings contribution ("wekeza" in Kiswahili). After three or four months of savings and contributions the study confirmed that VICOBA microcredits services of the groups participated in this study offer loans to support members' businesses and concur with services presented by [9]. For the first time members take short term loans of three months, later after gained competency in entrepreneurial skills they take long term loans of six months up to one year. Each loan is returned to the group basket with low rate of interest compare to commercial bank and formal financial institutions. These institutions enable members to mobilize savings, build capital, and receive training on how to allocate funds and expose them in business management skills. Moreover, they create chances of market network among members and building financial strength to the group and individual member of the group.

B. RQ2: What are the deficiencies of mobile devices usage in VICOBA microcredit services that affect members' participation?

Although, different studies pointed out challenges faced by most of VICOBA members in microcredit programme participation [49], our study wanted focused only on deficiencies on using mobile devices that affect members' participation. Our finding from participants confirmed that the services offered by VICOBA microcredit programme are limited and not automated, hence affect members' participation.

We confirmed that members of VICOBA at Ndungu village had access to a limited number of microcredits services offered through mobile devices. Services offered through mobile devices are such mobile payments, calls, short messages and online banking. Our findings confirmed that more sophisticated microcredit services, such as mobile training, business network are some of alarming deficiency associated with the current mobile devices usage. The

deficiency confirmed in our study indicated to affect members' participation in VICOBA microcredit programme.

Members, participation in VICOBA microcredit programme typically needed assistance of usage a sophisticated mobile devices application that can enhance the efficiency sharing of information among the players. We also confirmed that majority of VICOBA typically use mobile phone only for calls, short messages and money transfer. Additionally, there is high need of sophisticated platform that can assistance to increase number of services that can offered to reduce time and physical distance for VICOBA microcredit services.

Members seemed to be uncomfortable as using a mobile phone for microcredits services as there is no solution for a need of physical contact. Usage of mobile phone in VICOBA was more widespread in the villages members used mostly mobile money services, and almost exclusively for sending/receiving money. This is because relatively few rural Tanzanian had bank accounts reported accessing their money over-the counter in a bank instead of depositing or withdrawing funds by themselves at an ATM machine. However, lack of digital skills, unsophisticated number of services and lack of access to firsthand information were raised as reasons for discomfort using mobile technology. Consequently this resulted to heavy reliance on manual works for most of rural society.

C. RQ3: How mobile phone can effectively use to enhance members participation in VICOBA microcredit services?

From our findings we confirmed that there is a need to introduce a tool that can assist in service provision to facilitate VICOBA microcredit services. As we confirmed that most of services provided by VICOBA is limited to time and physical contact that affect members' participation. We confirmed that the current mobile usage in VICOBA are mainly on mobile calls, mobile money, short messages and some are online group savings. Normal calls used by VICOBA leaders to make follow up for meetings and request for weekly saving or loan repayment. Short messages used for information sharing, mobile money services for money transfer and online group savings is to make savings direct to group account maintained by commercial bank.

Despite of the high usage of mobile services, they proved not to motivate transparency and sophisticated information sharing. Therefore, calls for an attention of having a sophisticated system that can assist to provide online microcredit services programmes. Moreover, it confirmed that's usage of mobile devices to these programme will amount to increase number of services, data to be analysed, and sophisticated information and transparency.

Based on participants view it was revealed that technological application will encourage members to take position and participate in VICOBA programmes at large. Ubiquitousness of mobile technology and supervising tasks will save time and having automated operations will speed up the many activities which currently are done manually as observed by [48]. This will motivate members' participation in VICOBA microcredit programmes.

It was experimented that an effective use of mobile phone may bring a substitute to manual performance and information sharing. It is argued that, short messages that

contain report will assist leaders to be transparency and provide regular report through posting data in system and send information to members. Also automated texting will save time, and effort of movement to find information.

Also argued that receiving training through mobile phone may save time, space required and enhance training continuity. As pointed out by [9, 4, 50, 51, 52, 53] that among other roles, one of the big role of VICOBA is capacity building through training, so leveraging on that will bring a very fruitful outcome to low income earners and under privileged group in the society [50, 51, 52, 53, 54, 55]. Moreover, members participated in the study propounded that there is a need to introduce means that may assist women to receive training remotely at any time.

VI. CONCLUSION

Contemplations need to be kept in mind when considering usage of mobile devices in financial products and services for rural population in Tanzania. From the study it was realized that poor members' participation in VICOBA caused by lack of sophisticated information sharing. However, efficient usage of mobile devices usage to VICOBA among people in rural Tanzania may increase as long as there are specific applications that have been customized to fit their needs. Mobile phones are already a fixture in the sampled areas in Tanzania and can boost up members' participation in VICOBA. However, new products and services should be imbedded in the mobile devices that may enhance the members' participation.

Off-the-shelf products, including VICOBA, are unlikely to reach and appeal to all members of the society in rural areas. Applications need to take into account services required by particular society. Tanzanian rural need help to effectively use of a mobile devices that will enhance their participation to VICOBA microcredit programme. Currently very few VICOBA services are associated with usage of mobile device, therefore to assist effective usage of mobile devices, researchers and software engineers have to consider coming up with application that can offered different services to members.

Our study paved way for future studies that may aim at coming up with a tailor-made application on VICOBA microcredits services. However the consideration of any application should take into account understanding level of rural society, services required, cultural of rural society and environmental setting of rural society. It should have a user approach development. Most of participants claimed to have basic knowledge on mobile phone operation, moreover admitted on limited participation to VICOBA microcredit programme. To focusing findings, the study proposes researchers to consider proper design procedures that help to use potentials that available in mobile technology to enhance members' participation to community groups.

REFERENCES

[1] L. R. Reed, "State of the microcredit summit campaign," in Microcredit summit campaign, Washington DC, 2011.
 [2] M. Yunus, "Vision 2050: A Poverty Free World," The Journal of Social Business, vol. 1, no. 1, pp. 7 - 23, 2011.
 [3] J. J. Mmasa, "Determinants of Smallholder Women Farmer Access to Informal credit in Tanzania. A case of Singida and Chamwino Districts," vol. 3, no. 2, 2017.

[4] B. Verhan, R. Magesa and S. Akidda, "Mushrooming Village Community Banks in Tanzania: Is really making a difference," International Journal of Innovation and Scientific Research, vol. 6, no. 2, pp. 127 - 135, 2014.
 [5] B. Priva and P. Srivastava, "Scaling - up Microfinance for Indias Rural Poor," World Bank Policy Research Working Paper (WPS 3646), 2005.
 [6] M. Ahlen, "Rural Member - Based Microfinance Institutions - A field assessing the impact of SACCOs and VICOBA in Babati district, Tanzania," Sodertorn University, 2012.
 [7] G. N. Maleko, B. S. Liheta, D. Aikaruwa, A. Lukas and G. A. Sumari, "Women Participation in microfinance Institutions of Tanzania: The case of SAVings and credit cooperative societies (SACCOs)," Journal of Business Administration and Education, vol. 4, no. 2, pp. 139 - 175, 2013.
 [8] J. M. Muhoho and H. Mwinyimkuu, "To ascertain the contribution of the microcredit institutions in women poverty alleviation in Tanzania: The case of Kinondoni municipality, Dar es Salaam, Tanzania," International Journal of Innovation and Scientific Research, vol. 9, no. 2, pp. 410 - 419, 2014.
 [9] P. M. Kato and J. Kratzer, "Empowering women through microfinance: Evidence from Tanzania," ACRN Journal of Entrepreneurship Perspectives, vol. 2, no. 1, pp. 31-59, 2013.
 [10] W. C. Gomera, J. Suhonen, S. S. Oyelere and A. F. Kapinga, "Mobile Technology for Record Keeping by Women Entrepreneurs in Tanzania: User Requirement Assessment," in In International Conference on Social Implications of Computers in Developing Countries, Dar es Salaam, Tanzania, 2019.
 [11] F. A. Kapinga, C. S. Monterp and R. E. Mbise, "Mobile technology for women entrepreneurs in Iringa, Tanzania: User requirements and architectural design," in AFRICON, Cape Town, South Africa, 2017.
 [12] L. Dube' and G. Pare', "Rigor in Information Systems Positivist Case Research. Current Practices, Trends and Recommendations," Management Information System Quarterly Research Essay, vol. 27, no. 4, pp. 597 - 636, 2003.
 [13] M. B. Charyya, Mobile Phone: A new tool to empower rural women in India with Special reference to Bulpur - Santiniketan, International Association of Women in Radio and Television (IAWRT), 2015.
 [14] R. M. DeLancey, "Gendering Entrepreneurship and Technology: A mixed study of Retaolers in a Development Economy," vol. 2, no. 3, 2015.
 [15] M. Joseph, "The Potential of ICTs to Empower Rural Women," in IST-Africa 2011 Conference, 2011.
 [16] E. H. Diniz, M. Pozzebon, M. Jayo and E. Araujo, "The role of Information and Communication Technologies (ICT) in Improving Microcredit: The case of Correspondent Banking in Brazil," in PICMET, CAPE TOWN SOUTH AFRICA, 2008.
 [17] J. C. Aker and I. M. Mbiti, "Mobile phones and Economic Development in Africa," Journal of Economic Perspectives, vol. 24, no. 3, pp. 207 - 232, 2010.
 [18] M. Komunte, "Usage of Mobile Tecgnology in Women Entrepreneurs. A case study of Uganda," The African Journal of Information Systems, vol. 7, no. 3, 2015.
 [19] J. Kesanta and B. Andre, "Impact of Women Empowered through Community Savings Group on the Wellbeing of their Families: A Study from Mgbuwe Tanzania," Interdisciplinary of Best Practices in Global Development, vol. 1, no. 1, 2015.
 [20] H. Chipindula and J. E. Mwanga, "Assessing the role of Village Community Banks in social Empowerment of Women within households in rural communities of Mtwara District, Tanzania.," Research on Humanities and Social Sciences, vol. 5, no. 20, 2015.
 [21] T. Milonge, "Empowering People through Community Banks (VICOBA); A Case of Simanjiro VICOBA Group Poultry Project in Sombetini Ward Arusha - Tanzania," Dissertation. The Open University of Tanzania, 2014.
 [22] D. Fletschner and L. Kenney, "Rural Women's access to financial services credit, savings and insurance," Agricultural Development Economics Division: The Food and Agriculture Organization of the United Nation FAO, ESA working paper No. 11-07, 2011.
 [23] S. S. Temu and E. K. Ishengoma, "Financial Linkages and Perfomance of Rural microfinance Cooperatives: Tanzania Case,"

- Indian Journal of Economics and Business, vol. 9, no. 2, pp. 285 - 302, 2010.
- [24] S. Khandker, G. Koolwal and N. Sinha, "Benefits of improving young women labor market opportunities: Evidence from Group based credit program in rural Bangladesh," World Bank Development Report, Bangladesh, 2008.
- [25] M. Lindrert, D. Yazdanfar and H. Boter, "'Perceptions of financial sources among women entrepreneurs in Tanzania'," African Journal of Economic and Management Studies, vol. 6, no. 2, pp. 197 - 218, 2015.
- [26] J. Battilana and S. Dorado, "Building sustainable Hybrid Organizations: The case of Commercial Microfinance Organizations," Academy of Management Journal, vol. 53, no. 6, pp. 1419 - 1440, 2010.
- [27] S. R. Khandker, A. M. Khalily and H. A. Samad, "'Seasonal and extreme poverty in Bangladesh: Evaluating an ultra-poor microfinance project'," World Bank Development Report, 2008.
- [28] D. Hussain and M. Z. Yaqub, "'Micro - Entrepreneurs: motivations, success Factors and Challenges'," International research Journal of Finance and Economics, vol. 56, pp. 22-28, 2010.
- [29] M. Alam and I. R. Molla, "'The limitation of microcredit for promoting microenterprises in Bangladesh," Economi annals, vol. 57, no. 192, pp. 41-53, 2012.
- [30] U. Bhuiyan and A. Shah, "'A literature survey of value creation through relationship banking'," Africa Journal of Business Management, vol. 35, no. 5, pp. 13517 - 13522, 2011.
- [31] S. Gibb, "Microfinance's impact on education, poverty and empowerment. A case study from Bolivian," Altiplano Development Research working paper series, institute for advanced development studies, 2008.
- [32] N. Kabeer, "Between affiliation and autonomy: Navigating pathways of women's empowerment and gender justice in rural Bangladesh," Development and change, vol. 42, no. 2, pp. 499-528, 2011.
- [33] F. Girabi and A. E. G. Mwakaje, "Impact of Microfinance on smallholder farm productivity in Tanzania. The case of Iramba District," Asian Economic and Financial Review, vol. 3, no. 2, pp. 227 - 242, 2013.
- [34] P. J. Salia and S. J. Mbwambo, "Does microcredit make any difference on borrowers' business? Evidence from a survey of women owned microenterprises in Tanzania," International Journal of Social Science and entrepreneurship, vol. 1, no. 9, pp. 431 - 444, 2014.
- [35] X. Qin and B. O. Ndiege, "Role of financial Development in Economic Growth: Evidence from Savings and Credits Cooperative Societies in Tanzania," International Journal of Financial Research, vol. 4, no. 2, 2013.
- [36] N. Mori and D. Olomi, "The Effect of Broards on the Perfomance of Microfinance Institutions: Evidance from Tanzania and Kenya," REPOA - Policy Research for Development, 2012.
- [37] H. Mahabub, "Credit for alleviation of rural poverty," International Food Policy, 1988.
- [38] C. Brindley, "Barriers to women achiving their antrepreneurial potential," International journal of Entrepreneurial Behavior and Research, vol. 11, no. 2, pp. 144 - 161, 2005.
- [39] E. A. Bosompem, "Micro-Credit lending and Women's Empowerment: Experiance of Rural Kasoa Women," The Journal of Global Gender Studies, vol. 1, no. 1, 2013.
- [40] J. Rumanyika, M. Tedre, M. Apiola, S. S. Oyelere and N. R. Mramba, "Contemporary Challenges in Street Trader-Customer Interaction Through Mobile Devices in Dodoma, Tanzania," in In International Conference on Social Implications of Computers in Developing Countries,, Dar es Salaam, Tanzania, 2019.
- [41] J. Donner and M. Escobari, "A review of the research on mobile use by micro and small enterprises (MSEs)," in International Conference on Information and Communication Technology and Development (ICTD), 2009.
- [42] W. C. Gomera and A. Mikko, "Improve MFI-MB Interaction with Technology: an Explorative Study in Dar es Salaam, Tanzania," in IEEE, AFRICON 2015, Addis Ababa, 2015.
- [43] C. A. Chair, Mobile Phone for Development: How have women in the informal sector used their mobile phone to enhance themselves and their business? Cape Town: University of Cape Town, 2012.
- [44] C. Blair, "Women & Mobile: A global opportunity. A study on the mobile phone gender gap in low and middle income countries," Foundation for Women, 2015.
- [45] J. w. Creswell, Research Design - qualitative, quantitative and mixed methods approach 4th Edition, Croydon, CR04YY: CPI Group UK Ltd, 2014.
- [46] G. Walsham, "Doing Interpretive Research," European Journal of Information System, vol. 15, pp. 320 - 330, 2006.
- [47] J. N. Amaral, "About Computing Science Research Methodology," webdocs.cs.ualberta.ca/~c603/readings/research-methods.pdf.
- [48] W. C. Gomera, G. Oreku, M. Apiola and J. Suhonen, "Mobile training in micro business: Design Science Research for Frugal Innovation," Cape Town. South Africa, 2017.
- [49] S. Khavul, "'Microfinance: Crating Opportunities for the poor?'" The academy of management Perspective, vol. 24, no. 3, pp. 58 - 72, 2010.
- [50] L. P. Kyrgidou, "Developing women entrepreneurs' knowledge, skills and attitudes through e-mentoring support," Journal of Small Business and Enterprise Development, vol. 20, no. 3, pp. 548-566, 2013.
- [51] K. Naser and R. W. Mohammed, "Factors that affect women entrepreneurs: Evidence from emerging economy," International Journal of Organization Analysis, vol. 17, no. 3, pp. 225-247, 2009.
- [52] S. Shamsad, "Social - economic dynamics of microcredit programs in informal settlements: The case of Khulna City, Bangladesh," HKU Theses, Bangladesh, 2008.
- [53] D. Roodman and J. Marduch, "The impact of microcredit on the poor in Bangladesh - Revising evidence," Center of Global Development Working paper 174, June 2009.
- [54] S. Afrin, I. Nazrul and U. A. Shahid, "'A multivariate model of micro credit and rural women entrepreneurship development in Bangladesh," International Journal ofBusiness and Management, vol. 3, no. 8, 2008.
- [55] D. Kumar, A. Hossain and M. C. Gope, "Role of micro credit programme in empowering rural women in Bangladesh: A study on Grameen Bank Bangladesh limited," Asian Business Review, vol. 3, no. 6, 2013.
- [56] https://www.tcra.go.tz/images/headlines/TelCom_Statistics_March_2019.pdf

An Analysis of the Best Practices for the Implementation of Data Centers in Zambia

Kasonde Victor

*Department of Computer Science, Mulungushi University,
Kabwe, Zambia
vkasonde@cbu.ac.zm*

Christopher Chembe

*Department of Computer Science, Mulungushi University,
Kabwe, Zambia
cchembe@mu.ac.zm*

Abstract—The purpose of this study was to assess the best practices in the implementation of the data centres in Zambia. For this reason, a total of 120 respondents were randomly selected from various institutions to participate in the study. The results shows that the impact of new hardware technologies on the infrastructure of the data center facility in Zambia will result in speedy access to data. The results also shows that the location of the Data Centre can be a critical design considerations and best practices in the site selection and design of data center in Zambia. It was observed that availability of power and also security can be a critical design considerations and best practices in the site selection and design of data center in Zambia. In addition, the study showed that human capital is the most common cost elements associated with different levels of data center reliability and fault tolerance in Zambia among others. Based on these recognitions, the following recommendations were made: that deliberate measures should be put in place to acquire new hardware technologies for the data center facility in Zambia to ensure speedy access to data. It was further recommended that in trying to design new data centers, the selection of the location should be a critical consideration. Furthermore, it was revealed that the management of data center facility in Zambia should also be organized through capacity building.

Keywords-- Performance, public sector, data center, server. LANs, database, flexibility; security aspects, design consideration, data;

1. Introduction

According to Bell [14] the key to a successful data center facility is to have one that is sustainable in the long term and is to consider it as a receptacle for equipment and operations, as well as an integrated system, in which each component must be considered in the context of flexibility and scalability [1]. Despite these requirements, Data centers seldom meet the operational and capacity requirements of their initial designs. The principal goals in data center design are flexibility and scalability, which involve site location, building selection,

floor layout, electrical system design, mechanical design and modularity.

In the recent past, Zambia like many other countries in the region has embraced the use of Information and Communication Technology (ICT) in almost every sphere of its economy. The year 2011 saw the government enforcing a number of policies to promote ICT usage in government departments [17]. However, despite the attempt to implement a data center, most requirements pertaining to the creation of the data center have been almost unattainable making the process to take long.

In other countries, the advent of Data Centre Consolidation has brought efficiency both in the design and management of Information Communication Technology (ICT) infrastructure resources. It has also significantly reduced Data Centre Information Technology (IT) costs. In the United State of America for instance, the Government has witnessed an increasing demand for IT that has led to a dramatic rise in the number of its Data Centres (Zambia Business Review, Report, and 2016 Edition). The report added that operating such a large number of Data Centres with heterogeneous devices was inefficient and expensive. Nevertheless, the introduction of Data Centre consolidation in the Zambian Public Sector is inevitable. In addition, it represents one of the available solutions to most Data Centres challenges in the areas of design and management of ICT infrastructure resources.

It is against this background that this study attempts to analyse the best practice of the implementation of the data center in the country to consolidate the data from one central location.

According to Computer Economics Report (2016), Data Centre Consolidation is defined as a physical combination of two or more Data Centres into a single facility with the goal of reducing costs or improving performance. In addition, the report explained that consolidation may simply be a physical relocation of moving computer hardware and support personnel to a single location. Furthermore, the report indicated that it might be `a more aggressive effort with other

types of consolidation which include servers, applications, databases, storage, networks and operating systems. Based on the above definition and the potential benefits that consolidation would bring, this study is proposing the creation of Data Centres for the country to consolidate the data management across ministries. These Data Centres are expected to be exclusively customized for the Zambian Public Sector to enhance the design and management of ICT infrastructure resources to enhance improvement in performance as well as reducing operational cost [3].

In the current study, it is observed that, despite the fact that Zambia has proposed data centres for the purpose of consolidating data from across the ministries, the implementation process has not been a success (Zambia Business Review Report, 2016). The organizations that are tasked with the implementation of this new technology in the country include: ZICTA, SMART Zambia, CIDZ among others and all have raised issues regarding the implementation of data centres in the country. According to Zambia Business Review Report (2016), the advent of new technologies, such as blade servers, that require substantial incremental power and cooling capacity; the pressure to consolidate multiple data centres into fewer locations; the need for incremental space; changes in operational procedures; and potential changes in safety and security regulations converge to impose constant facilities changes on the modern data center are still yet to be addressed in the initial plan of the data center creation[2].

The overarching rule in data center facilities is to design for flexibility and scalability. This rule embraces several key principles in the site location, building selection, floor layout, electrical system design, mechanical design, and the concept of modularity that enables the data center facility to change and adapt as needed, with minimum renovation and change to basic building systems (Michael A. Bell, 2015), it is based on this background that the study was conducted to determine the best practices to the implementation of data centre in Zambia. The following research questions are explored in this work:

- What is the impact of new hardware technologies on the infrastructure of the data center facility in Zambia?
- What are critical design considerations and best practices in the site selection and design of data center in Zambia?
- What cost elements are associated with different levels of data center reliability and fault tolerance?
- How should the management of data center facility in Zambia be organized?

2. Literature Review

The study by Dai et al., (2014) focussed its attention on the equipment of the data centres. It established that a data centre consists of four main parts: power equipment such as power distribution units and batteries, cooling equipment (chillers and computer room air-conditioning (CRAC) units), IT equipment (servers, storage and network), and miscellaneous component loads (lighting and fire protection systems) [27]. Electronic component systems that arrange processing, storing and transmission of data is the main part of the data centre, according to [6] all of which and create a large amount of heat, which must be removed from the ICT components at a rate sufficient to avoid serious overheating problems and system failures. More than 30% of the heat removal costs of a typical data centre is used in IT equipment and cooling equipment.

Similarly, a study by Linda (2015) focused on security aspects of the data centres. Linda stated that securing the Data Centre against vulnerabilities and attacks had grown in importance in the past several years. However, he said many Data Centre modernisation initiatives had missed on the opportunity to improve security and reliability while pursuing cost-reducing IT objectives. Varcoe (2015) added that security ranks highly on any Data Centre Manager's list of priorities especially when considering the devastating impact that downtime or data theft can have on a business. In support of the above authors, there were also security challenges due to thousands of diverse ICT resources such as servers, switches, routers, networks, storage devices and applications hosted at various Government Data Centres in Namibia [1].

In a related study, McNamara [27] outdated approaches have left a gap between the management of data and the management of storage. Further, he stated that this has resulted in inefficient operations, with considerable duplication of effort and with frequent interruptions to the activities of highly interdependent administrative groups. In favour of the author, outdated methods in the operations of Data Centres have also left a gap which resulted in the inefficient management of ICT resources and regular disturbances to critical services in the Namibian Public Sector. Therefore, introducing Public Sector Data Center Consolidation (PSEDCC) would consider replacing outdated methods with new and efficient methods. As result, frequent interruptions would be decreased.

On the other hand, [8] described decision-making process within the public sector as one of the challenges that hinder most Governments from stopping Data Centre infrastructure sprawl. He argued that this is because the decision-making process can be complicated and lengthy. He further added that in many agencies it was not unusual to have up to fifteen (15) or twenty (20) people signing off on a decision. In support of the author, the situation is exactly the same in Namibia where the decision-making process on anything related to ICT management and operations projects is lengthy and complicated. In other words, it is difficult to identify participants in the Government decision-making process. As a result, time and game changing opportunities that could transform Data Centre management and operations would be frequently missed.

Despite having adequate resources at their disposal, major companies had incorporated Data Centre consolidation as their strategic weapon for saving costs and improving service delivery. The above remark is supported by Computer Economics Report of 2015 which recognized that major organisations were already reaping the benefits of consolidation. The Report further provided a list of some corporations and how they had consolidated their Data Centres and services.

Lui as cited by Matthews (2017) proposed that typically, a Data Centre is served by dedicated mechanical, electrical and fire protection infrastructure that was independent from the systems that served other portions of the building. He further stated that similar to other building types, Data Centre infrastructures were designed to meet local, state and federal building codes, as well as design standards and guidelines set forth by industry organizations. In addition, he said Data Centre design was often required to meet certain levels of redundancy, reliability, maintainability, fault tolerance, scalability and flexibility. Similarly, (PSEDCC) would meet the above characteristics by establishing a consortium of O/M/As devoted to maximising uptime for the consolidated Data Centre.

Another study by Rose Wellington (2019) stated that running a data center is a complex undertaking. In addition to maintaining the strict physical security measures and logical security protocols needed to secure customer data, facility personnel face an ongoing challenge of optimizing IT infrastructure to improve power efficiencies and maximize cooling capacity. Quality data center operations are a key differentiator for colocation customers and managed service providers (MSPs) looking for the best possible partner to house and manage their IT infrastructure solutions (Rose Wellington, 2019). This makes it critically important for a facility to implement a variety of best practices to improve its data center operations.

Rose (2019) further added that **few** things are more critical to data center operations best practices than an effective data center infrastructure management (DCIM) platform. Managing a data center without DCIM software is like trying to sail a boat in complete darkness (Rose Wellington, 2019). It's nearly impossible to know what's happening in the moment and even minor problems can be extremely disruptive because they take the facility by surprise.

Implementing DCIM tools provides complete visibility into the facility's IT infrastructure, allowing data center personnel to monitor power usage, cooling needs, and traffic demands in real time (Rose Wellington, 2019). They can also analyze historical trends to optimize deployments for better performance. With a DCIM platform in place, IT support tickets can be resolved quickly and customers can communicate their deployment needs without having to go through a complicated request process [5].

The study conducted at Phenomon Institute (2016) establish that Data centers necessarily use quite a lot of cable. Whether it's bulky power cables or fiber-optic network cables, the

facility must find ways to manage all that cabling effectively to make sure it all goes to the proper ports (Phenomon Institute, 2016). While messy, unstructured cabling might be a viable solution for a very small on-premises data room in a private office, it's completely unsuitable, and even dangerous, for even the smallest data centers. Cabling used in scalable infrastructure must be highly structured and organized if IT personnel are going to have any hope of managing it all.

Poorly organized cabling is not only messy and difficult to work with, but it can also create serious problems in a data center environment. Too many cables in a confined space can restrict air flow, putting more strain on both computing equipment and the facility's cooling infrastructure ([7] Inefficient cabling can also place unnecessary restrictions on deployments, which can make power distribution inefficiencies even worse.

Other studies established that there is fundamental conflict between the desire to reduce operational expense (OPEX) and the lack of implementation of higher cooling temperatures and the resulting benefits Keith (Klesner, 2017). The goal of the study was to give data center managers the tools to reduce the mechanical cooling power consumption through more efficient use of cooling units. Improving airflow management will allow for elevated return air temperatures and a larger across the cooling unit resulting in increased capacity of the cooling unit. This allows for the cooling to be performed by fewer cooling units to reduce mechanical energy consumption. Further cooling system best practices are discussed, including transition to supply air control, increasing chilled water temperatures, and refurbishment or replacement of fixed speed cooling units with variable speed capability.

Uptime Institute has consistently found that many site managers are waiting for the 'next big breakthrough' before springing into action (Klesner, 2017). The dramatic OPEX, subsequent environmental and cooling stability benefits of improved cooling air management, increased operating temperatures, variable speed cooling fans and other best practices are not being exploited by the data center industry at large.

3. *Literature Review Theoretical*

For the purpose of this study, the researcher used the Technology Acceptance Model (TAM). According to Davis (2015), TAM is a theoretical model aiming to predict and explain ICT usage behavior, that is, what causes potential adopters to accept or reject the use of information technology. Theoretically, TAM is based on the Theory of Reasoned Action (TRA). In TAM, two theoretical constructs, perceived usefulness and perceived ease of use, are the fundamental determinants of system use, and predict attitudes toward the use of the system, that is, the user's willingness to use the system. Perceived usefulness refers to "the degree to which a

person believes that using a particular system would enhance his or her job performance”, and perceived ease of use refers to “the degree to which a person believes that using a particular system would be free of effort.

Another theory that will be considered is the Theory of Reasoned Action (TRA). The theory originates from social psychology, and it is a special case of the Theory of Planned Behavior (TPB) [27] developed TRA to define the links between the beliefs, attitudes, norms, intentions, and behaviors of individuals. The theory assumes that a person’s behavior is determined by the person’s behavioral intention to perform it, and the intention itself is determined by the person’s attitudes and his or her subjective norms towards the behavior. The subjective norm refers to “the person’s perception that most people who are important to him think he should or should not perform the behavior in question”. The authors state that TRA is applicable, for example, when studying consumer behavior, women’s occupational orientations, or family planning behaviors.

From the Theoretical Framework and literature review, the researcher formulated the conceptual frame showing relationship of variables, where the independent variables included availability of new hardware technologies, site selection, cost elements and management of the data center facility and the dependent variable was the successful implementation of the data center. The conceptual framework is shown below:

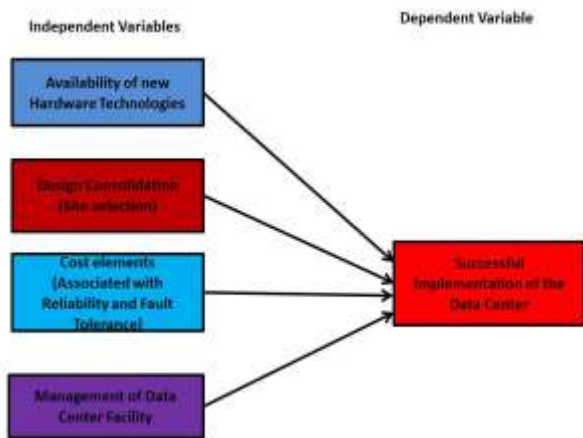


Figure 1.0 Conceptual Framework

•
4. Methodology

4.1 Overview

The study made use of a mixed-method research approach, where both qualitative and quantitative methods are considered. Quantitatively, a structured questionnaire was used to explore the current state of Data Centre best practices

in the Zambia Public Sector. Qualitatively, an interview schedule was also used to find in-depth information on the operations, designs and management of Data Centre ICT infrastructure resources in the public sector.

4.2 Study population

According to Mwion (2015) a population is an entire group of individuals, events or objects with some common observable characteristics. The study took into account employees of at least 12 ministries, at their head office to determine the implantation strategies. The population was approximately 870 employees. The sample size was 120 respondents drawn from various ministries. The sample size was calculated using Raosoft sample size calculator accessed on www.raosoft.com.

4.3 Methods of Data Collection

For the purpose of this study, the researcher used both primary and secondary data. Secondary data gathered from different published reports by government sources, and other countries that have managed to successfully switch from analog to digital broadcasting. Primary data was collected using questionnaires and unstructured interviews.

4.4 Sampling Size

Argyrous (2005) described a sample as a set of cases that does not include every member of the population. The sample of this study consisted of fifteen (15) (randomly selected) and two (2) (purposely selected) Data Centres. Furthermore, the study was made up of thirty four (34) respondents in total, of which thirty (30) represented fifteen (15) randomly selected Data Centres while four (4) out of thirty four (34) was from two (2) purposely selected Data Centres at the headquarters in Lusaka. Of these respondents, seventeen (17) included IT Managers/representatives and seventeen (17), System Administrators/IT operators who overseeing Data Centre operations in the public sector. In short, the study was made up of thirty four (34) participants and 17 Data Centres of different ministries-these included randomly and purposefully selected facilities. The total sample was 120 respondents.

4.5 Methods of Data Analysis

Field data was analysed with the help of IBM SPSS Version 20.0. Further Mega statistics software was used to test the hypothesis. Data was organized according to the stated objectives and checked for errors before it could be analysed. One sample t-test was the statistical method used to test the hypothesis. An assumption of the t-test was that, data is coming from normally distributed populations. The data on the variables was evaluated for normality using means, standard deviations, skewness and kurtosis statistics. More importantly skewness and kurtosis values within the range -2 to +2 were used to indicate no serious deviation from normality. Regression analysis was used to identify the relationships between variables of the study.

In addition the literature analysis revealed a variety of designs of both physical and logical architectures of data centers as depicted in figure 2.0 and figure 3.0 respectively.

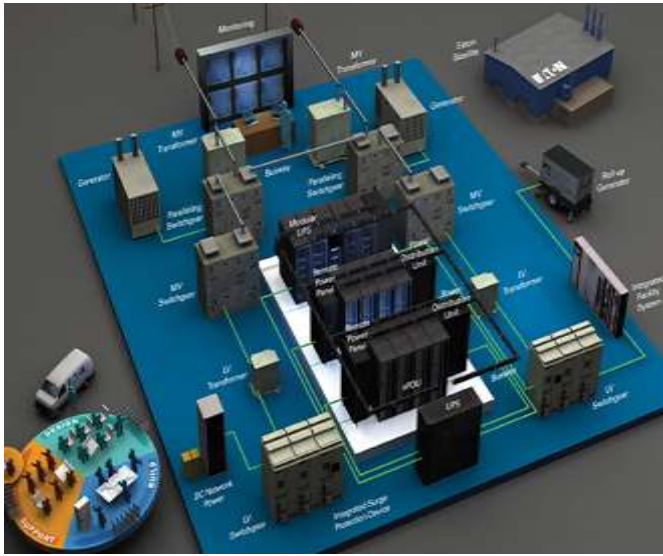


Fig 2.0 Physical Design of Data Center (source <https://www.apetus.com/solutions>)



Fig 3.0 logical design of data center (source https://e.huawei.com/es/tech-topic/en/hw_133535)

1. Results

The results obtained from the field study are represented using mean, Standard Deviation and Skewness. The results are presented using research questions.

1. **Question one:** What is the impact of new hardware technologies on the infrastructure of the data center facility in Zambia?

TABLE I. NEW HARDWARE TECHNOLOGIES

	Mean	Standard Deviation	Skewness
The impact of new hardware technologies on the infrastructure of the data center facility in Zambia will result in speedy access to data	1.18	0.389	1.658
The impact of new hardware technologies on the infrastructure of the data center facility in Zambia will result in Improved Storage of Data.	1.28	0.449	1.021
The impact of new hardware technologies on the infrastructure of the data center facility in Zambia will result in Improved Data Redundancy.	1.27	0.444	1.074

2. **Question Two:** What are critical design considerations and best practices in the site selection and design of data center in Zambia?

TABLE II. BEST PRACTICES

	Mean	Standard Deviation	Skewness
Location of the Data Centre can be a critical design considerations and best practices in the site selection and design of data center in Zambia	1.85	.791	1.296
Availability of power can be a critical design considerations and best practices in the site selection and design of data center in Zambia.	2.12	1.078	1.567
Availability of security can be a critical design considerations and best practices in the site selection and design of data center in Zambia..	3.22	1.696	-.086

3. **Question Three:** What cost elements are associated with different levels of data center reliability and fault tolerance?

	Mean	Standard Deviation	Skewness
Acquisition of New Equipments is the most common cost elements associated with different levels of data centre reliability and fault tolerance in Zambia.	2.12	1.144	1.348
Human capital is the most common cost elements associated with different levels of data centre reliability and fault tolerance in Zambia	3.14	1.243	-.060
Internet connectivity is the most common cost elements associated with different levels of data centre reliability and fault tolerance in Zambia.	2.11	1.012	1.739
Cost of servers' acquisition is the most common cost elements associated with different levels of data centre reliability and fault tolerance in Zambia.	1.51	.502	-.056

TABLE III. RELIABILITY AND FAULT TOLERANT

6. Multiple Discussion

7. RELIABILITY AND FAULT TOLERANT

5. Results Discussion

The study sought to determine whether the impact of new hardware technologies on the infrastructure of the data center facility in Zambia will result in speedy access to data. According to the results obtained and as depicted in Table 5.4 above, the statistics from the data on this variable were very close to the mean as reflected by a mean of 1.18. Showing that most of the respondents strongly agreed that the impact of new hardware technologies on the infrastructure of the data center facility in Zambia will result in speedy access to data. These results are supported by McNamara [16] who also established outdated approaches to data center have left a gap between the management of data and the management of hardware storage.

With regards to whether the impact of new hardware technologies on the infrastructure of the data center facility in Zambia will result in Improved Storage of Data, statistics from the data on this variable were slightly above the mean as reflected by a mean of 1.28. This means that majority of the respondents were agreeing that the impact of new hardware

technologies on the infrastructure of the data center facility in Zambia will result in improved storage of data at the data centers countrywide. Based on the findings, it can be deduced that, new hardware technologies can result in improved storage of data.

Conducted tours at Kitwe Zambia National Data Center (ZNDC) were data centers has been actualized revealed quite much in terms of design considerations. The focus was to assess whether the best practices have been adhered to in these initial implementations. It was observed that indeed much has been done in terms primary power and backup power, security from eventualities such natural disasters. However it has not been established yet whether reliability concerns will stand the test of time in terms of upgrading the hardware should the current become obsolete.

In a case of design consideration, the respondents revealed that the location of the Data Centre can be a critical design considerations and best practices in the site selection and design of data center in Zambia. For this reason, the respondents indicated the data centers are best suited to be located along the line of rail with better access to electricity and communication lines.

In addition, the study sought to determine whether or not the availability of security can be a critical design considerations and best practices in the site selection and design of data center in Zambia. According to the results obtained, the mean value of the response rate from the respondents was 3.2, indicating that the respondents could not state their position as to whether or not the availability of security can be a critical design considerations and best practices in the site selection and design of data center in Zambia.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df	df2	Sig. F Change
1	.730 ²	.533	.519	.270	.533	39.875	3	10	.000

Linda (2015) also stated that securing the Data Centre against vulnerabilities and attacks had grown in importance in the past several years. However, he said many Data Centre modernisation initiatives had missed on the opportunity to improve security and reliability while pursuing cost-reducing IT objectives.

The researcher also wanted to determine whether the acquisition of New Equipments is the most common cost elements associated with different levels of data center reliability and fault tolerance in Zambia. Statistics from the data on this variable were slightly above the mean as reflected by a mean of 2.12. This means that the respondents agreed that the acquisition of New Equipments is the most common cost

elements associated with different levels of data Centre reliability and fault tolerance in Zambia.

6. Conclusion

Based on the results and the discussions provided above, the following conclusions were made:

- That the impact of new hardware technologies on the infrastructure of the data center facility in Zambia will result in speedy access to data.
- That the location of the Data Centre can be a critical design considerations and best practices in the site selection and design of data center in Zambia.
- That the availability of power can be a critical design considerations and best practices in the site selection and design of data center in Zambia.
- That the acquisition of New Equipment's is the most common cost elements associated with different levels of data center reliability and fault tolerance in Zambia.
- That human capital is the most common cost elements associated with different levels of data center reliability and fault tolerance in Zambia.
- That providing for maintenance bypass and emergency shutdown to the data center can be the starting step in management of the data Centre facility in Zambia.
- That determining if equipment requires single phase or three-phase power to the data center can be the starting step in management of the data Centre facility in Zambia.

7. Recommendations

- That deliberate measures should be put in place to acquire new hardware technologies for the data center facility in Zambia to ensure speedy access to data.
- That in trying to design new data centers, the selection of the location should be a critical consideration. In addition, the availability of power is also another critical design considerations and best practices in the site selection and design of data center in Zambia.
- That reduction of cost associated with Acquisition of New Equipments, human capital and Internet connectivity should be reduced automating certain processes and this will in-turn ensure reliability and fault tolerance data Centre.
- That designers of data centers should strive to assess the overall power requirements to manage the data Centre facility in Zambia

- That there should be multiple utility feeds to the data center as a measure to manage the data Centre facility in the country

8. Acknowledgement

The authors would like to express their sincere gratitude to the Management of Mulungushi University, Kitwe Zambia National Data Center (ZNDC), Government Ministries and the Computer Science for their support to the completion of this study.

References

- [1] N. P. S. Chandra, "Strategic Operations Management in Data Centres," *An Empirical Investigation*, pp. 10(3), 129-136., 2010.
- [2] F. Chu, "VMware's App Makes the Cut. Week," pp. 20(8), 46., 2003.
- [3] J. Consoli, "DCIM Meeting the Need to Optimize the Data Center. Mission Critical," pp. 4(4), 60-63, 2011.
- [4] A. & K. R. Dasgupta, "Workload Management for Power Efficiency in Virtualized Data Centers," 2011, pp. 54(7), 131-141.
- [5] P. & N. R. Filani, "Dynamic Data Center Power Management," pp. 12(1), 59-67., 2008.
- [6] W. & B. M. Fox, "A Guide to Managing Research.," *Cape Town: Juta.*, 2007.
- [7] J. Garbani, "Steps for Better Data Centre Capacity Planning," *Networkworld Asia*, pp. 8 (3), 10-11., 2011.
- [8] Namibia., "Namibian Public Service eGovernment Strategic Action," *Plan: e-Government Readiness Report*, 2012.
- [9] S. Hassell, "Promises, Challenges, and Imperatives of Data Centre Infrastructure Management," *Mission Critical.*, pp. 5(3), 44-47., 2012.
- [10] M. Hirst, Effective Data Center Cooling Means Big Savings, *Mission Critical*, 2013.
- [11] A. A. & S. M. M. Khalid, Cloud Computing Technology, *Pakistan Journal Of Science.*, 2013.
- [12] K. Lim, Improving costs, Control and Compliance at the Data Center.
- [13] Y. Y. Lui, Waterside and Airside Economizers Design Considerations for Data Center Facilities, *ASHRAE Transactions.*,
- [14] S. Madaffari, Demystifying Data Center Design Options. *Mission Critical*, 2013.
- [15] S. Madaffari, "Making the Case for Data Center Consolidation with IT Spending Benchmarks," *Demystifying Data Center Design Options. Mission Critical*, pp. 6(2), 44-48., 2013.
- [16] "Computer Economics Report," *Making the Case for*

Data Center Consolidation with IT Spending Benchmarks, pp. 36(4), 1-15., 2014.

- [17] M. t. C. f. D. C. Consolidation., "Computer Economics Report," pp. 30 (8), 1-8., 2008.
- [18] C. G. & B. C. L. Malone, "Optimizing Data Center TCO," *Efficiency Metrics and an Infrastructure Cost Mod*, 114(1), 44-50..
- [19] S. L. Z. B. S. Z. J. & G. A. Marston, "Business Perspective," *Decision Support Systems*, pp. 51(1), 176-189., 2011.
- [20] M. McNamara, "Data Center Storage Trends, Challenges, & Solut. Siliconindia," pp. 12(7), 34-37., 2009.
- [21] B. Milligan, "Raised-floor Solutions Step Out of the Data Center," *Cabling Installation & Maintenance*, pp. 14(2), 24., 2006.
- [22] G. Mone, "Reducing Data Center Power Consumption with Green Storage Technology," *SDA Asia Magazine*, pp. 2(9) 36-39., 2009.
- [23] R. Angles, A comparison of current graph database models., IEEE, 2012, April.
- [24] A. R. F. & A. L. I. Boicea, MongoDB vs Oracle-- database comparison, IEEE, 2012, September, pp. pp. 330-335.
- [25] C. E.F., A Relational Model of Data for Large Shared Data Banks, 1970.
- [26] H. W. Lane David, "Web Applications Applications with PHP and MySQL," vol. 29, no. 104, 2009.
- [27] M. Valentin, Database comparison system and method., 2014.

An Assessment of the Challenges and Benefits of Digital Broadcasting Migration in Zambia

Davies Kintu
Mulungushi University
School of Science, Engineering and Technology

Christopher Chembe
Mulungushi University
School of Science, Engineering and Technology

Abstract— Digitization is the current trend in Television broadcasting, both in Zambia and the world over. The International Telecommunication Union (ITU), set 2015 as the year for the entire broadcasting stations in the world to go digital. Despite the advantages that come with digital migration, Zambia is yet to realize the full benefits of digital migration implementation. It's against this background that the study is to be conducted to assess the challenges of digital migration process and its benefits in Zambia. To this effect, a quantitative research design will be used on a sample of 300 stakeholders who have direct influence on the process to determine their views. The data collected shall be analysed using Statistical Package for Social Sciences (SPSS) to categorise the challenges of the migration process and its benefits. The study is aimed at assessing the digital terrestrial television migration implementation process; examine the challenges of switching from analogue to digital technology and its benefits.

Index Terms-- ICT, Digital Migration, Broadcasting, Digital, Terrestrial, Benefits and Challenges.

1.0 INTRODUCTION

In Zambia, just like in many other countries, Television is a popular source of information, a medium of education and entertainment (Musa, 2014). The analogue and digital technologies comprise the two forms of television broadcasting. The former has been in existence since the onset of television broadcasting in the world. However, due to rapid changes in information and communication technologies (ICTs), analog has been phased out by digital whose myriad features makes it more superior (Agona & Otim, 2011).

Digital Broadcasting Migration is a process by which broadcasting services offered on analog networks are transferred to digital networks over a specific period (Sam Agona and Sansa Otim Julliane, 2012). Zambia, like many other countries in the world has migrated from analog to Digital Television Transmission. The benefits of digital broadcasting stretch to the consumers, broadcasters as well as service providers. The transition from analog transmission to

digital broadcast system implies that limited radio frequency will be used in the transfer of television signals.

The presentation of numerous differentials is quite technical that the traditional terrestrial signal transmission lacks the capacity to provide. A digital signal provides viewers with options to choose their desired content, has high flexibility, besides a place and time of content use via any machine the user decides to use. For example, a portable terminal, a television or even a computer (Ihechu, 2012).

Digitalization has the potential to offer new openness coupled with diversity to the broadcast sector by developing multiple new channels, through downsizing of the amount of frequencies necessary for one station's broadcast, Zambia has finally migrated from analogue to digital.

This study is therefore aimed at assessing the challenges and benefits of digital migration in Zambia. The study focused its attention on the benefits and challenges associated with digital migration in Zambia.

1.1 STATEMENT OF THE PROBLEM

Over the years now, the analogue television system has been in existence in Zambia. However, the analogue broadcasting techniques has contributed its own quota to the development of broadcasting industry in Zambia and cannot be overemphasized. With the advent of the digital transmission technology which is considered to revolutionize the broadcasting industry in Zambia and set to be a technical solution to the many technical challenges inherent in the analogue terrestrial system. It is pertinent to disclose that digital technology has brought about qualitative signal processing, reception, reach and digital dividends which were never present in the analogue broadcasting era (B. Sanou, 2016). It is against this background that this study is conducted to assess the challenges and benefits of digital migration process in Zambia.

1.2 OBJECTIVES OF THE STUDY

1. The main research objective is to assess the challenges and benefits of digital broadcasting migration implementation process in Zambia.
2. To assess the main challenges associated with digital broadcasting migration in Zambia
3. To establish what benefits can be obtained as results of migrating from analogue to digital.
4. To assess the policy frame work which the Ministry of Information and Broadcasting has put in place after the successful digital migration.

1.3 RESEARCH QUESTIONS

1. What are the challenges and benefits of digital migration implementation in Zambia?
2. What are the main challenges associated with digital migration in Zambia?
3. What benefits can be obtained as results of migrating from analogue to digital?
4. What policy frame work has been put in place by the Ministry of Information and Broadcasting after the successful digital migration?

1.4 SIGNIFICANCE OF THE STUDY

The study has the potential to help other African countries which are yet to implement or in the process of digital migration implementation. It is also aimed at making suggestions to overcome the identified challenges.

2.0 REVIEW OF LITERATURE

Digital broadcasting explains the physical transfer of data in digital bit stream over a point-to-point, point-to-multipoint transmission medium or multi point to multipoint in discrete levels. This can be over media such as copper wire, optical fiber, wireless communication media, and storage media. In digital video transmission, the cameras, recorders and studio equipment are digital. In addition, digital files are smaller compared to analog due to more sophisticated compression techniques. A digital signal is a discontinuous signal that changes from one state to another in discrete steps (Jared Baraza, 2009).

In Kenya's experience with the digital television migration included the key challenges faced and the solutions implemented to highlight valuable lessons that can provide useful information and guidance to regulators and policymakers (Telecommunications Management Group Inc, 2017).

In a related study the, ITU Regional Office for Asia and the Pacific indicated that the costs, enormity of task, lack of government support and cooperation among stakeholders, and ever evolving technologies remain as some of the challenges in the migration from analog to digital in the region (Wisit Atipayakoon, 2017). He further noted that, not only is the setup of digital infrastructure expensive, also carrying out a simulcast service for a given period involves a lot of money and resources. Subsidizing set- top-boxes to make them affordable for the people is another additional cost.

Another study conducted by (Chang Jin, 2017) also established that many governments still do not consider the digital migration as a priority hence the move is not fully endorsed or supported by the authorities and the necessary initiative and push. He further added that the move to digital is not initiated as a collective effort by private broadcasters, regulators and others in many countries. It was further established that the digital terrestrial transmission technologies are matured with a number of providers and supporters readily available. However, there are still cases where some still feel it may be better to wait for the next technology or next standard.

The study conducted in South Africa revealed that digital migration in Africa has rarely been discussed or taken into account (SA Time Publications, 2016). The end of analog broadcasting and the production of dedicated equipment are likely to give rise to serious problems in Africa such as: problems of maintaining infrastructures which remain analog and the further 'Balkanisation' of African production, etc. In addition, the move to digital broadcasting brings with it other crucial challenges regarding regulation planning, pluralism of information, media development and access for all too new digital equipment, and increasing dependency beyond national borders in the sector (SA Time Publications, 2016).

3.0 RESEARCH METHODOLOGY

3.1 RESEARCH DESIGN

In this study, a quantitative strategy was taken in the design, data collection and data analysis. Particularly, a cross-sectional survey was conducted, on a sample size of 300 individuals selected using a snowball sampling technique on the basis of sub-sectors, size, and location. Data was collected using a survey questionnaire, to be tested for reliability and validity using coefficient, and confirmatory factor analysis respectively. Finally, the survey questionnaire was self-administered by the researcher using a drop-pick method, and data collected, analysed using descriptive statistics and hypotheses tested with SPSS.

3.2 METHOD OF DATA COLLECTION

For the purpose of this study, the researcher used both primary and secondary data. Secondary data gathered from different published reports by government sources, and other countries that have managed to successfully switch from analog to digital broadcasting. Primary data was collected using questionnaires.

3.3 SAMPLING DESIGN

Sampling is the procedure a researcher uses to gather people, places or things to study. It is a process of selecting a number of individuals or objects from a population such that the selected group contains elements representative of the characteristics found in the entire group (Azhar, D. 2002). The researcher used simple random sampling technique. The researcher will obtain a list of the 300 Zambia National Broadcasting Corporations (ZNBC) TV viewers and few selected people involved in the digital migration exercise.

3.4 METHODS OF DATA ANALYSIS

The collected qualitative data was analyzed using content analysis whereas quantitative data was analyzed using Statistical Package for Social Sciences (SPSS). These methods entail transcribing the recorded data and systematically identifying the major themes in the data and categorizing the data into common themes and testing it using statistical measures (Fox 2009).

4.0 RESULTS

The results obtained from the field study are represented using mean, Standard Deviation and Skewness. The results are presented using the objectives that were sent in section one of this study.

1. **Objective one:** To assess the main challenges associated with digital migration in Zambia

TABLE I. CHALLENGES OF DIGITAL MIGRATION

	Mean	Standard Deviation	Skewness
Does the country have adequate infrastructure for digital migration?	3.03	1.118	.694
Does the country have adequate skilled personnel to facilitate the digital migration process?	3.06	1.146	-.040
Are there adequate funds available for digital migration project?	2.84	1.127	.437
Does the country have enough digital migration equipment to facilitate the process?	3.24	1.234	-2.18

2. **Objective Two:** To establish what benefits can be obtained as results of migrating from analogue to digital.

TABLE II. BENEFITS OF DIGITAL MIGRATION

	Mean	Standard Deviation	Skewness
One of the benefits of digital broadcasting migration is quality of transmission	3.58	1.125	.025

Does digital migration help improving the content delivery than analog transmission?	4.71	1.266	1.089
I do understand that through digital migration, I can have more access to various TV channels.	4.97	1.378	.270
I do understand that through digital migration, I can have better picture quality on my TV set.	3.84	1.301	.409

3. **Objective Three:** To assess the policy frame work which the Ministry of Information and Broadcasting has put in place after the successful digital migration.

TABLE III. POLICY FRAMEWORK OF DIGITAL MIGRATION

	Mean	Standard Deviation	Skewness
Has the government put policies in place to ensure the transition process from Analogue to Digital broadcasting is a success?	3.58	1.278	.034
Does this policy framework for digital migration process involve subsidizing the acquisition of set top boxes or TV Antenna?	2.10	1.274	.095
Does the policy framework ensure that it creates public awareness with regards to the importance of digital migration process?	1.35	2.137	.407
Does the government policy attempt to make sure that the set top boxes or TV Antenna are readily available?	4.07	.036	.131

5.0 RESULTS DISCUSSION

As with regards to challenges of digital migration, the study wanted to establish whether the country had adequate infrastructure for digital migration. According to the results obtained and as depicted in Table I, most respondents were neutral on this item in such a way that some indicated that the country has inadequate infrastructure for digital migration were as others were of the view that there was adequate infrastructure. This correlate with the findings of Wisit Atipayakoon (2017) in Kenya which revealed that not only does the setup of digital infrastructure expensive, also carrying out a simulcast service for a given period involves a lot of money and resources.

The study also further sought to determine whether there are adequate funds available for digital migration project in the country. The statistics obtained from these questions showed that the respondents disagreed to that there was no adequate funding for the digital migration process. Because of this, many of the challenges being talked about could have resulted.

This is attributed to the mean scored which stands at 2.84 and standard deviation at 1.127 of the total response rate. The results correlate with those by Chang Jin (2017) who indicated that lack of funding is a barrier to a successful migration process.

The study further established that the country did not have enough digital migration equipment to facilitate the process. Statistics from the data on this variable were above the mean as reflected by a mean of 3.24 and standard deviation of 1.234.

At the same time the researcher sought to know whether one benefit of broadcast digital migration is quality transmission. The results obtained revealed the mean value of 3.58 and standard deviation of 1.125. The skewness was .025. This implies that the majority of the respondents stated that, quality transmission is one benefit of digital migration. The results correlated with those by Jared Baraza (2009). It was also observed that digital migration has the potential to improve content delivery.

The researcher sought to establish whether or not the government had put policies in place to ensure the transition process from Analogue to Digital broadcasting is a success. Statistics from the data on this variable was slightly above the mean as reflected by a mean of 3.58 and standard deviation of 1.278. The results showed that the respondents slightly agreed that the government had put policies in place to ensure the transition process from Analogue to Digital broadcasting is a success.

In the same vain, the researcher sought to establish whether or not the policy framework ensure that it creates public awareness with regards to the importance of digital migration process. According to the results obtained and as indicated by the mean value of 1.35 and standard deviation of 2.137. The respondents agreed that the policy framework ensure that it creates public awareness with regards to the importance of digital migration.

6.0 CONCLUSION

The study of the challenges and benefits of digital migration process in Zambia has revealed that inadequate infrastructures for digital migration had impacted negatively on the implementation process. However, other factors that were considered to have had impacted negatively in the process include the fact that the country was not technologically prepared for the digital migration in terms of the skilled personnel to handle the technology among other things.

The study found that inadequate funding to the process by the government also hindered the process of migration. It was found that, funding to the process of migration in the country was not well coordinated as it kept delaying. These delays impacted negatively on the procurement of digital migration equipments. Since the country lacked, digital migration equipments, it was difficult to implement the process on time. However, despite the challenges that have been outlined, the respondents revealed that there were a number of benefits associated with digital migration and prominent ones include, quality of transmission, improved content delivery, ability of

the viewers to view more than one TV channel at once thereby giving them a better viewing pleasure.

The findings also revealed that, the government had put in place policy framework to ensure a successful migration process. It was established that the policy framework put in place meant to create awareness among the people of the benefits of digital migration.

7.0 RECOMMENDATIONS

- i) That deliberate policies should have been put in place to ensure that the infrastructure needed for the digital migration are readily available.
- ii) Government policies should have been put in place to ensure that it creates public awareness with regards to the importance of the digital migration process to allow for more public participation.
- iii) Government should have ensured that the country has adequate skilled personnel to facilitate the digital migration process. This would have ensured that the process is completed on time and with fewer challenges.
- iv) Government should actualize the establishment of the content development fund to finance local content development. This will help create employment and ensure optimal utilization of channels.
- v) That in the future whenever policy framework is being set, it should be able to outline the funding for the Digital migration process.

8.0 FURTHER RESEARCH

Though there are a number of studies on the digital migration, continued research is needed to improve this study and to address its limitations. More research on digital migration in remote parts of the country must be explored. There is need to look into change of sources of funding for digital migration.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Management of Mulungushi University and School of Science, Engineering and Technology.

REFERENCES

- [1] B. Sanou (2016) Guidelines for the Transition from Analogue to Digital Broadcasting-Including the Asia-Pacific Region, Telecommunication Development Sector, ITU, Geneva, Switzerland,
- [2] B. Roland (2013), The Digital Dividend of Terrestrial Broadcasting, Springer, Dordrecht Heidelberg, Germany
- [3] J. Doeven (2012), Guidelines for the transition from analogue to digital broadcasting, ITU, New York.
- [4] B. Sanou, and F. Rancy (2012), Digital Dividend: Insight for spectrum decisions, ITU, Geneva, Switzerland,
- [5] SADC Report (2014), SADC Roadmap for Digital Broadcasting Migration, Johannesburg, South Afric.

- [6] S. Kekana (2014), Digital broadcasting state of play – SADC Opportunities, Challenges and Possibilities, Pretoria, South Africa, 2014.
- [7] Telecommunications Management Group inc, 2017– Publication Television Archiving -, ABU Malaysia
- [8] Wisit Atipayakoon (2017)(ABU) – Publication File Formats for sound programme exchange, Archives at Risk Alliance: www.archivesatrisk.org
- [9] Butcher me, “Digital Video Archives – facing the facts” EBU Technical Review – Geneva Summer 2017 11 pages. Dublin Core Metadata Initiative; <http://dublincore.org/>
- [10] Chang Jin (2017) Systems Architecture, Services and Functionality” Geneva December 2017 , page 111
- [11] EBU Project Group P/FTA (Future Television Archives), “Supplement to the EBU Report Future content Management Systems Architecture, Services and Functionality” Geneva December 2016
- [12] European Broadcasting Union (EBU), “Technical Report 003, Radio Archives: Conception and Practice, Geneva July 2017
- [13] European Broadcasting Union (EBU) “Recommendation R105-2008 Digitisation of programmed material in Audio Archives: Geneva March 2015.
- [14] Ministry of Information and Broadcasting Services (MIBS). (2014, February 15). Digital Migration Policy. Retrieved 3 July 2019, from <https://www.iba.org.zm/downloads/Digital-Migration-Policy.pdf>
- [15] Organization for Security and Co-operation in Europe (OSCE). (2010). Guide to the Digital Switchover: <https://www.osce.org/fom/73720?download=true> accessed on 07/08/2019
- [16] Ardizzone L and Ferrari R (2010). The Challenges of digitisation of broadcasting in Nigeria: Department of Mass Communication, Abia State University. [Online]. Available: <http://iiste.org/Journals/index.php/NMMC/article/viewFile/3114/3155>. Accessed on 20/08/2018
- [17] Bunshak, T. (2006). Digital broadcasting is now. NBC News. 8(2).
- [18] Burrows R. (2014). Media Law in South Africa. Kluwer Law International: New York.
- [19] Digital Migration Policy For Television Broadcasting in Uganda. 2011. [Report] Uganda by Ajzen (2014).
- [20] Fox Tewksbury. (2009) Qualitative Versus quantitative Methods: Understanding why Qualitative methods are Superior for Criminology and Criminal Justice. Journal of Theoretical and Philosophical Criminology, Vol 1. P.
- [21] Fred Davis (2014). Mapping Digital Media: Kenya. A Report by the Open Societies Foundation. Open Society Foundations, United Kingdom.
- [22] Hanson Edmunson. (2005). ICT Lawyers. [Online]. Available: <http://www.kerronedmunson.co.za/resources/dtt-article-ke-oct-2011.pdf>

An Insight of Smart Health Care As A Growing Application Need In Zambia-A Systematic Literature Investigation

Darius Bwalya
*Directorate of Information
 Communication Technology
 Copperbelt University
 Kitwe, Zambia
 Email:Darius.bwalya@cbu.ac.zm*

Dr. Christopher Chembe
*School of Science, Engineering
 and Technology
 Mulungushi University
 Kabwe, Zambia
 cchembe@mu.ac.zm*

Victor Kasonde
*Directorate of Information
 Communication Technology
 Copperbelt University
 Kitwe, Zambia
 Email:vkasonde@cbu.ac.zm*

Abstract:-With the current trend of world population exploding, there is a rising constraint on the provision of quality healthcare. This exactly describes Zambia in which according to the United Nations Department of Economic and Social Affairs: Population Division, the current population is now well over 18 Million against a limited number of medical staff, creating a situation where access to a doctor is characterized by long waits which escalate in hours in most cases. Add to that management of chronic illness. It is undeniably clear that there is a great need to explore other viable healthcare alternatives. Thanks to ICT technology, there is an emerging form of healthcare provision called Smart Health Care in which Cloud Computing and Internet of Things (IoTs) is expected to complement each other, creating a great possibility to automate health care. Before Zambia adopts Smart health Care, it must be understood and appreciated by the stakeholders. This paper is therefore a systematic literature investigation of an insight of Smart health Care provision. The paper highlights what the components of Smart Health Care (Cloud Computing and IoTs) are and how the two technologies can collaborate as well as the challenges arising from this. Finally the paper shows how this technology can be adopted in Zambia in its endeavor to improve provision of healthcare.

Keywords: *Cloud Computing, IoT, CloudIoT, Smart Health Care, Wireless Body Area Network(WBAN)*

1.0 INTRODUCTION

The population explosion currently being experienced world over is placing a great constraint on the provision of high quality health care for all. Zambia is a typical example whose current population as reported by United Nations Department of Economic and Social Affairs: Population Division, has exploded to over 18million currently against limited number of medical doctors.

According to the World Health Organization, the ideal doctor to patient ratio is 1 to 5,000 respectively whereas in Zambia the ratio is shockingly at 1 doctor per 12,000 patients. This situation results in an unacceptably long waits before having access to the doctor which in many cases is over an hour. Appointments can last even a year or more. Chronic illnesses that require constant management are increasingly common. Add to that the difficulties in life that comes with getting aged. It is undeniably clear that there is a great need to explore other viable healthcare alternatives. Thanks to ICT technology, there is an emerging form of healthcare provision called Smart Health Care in which Cloud Computing and Internet of Things (IoTs)(Botta, de Donato, Persico, & Pescapé, 2016) is expected to complement each other, creating a great possibility to automate healthcare. Before Zambia adopts Smart health Care, it must be

understood and appreciated by the stakeholders. This paper is therefore a systematic literature investigation of an insight of Smart health Care provision. The paper highlights what the components of Smart Health Care (Cloud Computing and IoTs) are and how the two technologies can collaborate as well as the challenges arising from this. Finally the paper shows how this technology can be adopted in Zambia in its endeavor to improve provision of healthcare. It is now imperative that important aspects of the research are explored to provide an insightful perspective these being: Cloud Computing, IoT, CloudIOT paradigm, Smart Health Care.

2.0 Insight on Cloud Computing

Cloud Computing revolves around servers on the internet providing services to clients in terms of infrastructure, Platform and software. The services highlighted are classified into service models as being: Infrastructure-as-a-Service(IaaS), Platform-as-a-Service(PaaS) and Software-as-a-Service(SaaS) respectively. Perhaps one of the all-encompassing definition of Cloud Computing is one provided by Mell & Grance (2011) when they defined it as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. The main service models are now briefly explained:

- (a) Infrastructure-as-a-Service (IaaS):- Under this model clients or customers are provided with infrastructure services by the service providers including networks, storage, computing allowing a client or cloud user to deploy and run application as well as system software. The benefit arising from this is that the cloud user does not need to control or manage the infrastructure but has full control over applications and system software.

Additionally the user will have nothing to do with the knowledge of where his or her information or application is stored and actually running(Transparency concept).

- (b) Software-as-a-Service(SaaS):-This service model implies that the cloud user or customer uses applications of the service provider running on the cloud. Mode of access of these applications could be via interfaces such as web browsers and other Application portals. Control and knowledge of servers, network, processing scalability involved, storage is not necessary which turns out to be the benefit.
- (c) Platform-as-a-Service(PaaS):-This service model according to Mell & Grance (2011) provides capability to the consumer by means of deploying consumer-created and or acquired applications created using programming languages, libraries and tools supported by the provider onto the cloud infrastructure. Similar benefits that apply to other models apply here also being hiding or separating from the user management and control of network, servers, OS or even storage apart from user deployed and perhaps configurations of the host environment.

The service models are now summarized in figure 1.0 below.

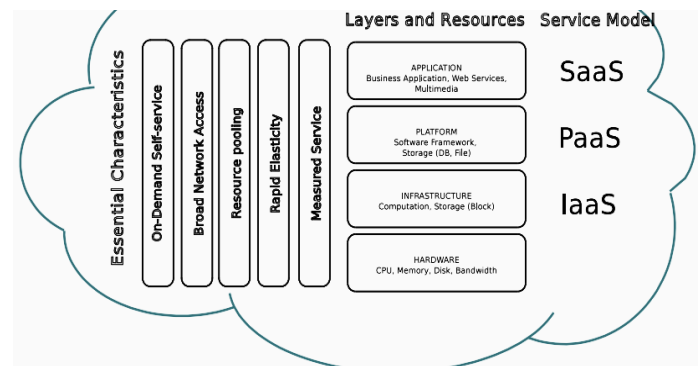


Fig 1.0 (Botta et al., 2016) an overview of cloud computing paradigm

Benefits and Barriers

(Avram, 2014) highlighted some of the benefits as follows:

- Great potential in lowering the cost of entry for smaller firms trying to benefit from compute-intensive business analytics that were hither to available only to the largest of corporations.
- It can provide an almost immediate access to hardware resources, with no upfront capital investments for users, leading to a faster time to market in many businesses.
- Makes it possible for emergent of new applications and services that were not possible before.

(Avram, 2014) however also noted some of the major concerns of going cloud, warning that there may be :

- Security and privacy concerns for maintaining privacy of an individual’s information.
- Reliability of connectivity: adoption depend on having optimum internet connectivity as well as power. This also implies that services must available 24/7 basis.
- Political issues to do with global boundaries: For cloud computing to continually evolve into a borderless and global tool, it needs to be separated from politics. Currently, some major global technological and political powers are making laws that can have a negative impact on the development of the global cloud.

Despite the challenges Avram (2014) noted that Cloud computing represents a huge opportunity to many third-world countries that have been so far left behind in the ICT revolution.

3.0 An Insight on Internet of Things (IoT)

This is an emerging technology that is receiving attention and which is rapidly progressing and a

good number of literature has been published on the same. Dimitrov(2016) defines IoT as a network of physical devices and other items, embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data. It is one of the technologies that is now referred to as disruptive technology (Weyrich & Ebert, 2016). It is becoming part of our lives and represents one of the most disruptive technologies giving rise to ubiquitous and pervasive computing scenarios (Botta et al., 2014).

4.0 Insight on the CloudIoT Paradigm

Cloud Computing and IoT are two separate technologies and much has been published about them separately. The CloudIoT paradigm seeks to combine these two technologies in ways that will bring about technological revolution. The major drivers to combining cloud computing and IoT have been well noted by many researchers. The main reasons for integration stems from the fact that the two technologies are complementary in nature that is, having capabilities that lack from each other. Table 1 below is a summery by Botta et al.(2014) of the main reasons for the integration:

Table 1: Complementary aspects of IoT and Cloud Computing

	IoT	Cloud Computing
Displacement	Pervasive	centralized
Components	Real world	Virtual resources
Computational Capacities	Limited Processing capability	Virtually unlimited Processing capability
Storage	Limited or none	Virtually unlimited storage
Role of the Internet	Point of Convergence	Service delivery
Big data	Big data source	Means to manage big data
Reachability	Limited reachability	Ubiquitous

From the table above it can be deduced that the major drivers for integration of IoT and Cloud

Computing being referred to as CloudIoT are as follows:

- **Storage:** The Things have limited or even none storage capabilities in which case they can benefit from the virtually unlimited storage provided on the cloud infrastructure.
- **Processing:** With limited computation capabilities, the Things can utilize the virtually unlimited processing capabilities provided on the cloud.
- **Power:** since the things have limited power they can still benefit from the power provided on the cloud.
- **Security:** This too can be one of the advantages that the things can benefit from the cloud.
- **Reliability:** Because cloud computing is a form of distributed system, there is a high level failure transparency because of pooled interconnected resources which seamlessly ensure reliability which the things can benefit from.

On the other hand the cloud can benefit from the CloudIoT by extending its scope to deal with real life objects in a more dynamic and distributed way and providing services for countless devices in assorted real life scenarios (Atlam, Alenezi, Alharthi, Walters, & Wills, 2017). The cloudIoT paradigm give rise to the possibility of smarthealth care application scenario which is just one among many other application scenarios that CloudIoT makes possible. In this paper the focus is on Smart health care which will now be discussed.

5.0 Smart Health Care As A CloudIoT Application Scenario

This involves management of a patient's condition by using ICTs application resulting in improved better quality of services. Stantchev et al. (2015), stated the benefits of this application scenario for instance, that smart health care

support patients to adhere to prescribed medication increasing chances of recovery. They add that smart healthcare makes it possible for better treatments by for instance providing of smart pill boxes with appealing features like opening at precise time and recording of pill intake time. There is too a promise of good hospital management. There is even an emerging network application in IoT called Wireless Body Area Network (WBAN). Khan et al. (2010) defines this form of technology as being a special purpose sensor network designed to operate autonomously to connect assorted sensors and appliances, located inside and outside of human body.

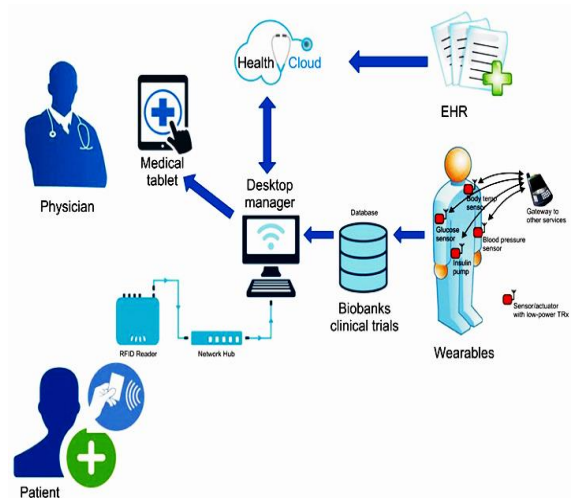


fig 1.1 Typical illustration of possible Smart Health application (Dimitrov, 2016)

Fig 1.1 shows how the things referred to as wearables can be used to collect various internal conditions such as blood pressure, glucose level, body temperature e.t.c and transmit that data to Biobanks for storage or to other gateway to some other services. The physician can then, using a mobile device or desktop PC, access or monitor patient's condition and advise accordingly or perhaps schedule and appointment as need arises. At times a patient could be alerted by means of the mobile devices using the information gathered from the wearables and be warned of possible worsening condition or of missing prescription at an appropriate time. IoTs can be used in what is now referred to as Ambient

Assisted Living(AAL)(El murabet, Abtoy, Touhafi, & Tahiri, 2018) for people with chronic diseases and the elderly. However there are a number of challenges that comes with the Smart healthcare and the next section discusses these.

6.0 Challenges of adoption of Smart Health as a CloudIoT Application Scenario

Because CloudIoT paradigm as already alluded to is a disruptive technology that brings about entirely new applications scenarios, it comes with it new fresh challenges to deal with which have a potential to impede successful implementation. Researchers in (Atlam et al., 2017) identified the many ones as being:

1. *Security and privacy*:- As data are transmitted from the things, implementation of only authorized users to have access to the confidential data which obviously is so important to patients must be guaranteed and additionally when data moves to the cloud trust issues arise (Atlam et al., 2017). Possibility of attacks must be put into consideration too.
2. *Heterogeneity*:- In the CloudIoT, there can be so much versatility when it comes to the types of devices, platforms, OSs and services that may exist.
3. *Big data*:- it is expected that there will be some 50billion IoT devices by 2020 (Atlam et al., 2017) and probably you would know what that implies as to the volume and the velocity(frequency of data production) that will be expected to be transmitted. This brings about storage constraints, bandwidth as well as analysis complexities.
4. *Large scale*:- since there is interaction of billion devices, this poses some new challenges difficult to deal with such as computation capabilities and the IoTs challenge of connectivity because of their wide distribution.
5. *Standardization*:- “Although a number of proposed standardizations have been

put forth by the scientific society for the deployment of IoT and Cloud approaches, it is obvious that architectures, standard protocols, and APIs are required to allow for interconnection between heterogeneous smart things and the generation of new services, which make up the Cloudbased IoT paradigm”(Atlam et al., 2017).

6. *Energy efficiency*:- Most IoTs have very limited amounts of power as table 1 highlighted especially that data is expected to be transmitted between the cloud and the things.

Concerning these challenges a number of research has been done to handle these challenges and now these will be the focus of discussion in the summery Table 2 below.

Table 2: Studies conducted to handle challenges of Cloud and IoT integration

Challeng e	Study	Problem	Solution
Security and Privacy			
	(Stergiou, Psannis, Kim, & Gupta, 2018)	Security issues in both cloud and IoT	Part of AES presented for improvement of security issue, resulting from integration .
	(Dorri, Kanhera, Jurdak, & Gauravaram, 2017)	Blockchain-based approaches provide decentralized security and privacy,	Proposed BC-based smart home framework was proved secure by thoroughly analysing

		yet they involve significant energy, delay, and computational overhead that is not suitable for most resource-constrained IoT devices	its security with respect to the fundamental security goals of confidentiality, integrity, and availability.
	(Mukherjee et al., 2017)	New security and privacy threats arise that were not present in centrally-managed cloud computing	Presents a summary of up-to-date research contributions and to outline future research direction to solve different challenges in privacy and security in the fog computing .
Heterogeneity			
	(Sourso et al., 2016)	The need to have a cross platforms for a unified access to and sharing of sensing or actuating resources	Description of IoT landscape revealing challenges which need to be address also authors presents a SymbloTe project which has

			focus on interoperability platforms.
	(Yaqoob et al., 2017)	Expected interconnection of trillion number of a variety of devices and sensors	An intensive exploration of IoT architecture is reported focusing on advances made then categorizing and classifying of IoT architectures based on applications, business objectives, architectural requirements, network topologies among others and then identify key future requirements.
	(Suciu, Halunga, Vulpe, & Suciu, 2013)	Success of IoT expected to depend so much on connecting and integration of existing resources.	The solution is a generic platform which demonstrates that IoT is all about interoperability, from connected cloud

			computing using RFID, NFC, M2M as well as digital content and context-aware services.
Large Scale Computation			
	(Hasan, Hossain, & Khan, 2015)	Requirement of localized computation in order to ensure performance, security, and lower costs in IoTs	Aura, an advanced localized IoT based cloud computing model is presented in which computations are done with high flexibility, giving rise to full control to start, stop, migrate, and restart computations in nearby devices as clients move between different physical locations.
	(Dey, Mukherjee, Paul, &	Gaining an understanding of challenge	By means of experiments the study revealed

	Pal, 2013)	services arise with running computational jobs on mobile devices using different types of workloads	that mobile devices offer free time slots to servers in the cloud based on data analytics system.
Standardization			
	(Banafa, A., 2016)	Most IoT service providers building manufacturer specific hardware devices to the relevant cloud services thereby lacking consistency. The need for a standard model to perform popular IoT backend tasks, such as processing, storage, and firmware updates, is becoming more significant.	IoT standardization model is explained with components Platform, connectivity Killer Applications, Business Model and Challenges of IoT are identified.

	(Al-Qaseemi, Almulhim, Almulhim, & Chaudhry, 2016)	Issues to do with a variety of languages, protocols and standards. Lack of single platform of standardization.	Review challenges of standardization and existing state of the art standards to highlight and overcome it by presenting some technical solutions.
	(Shan zhi Chen, Hui Xu, Dake Liu, Bo Hu, & Hucheng Wang, 2014)	Challenges on technologies, applications, and standardization a Chinese perspective	Proposes an open and general IoT architecture made up of three platforms to meet the architecture challenge as well opportunity and prospects of IoT.
Energy efficiency			
	(Yaqoob et al., 2017)	Frequent interchange of datasets among things impacting negatively on performance	Paper develops a novel multi-cloud IoT service composition algorithm (E2C2) that seeks to creating an energy-

			aware composition plan by searching for and integrating least possible number of IoT services.
--	--	--	------------------------------------------------------------------------------------------------

From summery table it is indeed clear that a number of strides have been made in seeking to address the main hindrances to cloudIoT, a basic foundation on which Smart Health Care application scenario is based.

7.0 Literature Search Methodology

The primary search engine utilized was google scholar it being one of the most popular and advanced search engines aggregating several famous databases of various research works. A search strategy focused on the keywords on which the research topic is centered, these being Cloud Computing, IoT, CloudIoT, Smart Health Care, Wireless Body Area Network. Queries on Scholar.google.com were done. For instance to make sure much of the relevant links are harvested query expressions such as “Smart Health Care” OR “smart healthcare”, “CloudIoT” OR “Cloud IoT” OR “Cloud Computing”. For IoTs terms like “Smart Objects” OR “Internet of things” OR “IoT” were used.

To aggregate the research works on the challenges of integrating cloud computing and IoT or CloudIoT the same last key phrases “.....challenges of Cloud IoT” were used for all six of them. For example “security and privacy challenges of CloudIoT” which would harvest the related relevant search results.

Any paper that never matched any key terms was rejected.

Databases explored were ieeexplore.ieee.org, Elsevier.com,

dl.acm.org, researchgate.com, emeraldinsight.com, link.springer.com, sciencedirect.com and grey sources from general www.google.com. Focus was placed on research work done between 5 to 10 years representing most recent research work.

7.1 Results

More than 25 research papers were accessed although the limiting factor was not having enough subscription to some other databases.

8.0 Conclusion

From this in depth understanding of Smart health care as an application scenario of a disruptive technology (CloudIoT paradigm) with a great potential to revolutionize management of not only diseases but also hospital management, it is expected that there will be a new dawn of high quality unprecedented health care. Much work has been published has already noted which Zambia can take advantage of as it does not have to be a trial method. For instance the concern of security and privacy which is a very significant impedance to integrating cloud and IoT which is a key in implementing smart health has attracted useful research highlighted in table 2 which should be a reason for having implementation confidence. Of most importance, smart health care will very much decongest hospitals and make illness management more personal, improve medication prescription adherence via alerts from smart objects or things and assist the aged in Ambient Assisted Living. It is definitely a technology worthy trying.

References

World Health Organization. (2000). *The world health report 2000: health systems: improving performance*. World Health Organization.

Al-Qaseemi, S. A., Almulhim, H. A., Almulhim, M. F., & Chaudhry, S. R. (2016). IoT architecture challenges and issues: Lack of standardization. In *2016 Future Technologies Conference (FTC)* (pp. 731–738). San Francisco, CA, USA: IEEE. <https://doi.org/10.1109/FTC.2016.7821686>

Atlam, H. F., Alenezi, A., Alharthi, A., Walters, R. J., & Wills, G. B. (2017). Integration of Cloud Computing with Internet of Things: Challenges and Open Issues. In *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (pp. 670–675). Exeter: IEEE. <https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.105>

Avram, M. G. (2014). Advantages and Challenges of Adopting Cloud Computing from an Enterprise Perspective. *Procedia Technology*, *12*, 529–534. <https://doi.org/10.1016/j.protcy.2013.12.525>

Botta, A., de Donato, W., Persico, V., & Pescapé, A. (2016). Integration of Cloud computing and Internet of Things: A survey. *Future Generation Computer Systems*, *56*, 684–700. <https://doi.org/10.1016/j.future.2015.09.021>

Dey, S., Mukherjee, A., Paul, H. S., & Pal, A. (2013). Challenges of Using Edge Devices in IoT Computation Grids. In *2013 International Conference on Parallel and Distributed Systems* (pp. 564–569). Seoul, Korea (South): IEEE. <https://doi.org/10.1109/ICPADS.2013.101>

Dimitrov, D. V. (2016). Medical Internet of Things and Big Data in Healthcare. *Healthcare Informatics Research*, *22*(3), 156. <https://doi.org/10.4258/hir.2016.22.3.156>

Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P. (2017). Blockchain for IoT security and privacy: The case study of a smart home. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (pp. 618–623).

- Kona, Big Island, HI, USA: IEEE.
<https://doi.org/10.1109/PERCOMW.2017.7917634>
- El murabet, A., Abtoy, A., Touhafi, A., & Tahiri, A. (2018). Ambient Assisted living system's models and architectures: A survey of the state of the art. *Journal of King Saud University - Computer and Information Sciences*.
<https://doi.org/10.1016/j.jksuci.2018.04.009>
- Hasan, R., Hossain, M. M., & Khan, R. (2015). Aura: An IoT Based Cloud Infrastructure for Localized Mobile Computation Outsourcing. In *2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering* (pp. 183–188). San Francisco, CA, USA: IEEE.
<https://doi.org/10.1109/MobileCloud.2015.37>
- Mukherjee, M., Matam, R., Shu, L., Maglaras, L., Ferrag, M. A., Choudhury, N., & Kumar, V. (2017). Security and Privacy in Fog Computing: Challenges. *IEEE Access*, 5, 19293–19304.
<https://doi.org/10.1109/ACCESS.2017.2749422>
- Shanzhi Chen, Hui Xu, Dake Liu, Bo Hu, & Hucheng Wang. (2014). A Vision of IoT: Applications, Challenges, and Opportunities With China Perspective. *IEEE Internet of Things Journal*, 1(4), 349–359.
<https://doi.org/10.1109/JIOT.2014.2337336>
- Soursos, S., Zarko, I. P., Zwickl, P., Gojmerac, I., Bianchi, G., & Carrozzo, G. (2016). Towards the cross-domain interoperability of IoT platforms. In *2016 European Conference on Networks and Communications (EuCNC)* (pp. 398–402). Athens, Greece: IEEE.
<https://doi.org/10.1109/EuCNC.2016.7561070>
- Stergiou, C., Psannis, K. E., Kim, B.-G., & Gupta, B. (2018). Secure integration of IoT and Cloud Computing. *Future Generation Computer Systems*, 78, 964–975.
<https://doi.org/10.1016/j.future.2016.11.031>
- Suciu, G., Halunga, S., Vulpe, A., & Suciu, V. (2013). Generic platform for IoT and cloud computing interoperability study. In *International Symposium on Signals, Circuits and Systems ISSCS2013* (pp. 1–4). Iasi, Romania: IEEE.
<https://doi.org/10.1109/ISSCS.2013.6651222>
- Yaqoob, I., Ahmed, E., Hashem, I. A. T., Ahmed, A. I. A., Gani, A., Imran, M., & Guizani, M. (2017). Internet of Things Architecture: Recent Advances, Taxonomy, Requirements, and Open Challenges. *IEEE Wireless Communications*, 24(3), 10–16.
<https://doi.org/10.1109/MWC.2017.1600421>
- Stantchev, V., Barnawi, A., Ghulam, S., Schubert, J., & Tamm, G. (2015). Smart items, fog and cloud computing as enablers of servitization in healthcare. *Sensors & Transducers*, 185(2), 121.
- Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.
- Avram, M. G. (2014). Advantages and challenges of adopting cloud computing from an enterprise perspective. *Procedia Technology*, 12, 529-534.
- Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare informatics research*, 22(3), 156-163.
- Weyrich, M., & Ebert, C. (2016). Reference architectures for the internet of things. *IEEE Software*, (1), 112-116.
- Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2014, August). On the integration of cloud computing and internet of things. In *2014 International Conference on Future Internet of Things and Cloud* (pp. 23-30). IEEE.
- Khan, J. Y., & Yuce, M. R. (2010). Wireless body area network (WBAN) for medical applications. In *New developments in biomedical engineering*. InTechOpen.
- Banafa, A. (2016). IoT standardization and implementation challenges. *IEEE. org Newsletter*.

Derhami, V., Khodadadian, E., Ghasemzadeh, M., & Bidoki, A. M. Z. (2013). Applying reinforcement learning for web pages ranking algorithms. *Applied Soft Computing*, 13(4), 1686-1692.

Multi-Objective Resource Scheduling for QoS-Based Cloud Computing

Hilaire NKUNZIMANA¹, Michele MUKESHIMANA¹, Jérémie NDIKUMAGENGE¹

¹ *Department of Information and Communication Technology, University of Burundi, Bujumbura City, BURUNDI*

¹ *michele.mukeshimana@ub.edu.bi, ¹hilairecn@gmail.com, ¹jeremie.ndikumagenge@ub.edu.bi*

Abstract— Cloud computing represents the new computing paradigm. Due to its essence and nature it has brought some challenges such as security and quality of service. In order to ensure good quality of service, one should carefully manage computing resources available in the cloud. In this paper, we propose a cloud computational resource scheduling algorithm based on a genetic algorithm. First, the genetic algorithm is improved and applied to solve the problem of virtual machine allocation in cloud computing. We describe the modelling process for chromosomes and genes that correspond to the virtual machine scheduling problems. This scheduling algorithm aims at optimal placement of virtual machines in order to make full usage of the CPU and hence curb the waste time and energy of idle machines.

Keywords— scheduling, genetic algorithm, virtual machine, quality of service

I. INTRODUCTION

Cloud computing can be defined as combination of leasing techniques of computer resources through Internet. In this endeavour, the cloud computer service provider must ensure quality of service imposed on him through service level agreement (SLA).

On the other hand, the service provider main concern is to reduce the overall running cost while delivering such quality of service. Therefore, reduction of energy consumption through adequate computer resources management within the cloud remains the main concern for the service provider while at the same time granting the best of its service.

This situation puts forward two antagonistic types of quality of service: reduction of energy consumption and high level of quality of service.

As a matter of fact, most of the time, to ensure high quality of service to the users requires many physical machines powered on which incurs high energy consumption on the provider side.

In order to put forward the best algorithm aimed at minimizing the energy cost, one need to understand first the underlying parameters of cloud computing service. We all know that providing services whether being software service or storage service, they all require that the servers must be powered on resulting in big amount of energy to be consumed. The question is to identify which parameter shall we act on so that we can lower down the energy consumption without affecting the service level agreement. Therefore, using multiple metrics provides a more complete picture of the state

of the total physical resources and provides to service providers with more analytics possibilities and a more suitable configuration solution. In addition, multi-criteria optimization can be used to analyse the effects of parameters relative to each other. This is to estimate how each parameter affects optimization, analyse the confrontation caused by the combined use of these indicators, and verify their relevance.

These differences are not of no importance, as it is not necessarily easy to select a set of QoS parameters that have two correlations for analysing system functions, and their optimization together with other parameters leads to a satisfactory compromise. Another interesting advantage of simultaneously analysing various QoS cloud parameters can reveal attributes that have not previously been proven to be a bulimia algorithm whose intrinsic behaviour does not allow direct optimization of the analysed metrics.

This section explains why and how to use the proposed QoS model in a multi-standard optimization approach. The parameters and defined metrics are designed to provide a broader, more accurate analysis of the issues highlighted in cloud computing efforts. Finding the optimal solution to an investment problem is not limited by polynomial time. The general case described in [3] is closer to the optimal solution than the specific case in [9,10] to obtain a satisfactory solution. In addition to the chosen quality of service parameters, virtual machine reconfiguration and DVFS are also part of the scheduling problem in this manuscript, which clearly enhances the latter's complexity. The contributions of this paper are summarized below:

- We define a multi-criteria enhancement method for QoS constraints in cloud computing.
- We introduce the GA-based scheduling algorithms into the cloud to place virtual machines and implement two different versions.

II. MULTI-CRITERIA ENHANCEMENT METHOD FOR QOS CLOUD COMPUTING

Four QoS parameters were chosen from all types. Energy consumption that allows for environmental problems to be taken into account, the response time in order to have a pure performance measure, robustness that reassures service providers and users about the probability of being reallocated due to a system failure, and the dynamism that ensures a certain reserve of performance in the event of a traffic peak. To measure and evaluate these metrics in the two approaches proposed in this section, some modifications have to be made

to the definitions of the metrics. The metrics, as used by the placement algorithms, are described below:

1) *Response time*: The response time metric is assimilated to the execution time of each physical machine. Thus, the Response Time (in seconds) is calculated as a function of the capacity (in MIPS) of each virtual machine and the number of instructions (in Million Instructions) that they Modelling must perform.

$$T_{h,k}(rep) = \min(NbrInstr w_{v,h}(f_{CPU}) \quad (1)$$

where NbrInstr is the number of instructions to be executed by the virtual machine V. This metric must be minimized in order to minimize the Response Time.

2) *Energy consumption*: Energy consumption is calculated by using the power formula given in the definition of the Energy Cost Parameter $P(F_i) = (P(F_i) - Pidle(F_i)) + Pidle(F_i)$.

To obtain the exact consumption value for each physical machine, this is calculated as a function of the power variations delivered by each physical machine over time (depending on the execution time of virtual machines) and the total execution time

The energy consumption metric (in Watt) is given by :

$$E = N_h \sum_{i=1} P(F_i) h_i \times T_{h,k}(rep) \quad (2)$$

This metric (total time of execution) must be minimized in order to minimize energy consumption

III. GENETIC ALGORITHM

Genetic Algorithm (GA) is an optimized metaheuristic algorithm that mimics natural evolution. Genetic algorithms use a group of individuals called a group, where each individual (also called a chromosome) represents a solution to the problem to be solved. The basic principle of genetic algorithms is to evolve this initial population over several generations to obtain a population that contains chromosomes that are better than at the beginning. From generation to generation, specific operators (genetic operators) are applied to everyone to explore new possible solutions. In each generation, after applying these genetic operators that generate new individuals, all chromosomes (initial and new) are classified according to their fitness. After calculating the objective function, each individual is assigned as core (fitness value). Each individual is evaluated based on this value, and only those with a fixed value are selected to be good enough to include those chromosomes in the group as a working group for the next generation population. Thus, in each generation, the operator's operation to create a new chromosome tends to cause the population to include individuals representing better solutions. Therefore, the genetic algorithm is a meta-heuristic algorithm that is applicable to several types of problems, and the population of the solution space becomes a possible operator by random application, and the population is evolved from the beginning. It is important to remember that the way genetic algorithms

perform search solutions is applied to explore the solution space as much as possible, but does not guarantee the quality of the solution found (for optimal solutions). In this section, the proposed genetic algorithm is dedicated to solving the problem of virtual machine allocation in the cloud. Considering that the solution to the problem to be solved is the result of multi-objective optimization, considering each selected QoS metric, calculating the finite value of each chromosome makes it possible to judge the quality of the solution. Modelling of chromosomes and genes corresponding to the problem of scheduling virtual machines on a set of physical machines is described in the following sections.

A. Modelling

A Genetic algorithm is defined as a meta-heuristic that can be adapted to many types of problems. A clear description of the model chosen in the context of this paper is therefore necessary.

- A chromosome is a virtual machine placement solution
- A generator represents a virtual machine.
- The value assigned to a generator is the number of the physical machine on which the virtual machine has been allocated.
- In parallel, the characteristics of virtual machines and physical machines are saved in order to be able to calculate the values of each of the metrics as well as the value of fitness of the chromosomes. Every chromosome is composed of N genes. Every gene is composed of a virtual machine and number of physical machine on which it is being allocated to.

1) *Operators*: one of the basic principles of the algorithm is to randomly apply operators on the chromosomes of a group to form new operators. The best people are protected as part of the next generation and the start the process again. The operators used and the operators described below are three typical operators of genetic algorithms:

A mutation operator applied to a fixed number of chromosomes randomly selects a gene and changes its value. This change in the gene value means that the virtual machine represented by the gene has been assigned to another physical machine. Therefore, the new chromosome obtained represents a new investment solution and is integrated into the current population. Crossover operators are also applied to a fixed number of chromosomes, inverting portions of two chromosomes, resulting in two new solutions. Use two intersections to do this. The operators of the elections can reduce the number of individuals in the population and temporarily expand the two operators by performing the above-mentioned operations to maintain the best condition and restore the population to its original size.

2) *Validation of chromosomes*: An important process of genetic algorithms is to ensure that the solution found after the application of the operator is an effective solution. In other words, the solution found must respect the constraints of the problem model to be solved. For the genetic algorithm used here, it includes verifying that each solution meets the

maximum utilization of each resource (CPU and memory). If the verification of the chromosome does not comply with these constraints and is therefore invalid, it is simply removed from the current population and a new individual is generated.

The termination of a genetic algorithm can be determined in two different ways: Define an improved threshold that allows the best generation of the current generation to be compared to the best generation of the previous generation. If the difference between the two chromosomes is less than the defined threshold, the GA stops, considering that the improvement between the two successive generations is not sufficient to make it interesting to continue the process. After the fixed algebra in the algorithm studied here, the first solution cannot be used. In fact, the chromosomes were evaluated as normalized values as described in the next section. This depends on the average of each indicator (calculated for all individuals in the population). These averages must be different between generations. Therefore, chromosome fitness values are not comparable between successive generations. This is why the second solution is adopted.

3) *Fitness metrics and values* Each metric calculation gives a value within the range that is essentially related to the relevant metric. In order to be able to correctly calculate the value of the objective function, involving a set of metrics whose values are not included in the same interval, the normalization of each of these metrics must be applied. This includes calculating a value called vstd normalization for each metric. Therefore, these values are comparable and then they can be added or subtracted between them in order to obtain a standardized fitness value. The normalization method used here is the Center-Reduced method, which gives a set of values with an average of 0 and a variance of 1.

$V_{std} = \frac{(v - \mu)}{\sigma}$ where V is the value of the metric to be normalized, μ is the mean of the metric over the whole population, and σ is the standard deviation. Thus, the standardized value of fitness is equal to a linear formula integrating all of these standardized values. This makes it possible to accurately and precisely compare each chromosome belonging to a generation as a function of their fitness value. The objective function used in this genetic algorithm is as follows:

$$F_{obj} = \alpha_1 E + \alpha_2 RespT + \alpha_3 Rob + \alpha_4 Dyn \quad (3)$$

where α_1 , α_2 , α_3 , and α_4 coefficients of the energy (E), the response time (RespT), the robustness (Rob), and the dynamics (Dyn). These coefficients can be modified (increased or decreased) to optimize one or more metrics. In this equation, the objective function, energy, response time, and robustness are metrics to be minimized, contrary to the dynamism metric, which must be maximized.

The first version of the genetic algorithm involves only the allocation of basic independent services, each of which is implemented in a virtual machine. Therefore, each virtual machine is a basic service with a departure date of $t = 0$. The solution to the problem involves looking up the investment in

the set virtual machine on the available physical machines while optimizing the QoS metrics

Obviously, this version of the genetic algorithm solves the very simplified service allocation problem compared to real cloud services. However, this simplified version of the service allocation allows for an optimization that focuses on the differences in selected QoS metrics. In fact, using this genetic algorithm as a heterogeneous investment tool can evaluate the quality of the latter's optimization, and can also prove the impact of one QoS on another, and then emphasize the importance of using multi-standard methods for service allocation. It is this version of the genetic algorithm used for other work

IV. EXPERIMENTS AND RESULTS

After modelling the quality of service based parameters, we used cloud sim to act on different variables and observe variations.

The result confirmed that evolutionary algorithm based solutions can reduce gradually the energy consumption in the cloud computing while granting at the same time the best quality of service.

The modelling helped us also to understand the correlations between each parameter and its effect on energy consumption and inspired us on how to deal with each parameter with bearing in mind our primary objective which is reducing the energy consumption.

V. CONCLUSION

Cloud computing is the new computing paradigm and is likely to be adopted by many businesses in the years to come. As this new of computing is greedy in terms of energy consumption, it is wise to think of the best deployment style of cloud computing aimed at keeping minimum the energy cost of the cloud service providers.

This is however must be done while ensuring that the service level agreement to the users is not compromised.

The study of the parameters supporting the quality of service in the cloud and their relevance to the energy consumption is worth to be studied and the present paper is in line with that objective.

As the study problem is an optimization issue, genetic algorithm is more indicated to deal with problems of this kind and the paper tried to harness the algorithm in order to reduce the energy consumption. The problem is a multi-objective optimization as multiple variables will be acted upon in order to reach the desired optimization.

REFERENCES

- [1] Hady S Abdelsalam, Kurt Maly, Ravi Mukkamala, Mohammad Zubair et David Kaminsky. "Analysis of energy efficiency in clouds". in Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, 2009. COMPUTATION WORLD'09. Computation World: IEEE. 2009, p. 416-421 (cf. p. 40).
- [2] B. Aksanli, J. Venkatesh et T.S. Rosing. "Using Datacenter Simulation to Evaluate Green Energy Integration". In Computer 45.9 (2012), p. 56-64 (cf. p. 49).

- [3] M. Alexandru, T. Monteil, P. Lorenz, F. Coccetti et H. Aubert. "Efficient Large Electromagnetic Problem Solving by Hybrid TLM and Modal Approach on Grid Computing". In: International Microwave Symposium, Montréal, Canada, 17-22 June 2012. 2012 (cf. p. 124).
- [4] Phillip E Allen et Douglas R Holberg. CMOS analog circuit design. Oxford Univ. Press, 2002 (cf. p. 34).
- [5] Cloud Security Alliance. Security Guidance for Critical Areas of Focus in Cloud Computing V3.0. <https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf> (cf. p. 73).
- [6] Amazon AWS. <http://www.aws-ug.fr/> (cf. p. 5).
- [7] Amazon S3. <http://aws.amazon.com/fr/s3/> (cf. p. 6).
- [8] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica et Matei Zaharia. "A view of cloud computing". In Commun. ACM 53.4 (avr. 2010), p. 50–58 (cf. p. 3, 59).
- [9] Remzi H. Arpaci-Dusseau et Andrea C. Arpaci-Dusseau. Operating Systems: Three Easy Pieces. Blackboard Books, 2012 (cf. p. 21).
- [10] Daniel Ashlock. Evolutionary computation for modeling and optimization. T. 103. Springer, 2006 (cf. p. 31).
- [11] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson et Dawn Song. "Provable data possession at untrusted stores". In Proceedings of the 14th ACM conference on Computer and communications security. ACM. 2007, p. 598–609 (cf. p. 70).
- [12] Nader Azizi et Saeed Zolfaghari. "Adaptive Temperature Control for Simulated Annealing: A Comparative Study". In Comput. Oper. Res. 31.14 (d'éc. 2004), p. 2439–2451 (cf. p. 30).
- [13] Slawomir Bak, Marcin Krystek, Krzysztof Kurowski, Ariel Oleksiak, Wojciech Piatek et Jan Waglarz. "GSSIMA tool for distributed computing experiments." In Scientific Programming 19.4 (2011), p. 231–251 (cf. p. 47)

Meter Reading Android Application for Water Utility Companies

Emmanuel Nyirenda¹ and Christopher Chembe²

*Mulungushi University, Department of Computer Science and Information Technology,
Box 80415, Kabwe, Zambia.*

¹*enyirenda37@gmail.com, ²cchembe@mu.edu.zm.*

Abstract

Water utility companies have a billing process which involves reading the meter and knowing the consumption of its customers. The major challenge faced by water utility companies in Zambia is inaccurate customer billing. Affecting hugely this inaccurate customer billing is the manual meter reading process. However, meter reading is a very complex task for water utility companies in Zambia. The current method for meter reading uses manual work of reading the meter and updating details of the customers. The readings are maintained in a record book and later entered onto the computer. This whole process creates the possibilities of human errors and a burden on the meter reader. The manual meter reading process has a lot of challenges like lack of analysis of readings, low productivity of meter readers, time consuming and loss of revenue by the water utility company. The other problem with manual meter reading is that both the customer and the water utility company have challenges in monitoring and analyzing of consumption usage. Although technologically advanced solutions exist, most water utility companies in Zambia have no budget for them. The proposed model comprised of the android mobile application for capturing readings in the field. A usability test was undertaken consisting of 36 respondents which comprised of 75% meter readers and 25% customers. Three user experience areas were used to undertake the survey namely, perceived ease-of-use, perceived flexibility and perceived information accuracy. The results obtained showed that the capturing, uploading, monitoring and analysis of readings were very good for the users and customers who participated in the study. Using the android application customers were able to take their own readings and view graphical usage at any time anywhere.

Keywords: Algorithms, Android, Android Application, Mobile Application.

I. Introduction

Most water utilities service providers in Zambia use manual process for taking meter readings. The readings are maintained in a record book or meter card. Customers have to wait for the meter reader to read and update the meter card for them to know their consumption monthly. This results into customers and the water utility not being able to monitor consumptions regularly. The manual meter reading process is prone to errors, time consuming and meter readers manipulate figures which result into customers complaining of inaccurate billing. This hugely affects water utility company NRW, Billing and collection efficiency. In [9], they explained that manual meter

reading process causes extra work load on person who is capturing meter readings

To solve this problem an android application was proposed. The android based meter reading application is used by both the meter reader and customers to take readings and monitor consumption usage graphically. At the start of the meter reading cycle, a meter reader goes in the field visiting different customer properties. When taking readings at a property, a meter reader logs onto the application, search for the property by name, meter number, house number or account number, then enters the current reading from the meter, and when saving the application opens the camera and forces the meter reader to capture the image of the meter. The readings are not saved if the picture is not taken. The GPS coordinates for the property are automatically picked by the application in the background. The captured meter readings are saved on the phone and can be uploaded to the server at any time. Thus this process reduces manual work, incorrect billing and provides proof of visiting the customer property. The objective of the study was to develop an android application for capturing, analysing and monitoring meter readings.

I. Literature Review

The major challenge faced by water utilities in Zambia is inaccurate customer billing caused by manual meter reading process. This section brings out some of the earlier used approaches to solve the problem at hand.

This study [3] explained that bill generation and meter reading is a very complex process in today scenario. The existing method of billing process uses manual work of taking reading and users are not allowed to track their consumption each second. In [4] they developed a method based on IoT platform which was more efficient and was used for keeping track of consumption. Users were able to get the reading on android any time. Also, a client was able to view the graphical usage for a particular duration. According to [2] it was explained that these days most of utility service provider companies use manual process for taking meter readings. Those readings are maintained on the meter card. To address this problem, an android application was built essentially to track the water usage and accessing the consumption data resulted in better management of water.

In [5], they stated that the maximum part of the billing process is a manual process which include calculating the bill and writing meter reading in the book. Due to manual

procedure incorrectness in bill generation occurred. The procedure created a lot of burden on meter reader and was also inconvenient for the customer to lodge any complaints. In this paper they suggested an Android application which was carried by meter reader and a web application for customer to interact with the utility company. Using the Android application meter reader captured the image and uploaded to the server. The server performed the pre-processing on image and OCR extracted Meter reading and Customer number.

This paper [6] introduced a system based on image processing to obtain efficiently and accurate reading of the digital meter. In this system the back camera of the mobile phone is used to acquire the image of the meter. The system then applies a sequence of image processing functions to automatically extract and recognize the digits of the meter reading image.

In [7], they introduced an automatic meter reading concept (AMR) which collected the readings automatically and then the system transfers them to the central database for billing.

Table 1 Comparison of meter reading technologies

Approach	Analysis of readings	Support for conventional meters	Gap	Reference
Meter Reading Application for Android Based Handheld Device.	YES	NO	OCR IS COMPLEX AND TAKES TIME TO PROCESS	(Mayur, Waje, 2016)
Automatic meter reading based on image processing	NO	NO	NO ANALYSIS AND MONITORING	(Elrefaie et al., 2015)
Android Based Meter Reading Using OCR	NO	YES	NO ANALYSIS AND MONITORING	(Dayama et al., 2014)

Android App for Meter Reading	NO	YES	NO ANALYSIS AND MONITORING	(Kotwal et al., 2015)
-------------------------------	----	-----	----------------------------	-----------------------

II. Existing System

Water utility companies use manual process for taking readings which are maintained in a record book or meter card. Customers have to wait for the meter reader to read their meter for them to know their monthly usage. Manual meter reading process is difficult to prove if a meter reader visited a property and it is time consuming as the readings in the book have to be entered into the billing system. Although technologically advanced solutions exist, water utility companies cannot invest a huge amount of money for a new advanced solution e.g. prepaid meters. According to customer point of view there is no facility provided by the water utility company of knowing their current consumption units.

III. Proposed System

The proposed system is to develop an android application that will be used by the water utility company and its customers to monitor consumption graphical usage. Customers can take their own readings and view their consumption history usage graphically on the phone. Water utility company is also able to monitor meter readers productivity and analyse consumption for its customers. The application is able to get the GPS coordinates, photo of the meter and time when the reading was captured.

IV. System Architecture

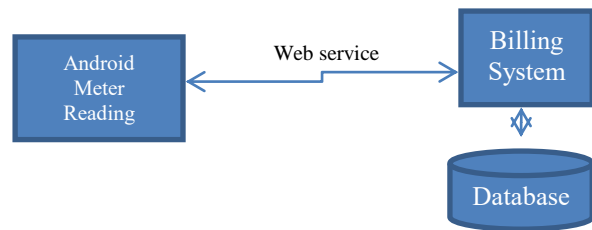


Fig 1: System Architecture

V. Research Methodology

The methodology adopted for this research is design science on the basis that it is particularly suited to the task of solving the manual meter reading problems more efficiently or more effectively. The process model employed is the Design Science Research Process (DSRP). According to [8] this

process has been widely utilised within information systems research. The organisational context is that of water utility companies in Zambia.

Process model for the research

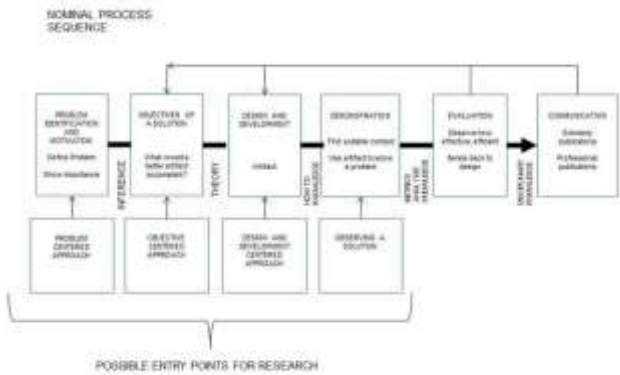


Fig 2 Design Science Research Process (DSRP) model after [8]

Applying the methodology to the research problem

This research has its focus on the problem of manual meter reading process for water utility companies in Zambia. The research is therefore based on a ‘problem centered approach’ and as such the entry point in the DSRP is at the first activity.

Problem Identification and Motivation

For the problem identification of this study refer to section I paragraph 1.

Objectives

For the objectives of this study refer to section I paragraph 2.

Design and Development

The Design and Development stage of the Pfeffers et al DSRP method has been adopted for this research. Below are the research questions.

1. How can meter readings for water utility company be captured using a mobile device and sent to the server for further analysis by both water utility company meter readers and its Customers?
2. How can water utility company and its customers analyze and monitor customer meter readings using an android application?
3. How can we assess and evaluate a meter reading mobile application?

Model Design

Model design will usually involve some data collection to inform the parameters for the model. This Research covered the case of Zambian water utility manual meter reading process.

Model Creation

The model built comprises of the following components:

1. Algorithm for Initialising Readings
2. Algorithm for capturing image and meter reading
3. Algorithm for Uploading Readings and Images

Demonstration

The DSRP followed in this research requires that the Demonstration activity involves the artefact to be used in some appropriate environment to solve the stated problem. The Demonstration activity was undertaken in order to identify any obvious errors/omissions.

Model Verification

Verification involved running and testing whether or not the artifact was working as it should and involved running several iterations of tests until the algorithms were able to produce usable results such as are represented in figure 3 to figure 6.

Android Application



Fig 3 Zonal summary report



Fig 4 Capturing readings



Fig 5 Readings list

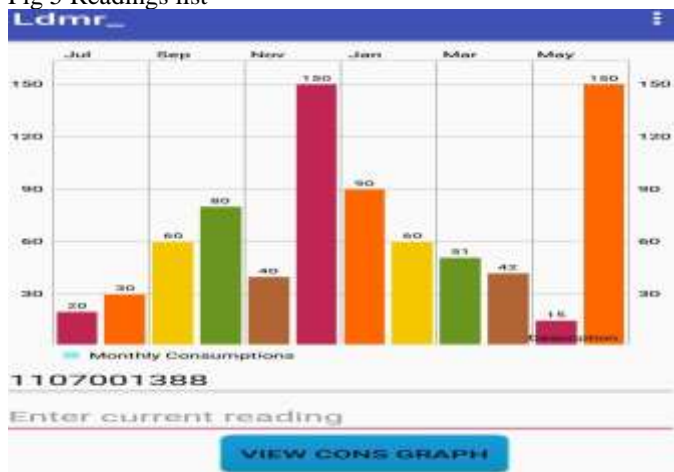


Fig 6 Customer monthly consumption

Evaluation

The usability evaluation was conducted over a period of 2 to 3 months and in 9 different districts. The users were allowed the freedom to perform the tasks first during the test phase and then during the pilot phase. In order to validate our model, users included a combination of different stakeholders among them customers and meter readers accessed the mobile application and completed an after-test questionnaire.

VI. Implementation

Heuristic for Uploading Readings and Images

- Step 1: Start
- Step 2: Declare variables
Readings, count, db, requestQueue, session Object;
- Step 3: Initialise variables and Read
db=new DatabaseHandler (this);
requestQueue=volley.newRequestQueue (this);
- Step 4: Create the post request (Android Volley)
- Step 5: For loop to loop through all the accounts
Loop
For (reading rd: db.getAllByUser (user) {
Save reading to the server

- Step 6: get the image to upload
- Step 7: if (img.Exists) {
- Step 8: Convert image to string
Save image to server
- Step 9: } else {
- No photo exists
}
- Step 10: Display message “photos/readings uploaded or there was an error”;

Heuristic for Initializing Readings

- Step 1: Start
- Step 2: Declare variables
Readings, count, db, requestQueue, sessionObject, URL;
- Step 3: Initialise variables and Read
db=new DatabaseHandler (this);
requestQueue=volley.newRequestQueue (this);
sessionObject=user.getUsername ();
URL=url.getUrl (sessionObject);
- Step 4: Create get request (Android Volley)
- Step 5: while loop to loop through the response
While (count<response.length ()) {
Get the accounts from server by sessionObject
Create the readings object and save the object onto the android phone database.
}
- Step 6: Display message “readings initialized or error occurred”;

Heuristic for Capturing Readings

- Step 1: Start
- Step 2: Declare variables
Readings, count, db, sessionObject, Camera, Image, GPS, button;
- Step 3: Initialise variables and Read
db=new DatabaseHandler (this);
image=image.createImageFolder ();
SessionObject=user.getUsername ();
Camera=camera.getRequest ();
- Step 4: button.click () {
- Step 5: Open the camera and get a photo of the meter
- Step 6: If (photo.IsSaved ()) {
Get data from form and save to readings object.
Display message “readings saved successfully/error occurred”;
- Step 7: } else {
Don't save the readings data
}
- Step 8: Go back to readings list and select another account to capture.

Usability Test

Usability testing is an important aspect of software development whose objective is to enhance user experience benefits, outlining choices and the limiting of the failure and oversights for users [9]. In order to achieve the goal of good usability, systems require building a suitable design. The outcome is that the users get important, viable, effective and satisfying application consequently [10].

VII. Results

The demographic information is as follows: Purposive sampling was used and we received 36 responses while conducting our study and were distributed as shown in (Figure 8: Responses by district). The response from target users of the usability test consisted of 75% meter readers and 25% customers as shown in (Figure 9: Responses by user type).

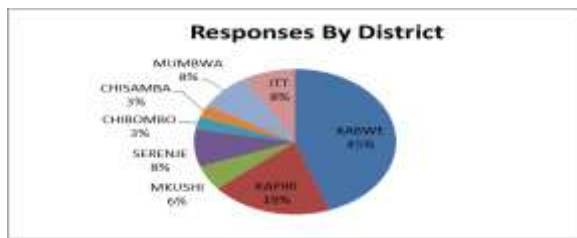


Fig 3 Responses by district

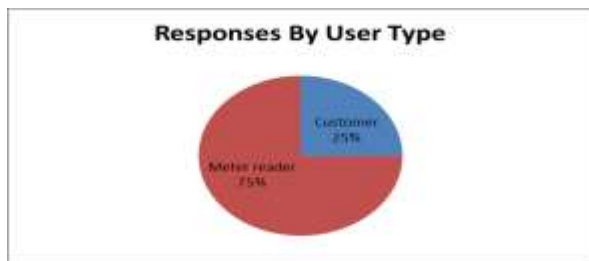


Fig 4 Responses by user type

Usability factors

The usability test was split into three (3) main areas namely: perceived ease-of-use (Ease-of-use), perceived flexibility (Flexibility) and perceived information accuracy (Information Accuracy) as shown in Table 2: Usability factors.

Table 2

Usability factors

Usability factor	Variables measured	Reference
Ease of use	Accessibility, Reliability and Availability	[9] [10]
Flexibility	Easy To Capture Readings, Easy To Update Readings, Easy To Search Properties, Easy To Upload and Easy To Analyse.	[10][10]
Information	Information clear ,	[10]

Accuracy	Information accuracy and	
----------	--------------------------	--

Perceived Ease-of-use

Most of the users found the mobile application to be very good, easy to access, highly reliable and available with the information needed for management decision (Figure 10:Ease of-use).

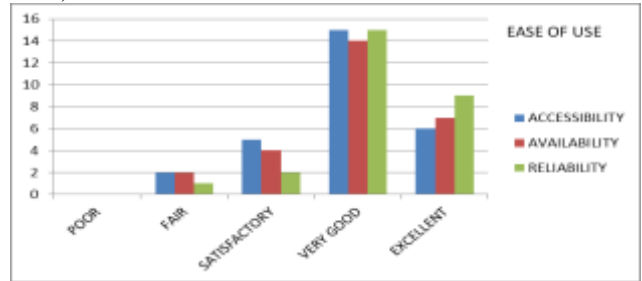


Fig 5 Ease of use

As shown above, majority of the participants agreed that the application was easy for them to use.

Perceived Flexibility

Majority of the users indicated they were happy with the flexibility of the mobile application for meter readings (Figure 11: Flexibility). This means that users were satisfied with the way the mobile application was designed.

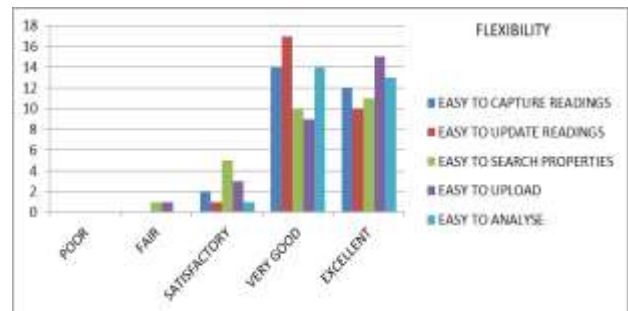


Fig 6 Flexibility

Perceived Information Accuracy

Majority of the users indicated they agreed that the mobile application for meter readings provided very clear and accurate information as shown in (Figure 12: Information Accuracy).

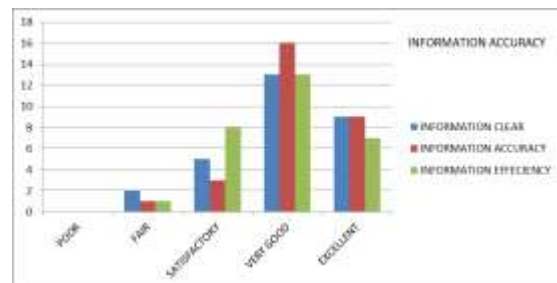


Fig 7 Information accuracy

VIII. Discussion

The main focus of this study was to develop a mobile application for meter reading for water utility companies in Zambia so as to improve efficiency and service delivery in the water sector. This study reviewed several other solutions that have been developed over time in the quest to find solutions for the manual meter reading problem and our literature review show a summary of our findings, which further fuelled the relevance of this study into finding a solution that could work in the context of Zambian water utility sector. This research designed a low cost solution for the manual meter reading problem customized to the local environment and easily integrates with the existing infrastructure for all water utility companies in Zambia. The water business requires regular monitoring of customers as they can engage in illegalities such as illegal reconnections.

With this application the meter reader is able to monitor customers properties in his/her area regularly when picking readings and provide proof of visitation through GPS and picture of the property or meter. The application can help water utilities in Zambia bill correctly, monitor staff productivity, increase customer satisfaction and easily resolve complaints as there is proof (GPS and Photo) for every reading taken.

IX. Conclusion

In this research, we have presented the problem of manual meter reading which is faced by all water utility companies in Zambia. The mobile application designed is able to provide proof of visiting the customer property through GPS coordinates, picture of the meter and time when the property was visited. In addition supervisors are able to monitor customer readings in real time and customers are also able to view their consumption graphical usage at any time anywhere in the comfort of their home. This study also evaluated the mobile application on three user experience main areas namely, perceived ease-of-use, perceived flexibility and perceived information accuracy. The results showed that the application provided very good user experience.

References

- [1] J. Docampo, S. Ramos, G. L. Taboada, R. R. Exposito, J. Tourino, and R. Doallo, "Evaluation of Java for general purpose GPU computing," in *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, 2013, pp. 1398–1404.
- [2] Mayur , Waje, "Meter Reading Application for Android Based Handheld Device," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. Volume 5, 2016.
- [3] R. Dayama, A. Chatla, H. Shaikh, and M. Kulkarni, "Android based meter reading using OCR," *Int J Comput Sci Mob Comput*, vol. 3, pp. 536–539, 2014.
- [4] P. Arora, M. Sharma, H. Jindal, and A. Mittal, "Efficient Way for Tracking Electricity Consumption with Meter Pulse Reader (MPR) Algorithm," in *International Conference on Smart Trends for Information Technology and Computer Communications*, 2016, pp. 331–339.
- [5] J. Kotwal, S. Pawar, S. P. M. Khopade, and P. Mahalunkar, "Android app for meter reading," *Int. J. Eng. Comput. Sci.*, vol. 4, no. 01, 2015.
- [6] L. A. Elrefaei, A. Bajaber, S. Natheir, N. AbuSanab, and M. Bazi, "Automatic electricity meter reading based on image processing," in *Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on*, 2015, pp. 1–5.
- [7] M. V. Shinde and P. W. Kulkarni, "Camera click energy meter reading system," in *2015 International Conference on Pervasive Computing (ICPC)*, 2015, pp. 1–4.
- [8] K. Pfeffers *et al.*, "The design science research process: A model for producing and presenting information systems research," in *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006), Claremont, CA, USA, 2006*, pp. 83–106.
- [9] J. C. Bastien, "Usability testing: a review of some methodological and technical aspects of the method," *Int. J. Med. Inf.*, vol. 79, no. 4, pp. e18–e23, 2010.
- [10] A. Hussain, E. O. Mkpojiogu, J. Musa, and S. Mortada, "A user experience evaluation of amazon kindle mobile application," in *AIP conference proceedings*, 2017, vol. 1891, p. 020060.

Using Cryptography to Secure Android based Mobile Application Systems

Michael Bwalya¹ and Christopher Chembe²

Mulungushi University, Department of Computer Science and Information Technology,
Box 80415, Kabwe, Zambia.

1. mikob87@gmail.com, 2. cchembe@mu.edu.zm

Abstract-Currently mobile applications are playing major role in many areas such as banking, social networking, financial apps, entertainment and many more. With this increasing number of applications, security is an important issue and thus focus should be given to the security. Mobile applications need to be assessed and ensure that secure coding practices have been followed during development. Data that is shared on an unsecured channel is vulnerable to attacks and to stop unauthorized access to this data, there is need to encrypt the data before it is sent to the server. In this research work, different cryptographic algorithms for encrypting data and secure data sharing in mobile applications across communications channels were examined. Simulation methodology was used to investigate a suitable cryptographic algorithm and to design a security framework for mobile applications to solve mobile application security problems. The proposed framework employs the use of Advanced Encryption Standard (AES) algorithm for encrypting meter readings data being exchanged between a smart phone and the server. The results obtained from the simulation of the security framework, showed that the four fields namely: Account number, image path, meter number and phone number on which AES encryption was applied were in an unreadable format (ciphertext), implying that the fields have been successfully encrypted. This solution allows application users to transfer (upload readings) data between a smart phone and database server in a secure manner without facing the problem of data attack. Data being uploaded to the server is encrypted before it is transferred and decrypted once it reaches the server side. This solution addresses android application security in the application and network communications layers and data transmission. The research paper ensures information security is guaranteed between an organisation and its customers.

Keywords: Encryption Algorithm; Cryptography; Information security; Mobile application; Attacks; Advanced Encryption Standard.

I. Introduction

Smartphones sales have achieved high number of records worldwide and going on to suppress other electronic technologies like notebooks and tablets[1]. Devices that use android have attracted huge market share due to the open architecture of Android, and the popularity of its Application Programming Interface (APIs) in the developer community. The expanded popularity of the Android devices and associated monetary benefits attracted the malware developers, resulting in big rise of the Android based security

issues [2]. The directness of these new environments will lead to new applications and markets and will enable greater interfacing with existing online services. However, as the importance of the data and services of smartphones increases, so too do the opportunities for vulnerability. It is important that this next generation of platforms provide a comprehensive and usable security infrastructure [3].

In view of the above, this research involves developing a security framework that employs the use of cryptography in a mobile application system. Cryptography is the technique and science of keeping messages secure [4]. Cryptography is an integral part of modern world information security making the virtual world a safer place [5] and it is the best way for encryption [6]. Cryptography converts the original data into ciphertext which is an unreadable form (Encryption) and again converts the ciphertext into original form (Decryption) [7]. The two widely accepted and used cryptographic methods are 1) Symmetric-key algorithms 2) Asymmetric-key algorithms. Symmetric algorithms utilize the same key for encryption and decryption. This is termed as secret key. By using the same key, messages are encrypted by the sender and decrypted by the receiver. It contains algorithms like Data Encryption Standard (DES), Advanced Encryption Standard (AES), Triple DES, Blowfish etc. Asymmetric algorithms use different keys. Public key is used for encryption and other private key is used for decryption. This is known as public key. The public key is known to the public and the private key is known to the user. It comprises various algorithms like Rivest, Shamir, & Adleman (RSA), Digital Signature Algorithm (DSA), Elliptic Curve(EC), Diffi-Hillman(DH) [8] [7].

The proposed solution uses symmetric algorithm as it has the speed and computational efficiency to handle encryption of large volumes of data. AES is the commonly used encryption technique nowadays, this algorithm is based on many substitutions, permutations and linear transformations, each executed on data blocks of 16 byte. As of today, there are no practicable attacks that exist against AES. Therefore, AES remains the preferred encryption standard for governments, banks and high security systems around the world [7], [9], [5].

This solution allows data to be encrypted before it is sent and decrypts the data once it reaches the server. In this regard, the solution addresses the concerns in android application security in the application and network communications layers and data transmission thereby achieving security goals

of data namely: Confidentiality, Integrity and Availability (CIA).

This study will potentially identify critical success factors for Companies planning to start using mobile technologies to protect the integrity, confidentiality and availability of information in today’s highly networked systems environment.

II. Literature Review

This section discusses the concepts and approaches that laid the groundwork on which the final conclusions and recommendations of the study were anchored.

Purpose of Cryptography

According to [10], Cryptography provides a number of security goals to ensure the privacy of data, non-alteration of data and so on. Due to the great security advantages of cryptography it is widely used today. Following are the various goals of cryptography [11], [12].

- i. Confidentiality: Information in computer is transmitted and has to be accessed only by the authorized party and not by anyone else.
- ii. Authentication: The information received by any system has to check the identity of the sender that whether the information is arriving from an authorized person or a false identity.
- iii. Integrity: Only the authorized party is allowed to modify the transmitted information. No one in between the sender and receiver are allowed to alter the given message.
- iv. Non Repudiation: Ensures that neither the sender, nor the receiver of message should be able to deny the transmission.
- v. Access Control: Only the authorized parties are able to access the given information.
- vi. Service Reliability and Availability: Since secure systems usually get attacked by intruders, which may affect their availability and type of service to their users. Such systems provide a way to grant their users the quality of service they expect.

Comparisons of different Cryptographic Algorithms

A comparative study of encryption algorithm is presented in Table 1. Encryption algorithms looked at are AES, DES, 3DES, RSA [6]. They are presented into fourteen factors, which are Algorithm type, Encryption, Decryption, Key length, Block size, No. of rounds, key used, Key for Ciphering and deciphering, Algorithm for Ciphering and deciphering, Algorithm scalability, Security, Vulnerability, Speed of simulation, Hardware and Software Implementation.

Table 1 Comparison of encryption algorithms

s/n	FACTOR	AES	DES	3DES	RSA
1	Algorithm type	Symmetric	Symmetric	Symmetric	Asymmetric
2	Encryption	Fast	Moderate	Slower	Slower

3	Decryption	Fast	Moderate	Slower	Slower
4	Key length in bits	128,192 & 256	56	56,112 or 168	1024 or greater
5	Block size in bits	128	64	64	512 or more
6	No. of rounds	10,12 & 14	16	16	1
7	Key used	Same for both encode and decode data	Same for both encrypt and decrypt data	Three different keys to encrypt, decrypt and again encrypt	Different keys to encrypt and decrypt data
8	Key for Ciphering and deciphering	Same key	Same key	Three different keys	Different
9	Algorithm for Ciphering and deciphering	Different Algorithm	Different Algorithm	Different algorithm	Same algorithm
10	Algorithm scalability	Not scalable	Scalable	Scalable	Not scalable
11	Security	Higher security	Not secure	Effective Security	Less security
12	Vulnerability	BFA (Brute Forced Attack)	BFA, Linear & differential cryptanalysis attack	BFA, sweet32 attack	BFA and Oracle attack
13	Speed of simulation	Faster	Faster	Faster	Faster
14	H/w and S/w Implementation	Fast implementation	As compared to software better in hardware implementation	Efficient in hardware but not in software	Not efficient

Related Works

In a study on comparative analysis of encryption algorithms for data communication, The study by [13], analyzed the performance of encryption algorithm considering the following parameters like Computation Time, Memory usage and Output Bytes. The simulation results showed the comparison of three algorithms which are AES, DES and RSA using same text file for five experiments, output byte for AES and DES is same for different sizes of files. It was noticed that the RSA has very smaller output byte compare to AES and DES algorithm. The time taken by RSA encryption algorithm is higher compared to the time taken by AES and DES algorithm. Based on the text files used and the simulation results obtained, it was concluded that DES algorithm consumes least encryption time and AES algorithm has least memory usage while encryption time is very minor in case of AES algorithm and DES algorithm. RSA takes the longest encryption time and memory usage is also very high but output byte is least in case of RSA algorithm.

It was concluded in [14] that AES is faster and more efficient than other encryption algorithms. There is insignificant difference in performance of different symmetric

key schemes when the transmission of data is considered. Even in scenarios involving data transfer it would be advisable to use AES in case the encrypted data is stored at the other end and decrypted multiple times. Another study by [15] discussed the principal advantages of AES with respect to DES, as well as its limitations. They indicated that AES can be quite comfortably implemented in high level or low level languages.

III. Research Methodology

This section is devoted to a discussion of methodological aspects including simulation approaches that were adopted and implemented in the study.

Computer simulation research involves running the executable simulation model and conducting many virtual experiments by varying the parameter space. The resulting simulated data can then be compared with both theoretical and empirical assessments. This is the first step in generating new research insights that will help improve existing theories and eventually create new ones [16]. According to [17], the purpose of simulation is to better understand an interesting phenomenon, such as the technology diffusion process, in order to predict variables of interest. In the same vein, [18] indicated that simulation allows modelers and problem owners to gain better understandings of complex systems as it allows modelers to assess the problems from different dimensions.

A model should be developed for a specific purpose or use and be as simple as possible but should meet its purpose. A simulation model is a structural model, which implies the model contains logical and causal relationships that occur in the systems. Developing a valid simulation model is an iterative process where several versions of a model are developed prior to obtaining a valid model [19]. Figure 1 below shows several steps involved in simulation method.

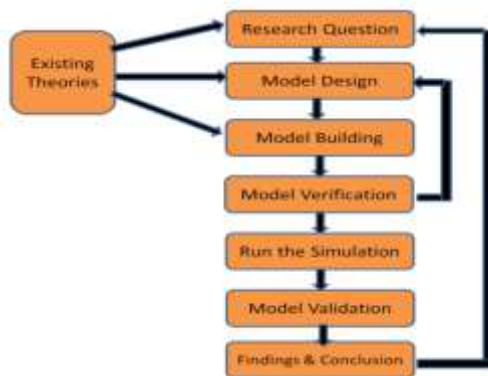


Figure 1: Simulation Steps

In order to validate our model, the following setup was built in which fiddler was installed on a windows machine to act as Man-in-the Middle to capture traffic and intercept data being transferred from the mobile application. The mobile app was installed on an Android phone.

Step by step procedure is given below:

- i. Install fiddler on the windows machine and connect that machine to DSL modem.
- ii. Connect the Android phone with an installed application to the same DSL modem.
- iii. Configure the proxy settings on the Android Phone with network settings on the windows machine.
- iv. Start fiddler on windows machine. Now all traffic of the android phone is routed through windows machine.
- v. Launch the mobile application from the android phone to upload meter readings and communicate with the server. The application included both the encrypted and decrypted version.
- vi. Traffic was captured on windows machine and analyzed.

These steps were carried out with the setup shown in figure 2 below:



Figure 2: Network setup using fiddler as a proxy on windows machine

The following observations were made from the captured traffic.

- I. All the meter reading details were showing when the unencrypted mobile application was installed on the application and used to upload readings.
- II. All the data in the fields (account number, image path, meter number and phone number) that have been encrypted using AES were in an unreadable format (ciphertext).

IV. Results

In this Section, we present the results generated from the simulation that were carried out. Figure 3 shows the results obtained from fiddler after capturing and uploading readings using the unencrypted mobile application. It can be seen that all the data in the fields is in plain text and readable.

On the other hand, Figure 4 clearly shows the results obtained when the readings are captured and uploaded using the encrypted mobile application. AES encryption has been applied on four fields namely: Account number, image path, meter number and phone number. It can be seen that data in these fields is in an unreadable format (cyphertext).

Figure 5 shows unencrypted meter reading data generated from a web API, this data consist of four different customer details. Figure 6 shows the same data as in figure 5, but this time the meter reading data (account number, image path, meter number and phone number) that is generated from a web API is encrypted. This is achieved when encrypted mobile application system is used.

In order to apply a suitable cryptography algorithm to an application, it is important to have knowledge regarding performance, strength and weakness of the algorithms. In efforts trying to find the suitable cryptographic algorithms, several different tests were carried out using a C# program to compare 3DES and AES in terms of encryption time and decryption time. The algorithms were implemented in C# using Visual Studio IDE. Files of sizes 50kb, 1mb, 5mb and 10mb consisting of text were used as input for encryption and decryption. For sake of comparison same input files were used for all algorithms throughout the experiment. All block cipher algorithms were set in the same mode ECB which is the default. The results obtained from the experiment are shown in table 2 and 3 below, additionally; these results have been represented using a line chart and a bar chart to clearly show that AES takes less encryption and decryption time. Also shown are simplified algorithms for encrypting and decrypting using java.

```

{
  "Sid": 1,
  "Account": "1100000054",
  "Current": 259,
  "Previous": 250,
  "Remarks": "Prepaid",
  "Meter": "C-ILC-3045",
  "StreetName": "GHANA AVENUE",
  "StreetNo": "14",
  "Dept": "KABWE",
  "ConsCode": "Chomba12",
  "Ward": "2019-06-19T11:27:39",
  "Suburb": "TOWN AREA",
  "Name": "LUBINDA LISHOMWA",
  "Latitude": -14.43679,
  "Longitude": 28.44828,
  "ImagePath": "/MeterImages/1100000054.jpeg",
  "Meter": "C-ILC-3045",
  "Name": "LUBINDA LISHOMWA",
  "Previous": 250,
  "Remarks": "Prepaid",
  "StreetName": "GHANA AVENUE",
  "StreetNo": "14",
  "Suburb": "TOWN AREA",
  "Ward": "2019-06-19T11:27:39"
}
    
```

Figure 3: Result generated unencrypted meter readings.

```

{
  "Sid": 1,
  "Account": "uM82Qv9UGDE8Hk+y9U31A==",
  "ConsCode": "Chomba12",
  "Current": 259,
  "Dept": "KABWE",
  "ImagePath": "+3FNn/QnU44gdKocPLVT0x@fPwhGJOW2R5VhdU30c=",
  "Latitude": -14.43679,
  "Longitude": 28.44828,
  "Meter": "Ave6eVF709GFuzvDz8jxg==",
  "Name": "LUBINDA LISHOMWA",
  "Previous": 250,
  "Remarks": "Prepaid",
  "StreetName": "GHANA AVENUE",
  "StreetNo": "14",
  "Suburb": "9VmsAJVCHBzFo9ubBqFA==",
  "Ward": "2019-06-19T11:27:39"
}
    
```

Figure 4: Result generated encrypted meter readings.

```

{"Sid": "1", "Account": "1100000054", "Current": 259, "Previous": 250, "Remarks": "Prepaid", "Meter": "C-ILC-3045", "StreetName": "GHANA AVENUE", "StreetNo": "14", "Dept": "KABWE", "ConsCode": "Chomba12", "Ward": "2019-06-19T11:27:39", "Suburb": "TOWN AREA", "Name": "LUBINDA LISHOMWA", "Latitude": -14.43679, "Longitude": 28.44828, "ImagePath": "/MeterImages/1100000054.jpeg"},
{"Sid": "2", "Account": "1100000055", "Current": 918, "Previous": 918, "Remarks": "Borehole", "Meter": "97-012587", "StreetName": "JAMESON AVENUE", "StreetNo": "64", "Dept": "KABWE", "ConsCode": "Chomba12", "Ward": "2019-06-20T09:32:36", "Suburb": "TOWN AREA", "Name": "JEFF RUTH", "Latitude": -14.43894, "Longitude": 28.45165, "ImagePath": "/MeterImages/1100000055.jpeg"},
{"Sid": "3", "Account": "1100000787", "Current": 7783, "Previous": 7652, "Remarks": "Active", "Meter": "3683", "StreetName": "MULLINGUSHI ROAD", "StreetNo": "2", "Dept": "KABWE", "ConsCode": "Chomba12", "Ward": "2019-06-21T12:41:34", "Suburb": "TOWN AREA", "Name": "LIFE TRUST", "Latitude": -14.43803, "Longitude": 28.43945, "ImagePath": "/MeterImages/1100000787.jpeg"},
{"Sid": "4", "Account": "1100000788", "Current": 5250, "Previous": 5230, "Remarks": "Active", "Meter": "97-012587", "StreetName": "PICKARD CRESCENT", "StreetNo": "1", "Dept": "KABWE", "ConsCode": "Chomba12", "Ward": "2019-06-23T09:47:15", "Suburb": "TOWN AREA", "Name": "DESALIM", "Latitude": -14.4405683, "Longitude": 28.43945, "ImagePath": "/MeterImages/1100000788.jpeg"}
    
```

Figure 5: Unencrypted Meter readings generated from web API data

```

{"Sid": "1", "Account": "uM82Qv9UGDE8Hk+y9U31A==", "StreetName": "GHANA AVENUE", "StreetNo": "14", "Dept": "KABWE", "ConsCode": "Chomba12", "Ward": "2019-06-19T11:27:39", "Suburb": "9VmsAJVCHBzFo9ubBqFA==", "Name": "LUBINDA LISHOMWA", "Latitude": -14.43679, "Longitude": 28.44828, "ImagePath": "/MNoQdI44pKocPLVT0x@fPwhGJOW2R5VhdU30c="},
{"Sid": "2", "Account": "AbDHM2N0xCI1Mc08v8w==", "StreetName": "JAMESON AVENUE", "StreetNo": "64", "Dept": "KABWE", "ConsCode": "Chomba12", "Ward": "2019-06-20T09:32:36", "Suburb": "9VmsAJVCHBzFo9ubBqFA==", "Name": "JEFF RUTH", "Latitude": -14.43894, "Longitude": 28.45165, "ImagePath": "/MNoQdI44pKocPLVT0x@fPwhGJOW2R5VhdU30c="},
{"Sid": "3", "Account": "QJsd0ub0++5WRQPIKVA==", "StreetName": "MULLINGUSHI ROAD", "StreetNo": "2", "Dept": "KABWE", "ConsCode": "Chomba12", "Ward": "2019-06-21T12:41:34", "Suburb": "9VmsAJVCHBzFo9ubBqFA==", "Name": "LIFE TRUST", "Latitude": -14.43803, "Longitude": 28.43945, "ImagePath": "/MNoQdI44pKocPLVT0x@fPwhGJOW2R5VhdU30c="},
{"Sid": "4", "Account": "EzhdprXvryIEFVMS1Sg==", "StreetName": "PICKARD CRESCENT", "StreetNo": "1", "Dept": "KABWE", "ConsCode": "Chomba12", "Ward": "2019-06-23T09:47:15", "Suburb": "9VmsAJVCHBzFo9ubBqFA==", "Name": "DESALIM", "Latitude": -14.4405683, "Longitude": 28.43945, "ImagePath": "/MNoQdI44pKocPLVT0x@fPwhGJOW2R5VhdU30c="}
    
```

Figure 6: Encrypted Meter readings generated from web API data

Table 2 represents the different sizes of the text files and corresponding encryption and decryption execution time taken by 3DES algorithm in milliseconds for the windows operating system.

Table 2: 3DES Encryption and Decryption time in milliseconds for different data size in windows

Triple Data Encryption Standard (3DES)			
FILE SIZE	Start Time (ms)	End Time (ms)	Difference (ms)
10mb	345	800	455
5mb	562	789	227
1mb	463	615	152
50kb	123	235	112

Table 3 represents the different sizes of the text files and corresponding encryption and decryption execution time taken by AES algorithm in milliseconds for the windows operating system.

Table 3: AES Encryption and Decryption time in milliseconds for different data size in windows

Advanced Encryption Standard (AES)			
FILE SIZE	Start Time (ms)	End Time (ms)	Difference (ms)
10mb	364	512	148
5mb	423	588	165
1mb	392	575	183

50kb	268	468	200
------	-----	-----	-----

Figure 7 has been shown using a bar chart, this shows the encryption and decryption time taken by AES and 3DES for different sizes of text files. The results reveal that AES takes less encryption and decryption time on files of larger sizes as compared to 3DES.

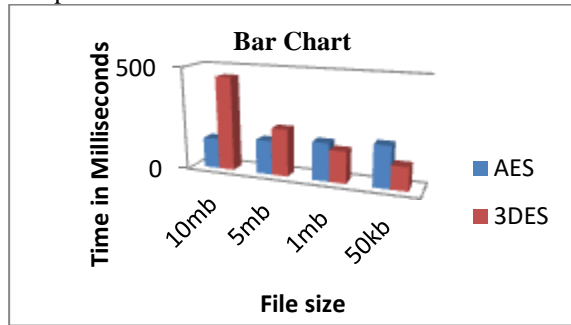


Figure 7: Bar chart showing Encryption and Decryption time for different file sizes.

Algorithm for Encrypting Java

```

Step 1: Start
Step 2: Declare variables
    IV, key, salt;
Step 3: Initialise variables
    IV = "IV_VALUE_16_BYTE";
    key = "PASSWORD_VALUE";
    SALT = "SALT_VALUE";
Step 4: Method : Encrypt (String raw)
    {
    Cipher c = getCipher(Cipher.ENCRYPT_MODE);
    byte[] encryptedVal = c.doFinal(getBytes(raw));
    String s = getString(Base64.encodeBase64(encryptedVal));
    Return s
    }
Step 5: Throws Exceptions
Step 6: Creating Cipher using AES CBC mode
    {
    cipher c = Cipher.getInstance("AES/CBC/PKCS5Padding");
    byte[] iv = getBytes(IV);
    c.init(mode, generateKey(), new IvParameterSpec(iv));
    return c;
    }
Step 7: Generate ciphertext and secret key
    
```

Algorithm for Decrypting Java

```

Step 1: Start
Step 2: Declare variables
    IV, key, salt;
Step 3: Initialize variables
    IV = "IV_VALUE_16_BYTE";
    key = "PASSWORD_VALUE";
    SALT = "SALT_VALUE";
Step 4: Method : Decrypt (String encrypted) throws Exception
    {
    byte[] decodedValue =
    Base64.decodeBase64(getBytes(encrypted));
    Cipher c = getCipher(Cipher.DECRYPT_MODE);
    byte[] decValue = c.doFinal(decodedValue);
    return new String(decValue);
    }
Step 5: Throws Exceptions
Step 6: Instantiating Cipher using CBC mode
Step 7: Generate plaintext using secret key
    
```

V. Discussion

This research employed the use of an established cryptographic algorithm in the development of a security framework for mobile application systems to protect data being transmitted, as this would improve a high degree of privacy and security of data. The symmetric cryptographic algorithm used is Advanced Encryption Standard (AES). According to [20], this technique is based on several substitutions, permutation and transformation. AES was preferred because of the strength that it possesses. Unlike 3DES and RSA, AES is strongly resistant to differential, truncated differential, linear, interpolation and Square attacks. Based on the text files used and the experimental result it was concluded that AES algorithm consumes less encryption and decryption time compared to 3DES. The encryption process of using 3DES is much longer than AES, because there is need to repeat the same encryption process three times in 3DES and this takes much time when compared to the AES encryption process which is much faster [20]. The encryption and decryption time is high in RSA. RSA being a public key cryptosystem uses one way function which is hard to invert using prime numbers. The use of modular exponentiation, multiplication inverse and two keys public and private key makes RSA slow compared to symmetric key algorithms. It is in this regard that AES was preferred.

This study designed a system that allows the users to transfer (upload readings) data between a smart phone and server database in a secure manner without facing the problem of data attack. Data being uploaded to the server is encrypted before it is transferred and decrypted once it reaches the server side, the idea behind this is to protect the data while being transmitted across communication networks, therefore, any unauthorized access to the data will be rendered useless since every bit of data being transferred is encrypted. As indicated by [21], the best method for securing the data being transmitted is to apply encryption before the transmission and uploading process of the data to ensure that the data remain protected in the servers against any authorized access at all times. This study further demonstrates that the use of cryptographic algorithms increases the level of security of encrypted mobile data.

This research would have been even more enlightening if the validation included even other mobile platforms like iOS, Windows, Symbian etc. However, with the results obtained from this research study, it is strongly felt that the developed framework can be used to protect and secure data in different mobile applications.

VI. Conclusion

The challenges of security and privacy of data being exchanged between mobile applications and servers were presented in this paper. For applications where the sensitive data is transmitted, security is very essential. Security of data is a key area of concern in the acceptance of mobile application systems. The designed security framework for mobile application addresses several challenges posed on

meter readings data such as unauthorized access to data, data manipulation and data leakages. These security flaws do not only affect the business and reputation of the company using a mobile application but it can also lead to loss of data, for example vulnerabilities in an app can allow an attacker to intercept data being transferred and if data is not encrypted manipulations can easily be made on the data. This study made use of meter readings data, from the captured meter readings that were uploaded using the encrypted mobile application, AES encryption was applied on four fields namely: Account number, image path, meter number and phone number. It was seen from the experimental results that data in these fields was in an unreadable format (ciphertext) indicating a successful encryption. The framework employed the use of AES cryptography which has a complex encryption process and has proven to be the best encryption method. The simulation results between 3DES and AES concluded that, AES is faster than 3DES in terms of execution time. By applying this encryption technique, this study made sure that the security goals of data Confidentiality, Integrity and Availability (CIA) were met.

Future Work

As a future work, the security framework will further be tested on different mobile platforms. Hybrid cryptographic algorithms will be utilized to efficiently secure the data and information which is transmitted through systems. Several simulations will be carried out to assess and mitigate the vulnerabilities that are exploited by attackers on different mobile platforms such as stealing users' confidential information without their consent.

References

- [1] A. Hayran, M. İğdeli, A. YILMAZ, and C. Gemci, "Security Evaluation of IOS and Android," *Int. J. Appl. Math. Electron. Comput.*, vol. 4, no. Special Issue-1, pp. 258–261, 2016.
- [2] P. Faruki *et al.*, "Android security: a survey of issues, malware penetration, and defenses," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 2, pp. 998–1022, 2015.
- [3] W. Enck, M. Ongtang, and P. McDaniel, "Understanding android security," *IEEE Secur. Priv.*, no. 1, pp. 50–57, 2009.
- [4] N. D. Jorstad and T. S. Landgrave, "Cryptographic algorithm metrics," in *20th National Information Systems Security Conference*, 1997, pp. 1–38.
- [5] P. Patil, P. Narayankar, D. G. Narayan, and S. M. Meena, "A comprehensive evaluation of cryptographic algorithms: DES, 3DES, AES, RSA and Blowfish," *Procedia Comput. Sci.*, vol. 78, pp. 617–624, 2016.
- [6] J. Kaur, S. Sharma, and M. Tech, "Secure image sharing on cloud using cryptographic algorithms: survey," *Int J Future Revolut Comput Sci Commun Eng*, vol. 4, no. 2, pp. 319–325, 2018.
- [7] V. R. Pancholi and B. P. Patel, "Enhancement of cloud computing security with secure data storage using AES," *Int. J. Innov. Res. Sci. Technol.*, vol. 2, no. 9, pp. 18–21, 2016.
- [8] Y. Kumar, R. Munjal, and H. Sharma, "Comparison of symmetric and asymmetric cryptography with existing vulnerabilities and countermeasures," *Int. J. Comput. Sci. Manag. Stud.*, vol. 11, no. 03, 2011.
- [9] M. N. A. Wahid, A. Ali, B. Esparham, and M. Marwan, "A Comparison of Cryptographic Algorithms: DES, 3DES, AES, RSA and Blowfish for Guessing Attacks Prevention," 2018.
- [10] E. Thambiraja, G. Ramesh, and D. R. Umarani, "A survey on various most common encryption techniques," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 2, no. 7, 2012.
- [11] J. Thakur and N. Kumar, "DES, AES and Blowfish: Symmetric key cryptography algorithms simulation based performance analysis," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 1, no. 2, pp. 6–12, 2011.
- [12] Adedeji Kazeem B. & Ponnle Akinlolu.A, "A New Hybrid Data Encryption and Decryption Technique to Enhance Data Security in Communication Networks: Algorithm Development.," *Int. J. Sci. Eng. Res.*, vol. 5, 2014.
- [13] S. M. Seth and R. Mishra, "Comparative analysis of encryption algorithms for data communication 1," 2011.
- [14] S. A. Hirani, "Energy consumption of encryption schemes in wireless devices," PhD Thesis, University of Pittsburgh, 2003.
- [15] N. Penchalaiah and R. Seshadri, "Effective Comparison and evaluation of DES and Rijndael Algorithm (AES)," *Int. J. Comput. Sci. Eng.*, vol. 2, no. 05, pp. 1641–1645, 2010.
- [16] M. Ihrig and K. G. Troitzsch, "An extended research framework for the simulation era," in *Proceedings of the Emerging M&S Applications in Industry & Academia/Modeling and Humanities Symposium*, 2013, p. 12.
- [17] N. Gilbert and K. Troitzsch, *Simulation for the social scientist*. McGraw-Hill Education (UK), 2005.
- [18] J. Zulkepli and T. Eldabi, "Towards a framework for conceptual model hybridization in healthcare," in *2015 Winter Simulation Conference (WSC)*, 2015, pp. 1597–1608.
- [19] R.G. Sargent, "AN INTRODUCTORY TUTORIAL ON VERIFICATION AND VALIDATION OF SIMULATION MODELS," *Proc. 2015 Winter Simul. Conf.*, 2015.
- [20] N. Aleisa, "A Comparison of the 3DES and AES Encryption Standards," *Int. J. Secur. Its Appl.*, vol. 9, no. 7, pp. 241–246, 2015.
- [21] Sakinah Ali, Sa Wail Abdo Ali Alhiage, and Farida Ridzuan, "Mobile Application Design for Protecting the Data in Cloud Using Enhanced Technique of Encryption," *Int. J. Eng. Technol.*, 2018.

The Adoption Footprints of Koha as a Library Management System (LMS) in Higher Education Institutions (HEIs) in Zambia

¹Tuesday Bwalya, ²Akandelwa Akandelwa, ³Edward, C. Mwalimu

^{1,2,3}Department of Library and Information Science, University of Zambia, Lusaka, Zambia

¹bwalya.tuesday@unza.zm, ²akakandelwa@unza.zm, ³edward.mwalimu@unza.zm

Abstract- Koha is one of the widely used free and open source library management systems (FOSLMS) in libraries. Higher education institutions (HEIs) libraries in Zambia like any other libraries in other countries have been automating their operations using Koha. This study therefore thought to investigate the adoption footprints of Koha among HEIs in Zambia. The study was quantitative in design and cross-sectional survey was used. A complete census method was used in which all the 54 libraries that have automated their operations in HEIs were surveyed. The research findings show that 41 (76%) of HEIs surveyed have automated their libraries using Koha. The study further shows that, colleges of education and universities in Zambia have adopted Koha more than any other type of HEIs. It has been also established that public owned HEIs have adopted Koha more than the privately owned HEIs. The study has further shown that there is no association between the adoption of Koha in HEIs and demographic variables; there is no association between gender and adoption of Koha ($\chi^2=1.612$; $df=1$; $p=0.204$), between categorized age and Koha adoption ($\chi^2=2.614$; $df=1$; $p=0.181$), academic qualification of librarian and Koha adoption ($\chi^2=5.61$; $df=1$; $p=0.517$ and between work experience and adoption of Koha ($\chi^2=4.160$; $df=2$; $p=0.124$).

Keywords—component, formatting, style, styling, insert (key words)

I. INTRODUCTION

The birth of free and open source library management systems (FOSLMS) such as Koha has excited many libraries especially in developing countries that have not been able to purchase proprietary turnkey library management systems to automate their operations. As Mutula [1] observed that for many years, libraries in Sub-Saharan Africa remained the laggards as regards automation and modernization of library operations due to among other factors, budgetary constraints and high cost of ICT facilities. Like many other libraries in Sub-Saharan Africa, libraries in higher education institutions (HEIs) in Zambia have for long failed to automate their operations; automation in this sector has been largely limited to a few university libraries such as the University of Zambia and Copperbelt University libraries [2]. However, in the late 1990s, the library automation landscape in Zambia began to change with the coming of a free library system called Computerised Documentation Services/Integrated Set of Information Systems (CDS/ISIS) which was sponsored by UNESCO. CDS/ISIS was widely used in the late 1980s and 1990s by libraries in Sub-Saharan African countries such as Zambia [1]. The dawn of the 2000 decade saw the birth of robust FOSLMS such as Koha, Evergreen, ABCD and OpenBilio, thus, providing opportunities and choices for libraries in HEIs in Zambia to automate their operations.

In many countries, libraries have adopted Koha. According to Small Business Software Reviews [3], Koha was the best performing and widely used FOSLMS in 2018. Further, a study conducted by Adera [4] in Kenya in 2012 revealed that 67% of the surveyed libraries in Kenya were using Koha. Koha was started in New Zealand by Harowhenua Library Trust; a group of rural libraries in Harowhenua District of New Zealand because it became unsustainable for them to pay for commercial software [5]. Katipo Communications Limited developed Koha for Harowhenua Library Trust in 1999. Later in 2000, Koha was released under General Public License (GPL). According to Kummar and Jasimudeen [6], Koha is the first full-fledged integrated library management system to be developed; it has all modules needed for the operations of a library. Further, Koha conforms to cataloguing standards which include metadata standards such as MARC and ISO 2709 standard for information coding and exchange [7].

In a bid to accelerate the process of library automation using Koha in HEIs in Zambia, the Indian Government through its Indian Technical and Economic Cooperation (ITEC) has for many years been offering training to librarians from HEIs in Zambia and other developing countries on how to use Koha to automate library operations. Further, the International Network for the Availability of Scientific Publications (INASP) has also been training librarians from HEIs in Zambia on how to use Koha to automate their libraries [7]. At national level, between 2008 and 2012, the Ministry of General Education in partnership with the Flemish Association for Development Cooperation and Technical Assistance (VVOB), organized training workshops for librarians in all government colleges of education on how to use Koha in automating libraries operations [8]. The Department of Library and Information Science at the University of Zambia since 2013 also has been teaching library automation using Koha to undergraduate students in order to speed up the process of library automation in Zambia. It is against this background that a comprehensive study was conducted to establish the extent to which Koha was being used and characterize the libraries in HEIs that were using it in automating their operations.

A. Statement of the Problem

As already articulated, Koha is one of the free and open source library management systems available on the Internet; its availability has made it possible for the financially challenged libraries in Zambia to automate their operations. In addition, several organizations (both local and international) such as VLIR/UOS, ITEC and VVOB have been offering training to librarians in HEIs in Zambia on

how to automate their libraries using Koha [8]. As a result, of the availability of Koha and the provision of training to librarians in some HEIs in Zambia, it is expected that by now many libraries in HEIs have automated their libraries using Koha. It is however not known how wide spread is the adoption and use of Koha in libraries in HEIs in Zambia. This is because there has never been a study conducted to determine the adoption footprints of Koha. Furthermore, lack of research has made it impossible to know the types of HEIs that use Koha more than others in their library automation. It was therefore important that a nationwide research on the adoption footprints of Koha in HEIs in Zambia be conducted.

B. Objectives of the study

The study thought to investigate the use of Koha among HEIs in Zambia. In doing so, the study sought to: -

1. establish the extent to which libraries in HEIs use Koha,
2. determine the types of HEIs that use Koha more than others,
3. establish the library functions that have been automated using Koha,
4. determine if demographic variables such as age, gender, level of education and work experience of librarian influence the adoption of Koha in HEIs.

II. LITERATURE REVIEW

A. Koha adoption in some selected countries

Libraries in HEIs across the globe have continued to adopt Koha in a bid to do among other things; avoid the cost of commercial software. Koha is an award winning FOSLMS; it is used world-wide by all types of libraries. Koha, which is developed using open source technologies, has a huge group of users and developers thus, making it attractive to more libraries that wish to migrate from commercial library systems or embarking on library automation for the first time [7]. In Southern Africa and Zambia in particular, there is increased trail of evidence indicating that libraries in HEIs are adopting Koha. According to MESVTEE and VVOB [8], all public colleges of education librarians in Zambia had been trained in Koha library management system with a view to automating the operations of their libraries. As a result of the efforts of MESVTEE and VVOB, many government run colleges of education in Zambia have automated their libraries using Koha. In this regard, Charles Lwanga college of education in Zambia is reported to have pioneered the use of Koha in public colleges of education [8]. Koha is also reported by Marshall Breeding [9] to be used in Zambia by the Institute of Business Studies and Industrial Practice (ZIBSIP).

In Zimbabwe, Koha is reported to be used by many HEIs which include Theological College of Zimbabwe, Harare Theological college, Zimbabwe Christian College and Domboshawa Theological College [10]. Other users of Koha in Zimbabwe include Midlands State University and Mutare Teachers College. In Malawi, Koha is said to be used by Mzuzu University library, one of the public universities in Malawi [11].

The use of Koha is rife too in Kenya. According to Adera [4], Koha is the mostly used open source library management system in Kenya as 67% out of 200 libraries surveyed indicated using it (fig.1).

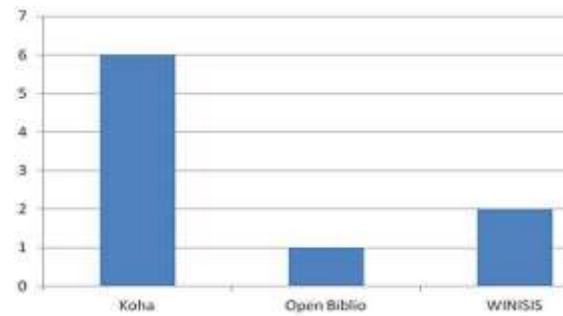


Fig. 1. Distribution of FOSLMS used in libraries in Kenya [4]

Other studies in Kenya have revealed that Koha is widely used in HEIs; which include Kenyatta University, Technical Institute of Kenya, Pwani University College, Technical University of Mombasa and Masinde Muliro University [9] and [12]. In Ethiopia, Koha has become popular; it has a huge user community among libraries in HEIs. These include Addis Ababa and Adma Universities [9]. Further, ETHIOKOH [13] shows that many government run universities use Koha more than other types of HEIs. These include Arba Minch University, Bahr Bar University, and Gonder University.

In Nigeria, Koha has quickly established itself as a choice of many libraries for automation; 24 (66.7%), out of the 36 academic libraries surveyed were found to be using Koha [14]. Further, Edem [15] reported that Koha was widely used software with 28 university libraries out of 58 universities surveyed were found to be using it.

In India, Koha is widely used; it was first installed in India at St Joseph’s College, Devvagiri in 2010. Since 2010, Koha has attracted small and big libraries in India [6]. Many prestigious library automation projects in India have adopted Koha due to its capability to handle Indian languages. Marshall Breeding [9] revealed that there were more than 199 libraries in India in 2017 that were using Koha. Further, Koha Community of users’ website as at 05/08/2017 revealed that there were 294 registered libraries in India that were using Koha. Among many HEIs that were reported to be using Koha in India include Alliance University, Indian Institute of Space Science and Technology, Madras Institute of Technology, Mahatma Gandhi University, Mysore University and South Asian University.

In United Kingdom (UK), Koha has been installed by PTFS Europe (a consultancy company) in more than 100 libraries by the November 12th, 2019 and that a large percentage of libraries were from HEIs [16]. Koha has been adopted by medium size academic libraries which include University off Hertfordshire, Blackpool and the Fylde College, University of Arts London, UK Wiltshire College, Belfast Metropolitan College and St Helens College. It is clear from this data that more privately owned colleges in UK use Koha.

Koha has also some adoption footprints in France; with French nationals such as Paul Poulain actively involved in its development since 2002. A quick check on Koha community on 11th November, 2019 revealed that many HEIs were using Koha in France. These include University Lyon3, University Rennes 2, and Aix-Marseille University [17].

Koha has continued to grow its popularity among small and medium academic libraries in the United States of

America (USA). According to Marshall Breeding [9], Koha continues to attract a diverse demographic of small to mid-sized academic, school and public libraries in the USA and its implementation is mainly supported by a private firm called ByWater Solutions which has several contracts with various libraries. ByWater Solutions provides consultancy services on Koha to over 225 libraries in 2018 [9]. HEIs that were reported to be using Koha in the USA by Marshall Breeding include Antioch University, Samuel Merritt University, Davis College and New York Universities of Health Sciences.

B. Theoretical framework

The study was guided by a theory called Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) which was developed by Thong and Xu in 2012. According to UTAUT2, there are seven (7) constructs that explain the adoption and use of technology by consumers; namely:-

1. performance expectancy- degree to which using new technology benefits consumers in performing certain activities,
2. effort expectancy- degree of ease associated with the use of a new technology by a consumer,
3. social influence- extent to which consumers perceive that other important people to them think that they should use the new technology,
4. facilitating conditions- consumers’ perceptions of the availability of resources and support to use the new technology,
5. hedonic motivation- the fun or pleasure derived from using technology,
6. price value- consumers' cognitive tradeoff between the perceived benefits of technology and the monetary cost of using it, and
7. experience and habit- the extent to which people tend to perform behaviors automatically because of learning.

However, Venkatesh, Thong and Xu, [18] observed that the seven factors are moderated by three (3) variables namely; age, gender and experience (figure 2).

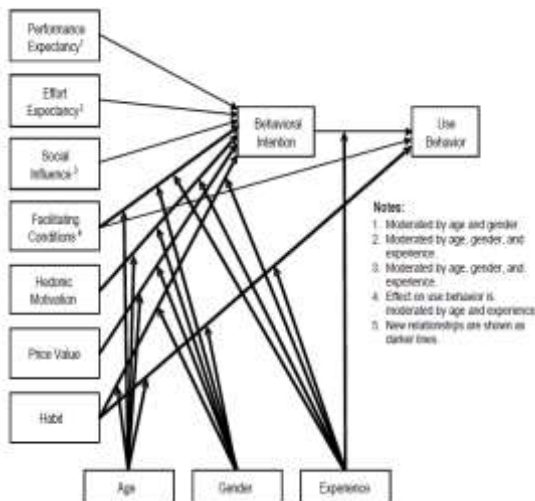


Fig. 2. UTAUT2 Research Model

Several studies have been conducted to demonstrate how demographic variables such as age, gender and experience influence the adoption of technology. Indrawati and Marhaeni [19] in their study of the adoption of instance

services applications in Indonesia observed that younger customers were interested in the performance expectancy of technology. Similarly, the Venkatesh, Thong and Xu [18] discovered in Hong Kong that younger consumers were likely to adopt technology if it improves their work. This implies that young librarians in HEIs in Zambia are more likely to adopt Koha compared to the older ones even if supporting conditions were not adequate in libraries.

Further, [19] observed in Indonesia that men were concerned with hedonic aspects of technology while women are concerned with the price value of technology. The other studies in Hong Kong discovered that old female Mobile Internet consumers were sensitive to price value of technology; they were seen to adopt technology which was perceived to be cheaper [18]. This implies that more female librarians are likely to adopt Koha as it costs less than commercial library management systems.

It was also observed that greater experience can lead to greater familiarity with technology and better knowledge structures to facilitate user learning which reduces dependence on external support [18]. AUTAUT2 has also underwritten the proposition that customers who have some experience with a new technology have higher probability of adopting it.

III. METHODOLOGY

The study was quantitative-descriptive in design; a cross-sectional survey method was employed. Data was collected at a single point in time from 54 HEIs libraries that have automated their operations. According to Bwalya [7] many HEIs in Zambia exist without libraries; only 142 were found to have functioning libraries and that only 54 had automated their operations. Considering the fact that the population (54) was small, all the 54 HEIs libraries were surveyed; a complete census method was employed. Chief librarians or any competent staff from the libraries was asked to answer a self-administered questionnaire. Further, telephone interviews were conducted for librarians whose libraries are found in far reaching areas of Zambia. The questionnaire was piloted on thirty (10) libraries in HEIs in order to perfect it.

The collected data was quantitatively analysed using software called Statistical Package for Social Science (SPSS) version 22 and descriptive statistics were used to present the findings. Further, non-parametric statistical tool (Chi-Square) was used to test whether a relationship exist between the four (4) demographic variables namely; gender, age, academic qualification, and work experience of librarian and the adoption of Koha. In this regard, the following hypotheses were tested using Chi-Square (Fisher’s Exact Test) at 0.05 Alpha, 95% level of significance.

TABLE I. RESEARCH HYPOTHESIS

Null Hypothesis	Alternative Hypothesis
H0: The categorized age of a librarian is independent of the use of Koha in HEIs in Zambia	H1: The categorized age of a librarian is not independent of the use of Koha in HEIs in Zambia
H0: The academic qualification of librarian is independent of the use of Koha in HEIs in Zambia	H1: The academic qualification of librarian is not independent of the use of Koha in HEIs in Zambia
H0: The gender of a librarian is independent of the use of Koha in HEIs in Zambia	H1: The gender of a librarian is not independent of the use of Koha in HEIs in Zambia
H0: The work experience of a librarian is independent of the use of Koha in HEIs in Zambia	H1: The work experience of a librarian is not independent of the use of Koha in HEIs in Zambia

The key five (5) variables in this study have been operationally defined as shown in table 2.

TABLE II. DEFINITION OF VARIABLES

Variable Name	Variable Definition	Measurement level
Adoption of Koha	The use of Koha in the management of library operations	Nominal
Age	The state of being young or old; below 35 years will mean young while above 35 years will simply be old [20]	Nominal
Education Qualification	Academic qualification obtained; categorised as Diploma and below and Bachelor's degree and above	Nominal
Gender	The state of being female or male	Nominal
Work experience	The categorised number of years one has been working in a library	Nominal

IV. RESEARCH FINDINGS

A. Characteristics of HEIs and respondents

As can be seen in fig.3, more universities (20) participated in the study; this is because many universities in Zambia have automated their operations

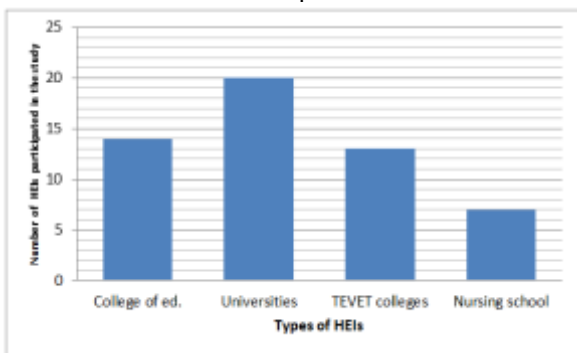


Fig. 3. Participation of HEIs in the study

Majority (61.1%) of the respondents were female while 38.9% were male. In terms of age, many (64.8%) were old, aged 35 years and above (table 3). Furthermore, majority (68.52%) of the participants were said to hold Bachelor's degree or better qualifications in Library and

Information Science (LIS). In terms of the number of years respondents had been working in the library, a considerable high percentage (37%) of respondents had been working in the library for a period of less than 3 years.

TABLE III. DEMOGRAPHIC CHARACTERISTICS OF RESPONDENTS

Variable	Value	Frequency	%
Gender	Male	21	38.9
	Female	33	61.1
Age	Young (15-35 years)	19	35.2
	Old (Above 36 years)	35	64.8
Qualification in LIS	Diploma and below	17	31.48
	Bachelor's degree and above	37	68.51
Work experience(yrs.)	Less than 3 years	20	37
	4-7 years	17	31.5
	Above 8 years	17	31.5

B. Adoption of Koha in HEIs

The respondents were asked to indicate whether or not they were using Koha as a library management system to automate the operations of their libraries; 41 (76%) of the libraries were reported to be using Koha while 13 (24%) were not using Koha but other library managements systems (fig. 4).

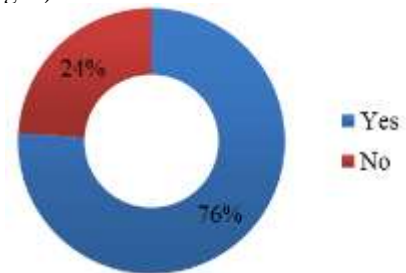


Fig. 4. Adoption and use of Koha by HEIs

The study has further shown that the use of Koha was more pronounced in colleges of education and universities respectively; all the 14 (34.14%) libraries surveyed in colleges of education were reported to be using Koha and 13 (31.70%) of university libraries indicated using Koha.

TABLE IV. THE USE OF KOHA BY TYPE OF HEIS

HEI		Adoption of Koha		Total
		Yes	No	
HEI	TEVET College	10 (24.39%)	3 (23.07%)	13(24.07%)
	College of Ed.	14 (34.14%)	0 (00.00%)	14 (25.92%)
	School of Nursing	5 (12.19%)	2 (15.38%)	7 (12.96%)
	University	13 (31.70%)	7 (53.84%)	20 (37.03%)
Total		41 (100%)	13 (100%)	54 (100%)

The study has further shown that the use of Koha is more pronounced in libraries in HEIs that are exclusively run by government; 20 (48.78%) of government run HEIs libraries

reported using Koha; 10 (24.39%) of quasi-government were found to be using Koha and 11 (26.82%) of private HEIs libraries reported using Koha (fig. 5).

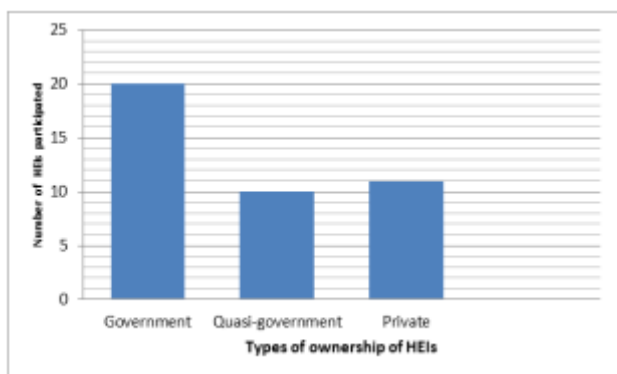


Fig. 5. Use of Koha by ownership of HEIs

The research findings further show that the use of Koha in HEIs in Zambia is on the upswing with more libraries adopting it. As depicted by the fig. 6 below, in 2009, Koha was being used by one (1) HEI; in 2010, two (2) libraries among the surveyed HEIs adopted Koha. However, 2011 did not register any HEIs adopting Koha. From 2012 to 2018, more libraries in HEIs adopted Koha and the peak was 2016 in which ten (10) HEIs adopted it.

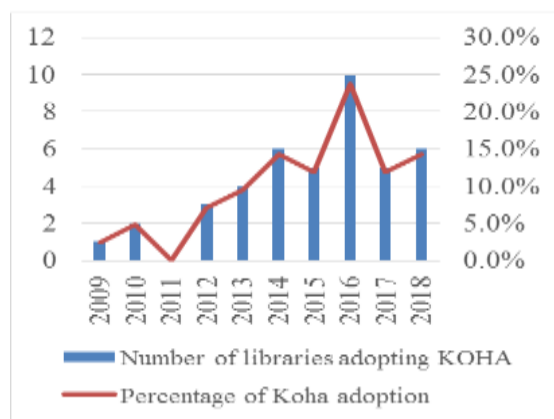


Fig. 6. Koha adoption from 2009-2018

The research has further shown that many library operations have been automated using Koha. Cataloguing functions, followed by circulation and report generation are some of the library functions that have been automated by many libraries (table 4).

TABLE V. LIBRARY ACTIVITIES THAT HAVE BEEN AUTOMATED WITH KOHA

Library activity automated	Yes	No	Total
Cataloguing	40 (97.56%)	1 (2.44%)	41 (100%)
Circulation	37 (90.24%)	4 (9.75%)	41(100%)
Report generation	33 (80.48%)	8 (19.52%)	41(100%)
Acquisition	14 (34.14%)	27 (65.85%)	41(100%)
Serial management	15 (36.58%)	26 (63.41%)	41(100%)
Budget tracking	9 (21.95%)	32 (78.04%)	41(100%)

C. Testing the association between demographic variables and the adoption of Koha

In this section, the researcher sought to determine if there is an association between the adoption of Koha and demographic variables such as categorized age, gender, education qualification and work experience of the librarians in HEIs. As you can see from table 5 below, in all the four tests, the P- values obtained are much higher than 0.05. This implies that all null hypotheses were not rejected.

TABLE VI. HYPOTHEIS TEST RESULTS

Variable	X2	df	P- value	Comment
Age	2.614	1	0.181	Not significant
Education	5.61	1	0.517	Not significant
Gender	1.612	1	0.204	Not significant
Work experience	4.160	2	0.124	Not significant

The above Fisher’s Exact tests show that there is no association between the adoption of Koha in HEIs and the demographic variables such as age, education qualification, gender and work experience of the librarians or library staff in charge of libraries in HEIs in Zambia. This entails that the adoption of Koha in HEIs is independent of the above demographic variables of librarians.

V. DISCUSSION OF RESEARCH FINDINGS

The presented research findings show that the use of Koha in HEIs in Zambia is widely spread; considering the fact that 41 (76%) out of the 54 HEIs surveyed indicated having automated their operations using Koha. This implies that only a small percentage (24%) of libraries in HEIs is using other library management systems. This is similar to the research findings by Adera [4] on the adoption of Koha in Kenya in which 67% of academic libraries surveyed were found to be using Koha. Further, the findings of this study are similar to the findings of the study conducted in Nigeria by Kari and Baro in 2015 among 36 academic libraries on the use of Koha for library automation in which the results showed many libraries were using Koha; (66.7%) of the surveyed libraries were found to be using Koha [14].

The findings have also shown that there is no association between the adoption of Koha and the four demographic variables namely; academic qualification, categorized age, gender and work experience (table 5). The adoption of Koha in HEIs in Zambia is mainly spurred by the fact that Koha is free and open source library management systems; its adoption cost is much lower than those of commercial library systems hence more libraries adopting it. External factors such economic reasons drive many libraries in developing countries such as Zambia to adopt open source solutions such as Koha not demographic factors of librarians responsible for the management of libraries [6].

As captured in table 5, Koha is more widely used among libraries from colleges of education; followed by universities. For example, 14 and 13 colleges of education and universities libraries respectively reported having automated their basic functions using Koha. This is contrary to the studies conducted by [12] and [9] on the use of Koha in Kenya where universities not colleges of education dominated the use of Koha in library automation. Similarly,

studies on the use of Koha in Ethiopia show universities leading other types of HEIs in the use of Koha [9]; [13].

A detailed analysis of the study findings on the type of HEIs that use Koha more than the others reveals a pattern where HEIs that are fully owned by government were dominating the use of Koha than other types of HEIs (fig.5). This pattern was observed in Ethiopia where majority of the HEIs that were found to be using Koha were run by government [13]. The high number of government run HEIs using Koha could be attributable to many reasons. These include the fact public HIEs are the most automated types of HEIs in Zambia hence using Koha more than others. Secondly, government or quasi-government colleges are privileged in Zambia to receive training support by cooperating partners on how to use free and open source library management systems such as Koha with a view to improving service delivery. For instance, for many years now, many librarians from government or quasi-government HEIs have been receiving scholarships from the Indian government through ITEC programme and the now defunct VLIR/UOS programme on how to use Koha and ABCD to automate library operations, hence many colleges of education that are publicly owned automating their operations using Koha than their counter parts which are privately owned.

As captured in figure 6, the peak of Koha adoption in HEIs in Zambia was in 2016; 10 HEIs adopted Koha in that year. The reduction now in the number of HEIs adopting Koha could be attributable to many factors; including the economic slowdown the country is facing. Prior to 2016, the Zambian economy was growing above 4% Gross Domestic Products (GDP) and government had money to spend on social sectors such as libraries. After, 2016, the economy has been shrinking and in 2019, the International Monetary Fund (IMF) reported that the Zambian economy would grow at 2% and that the debt stork for the country had risen to USD 18 billion [21]. This has squeezed public spending; budgets have been cut in many sectors of the economy including libraries. Even if Koha is free and open source library system, for a library to automate using it, there is need to acquire the necessary infrastructure such as servers. Further, in many cases, libraries need to engage a consultant to help install, customize Koha and train workers how to use it. Since there is little money in circulation, very few libraries are now implementing library automation in HEIs in Zambia.

V. CONCLUSION AND RECOMMENDATIONS

There is wide spread use of Koha among HEIs in Zambia; 76% of HEIs surveyed were found to be using Koha. The use is more pronounced in colleges of education followed by universities. Further, the adoption of Koha was high in 2016; 10 HEIs implemented library automation with Koha. The findings have also shown that there is no association between demographic variables such as academic qualification, categorized age, gender and work experience of librarians in HEIs; the adoption of Koha by more HEIs is attributable to external factors such economic factors. In view of findings above, the following could be recommended to stake holders in HEIs:-

1. Conduct further research on the possible challenges libraries in HEIs that have automated their operations with Koha are facing,

2. Allow librarians from privately owned HEIs libraries to benefit from international training scholarships by

ITEC and other cooperating partners on the use of Koha for library automation.

REFERENCES

- [1] Mutula, S.M. (2012) "Library automation in sub Saharan Africa: case study of the University of Botswana", *emeraldinsight*, vol. 46, issue: 3, pp.292-307. Available: <https://doi.org/10.1108/00330331211244832>
- [2] Nkhoma, B. P. (2003). "A Critical Analysis of Library Computerization at the Copperbelt University Library". *African Journal of Library, Archives and Information Science*, vol.13, issue: 2, pp. 133-153. Available: <https://www.africabib.org/rec.php?RID=269418164>
- [3] Small Business Software Reviews Services, Insight and Resources (2019, June.12). *Best Free and Open Source Library Management Software*. [Online]. Available: <https://cloudsmallbusinessservice.com/blog/best-free-and-open-source-library-management-software.html>
- [4] Adera, A.B (2013). "Feasibility of the adoption of open source ILS for libraries in Kenya: A practical evaluation". *emeraldinsight*, vol., 31 issue: 5, pp. 608 - 634. Available: www.emeraldinsight.com/0264-0473.htm.
- [5] Umahi, B. O. (2018, December). "Use of Koha Open Source Software and Library Automation in University Libraries in South-South, Nigeria". *International Journal of Scientific Research in Education*, Vol. 11, issue 6, pp.1118-1129. Available: [http://www.ij sre.com/assets/vol.%2C-11\(6\)-umahi-bassey.pdf](http://www.ij sre.com/assets/vol.%2C-11(6)-umahi-bassey.pdf).
- [6] Kumar,V.V and Jasimudeen. S (2012, December). "Adoption and user perceptions of Koha library management system in India". *Annals of Library and Information Studies*, vol. 59, pp. 223-230. Available: <https://pdfs.semanticscholar.org/bea1/1c0cfa85ec143fcf3436dd68434173236302.pdf>
- [7] Bwalya, T., "The use of free and open source library management systems (FOSLMS) in Higher Education Institutions (HEIs) in Zambia: Adoption footprints, opportunities and challenges". PhD dissertation, Department of Library and Information Science, The University of Zambia, Lusaka, 2019 [Unpublished]
- [8] MESVTEE and VVOB. (2013). *MESVTEE and VVOB Teacher Training Support Programme 2008-2013: Our Stories of Change*. [Online] Available: <http://www.vvob.org.zm>
- [9] Marshall Breeding. (2018). *Libraries.org: A directory of libraries throughout the world*. [Online] Available: <https://librarytechnology.org/diglib-displayindex.pl?>
- [10] Bibliotech. (2017). *Automate with Koha*. [Online] Available <https://www.bibliotechzw.com>.
- [11] EIFL. (2010). *Malawi Koha ILS pilot: The case study explores how Mzuzu University Library opted to pilot Koha*. [Online] Available: www.eifl.net/resources/malawi-koha-ils-pilot.
- [12] Makori, E.O and Mauti, N.O (2016). "Koha Enterprise Resource Planning System and its Potential Impact on Information Management Organizations". *Library Philosophy and Practice (e-journal)*. Online Available: https://digitalcommons.unl.edu/libphilprac/?utm_source=digitalcommons.unl.edu%2Flibphilprac%2F1378&utm_medium=PDF&utm_campaign=PDFCoverPages
- [13] ETHIOKOHA. (2011). *7-Day Advanced Training KOHA Library Management Software completed*. [Online] Available: <https://ethiokoha.wordpress.com/2011/06/27/7-day-advanced-training-koha-library-management-software-completed/>.
- [14] Kari, K.H and Baro, E.E., "The use of library software in Nigerian university libraries and challenges", *Library Hi Tech News* no. 3, pp. 15-20, 2014.
- [15] Edem, M. B. (2016). "Adoption of Software Packages in University Libraries in Nigeria". *Library Philosophy and Practice (e-journal)*. [Online] Available: <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=3654&context=libphilprac>
- [16] Calyx. (2019). *Who uses Koha?* [Online] Available: <https://calyx.net.an/services-and-solutions/koha>. Accessed on [31/12/2016]
- [17] Koha-community. org. (2019). *KohaUsers/Europe*. [Online] Available: <https://wiki.koha-community.org/wiki/KohaUsers/Europe>

- [18] Venkatesh, V., Thong, J.Y.L. and Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. [Online] Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2002388
- [19] Indrawati and Marhaeni, G. A.M. (2015). Predicting Instant Messenger Application Adoption Using a Unified Theory of Acceptance and Use of Technology. [Online] Available: <https://pdfs.semanticscholar.org/420d/a734875baa7e9d15d2f78ceca60f0649307e.pdf>
- [20] Central Statistical Office. (2017). 2017 Labour Force Survey Report. [Online] Available:
- [21] <https://www.zamstats.gov.zm/phocadownload/Labour/2017%20Labour%20Force%20Survey%20Report.pdf>
- [22] Bwalya, T. "Zambian needs a leadership that understands how the economy works". The Mast News Paper, August, 2019. [Online] Available: <https://www.themastonline.com/2019/08/16/zambia-needs-a-leadership-that-understands-how-the-economy-works/>

Development of a Mobile-based Geographical Information System for Managing Boreholes in Chilenje and Chalala Townships

Ruth Mulabita¹, Monica M. Kalumbilo²

^{1,2}Department of Computer Science, School of Natural Sciences, The University of Zambia

¹*rmchilufya@gmail.com*, ²*monde.Kalumbilo@cs.unza.zm*

Abstract—This study focuses on the development of a Web-based geographical information system for the management of boreholes in Chalala and Chilenje Townships. A mobile phone GPS facility was used to map the boreholes. This connects to the geographical information systems (GIS) database that stores all spatial data for the operational areas. 100 households in Chilenje and Chalala were visited for this survey. Another 20 from different occupations were also interviewed, this included employees from WARMA, Borehole Drillers and Farmers. The involvement of the residents in the registration process ensures ownership and responsibility of the all process. Further review of literature related to this study was carried out. The study proved that there were a number of challenges faced in the collecting of data on borehole information. These challenges can be overcome by developing the Web-based system and using data collection tools that send the information to the server. The Geo-ODK is open source software. The system is based on open source application and PostgreSQL is used for database management. The system has to be able to r5yconnect with other GIS application for future enhancements. There is need to enhance this system in future to enable it provide real time data. A Web-based Geographical information system if implemented will make it easier for all stakeholders to monitor the boreholes that are already drilled and the yet to be drilled boreholes.

Keywords – Boreholes, Information, Geographical Information System, Web-based

I. INTRODUCTION

Domestic Domestic and Commercial groundwater usage is inevitable in the absence of the right amount of water supply in terms of availability, quality, cost and pressure. This leads to people accessing ground water via the use of boreholes. They find boreholes to be self-sufficient save time and at the same time save money. In some areas, those who can afford for business purposes such as hand pumps, wells or solar pump drill boreholes. There are different mechanisms for

drilling these boreholes. The massive borehole drilling has a great effect on ground water. Water is in most cases contaminated as these boreholes are drilled right next to septic tanks. An improperly designed, located, constructed or maintained septic system can leak bacteria, viruses, household chemicals, and other contaminants in into the ground water causing serious problems. In order to control and manage the usage of ground water responsible authorities require the necessary type of information to support decision-making and access to the resource. In today’s age “information is power”, any individual or organisation requires information to be able to make an informed decision rather than guess work.

The aim of the study was to develop a web-based geographical information system for managing boreholes in Chilenje and Chalala townships. The system will enable the Water Resources and Management authority, capture details of all boreholes that have been sunk in the study area and the country at large. The Water Resources and Management Authority is an autonomous body established by the water resources management Act, No.21 of 2011[1]

The Water Resources and Management Authority have no way of providing updated information relating to borehole numbers, depths and locations in the country. With regard to domestic users, boreholes were drilled anyhow and without permission from any authority. Domestic users were free to drill boreholes in their premises without getting authorisation from the authorities. The commercial users were required to fill in a form and pay a fee of five kwacha for them to be allowed to drill the boreholes in their premise. The hydro-geologist does not always visit the site where the boreholes are to be drilled as the number of applications may be too many and some areas are difficult to reach. They end up giving out permits without doing physical site visits. WARMA is in need of a system that will support the new law.

II. LITERATURE REVIEW

A. Geographical Information Systems

A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data[2].



Figure 1: Applied GIS Platform

The key word to this technology is Geography – this means that some portion of the data is spatial[3]. Spatial data, also known as geospatial data, is information about a physical object that can be represented by numerical values in a geographic coordinate system[4].

B. Geo Odk Collect

Geo ODK collect is an android app used for data collection. It does not only deal with object related data like the hard copy questionnaires will do. This application also has a GPS tracker to get the actual location of the points relevant for the research. GeoODK collect is a mobile application and runs on android powered smartphones. This is based on the ODK Collect, but has the capability of extended with offline/online mapping functionalities, the ability to have custom map layer, as well as new spatial widgets, for collecting point, polygon and GPS tracing functionality. Looking at the spatial arrangement of the project area in relation to development of the best-suited system is one that will be working offline but be able to collect the necessary data. Yes they are dozens of other apps that has been developed to working offline like Geospatial data collection app but it is disadvantaged because of the cost of acquiring it as the app is sold per license of five[5]. Mobile devices and the GeoODK is a free app. Therefore, considering the cost and benefit analysis GeoOdk Collect beat Geospatial.

C. Related Works

A similar research was done in Kenya at the University of Nairobi. In their case Geo, technology was used to display borehole information. The information that was produced was to be used by borehole drillers and the public. The Web based system did not allow data collected in the field to be populated to the database. There was also no option for adding information that was already available in hard copy to the system[6].

In Ghana, a research was done to advocate the use of GIS and Remote Sensing Techniques in the sitting of boreholes for

maximum yield[7]. Ghana is not that different from Zambia as the sitting of boreholes is done without recourse to future development [7]. This causes most wells to dry up, most of their community members complain of less water or no water in boreholes at all [7]. The report further explains how the sitting of these boreholes without recourse to future developments and other environment considerations, poses a big danger to the community [7]. In the same research, it was concluded that the sitting and developing of boreholes in shallow sand and gravel with water tables less than three metres was to be avoided [7]. They also advised those drilling boreholes to concentrate on areas where there is silt or clay soils as these have low yields [7].

III. METHODOLOGY

A. Baseline Study

A baseline study was conducted to come up with ways that will improve the data collection mechanisms on borehole information in the study areas. The study employed a mixed methods research. Mixed methods research employs both qualitative and quantitative research methods. The baseline study was focused on improving ways of data collection in the study area through the developed offline interface.

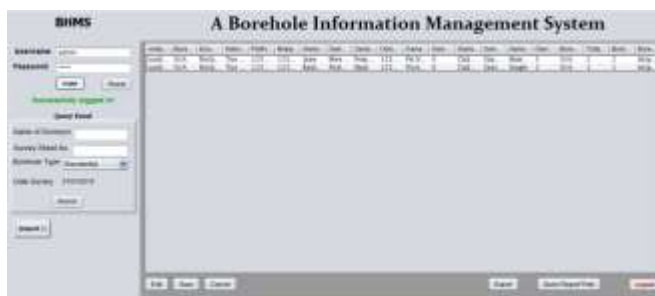


Figure 2: Developed Interface

The data was then used by WARMA to produce different reports as shown in figure 1. These reports include the population density of boreholes in the areas, rate at which the two study areas is fetching the aquifer and many other reports.

B. Target Group

The target group included organisations that deal in water management in Lusaka and the country at large. These included the water resources and management authority WARMA, Lusaka water and sewerage company, Millennium challenge account and the ministry of Sanitation and water resources. Households in Chilenje and Chalala were also visited to carry out this research.



Figure 3: Study Area Map

C. Sampling

The sample size was considered a reasonable size a total of 100 households were visited to carry out the study.

The sampling technique used was probability-sampling technique, which involved simple random sampling. A sampling technique is the process of drawing a sample from the population. This study used a random sampling to draw a sample, which represents the characteristics of the population. The sampling technique was applied to the members of the population comprising-water sector institutions, private water drilling companies, residents, local government institution, households and the Water Resources and Management Authority (WARMA).

Fi

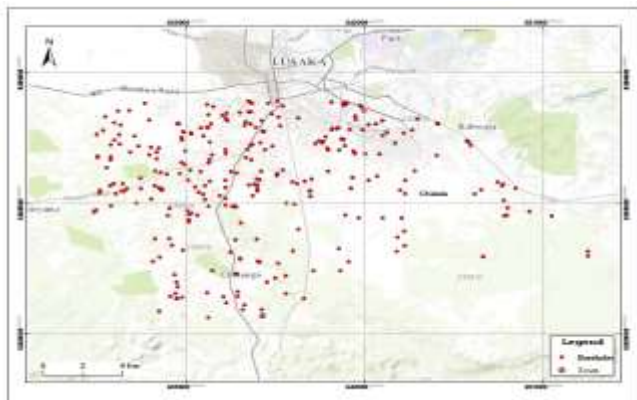


Figure4: Boreholes in Chalala and Chilenje

Random sampling involves the selection of people or events literally ‘at random’. Behind the use of random sampling lies the assumption that:

- If there are a sufficiently large number of examples selected and
- If their selection had genuinely been at random.

D. Data collection

The different methods used in Data Collection were Interviews, Questionnaires and Document Review

1) Interviews

In this research, workers at the Water Resources and Management Authority were interviewed and gave information on the way they were managing boreholes. It

became known that from the implementation of the SI on borehole registration for homeowners, a number of people from the society has come forward to have their boreholes registered. The authority was depending on the manual system. This has led to some households not registering for their boreholes.

2) Questionnaires

Geo ODK collect was used to collect data from the respondents. Geo ODK does not only deal with object related data like the hard copy questionnaires would do. This application also has a GPS tracker to get the actual location of the points relevant for the research.

3) Document Review

Document review was used as another source of data for this study. The term document review refers to a systematic procedure of evaluating documents as a way of obtaining evidence to gain understanding on the phenomena being studied. It can also be used in combination with other quantitative research methods as a means of triangulation to collaborate findings across data sets to reduce the potential impact of bias that can exist in a qualitative study. The random sampling technique was employed to identify and select various public documents to be reviewed

E. Current Business Process

The Water Resources and Management Authority depend on the borehole owners to come forward and have their boreholes registered. The same form is used for both existing and new borehole owners. They pay a one off fee of K250.00. The officers will then check if the borehole position is ok for them to drill before they are issued with a permit.



Figure 5: Borehole registration process [8]

F. Proposed System

The proposed system sought to automate the data collection method. The current system at WARMA requires the borehole owners to register their boreholes. The proposed system will see enumerators visiting people’s homes and collecting the borehole data.

One data has been collected the borehole details are then extracted into an excel file which is then populated to the

database for querying and producing of reports by field officers.

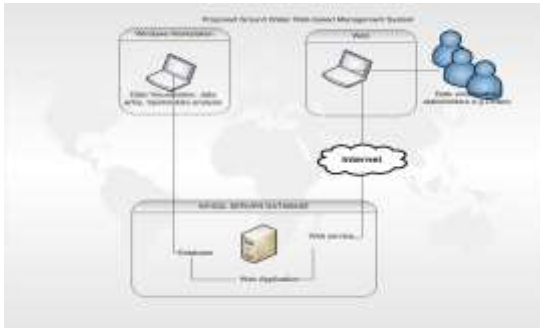


Figure 6: Proposed Borehole Management System

G. System Architecture

The three-tier architecture was used for the implementation of the new system.

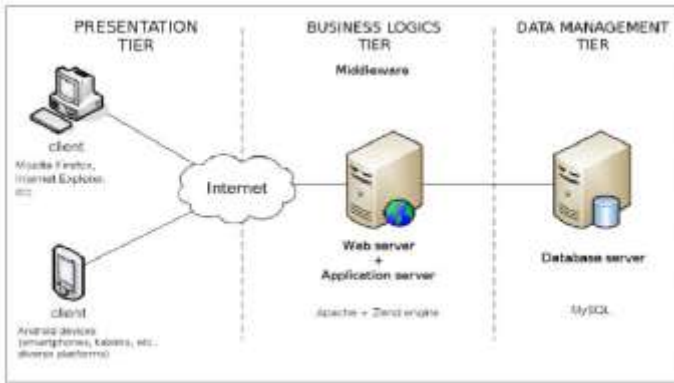


Figure 7: Three tier architecture [9]

The above architecture comprises the presentation tier, the Application tier and the Database tier. The Presentation tier represents the front end of the system. The Data collector and the office personnel are part of this tier. The Application tier is the middle ware. It contains all the activities that are to be carried out by the system, finally the Data Tier that stores the actual database of the system.

H. System Modelling and Design

Use cases are easier for the users of the system to understand although there will be need to break them further down into more detailed user requirements. A use case diagram shows the functionality provided by the system to its users. It shows what actors interact with what use case. The Actor can be a person or another system while the Use cases show the functionality provided by the system to its users.

The data Administrator will perform the following activities in the borehole management system.

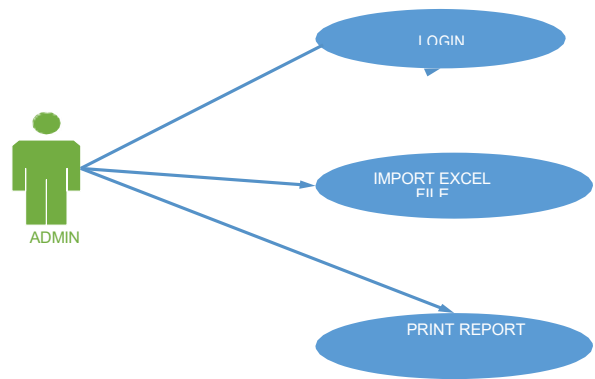


Figure 8: Administrator use case

IV. RESULTS

The results are based on the results of the baseline study. The main purpose of the baseline study was to overcome the challenges faced by WARMA officers when collecting borehole data. The results are collected using Questionnaires, oral interviews and document reviews. The collected data was analysed using charts and tables.

The findings confirmed the challenges that were being faced by WARMA due to not having sufficient data collection mechanisms. 100 households were visited out of which 82 were positive responses, giving a response rate of 82%. The response was good as data was collected from people’s doorsteps.

V. DISCUSSIONS

The results of the study proved that there were a number of challenges faced in the collecting of data on borehole information. These challenges can be overcome by developing the Web-based system and using data collection tools that send the information to the server. The web-based Geographical information system will enable the capturing and storing of all borehole data effectively. Most of the respondents have not registered their boreholes due to lack of sensitisation by the relevant authorities.

The proposed Borehole system will cut down on the issues of having residents walk to the WARMA offices to have their boreholes registered. By registering boreholes and the drilling companies, people will be cautious of the areas where they sit their boreholes.

VI. CONCLUSIONS

The advancement in technology has made the work easier for people involved in water management. A web-based geographical information system will make it easier for all to monitor the boreholes that are already drilled and yet to be drilled.

The web-based geographical information system could be improved so that it is able to capture real-time data. The borehole coordinates after being collected will automatically

update in the system. The system can also have features were customer's login from home and see the information in the database.

ACKNOWLEDGMENT

The authors wish to thank the main contributors to the research, the staff and management of WARMA and not forgetting the residents of Chalala and Chilenje townships who gave up their time and assisted with resources

REFERENCES

- [1] "The introduction of groundwater regulation in Zambia -," 2018. Available: <http://www.warma.org.zm/press-release/the-introduction-of-groundwater-regulation-in-zambia/>. [Accessed: 15-Oct-2019].
- [2] Shimonti Paul, "Geographic information system - Geospatial World," 2018. [Online]. Available: <https://www.geospatialworld.net/entity/geographic-information-system/>. [Accessed: 01-Oct-2019].
- [3] Omnie Solutions, "Geographic Information Tracking Solutions - Omnie Solutions," 2019. [Online]. Available: <https://www.omniesolutions.com/geographic-information-system.html>. [Accessed: 02-Oct-2019].
- [4] Patel Nikhilesh, "Spatial data types in SQL Server," 2018. [Online]. Available: <https://www.sqlshack.com/spatial-data-types-in-sql-server/>. [Accessed: 02-Oct-2019].
- [5] "Mobile Data Collection App - Try it for Free | GIS Cloud." [Online]. Available: <https://www.giscloud.com/apps/mobile-data-collection>. [Accessed: 10-Oct-2019].
- [6] N. S. Sande, "University of nairobi," p. 2726300, 2012.
- [7] Y. D. Twumasi and C. Kwang, "GIS Aided Prospect Siting of Boreholes for Maximum Yield," vol. 1, no. 9, pp. 4–8, 2017.
- [8] "Borehole Registration - WARMA Zambia Groundwater Regulations." [Online]. Available: <http://www.warma.org.zm/permits-and-applications/borehole-registration/>. [Accessed: 04-Oct-2019].
- [9] S. Caballé, F. Xhafa, J. Raya, L. Barolli, and K. Uchida, "Building a software service for mobile devices to enhance awareness in web collaboration," *Proc. - 2014 Int. Conf. Intell. Netw. Collab. Syst. IEEE INCoS 2014*, no. September, pp. 369–376, 2014.

A Comparative Study of Databases for Web Applications Using a Simple but Comprehensive Framework

Kasonde Victor

Copperbelt University
 Directorate of Information
 Communication and Technology
 Postal Address, 21692, Kitwe, Zambia
 Email: vkasonde@cbu.ac.zm

Bwalya Darius

Copperbelt University
 Directorate of Information
 Communication and Technology
 Postal address, 21692, Kitwe, Zambia
 Email: darius.bwalya@cbu.ac.zm

Josephat Kalezhi

Copperbelt University
 School of Information Communication and
 Technology
 Postal address, 21692, Zambia
 Email: kalezhi@cbu.ac.zm

Abstract - The purpose of this study was to compare database technologies for Web Applications with regards to their performance, security, cost, usability, support and scalability. The database technologies considered included MySQL, PostgreSQL, Microsoft SQL Server database, Oracle, MongoDB, SAP Hana, MariaDb. The results indicated that MySQL provided many performance-features that helped web-based applications perform at a rate that modern businesses demand. The study further revealed that MySQL is user-friendly and provides many security features for its implementation than other database technologies. Other features that makes MySQL more usable than others included the fact that it is readily available and can be customized to suit a particular user need besides the high level of security capabilities that it provides. The study recommended that web developers have to consider the benefits of MySQL database technologies for possible implementation and it is hoped that this research will be of great value to the web developers in determining the features that their web application need and ultimately the right database to implement.

Index Terms - Performance, Scalability, Security, database, Cost, Usability, Three-level architecture..

I. INTRODUCTION

The demands for instant access to information is driving changes that impact IT processes and infrastructures, as well as how information is analyzed and the way applications and databases are developed and delivered [1]. The continued increase by most organizations to have their presence only has necessitated web developers to begin considerations of which database is more secure and can deliver the best performance than others. Databases are logically modeled

storage spaces for all kinds of different information [1]. Each database, other than schema-less ones, have a model, which provide structure for the data being dealt with. Database management systems are applications (or libraries) which manage databases of various sizes, and sorts. As more companies begin to transition their mission critical systems to online platforms, it has become increasingly important for web developers to choose the right database solution that offers agility and elasticity but also provides maximum security, high availability and performance [2]. Web application developers are often caught in the middle of choosing the best database to use in while ensuring that performance is not compromised as well as the security of their applications.

According to [3] most databases available are relational in nature. [3] stated that relations in the database can be considered as mathematical sets that contain series of attributes which collectively represent the database and information being kept. These types of identification and collection method, allow relational databases to work the way they do. According to [3], when defining a table to insert records, each element forming a record (i.e. attribute) must match the defined data type (e.g. an integer, a date etc.). Different relational database management systems implement different data types which are not always directly interchangeable. [3] gave examples of MySQL 4.1 and SQLite as being relational.

Most web developers prefer to use relational databases because of the flexibility they provide in terms of query management. Web developers often face difficulties in trying to determine which database technology should be used when and where. There is need to provide a comparative study on various database technologies available in terms of their performances, security, cost, usability, support and scalability. It is against this background that the study was conducted to compare the performance, security, cost, usability, support

and scalability aspects of MySQL, PostgreSQL, Microsoft SQL Server database, Oracle, MongoDB, SAP Hana, MariaDb.

Therefore, this paper is organized as follows: Section one reviews of literature on similar studies. In Section two the methodologies that were considered for this study are introduced. We then examine the results of the comparative study in section three. Section four present the conclusions and recommendations of this work. Lastly section five future directions of a research are presented.

1.0 PROBLEM STATEMENT

Web developers usually face challenges in choosing the database to use for a particular web application. The challenges arise due to failures to clearly determine which database provides the best features in terms of performances, security, cost, usability, support and scalability.

1.2 OBJECTIVE OF THE STUDY

The objectives of the study included the following:

- 1.2.1 Identify the suitable framework for comparative analysis of the databases.
- 1.2.2 Use the framework to compare the various databases based on identified features.
- 1.2.3 To determine the database that is more secure and efficient for web developer to use in the web application development.

1.3 RESEARCH QUESTIONS

From the objectives above, the following research questions were derived:

- 1.3.1 What is the suitable framework for analysis of features of databases in terms of performance, security, cost, usability, support and scalability?
- 1.3.2 Does the framework identify the database suitable for application development by software developers?
- 1.3.3 Is there any feature that appears to be present in all and hence very important for every database?

1.4 SIGNIFICANCE OF THE STUDY

The findings of this study have the potential to help web developers to have an understanding of the challenges that exists when it comes to choosing a database that can be used

in developing a web application. The web developers shall be able to understand the performance and security issues that a particular database could pose and therefore will be able to make the right decision.

1.5 LIMITATION OF THE STUDY

The study has considered relational databases. Other database types have not been considered.

1.6 JUSTIFICATION OF THE STUDY

Relational databases were considered as they are popular. Examples of database management system considered in the study includes MySQL, Microsoft SQL Server database, Oracle, MongoDB, SAP Hana, MariaDb and PostgreSQL because they happen to be the most frequently used by developers.

2.0 LITERATURE REVIEW

There are several reports [18] [19] [20] in the literature on comparison of database technologies. For instance the study by Oracle Corporation [4] attempted to investigate the core differences of some of the most commonly used and popular relational database management systems (RDBMS). The study explored their fundamental differences in terms of features and functionality, how they work, and when one excels over the other in order to help developers with choosing a RDBMS. The RDBMS considered were SQLite and MySQL. The study revealed that MySQL 4.1 was the most popular and commonly used RDBMS because it is a feature rich, open-source product that powers a lot of websites and applications online. Getting started with MySQL 4.1 is relatively easy and developers have access to a massive array of information regarding the database on the internet.

In a related study, Ward [5] conducted a research on the performance of PostgreSQL, it was revealed that Compared to other RDBMSs, PostgreSQL differs itself with its support for highly required and integral object-oriented and/or relational database functionality, such as the complete support for reliable transactions, i.e. Atomicity, Consistency, Isolation, Durability (ACID). Ward(1993) indicated that, due to the powerful underlying technology, PostgreSQL is extremely capable of handling many tasks very efficiently. He added that PostgreSQL is highly programmable, and therefore extendible, with custom procedures that are called "stored procedures". These functions can be created to simplify the execution of repeated, complex and often required database operations. Ward(1993) also established that, although this DBMS does not have the popularity of MySQL, there are many third-party tools and libraries that are designed to make working with PostgreSQL simple, despite this database's powerful nature.

Furthermore, Richards (2007) also conducted a study to compare MongoDB with MySQL 4.1 with regards to data representations. It was revealed that there was a significant difference in the representation of data in the two databases. In MongoDB, data is represented in a collection of JSON documents while in MySQL, data is in tables and rows. Richards (2007) added that MongoDB's performance is better than that of MySQL 4.1 and other relational DBs. This is because MongoDB sacrifices JOINS and other things and has excellent performance analysis tools. MySQL 4.1 is criticized for poor performance, especially in ORM application. However, one is unlikely to have an issue if they do proper data indexing and is using a database wrapper. The study further reviewed that MySQL 4.1 was more secure than MongoDB and was more reliable.

In a related study by Kurniawan and Xue [6] it was revealed that Servlet technology has been the most widely used technology for building scalable Web applications. The study indicated that, there are four design models for developing Web applications using the Java technologies: Model 1, Model 2, Struts, and JavaServer Faces (JSF). Model 1 employs a series of JSP pages; Model 2 adopts the Model-View-Controller pattern; Struts is a framework employing the Model 2 design model; and JSF is a new technology that supports ready-to-use components for rapid Web application development. Based on this fact, Kurniawan and Xue (2004) compared and evaluated the ease of application development and the performance of the three design models (Model 2, Struts, and JSF) by building three versions of an online store application using each of the three design models.

Sun et al. [7] presented a comparative study between relational and non-relational database models in a web-based application, by executing various operations on both relational and on non-relational databases thus highlighting the results obtained during performance comparison tests. This study was based on the implementation of a web-based application for population records. For the non-relational database, they used MongoDB and for the relational database, they used Microsoft SQL 2014.

Finally [8] in his study also attempted to compare the performance of SQL to that of MySQL. He revealed that SQL is practically synonymous with a client/server database model where data is consolidated on, and access is co-ordinated by, a single server or a tightly coupled cluster of servers. This entails a set of requirements pertaining to the atomicity, consistency, isolation and durability (ACID) of data across transactions (potentially with many of these occurring concurrently). He added that, MySQL 4.1 is a particular implementation of an SQL RDBMS software system. It's free (and open source) software and is the most popular server based RDBMS for Internet start-ups and for use with open source web frameworks and many other software systems.

The literature above have attempted to provide a comparison of databases such as MongoDB, MySQL, and SQL with regards to performance. This study attempts to evaluate the

most commonly used database technology by web developers. These comparisons included the performances, security, cost, usability, support and scalability. This area has not been covered in the above literature.

Basically whatever database it may be at the technical level it shares the basic architecture. Figure 1.0 shows the ANSI-SPARC three-level architecture model on which most databases are generally designed.

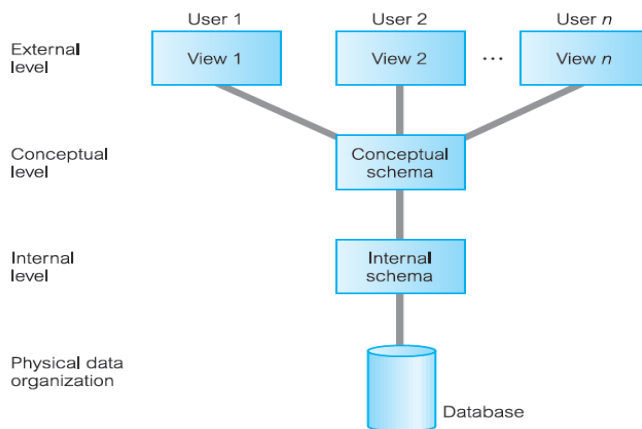


Fig. 1 The ANSI-SPARC three-level architecture.

Fig 1.0 illustrate the three levels of abstraction identified by Begg and Connolly in(Begg & Connolly, 2017) being external schema, conceptual schema, internal Schema and hysical database which are now described. The external schema is users' view of the database. This level describes that part of the level database that is relevant to each user. The conceptual schema is the community view of the database. This level describes what data level is stored in the database and the relationships among the data. The internal schema is the physical representation of the database on the computer. This level describes how the data is stored in the database.

3.0 RESEARCH METHODOLOGY

This section explains the research design used in the study. The methodology was to focus on the criteria for comparing the databases namely performance, security, cost, usability and support. Next we justified why each criteria was significant in the comparisons.

The conceptual framework formulated in this study has been depicted in the following model in figure 2.

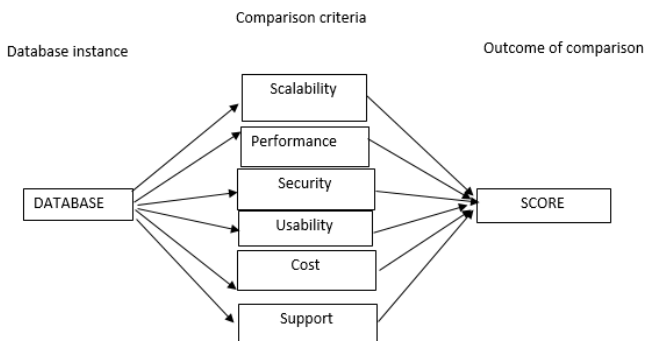


Fig. 2 Conceptual framework.

3.1 Selection And Justification Of Comparison Criteria

A. *Performance*: - Basically database performance [9] may be defined as the rate at which a database management system (DBMS) responds to users or application request. This measure is so significant in that most applications may depend on one another especially in distributed systems and so if the performance is the performance is slow on one applications set of tables, it may affect the efficiency on other applications.

B. *Security*: - Database security implies all the mechanisms employed to protect and secure a database from unauthorized use, malicious attacks and threats. Because users of the data may hold so valuable a resource, therein lies the importance of this measure which is why it was considered [7].

C. *Scalability*: - When a database is said to be scalable it refers to the ability of the database to grow in terms of size and support of various applications [10]. This criterion came under consideration because as the database grows, the main objective is to continue to meet business demands of the organization.

D. *Usability*: - Usability of a database determines how easy it will be for its users to interact with it and therein lies its importance as a feature of databases [11] [12]. Different user groups will be working with the DBMS. There are the administrators, IT and Database admins, application integrators and data consumers. All these different roles need an easily understandable query language and intuitive UI to use the DBMS system efficiently. The easier it is for the user to work with the DBMS, the lower the cost will be for people.

E. *Cost*: - The availability of documentation and support needs to be taken into account as part of the implementation plan including the Total Cost of Ownership (TCO) [13]. A full understanding of these needs and costs will facilitate the choice of the

right tool at reasonable cost in relation to the business needs. Therefore this criterion was picked.

3.2 RESEARCH DESIGN

The design of a study is basically the overall approach used to investigate the problem of interest, i.e., to shed light on, or answer the question of interest. It includes the method of data collection and related specific strategies. We employed a scoring database rating based on their own features identified from literature as shown in the TABLE I below.

TABLE I DATABASE COMPARISON [14]

COMPARISON CRITERIA(taking 1 value for presence of a feature)							
NO.	Scalability	Performance	Security	Cost	Usability	Support	
1	Size scalability	Throughput	Good access control	Freeware	Easy configurability	Updates	
2	Administrative scalability	Support for concurrent transaction	Encryption	Payware	User Interface	Customer support	
3	Geographical Scalability	Quick response	Auditability		Easy of learning	documentation	
4		Capacity	SQL injection and query protection		User success rate		
5		Few query conflicts during multi-user access	Data encryption and masking		Versatility/intergrability		
6		Resource utilization					
ACTUAL FEATURE SCORES OF EACH DATABASE							DB Total Score
MySQL	3	6	4	1	5	3	22
MS SQL	3	6	5	0	3	3	20
PostgreSQL	3	0	3	1	3	2	12
MariaDB	3	6	5	0	2	3	19
MongoDB	0	6	2	0	5	2	15
DB2	3	6	1	0	2	3	15
Oracle	3	6	5	0	2	3	19
SAP HANA	3	6	3	0	3	3	18
Totals							

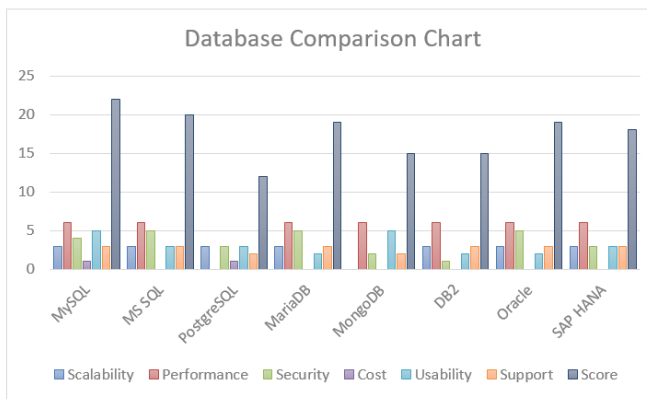


Fig. 3 Comparison chart

3.0 RESULTS AND DISCUSSION

[2]From the results that have been obtained by running each database through each of the identified comparison criteria, it has been noted from table 1.0 and figure 1.1 that MySQL scored highest followed by MSSQL in the overall comparisons. The noteworthy comparison too is the cost aspect criterion which clearly shows that most databases are

commercial as the 0 value indicates with only two databases MySQL and PostgreSQL being the freeware. We also noted the performance aspect of the databases which from the results has indicated that there has been a strong emphasis of performance design consideration by almost all database design engineers with all databases scoring favorably in this feature. This shows that performance aspect of a database is very significant in the choice of any possible use of a database [1].

The support aspect too caught our attention with most database generally having a very good support with the exception of PostgreSQL and MongoDB. With PostgreSQL for instance "Documentation can be spotty, so you may find yourself searching online in an effort to figure out how to do something" [14].

4.0 CONCLUSION

The objective of this study was to conduct a comparative study on database technologies in order to recommend the web servers that provides the best performances, security, cost, usability, support and scalability. The database technologies considered included MySQL, PostgreSQL, Microsoft SQL Server database, Oracle, MongoDB, SAP Hana, MariaDb, Ingres and Hadoop. The results indicated that MySQL, Oracle and Microsoft SQL Server provides many performance- features that help web-based applications perform at a rate that modern businesses demand. MySQL, Microsoft SQL and Oracle were found to be user-friendly and provides many security features for its implementation. Other features that make these web servers more usable than others include the fact that they are readily available and can be customized to suit a particular user need. The study also concluded that, MySQL, MS-SQL Server and Oracle offers exceptional security features that ensure absolute data protection than other database technologies considered in this study.

References

- [1] H. W. Lane David, "Web Applications Applications with PHP and MySQL," vol. 29, no. 104, 2009.
- [2] C. E.F., A Relational Model of Data for Large Shared Data Banks, 1970.
- [3] P. John, "A Comparison Study of Impact Factor in Web of Science and Scopus Databases for Engineering Education and Educational Technology Journals," *Informing Science and Information Technology*, vol. 9, pp. 187-194, 2012.
- [4] Oracle, *Oracle Database Advanced Application Developer's Guide 11g Release*, Corporation, Oracle, 2009.
- [5] A. M. K. John Ward, *Bridging relational databases and object-oriented programming*, Washington: Conf. Management of Data, 1993.
- [6] B. & X. J. Kurniawan, "comparative study of web application design models using the Java technologies," *In Asia-Pacific Web Conference*, pp. 711-721, ND.
- [7] Y. Z. J. X. Y. Z. G. Sun, "Data security and privacy in cloud computing," *International Journal of Distributed Sensor Networks*, vol. 10, no. 7, 2014.
- [8] J. Dennis, "Bridging relational databases and object-oriented programming," *Conf. Management of Data*, 1993.
- [9] D. U.-B. P. G.-V. A. G.-V. S. M.-B. N. L.-P. J. L. Dominguez-Sal, "Survey of graph database performance on the hpc scalable graph analysis benchmark.," *In International Conference on Web-Age Information Management*, pp. 37-48, 2010.
- [10] J. Pokorny, "NoSQL databases: a step to database scalability in web environment," *International Journal of Web Information Systems*, vol. 9, no. 1, pp. 69-82, 2013.
- [11] R. B. K. C. N. L. D. A. v. T. G. E. S. L. Mathur, "Completeness and usability of ethnicity data in UK-based primary care and hospital databases," *Journal of public health*, vol. 36, no. 4, pp. 684-692, 2013.
- [12] N. Bevan, "Usability," *Encyclopedia of Database Systems*, pp. 3247-3251, 2009.
- [13] M. B. T. M. J. B. K. G. R. A. V. Salman, "Use, cost, complications, and mortality of robotic versus nonrobotic general surgery procedures based on a

nationwide database," *The American surgeon*, vol. 79, no. 6, pp. 553-560, 2013.

- [14] C. Arsenault, "Pros and Cons of 8 Popular databases," 2017. [Online]. Available: <https://www.keycdn.com/blog/popular-databases>. [Accessed 15 07 2019].
- [15] D. L. Williams Hugh, *Web Database Applications with PHP and MySQL*, vol. 29, London: University of London, 2009.
- [16] E. Codd, "A Relational Model of Data for Large Shared Data Banks," *Communications of the ACM*, vol. 6, pp. 377-387.
- [17] J. Richards, "A Implementation of Object Oriented Database Security," *5th ACIS International Conference on Software Engineering Research*, 2007.
- [18] A. R. F. & A. L. I. Boicea, "MongoDB vs Oracle-- database comparison," *IEEE*, 2012, September, pp. 330-335.
- [19] R. Angles, "A comparison of current graph database models," *IEEE*, 2012, April.
- [20] M. Valentin, "Database comparison system and method," 2014.

MULTIPLE FEATURE FUSION FOR UNIMODAL EMOTION RECOGNITION

Michele Mukeshimana, Hilaire Nkuzimana, Jeremie Ndikumagenge and Abraham Niyongere

*Faculty of Engineer Science
Department of Information and Communication Technologies
University of Burundi
Bujumbura, Chaussée du Prince Louis Rwagasore 164,
BURUNDI*
{michele.mukeshimana, hilaire.nkuzimana,
jeremie.ndikumagenge&abraham.niyongere}@ub.edu.bi

Xiaojuan Ban

*School of Computer and Communication Engineering
Department of Computer Science and Technology
University of Science and Technology Beijing
Haidian, Xueyuan Road 30, Beijing, China*

banxj@ustb.edu.cn

Abstract – The increasing use of the computing devices and applications in human daily life triggers the need of a natural human computer interaction. Emotion Recognition using multiple features using semi-serial fusion method is proposed. The study analyses the impact of the feature combinations in enhancement of the recognition enhancement. The Unimodal Audio and Facial Emotion Recognition using multiple features using semi-serial fusion method is proposed. The paper presents the use of the multi-view learning principle to fusion multiple features in audio and facial expression-based for emotion recognition. The results show that the Learning Using Privileged Information in unimodal case is effective when the size of the feature is considerable. In comparison to other methods using one type of features or combining them in a concatenated way, this new method outperforms others in recognition accuracy, execution reduction, and stability.

Index Terms – Emotion recognition, multiple feature fusion, Learning using privileged information(LUPI), Machine learning, Human Computer Interaction(HCI).

I. INTRODUCTION

Users interact with the computer through different ways. Human interaction integrating emotional aspect, is more significant and interesting. Due to the individualism of human emotion expression, emotion expression modelling requires to collect many features. Thus, an Automatic recognition, and expression application needs a learning and adaptation based on a combination of information from multiple source and of multiple forms. Modalities used to express emotions, include face, sound, and physiological signals. According to the theory of Mehrabian’s theory [1] [2], as illustrated on Fig.1, the human emotion communication is more than 90% conveyed by the body expression and voice or tones; it is at 93% non-verbal.

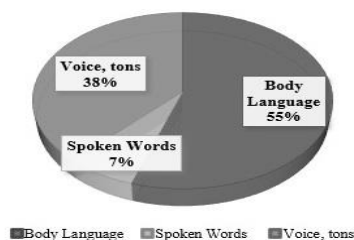


Fig. 1 Mehrabian’s theory about human emotion communication

On Fig.1, human being’s emotion communication information is at 55% from the body language, 38% from the voice or tone of pronunciation and only 7% from the words they speak. Therefore, a visual expression such as facial expression, represent at least 55% of the emotional information. Added to the voice, it makes more than 90% of the emotional information. This rule is applicable to the emotion communication. Explicitly, when one person is expressing her emotion, 55% of the emotion expression information come from the body language, other 38% come from the voice and tons and the last 7% come from the text or spoken words. This consideration should not be confused with the human communication because this last involves more than the emotion expression.

Additional to this rule's support, the facial expression and sound signals are extensively studied. Actually, the audio-visual acquisition is easier than the bio-physiological signal or other signals, which require human body physical connection. The connection on human body generates an obstruction, and embarrassment to the subject. Therefore, the present work proposes a study on Audio and Facial (video) emotion recognition.

Emotion recognition using one modality consists in capturing and learning from information provided by only one modality. A modality is similar to channel by which someone expresses emotion such as face [3] [4] [5], text [6] [7], sound or speech [8] [9], body gesture and movements [10], physiological signal [11], etc. These modalities’ application in Human Computer Interaction means to provide a computer with ability to express or recognize the user’s emotion and reacting accordingly. Due to the complexity of the human behavior, the design of a full-completed system still a point of study. However, some experimental works have attempted to give valuable results.

The present work focuses on coupling multiple features of facial expression and speech modalities to increase the recognition rate and propose the use of Learning Using Privileged Information paradigm to speed up and enhance the recognition from a video stream. The paper endeavours to answer one of the current issues associated with information

fusion in emotion recognition; which is the best way to combine modalities in order to get the best performance in the recognition process?

The paper is arranged as following: after the brief introduction of the work, the second point relates the proposed information fusion method, the third section relates the experimental results and discussion, the last part presents the conclusion and recommendations.

II. PROPOSED METHOD

Multiple feature fusion is pratique targeting to exploit various features extracted from one same modality in order to improve the recognition rate, by gathering more information. Actually, the feature extraction pursues to gather as much as possible salient characteristics representing the pattern of study. The increase of the number, rises the recognition rate, but can tend to overfitting.

Actually, a virtuous recognition depends on the feature extraction (or information gathered) and the classification method exploited. Multiple features fusion, supposes utilizing information gathered from one modality (source), with different views. These features present a complementarity relationship. In the literature, the mostly proposed methods of fusion is the concatenation of the multiple features in one whole set. This serial fusion consists in combining the two different types of features by a horizontal concatenation. The serial fusion technique, improves the recognition rate. The improvement relies on the complementary property of different features. However, the serial fusion increases the dimension of the new feature-spaces. That increase affects the learning process, by increasing the execution time.

The semi-serial fusion method means the combination in the beginning stage, followed by a distinct subdivision into training and testing set for the learning process. It combined the multiple features into one concatenated feature in building the training set, the testing set remains the subset of the standard information set, during the learning session. The Fig.2 depicts the flow of the learning process preparation.

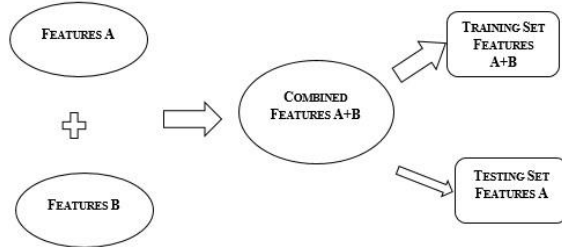


Fig. 2: Flow of the semi-serial fusion method

On Fig.2, the Features A and Features B stand for the proposed multiple features. They are combined to form the full combined features space with a dimension of sum A and B. The proposed method considers that multiple features from one modality, view the same data (facial expression or vocal signal) from different perspectives. Thus, one type of feature can serve as additional information to the other, considered as standard feature set.

This semi-serial fusion method proposes the fusion of multiple features using the Learning Using Privileged Information (LUPI) model. Actually, the semi-serial fusion method, supposes one type of features set as the standard information space and the other type of features set as the Privileged Information for training, and the testing set contains only the testing samples of the standard information set.

Set two different features spaces, A and B, defined on V pattern sample space.

$$A = \begin{pmatrix} a_{11} & \dots & a_{1m} & y_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \dots & a_{nm} & y_n \end{pmatrix} \text{ et } B = \begin{pmatrix} b_{11} & \dots & b_{1p} & y_1 \\ \vdots & \ddots & \vdots & \vdots \\ b_{n1} & \dots & b_{np} & y_n \end{pmatrix} \quad (1)$$

Where a_{ij} stands for the j^{th} feature of the i^{th} sample in the A type of features, and b_{ij} stands for the j^{th} feature of the i^{th} sample in the B type of features, and y_i stands for the class of the i^{th} sample. The row number correspond to the number of the samples and is similar to the two sets of features (n in the example), they derived from a same pattern sample space. The number of columns corresponds to the number (dimension) of the features in the feature space. It is usually distinct from one to another, m for the A type and p for B type in the given example. The new concatenate sample space, C matrix is:

$$C = \begin{pmatrix} a_{11} & \dots & a_{1m} & b_{11} & \dots & b_{1p} & y_1 \\ \vdots & \ddots & a_{im} & b_{i1} & \dots & \vdots & \vdots \\ a_{n1} & \dots & a_{nm} & b_{n1} & \dots & b_{np} & y_n \end{pmatrix} \quad (2)$$

Supposing that A type features set serves as Standard Information set, and the B type feature set serves as the Privileged Information set, the learning process follows the flow represented on Fig. 2.

The training feature set is a partition chosen from the new formed combined features space C, and has the same dimension as the concatenation matrix C. The testing set is a remaining partition, but it only contains the feature of the Standard Information set A. The Fig.3 represents the new proposed learning method.

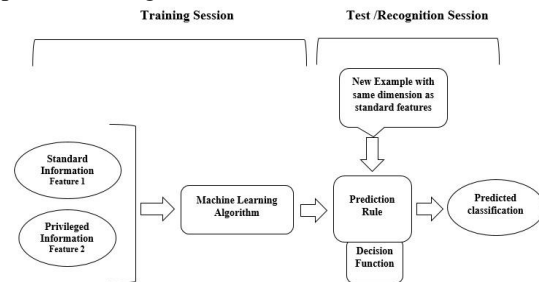


Fig. 3: The multiple feature semi-serial combination diagram

On Fig.3, the labelled samples space, comprehends multiple features, extracted from same original pattern. The training samples space is a subset of the labelled samples set, they have same dimension. The new example is a new sample; whose features are as the same dimension as the standard features space. This proposed testing modification reduces the

complexity of the learning from $O((m+p)\log^{(m+p)})$ to $O(m)\log^{(m)}$.

Actually, the reduction of the dimension for new examples, reduces the storage memory usage, and the execution time. In addition, the recognition rate is improved through the knowledge transfer.

Indeed, Privileged Information reinforce the learning accomplishment by transferring its knowledge to improve the prediction rule. The semi-serial method, which is proposed here, can be applied with any linear based machine learning method. In the present work, the applied method is the Sparse Extreme Learning Machine – Learning Using Privileged Information [12].

During learning using the privileged information (LUPI) [17-18], the training set is composed of triplets, i.e. standards variables X, privileged variables X*, and their corresponding label Y, but the testing set comprehends standard variables X and the labels only. The vector X of the standard information is mapped into the hidden-layer feature space by h(x) and the vector X* of the privileged information is mapped into the hidden-layer correcting space by h*(x*). Actually, the two kernel functions h(x) and h*(x*), can be different or the same.

The introduction of the LUPI model in S-ELM (Sparse Extreme Learning Machine) [13-16] finds origins in the optimization-based ELM method; the slack variables are unknown to the learner. If there is an oracle who can give more information, they can be estimated by a correcting function defined by that additional information. The correcting function, which estimates the slack value, is computed as following:

$$\xi(x) = \phi(x^*) = h^*(x^*)\beta^* \quad (3)$$

Hence, the problem of optimization becomes as follows:

$$\min_{\beta, \beta^*} L_p = \frac{1}{2} \left[\|\beta\|^2 + \gamma \|\beta^*\|^2 \right] + C \sum_{i=1}^N h^*(x_i^*) \cdot \beta^* \quad (4)$$

$$\begin{aligned} s.t.: & y_i(\beta \cdot h(x_i)) \geq 1 - (h^*(x_i^*) \cdot \beta^*) \\ & h^*(x_i^*) \cdot \beta^* \geq 0 \end{aligned} \quad (5)$$

Where β^* is the correcting weight connecting the i^{th} hidden node to the out-put node in the correcting space and γ is introduced for the regularization.

To minimize the above functional (4) subject to constraints (5), The Lagrangian function is:

$$\begin{aligned} L_D = & \frac{1}{2} \left[\|\beta\|^2 + \gamma \|\beta^*\|^2 \right] + C \sum_{i=1}^N h^*(x_i^*) \cdot \beta^* - \sum_{i=1}^N \alpha_i [y_i(\beta \cdot h(x_i)) - (1 - (h^*(x_i^*) \cdot \beta^*))] \\ & - \sum_{i=1}^N \mu_i (h^*(x_i^*) \cdot \beta^*) \end{aligned} \quad (6)$$

Then the decision function is computed as follows:

$$f(x) = h(x) \cdot \beta = \sum_{i=1}^N \alpha_i y_i h(x_i, x) \quad (7)$$

And the corresponding correcting function is:

$$\phi(x^*) = h^*(x^*) \cdot \beta^* = \frac{1}{\gamma} \sum_{i=1}^N (\alpha_i + \mu_i - C) K^*(x_i^*, x^*) \quad (8)$$

The proposed method has the following algorithm:

Algorithm1: Multiple Feature Fusion using Privileged Information

Input: Standard Feature Set C, Privileged Information D, hidden number L and activation functions g, and g*

Output: The prediction of the approximated function f(c).

Procedure:

1. start
2. Random generation of the respective input weights
3. Calculate the hidden node output matrices H and H*
4. Solve the dual optimization computation equation (4)
5. Compute the output weight β
6. Compute the prediction function f(c)
7. end

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Description of the Dataset

The eINTERFACE'05 data set is a multimodal emotion recognition data set. It consists in recording from 46 individuals, expressing six different categories of affective states, i.e. anger, disgust, fear, happiness, sadness and surprise. These individuals are from different countries. The eINTERFACE'05 data set, belongs to the category of the induced data set, the actors are given stories and sentences to pronounce according to the situation. The data set contains video records with images and sound components. It is chosen because of this opportunity of collecting sound and facial features from the same set.

B. Feature extraction

Multiple features are extracted from the audio component and from the facial (images) component. The audio features are extracted using the PRAAT tool [21]. For the experiments, sound related features extracted, are of four types, i.e. features relating to pitch (11), sound (20), spectrum (13) and Intensity (9), in total 53 features.

Facial features extracted, are of three sorts, i.e. Local Binary Pattern (LBP) based features, Local Description Number (LDN) pattern and the Edge Orientation Histogram (EOH) patterns. These features, define local descriptors, but the information they provide differ from each other and are complementary.

C. Experiment setup

The pre-processed dataset is subdivided into five equal sub-data sets. Four sub-data sets are used for training and validation. Then, the remaining sub-data set is used for testing. All the samples can equally participate in the experiments as training, validation and testing sample.

In addition, the test is done on new samples, which are not used for the classifier training. Each experiment, using same parameters, repeats four (04) times and the average is recorded for all the pursued results, i.e. training and testing time, training and testing accuracy.

The simulation executive program is written and run using MATLAB R2014a version, running on Intel® Core™ i5-4590 CPU @ 3.30Ghz with 4.00GB RAM. The results are recorded and processed using Microsoft Excel 2016. The execution

(training/testing) are recorded in seconds and the accuracy (training/testing) is recorded in percentage or in the interval between to 0 and 1, corresponding respectively to 0 and 100 percent. Classification exploits Sparse ELM-Learning Using Privileged Information. Privileged information set is constituted using one group of features (intensity, pitch, spectral and global sound features for the audio component and LBP, LDN and EOH for the facial component) as privileged information and the remaining features as standard information vice-versa.

D. Experiment results and discussion

The first part of the experiment aims to evaluate the performance of the proposed method. The tables 1 and 2 represent the results of the audio and facial based recognition respectively.

TABLE I
AUDIO-BASED PERFORMANCE RESULTS

Dataset	Basic	Privileged	Train. Time	Test Time	Train Accu. %	Test Accu. %
Intensity	Privileged		0.111979	0	83.868353	85.3437095
			0.183594	0.0026042	83.868353	85.3437095
	Training		0.101563	0	83.868353	85.3437095
			0.1875	0	83.868353	85.3437095
Pitch	Privileged		0.104818	0	83.868353	85.3437095
			0.167969	0	83.868353	85.3437095
	Training		0.109375	0	83.868353	85.3437095
			0.160807	0	83.868353	85.3437095
Sound	Privileged		0.111979	0.0019531	83.868353	85.3437095
			0.171224	0	83.868353	85.3437095
	Training		0.111328	0.0026042	83.754864	85.2140078
			0.176432	0.0045573	83.754864	85.2140078
Spectrum	Privileged		0.104167	0	83.868353	85.3437095
			0.16862	0.0045573	83.868353	85.3437095
	Training		0.110677	0	83.868353	85.3437095
			0.172526	0	83.868353	85.3437095

TABLE II
FACIAL BASED PERFORMANCE RESULTS

Dataset	Basic	Privileged	Train. Time	Test Time	Train Accu.	Test Accu. %
LBP	EOH		0.5313	0.0547	82.49	77.82
			0.6445	0.0586	82.29	79.37
LDN	EOH		0.4922	0.0156	84.63	86.38
			0.6719	0.0547	82.29	79.37
EOH	LDN		0.5078	0.0156	82.49	77.82
			0.5469	0.0625	84.09	86.12

From the results, the average recognition rate is around 85.34% for the audio and 81.14% for the facial features which are acceptable results and prove that the method is effective for emotion recognition when there is only one modality. Concerning the execution time, the testing time is shorter than the training time, which is one of the objectives of the proposed method. The execution time depends on number of hidden nodes as shown on Fig.4.

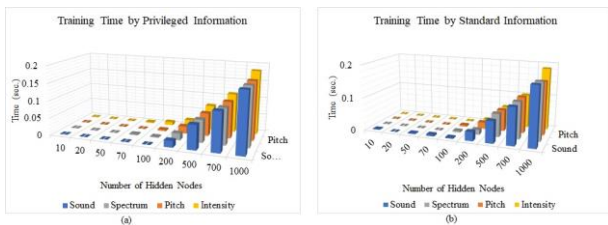
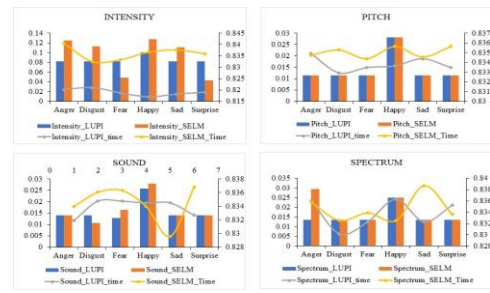
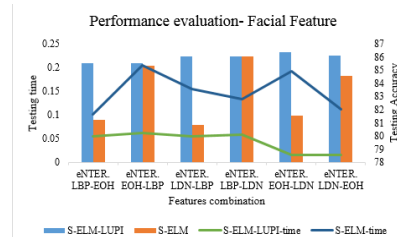


Fig. 4: Execution time by number of hidden nodes



(a)



(b)

Fig. 5: Comparison to concatenated methods: (a) audio (b) facial

Compared to other methods, multiple feature considered in a multiview learning problem yield better results than a concatenation learning method, on Fig.5 it is compared to Sparse Extreme Learning Machine. The recognition rate is improved because of the use of multiple features in a multiview manner. the execution time is reduced because of the reduced dimension of the testing data set.

Table III
COMPARISON TO OTHER WORKS FOR AUDIO FEATURES

Datasets	S-ELM-PLUS	SVM ^[22]	TTI-Based SVM ^[23]
Feature	Test.Accu.	Test.Accu.	Test. Accu.
Intensity, Pitch, Sound, Spectrum	85.344	44.730	73.060

Table IV
COMPARISON TO OTHER METHODS FOR FACIAL FEATURES

Datasets	S-ELM-LUPI	S-ELM	ELM BASIC	ELM PLUS	SVM
eNTER.LBP-EOH	85.5383	81.24	50.0778	68.1323	66.148
eNTER.EOH-LBP	85.5383	85.33	49.2607	58.7938	61.868
eNTER.LDN-LBP	86.0571	80.82	49.2607	58.7938	64.202
eNTER.LBP-LDN	86.0571	86.03	46.515	68.5019	64.202
eNTER.EOH-LDN	86.3813	81.57	48.716	63.3268	66.148
eNTER.LDN-EOH	86.1219	84.55	48.4825	68.5019	61.868

The comparison to the state-of-the art, the proposed method still performs better. eNTER. stands for Enterface05 data set. The features are extracted from it.

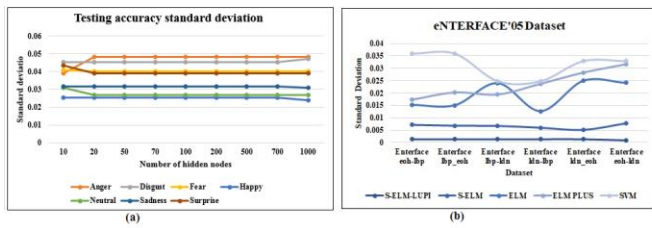


Fig. 6: Standard deviation of the Testing accuracy

On Fig.6, the recognition rate standard deviation is less than 0.05 thus the method is more stable. This stability is due to the use quadratic optimization which stabilizes better than other methods. On Fig.7 the proposed method is more stable compared to other neural network based methods.

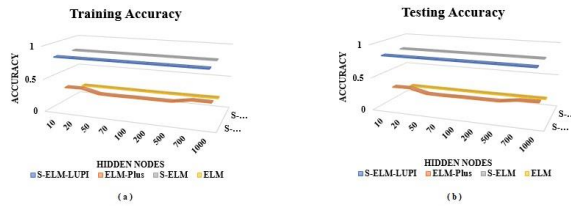


Fig. 7: Comparison to other neural network methods

IV. CONCLUSION.

This paper presents a new semi-serial fusion method of multiple feature based on Learning Using Privileged Information (LUPI) model. LUPI paradigm permits the improvement of the learning accuracy and its stability, by additional information and computations using optimization methods.

The execution time is reduced, by sparsity and dimension of testing feature. Multiple features considered in a multiview learning problem yield better results than a concatenation learning method. Thus, the proposed method is candidate to real-time problem by recognition time reduction and real-life problems by its stability in recognition.

REFERENCES

[1] Van Vliet V. Communication Model by Albert Mehrabian. 2012.
 [2] Mehrabian A. Communication without words [J]. Psychology Today 2 (4), 1968:53–56
 [3] Tian Y-L, Kanade T, Cohn J-F. Recognizing AUs for Facial Expression analysis [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, 23(2): 91-116.
 [4] Saragih J M, Lucey S, Cohn J F. Face alignment through subspace constrained mean-shifts [C]. // Proceedings of IEEE 12th International Conference of Computer Vision (ICCV). Kyoto, Japan, 2009:1034-1041
 [5] Akputu K O, Seng K P, Lee Y L. Facial Emotion Recognition for Intelligent Tutoring Environment [C]. // Proceedings of the second International Conference on Machine Learning and Computer Science (IMLCS'2013). Kuala Lumpur, Malaysia, 2013:9-13
 [6] Poria S, Cambria E, Howard N, et al. Fusing audio, visual and textual clues for sentiment analysis from multimodal content [J]. Journal of Neurocomputing 174 (A-22), 2015:50-59.
 [7] [17] Poria S, Cambria E, Hussain A, et al. Towards an intelligent framework for multimodal affective data analysis [J]. Journal of Neural Networks 63, 2014:104-116.

[8] Quek F, McNeill D, Bryll R, et al. Multi-modal human discourse: gesture and speech [J]. ACM Transactions on Computer-Human Interaction (TOCHI), 2002, 9(3):171-193.
 [9] Scherer K R. Vocal communication of emotion: A review of research paradigms [J]. Speech Communication, 2003, 40:227-256.
 [10] Navarretta C. Individuality in Communicative Bodily Behaviors [M]. Esposito A, Esposito A, Vinciarelli A, et al. (Eds), Cognitive Behavioral Systems of Series LNCS, 2011, 7403:417-423.
 [11] Knapp B R, Kim J, André E. Physiological Signals and Their Use in Augmenting Emotion Recognition for Human-Machine Interaction [M]. Petta, P., Pelachaud, C., Cowie, R., (Eds), Emotion-Oriented Sytems: The HUMAINE Handbook with 104 Figures and 35 Tables, 2011:133-161.
 [12] Michele Mukeshimana, Xiaojuan Ban, Nelson Karani. Sparse Extreme Learning Using Privileged Information. Communications in Computer and Information Science, 2017, 710: 205-213 Vapnik V, Vashist A. A new learning paradigm: Learning Using Privileged Information [J]. Neural Networks, 2009, 22(5-6):544-557.
 [13] Vapnik V, Izmailov R. Learning Using Privileged Information: Similarity Control and Knowledge Transfer [J]. Journal of Machine Learning Research, 2015, 16:2023-2049. Huang G B, Ding X-J, Zhou H-M. Optimization method based extreme learning machine for classification [J]. Neurocomputing, 2010, 74:155-163.
 [14] Cao L-L, Huang W-B, Sun F-C. Optimization-based Extreme Learning Machine with Multi-Kernel Learning Approach for Classification [C]. Proceedings 22nd International Conference on Pattern Recognition (ICPR). Stockholm, Sweden, 2014:3564-3569.
 [15] Huang G, Huang G-B, Song S-J, et al. Trends in extreme learning machine: A review [J]. Neural Networks, 2015, 61:32-48.
 [16] Bai Z, Huang G-B, Wang D-W, et al. Sparse extreme learning machine for classification [J]. IEEE Trans. Cybern., 2015,44(10):1858-1870.
 [17] Navarretta C. Individuality in Communicative Bodily Behaviors [M]. Esposito A, Esposito A, Vinciarelli A, et al. (Eds), Cognitive Behavioral Systems of Series LNCS, 2011, 7403:417-423.
 [18] Knapp B R, Kim J, André E. Physiological Signals and Their Use in Augmenting Emotion Recognition for Human-Machine Interaction [M]. Petta, P., Pelachaud, C., Cowie, R., (Eds), Emotion-Oriented Sytems: The HUMAINE Handbook with 104 Figures and 35 Tables, 2011:133-161.
 [19] Darwin C. The expression of the emotions in man and animals. New York: Oxford University. (Original work published in 1872).
 [20] Bachorowski J-A, Owren M J. Vocal Expressions of Emotion [M]. Handbook of Emotions, third Edition, Michael Lewis, Jeannette M Haviland-Jones, Lisa Feldman Barrett (eds). The Guilford Press, New York London.
 [21] Boersma P, Weenink D. Praat, a system for doing phonetics by computer, version 3.4 [R]. Institute of Phonetic Sciences of the University of Amsterdam, (1996). Report 132.
 [22] Zhang S-Q, Li L-M, Zhao Z-J. Audio-Visual Emotion Recognition based on Facial Expression and Affective Speech [C]. //Proceedings of the second International Conference CMSP 2012, CCIS, 346:46-52
 [23] Wang K-C. Speech Emotional Classification Using Textura Image Information Features [J]. International Journal of Signal Processing Systems, 2015, 3(1):1-7.

Exploring the Need for Context-Awareness Mobile Applications-A literature Analysis Perspective

Bwalya Darius

Copperbelt University
 Directorate of Information
 Communication and Technology
 Postal Address, 21692, Kitwe, Zambia
 Email: darius.bwalya@cbu.ac.zm

Kasonde Victor

Copperbelt University
 Directorate of Information
 Communication and Technology
 Postal address, 21692, Kitwe, Zambia
 Email: ykasonde@cbu.ac.zm

Nchimunya Chaamwe

Copperbelt University
 Directorate of Information Communication
 and Technology
 Postal address, 21692, Zambia
 Email: chimz@cbu.ac.zm

Abstract - The prevalence of smart mobile devices has brought about the need for applications to be location aware, content-aware, activity-aware, environmental and user-aware, a technological phenomenal now being referred to as Context-awareness mobile application. Great strides have been made in contextualization by means of Global Positioning System (GPS) enabled applications and Pervasive Computing. In this paper, we aim to understand mobile application context-awareness and why its application scenario should be significant in the Zambian context. We begin by examining literature on context-awareness to fully appreciate its meaning, identify factors or components of context and group into location-aware, Activity-aware, Environment-aware and User-aware and we further show why Zambian mobile app developers, business people including medical experts should think of implementing and utilize this form of smart dynamic technology to present more relevant information, functionalities to their clients depending on circumstances.

Index Terms - context-aware Application, location-aware, context types, Pervasive Computing, user-aware, GPS

I. INTRODUCTION

The advent of smart mobile devices has not only brought about the need for mobile applications to be location-aware, content-aware and user-aware which is the technological phenomena now known as context-awareness systems [1] wherein the behaviour of the application is dynamic depending on the context for instance location. A modern smart phone comes with it a sophisticated collection of sensors on-board that prove useful in recognition of human activities and monitoring, making them ideal gateways into capturing of metrics of interests such as physical activities, physiological activities including many other factors that can define daily pattern of routines[2]. Perhaps one of the all-encompassing definition of context is one that was made in [1] “Context is any information that can be used to characterize an entity, its condition, or its surrounding situation, if the information is

considered relevant to the interaction between a user and an application”. [1] further on defines an entity as a “a person, place, material object or immaterial state, including the user and the application themselves”. The definition of context reveals much about how dynamic context can be achieved by means of for instance sensors[3]. Sensors can be used to define the environment usually in a real time manner. The focus of context is not one sided as in providing context only to users but by extension something that becomes relevant to the interaction of both the user and the application dynamically to determine context[4]. For instance [5] noted that human beings are creatures that adapts to particular pattern of life in regard to the interaction with mobile devices and consequently such things as sites we browse, where we login which could either be work intranet or extranet, checking online banking with appropriate security or having a routine lunch break place where there is login based hot spot and many other cyber interaction routines. A smartphone may create a profile of this data and with the possibility of using the power of cloud services[6] data may be analyzed and contexts relevant to the user may be formed dynamically. Having defined what context is all about, it is important to investigate literature on the topic, then move on to identify what attributes defines context, classify the context types and finally show the practical value of why Zambians, especially mobile app developers, business people including medical experts should think implementing and utilize this form of smart dynamic technology. The paper now begins by focusing on what attributes defines context.

A. Problem Statement

The challenge is to understand and show the mobile context awareness in a Zambian context in which there is little if nothing that has been published on the topic.

B. Objective of the study

The main aim of the study was to understand mobile application context-awareness and why its application scenario should be significant in the Zambian context.

The specific objectives of the study included the following:

- ✓ To fully appreciate the meaning of context awareness in mobile application.
- ✓ To identify factors or components of context.
- ✓ To show why Zambian mobile app developers, business people including medical experts should think of implementing and utilize this form of smart dynamic technology

C. *Research Questions*

From the objectives above, the following research questions were derived:

- ✓ What is context awareness mobile application?
- ✓ What factors or components define context?
- ✓ Why should Zambian mobile app developers, business people including medical experts think of implementing and utilize this form of smart dynamic technology?

D. *Significance of the Study*

This literature analysis has the potential to help mobile app developers, business people including medical experts present more relevant information, functionalities to their clients depending on circumstances and we are confident that this is first research work to be undertaken in a Zambian context and we are sure the topic will gain further attention by scholars.

E. *Limitation of the Study*

The scope of this research was on aggregating research work from a number of sources with focus on understanding context-awareness in mobile applications, examining proposed models of its applications in those works and then show application areas of the technology especially in Zambian context.

F. *Justification of the Study*

Context-aware mobile application is a good research area gaining particular attention. It happens to have so many areas of applications cutting across several technologies as Smart health care GPS, Collaborative filtering, Smart Environment, Smart grid, biometrics authentications among others.

II. LITERATURE REVIEW

In this section we seek to focus on the research work done by scholars with particular focus on factors and components of Context-awareness and how this takes place. We also identify from literature some major application scenarios.

A. *Components of a Context awareness*

There are a number of factors which characterize context and herein these will be referred to as classes and much has been published by scholars as noted by Rosenberger & Gerhard in [1]. They agreed as to what constitute the basic classification being identity location and activity. Other grouping of factors that define context are classified into active group, that is context that are needed to identify the current entity and its conditions and passive group that is all other contexts. The authors of [1] argued that this is the more useful way to classify context. Table I shows a summarized classification of context. This is very significant when developing context dependent applications

TABLE I SUMMARY OF CONTEXT TYPES A PROPOSED FRAMEWORK IN [1]

Contextual category	Context type	Data source	Base type	Durability	Change	Usage
User	Personal information	External	Physical	Static	-	High
	Personal condition	Sensor	Physical	Dynamic	Continuously	Low
	Location	Sensor	Physical	Dynamic	Continuously	High
	Task	External	Virtual	Dynamic	Instantly	High
Environment	Date & time	Internal	Virtual	Dynamic	Continuously	High
	Resource	External	Physical	Dependent	Continuously	Medium
	Resource condition	Sensor	Physical	Dynamic	Dependent	Medium
System	Physical condition	Sensor	Physical	Dynamic	Dependent	Medium
	Software	Internal	Virtual	Dependent	Instantly	Low
	Device	Internal & External	Dependent	Dependent	Dependent	Low
	Network	Internal	Dependent	Dependent	Dependent	Low
Information retrieval	Server	Internal	Dependent	Dependent	Dependent	Low
	Content	Internal	Virtual	Dynamic	Instantly	Medium
Pattern recognition	Search	Internal	Virtual	Dynamic	Instantly	Medium
	Historical data	Internal	Virtual	Static	-	Medium
	Pattern	Internal	Virtual	Dynamic	Instantly	Medium

From Table I above it is clear that context depends too on sensors as inputs into defining what the circumstance is. It is expected that the possible data sources depicted will help a context-aware mobile application to intelligently add or remove functionalities.

This means that with little or without explicit human intervention [7] the mobile application will undergo the necessary reconfigurations and adapt at runtime and still expect that human users are satisfied [8].

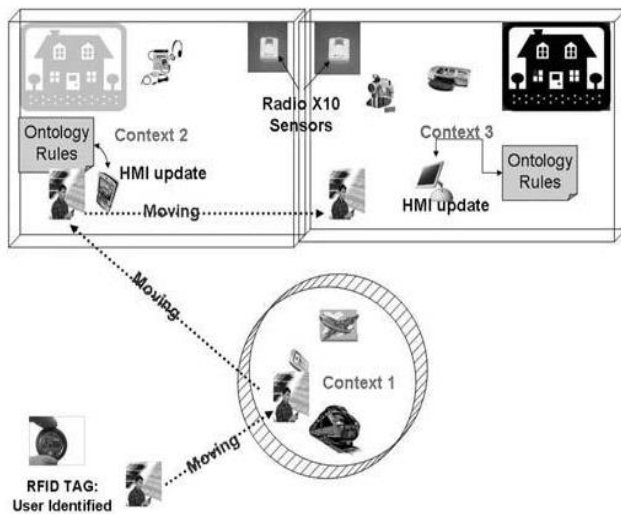


Fig 1.0 illustrating dynamic reconfiguration of context [9]

In fig 1.0 above, HMI updates are based on the services available and preferable in the abstract environment. To achieve this desirable functionality, advances have been made in the emerging form of technology called Pervasive Computing [10], in which nodes are small mobile and usually embedded in a larger system, characterized by the expectation that the systems must adapt to their surroundings [11]. In this form of distributed system the devices or microprocessors are intelligent, autonomy and effects contextualization.

A. Classification Context Types and Their Application Scenarios

There are a number of application scenarios of this form of technology and these will be grouped into the most useful factors of contextualization namely time, location or environment, user identity, activity[8] [12]. Table II summarises the contexts types reviewed and their application scenarios.

TABLE 2: CONTEXT TYPES AND THEIR APPLICATION SCENARIOS

Factor	Technology	Effect on the service	Application Example	Scenario
Location	GPS [13]	Loading of functionality based on location	Recommend nearby hotels, restaurants, listing of local authorities offices, Traffic checkpoints, recommending vacant parking slots, Vehicle tracking	
Activity	IoT[14] Sensors	Sending of detected conditions to connected mobile devices	Sending of distress messages and using GPS to provide location for example	

	Collaborative filtering [15]	Based on the user preference profile during interaction with the mobile application relevant similar information is loaded.	Recommendation of users with similar preferences and taste or services or products of similar functionality.
	Recommender Systems(RS)[21]	User preferences are learnt and are used to predict the services and other similar suggestion	Application in ecommerce, tourism
User Identity	Biometrics authentication [16]	Dynamically determining rights, privileges to resources	In Internet of Military Things (IoMT) [17], In Information Systems Security[18] Iris Scan technology and voice recognition technology [19]
Environment	Smart Environment[9] [20]	Large-scale context-awareness combining IoT and Big Data resulting in the creation of smarter application ecosystems in diverse vertical domains	Smart healthcare, smart cities and grids, smart transportation, smart industry 4.0
	Smart Grid[21]	Intelligent power management	Dynamic control and consumption of power by power utility firm.

III. METHODOLOGY

The approach taken in data collection was qualitative and so a number of scholarly articles were retrieved and a detailed study of them was done. The articles databases which were mined included sciencedirect.com, emeraldinsight.com, *EEEXplore*, *ScienceDirect*, *Springer*, *Google Scholar*, *Taylor and Francis*, *ACM Digital library*, *Kluwer*, *Wiley*, *DOAJ*, *IGI-Global*. The strategy was to use the keywords at the outset, a combination of the research terms, context-awareness concepts. For instance to get the whole concept of contextualization topic phrases like “context awareness mobile application” OR “context-awareness mobile application”, “classification of context types” were used. “Use cases of context-aware mobile application” was pushed in all the used article repositories.

A. Inclusion Criteria

Keywords and concepts:-The papers that contained the keywords at the outset including IoT, smart environment, Biometrics authentication, collaborative filtering and context-aware mobile application examples formed the primary criterion resources.

Time period of published literature: - Besides content being the article inclusion criterion, the time period was significant in the research focusing much on the research done 10 years to date with the confidence that research work is progressive and becomes even more refined. Grey areas on the Google web search engine too was considered but received less attention than the primary databases mentioned.

B. Exclusion Criteria

Automatic exclusion was made for any article that never contained any of the keywords or context-awareness concept and application examples. Primary sources chosen were journals because they reflect current research than books do, then books and internet gray areas were used as secondary sources.

IV. RESULTS

The most notable results are summarized in table III below

TABLE 3: AGREGATING SUMMERY OF RESULTS

<i>Search Term/Phrase</i>	<i>Article Database</i>	<i>Results (publications)</i>
<i>context awareness applications</i>	<i>Emeraldinsight</i>	<i>61601</i>
	<i>Researchgate</i>	<i>>100</i>
	<i>Sciadirect</i>	<i>216,579</i>
<i>classification of context types</i>	<i>Researchgate</i>	<i>100</i>
	<i>Emeraldinsight</i>	
	<i>Sciadirect</i>	<i>319,814</i>
	<i>dl.acm.org</i>	<i>525,594</i>
	<i>IEEXPLORE Digital library</i>	<i>957</i>
<i>Context-aware</i>	<i>Igi-global</i>	<i>26,357</i>
	<i>Wiley online library https://onlinelibrary.wiley.com</i>	<i>322,737</i>
	<i>Pervasive Computing</i>	
	<i>Researchgate</i>	<i>>100</i>

	<i>Emeraldinsight</i>	
	<i>Sciadirect</i>	<i>32,424</i>
	<i>dl.acm.org</i>	
	<i>IEEXPLORE Digital library</i>	
	<i>Igi-global</i>	
<i>smart grid context-aware</i>		
	<i>Sciadirect</i>	<i>5,989</i>
	<i>Researchgate</i>	<i>10</i>
	<i>Emeraldinsight</i>	<i>402</i>
	<i>Igi-Global</i>	<i>772</i>
	<i>Wiley online library</i>	<i>2,849</i>

V. DISCUSSION AND OPEN ISSUES

This literature investigation even as the results show, has revealed how much strides researchers have made in context-awareness technology though there is still more room for research especially on the factors that should define a particular context other than user, activity, location, environment [21]. This is so because there is virtually no limit as to what can characterize context. For instance customs, ethnicity, attitude, changed physical status from normal to loss of sight or combination of factors remain an open issue. To illustrate, a manager might feel that certain content on the mobile device should not be accessed by his subordinate or any other person automatically without his active involvement or worrying what might accessed whenever the device is held by other persons.

However going towards adopting this technology is still recommended even at the level it has reached in the Zambian context as already noted in the number of useful applications now possible as highlighted in table II above.

In Zambia, this technology can be used various ways such as in:

A. Smart health care provision

Personalized health care provision is now possible to assist the elderly and providing guidance to care givers “in a timely and appropriate manner by adapting the healthcare technology to fit in with normal activities”[12]. Additionally it is possible to use recommender systems in conjunction with wearables [22] wherein for instance a detection of a rise in body temperature alerts the personal physician or recommends nearby health care locations.

B. *Research*

Researchers can publish and accomplish so much in Zambia than what is obtaining [25] especially if collaborative filtering is implemented among scholars of similar research interests and their institutions.

C. *Tourism*

This is one of the most important industry of Zambia with many regular and new tourists streaming in every year. The challenge is for the new ones to get real time hospitality services ubiquitously.

D. *Efficient dynamic utilization of Network bandwidth*

Most application used in Zambia and of course world over require updates for their effective operation. However the challenge is mostly updating at the expense of deteriorating bandwidth. Kolomvatsos in [23] proposed a model for nodes to be deciding intelligently on the optimal time when they should carry out the update process balancing the network performance and update process all the while without user intervention.

E. *Power Management*

Zambia faces a great challenge of power demand while at the same time consumers tend to waste it. Context-aware application such as smart grid have now the capability of intelligently, depending on context information gathered from sensors make future best energy saving decisions [24]. It is hoped that the context-aware smart grid technology is implemented locally to save much needed power energy.

VI. CONCLUSION

It is clear from this literature investigation that context-aware mobile application is a good research area gaining particular attention. It happens to have so many areas of applications cutting across several technologies as shown in Table 2 such as Smart health care GPS, Collaborative filtering, Smart Environment, Smart grid, biometrics authentications among others. The aggregated literature from the famous publication repositories have revealed quiet a wealthy of information on applications of context-awareness which this study has undertaken. It has led to discoveries of many beneficial application scenarios locally by business people in their advertisement, mobile application developers, health care professionals.

REFERENCES

[1] P. Rosenberger and D. Gerhard, "Context-awareness in industrial applications: definition, classification and use case," *Procedia CIRP*, vol. 72, pp. 1172–1177, 2018.
 [2] J. Wannenburg and R. Malekian, "Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing,"

IEEE Trans. Syst. Man Cybern. Syst., vol. 47, no. 12, pp. 3142–3149, Dec. 2017.
 [3] C. B. Anagnostopoulos, A. Tsounis, and S. Hadjiefthymiades, "Context Awareness in Mobile Computing Environments," *Wirel. Pers. Commun.*, vol. 42, no. 3, pp. 445–464, Jun. 2007.
 [4] Y. Vanrompay, T. Smits, and Y. Berbers, "Efficient Adaptation Decision Making Algorithms for Context-Aware Applications," *Int. J. Adapt. Resilient Auton. Syst.*, vol. 1, no. 3, pp. 42–58, Jul. 2010.
 [5] K. Benzekki, A. El Fergougui, and A. ElBelrhiti ElAlaoui, "A Context-Aware Authentication System for Mobile Cloud Computing," *Procedia Comput. Sci.*, vol. 127, pp. 379–387, 2018.
 [6] R. Hasan, Md. M. Hossain, and R. Khan, "Aura: An IoT Based Cloud Infrastructure for Localized Mobile Computation Outsourcing," in *2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, San Francisco, CA, USA, 2015, pp. 183–188.
 [7] A. Moon, H. Kim, H. Kim, and S. Lee, "Context-Aware Active Services in Ubiquitous Computing Environments," *ETRI J.*, vol. 29, no. 2, pp. 169–178, Apr. 2007.
 [8] H. Mcheick, "Modeling Context Aware Features for Pervasive Computing," *Procedia Comput. Sci.*, vol. 37, pp. 135–142, 2014.
 [9] M. A. Feki, M. Mokhtari, and I. Khalil Ibrahim, "A Novel Approach for Ontology Distribution in Ubiquitous Environments," *Int. J. Web Inf. Syst.*, vol. 2, no. 3/4, pp. 225–231, Sep. 2007.
 [10] D. Guessoum, M. Miraoui, and C. Tadj, "Contextual case-based reasoning applied to a mobile device," *Int. J. Pervasive Comput. Commun.*, vol. 13, no. 3, pp. 282–299, Sep. 2017.
 [11] Deutschland, Ed., *Pervasive computing: trends and impacts*. Ingelheim: SecuMedia, 2006.
 [12] B. Yuan and J. Herbert, "Context-aware hybrid reasoning framework for pervasive healthcare," *Pers. Ubiquitous Comput.*, vol. 18, no. 4, pp. 865–881, Apr. 2014.
 [13] R. Bajaj, S. L. Ranaweera, and D. P. Agrawal, "GPS: location-tracking technology," *Computer*, vol. 35, no. 4, pp. 92–94, 2002.
 [14] A. Bröring, S. K. Datta, and C. Bonnet, "A Categorization of Discovery Technologies for the Internet of Things," in *Proceedings of the 6th International Conference on the Internet of Things - IoT'16*, Stuttgart, Germany, 2016, pp. 131–139.
 [15] S. Schelter, C. Boden, and V. Markl, "Scalable similarity-based neighborhood methods with MapReduce," in *Proceedings of the sixth ACM conference on Recommender systems - RecSys '12*, Dublin, Ireland, 2012, p. 163.
 [16] M. Nappi, S. Ricciardi, and M. Tistarelli, "Context awareness in biometric systems and methods: State of the art and future scenarios," *Image Vis. Comput.*, vol. 76, pp. 27–37, Aug. 2018.
 [17] A. Castiglione, K.-K. Raymond Choo, M. Nappi, and S. Ricciardi, "Context Aware Ubiquitous Biometrics in Edge of Military Things," *IEEE Cloud Comput.*, vol. 4, no. 6, pp. 16–20, Nov. 2017.
 [18] A. Sarkar and B. K. Singh, "A cancelable fingerprint biometric based session key establishment protocol," *Multimed. Tools Appl.*, Mar. 2019.
 [19] M. Pearce, S. Zeadally, and R. Hunt, "Assessing and improving authentication confidence management," *Inf. Manag. Comput. Secur.*, vol. 18, no. 2, pp. 124–139, Jun. 2010.
 [20] D. Preuveneers and E. Ilie-Zudor, "Special Issue: Big Data for context-aware applications and intelligent environments," *Future Gener. Comput. Syst.*, Nov. 2018.
 [21] M. Donohoe, B. Jennings, and S. Balasubramaniam, "Context-awareness and the smart grid: Requirements and challenges," *Comput. Netw.*, vol. 79, pp. 263–282, Mar. 2015.
 [25] Phiri, Lighton, "Research Visibility in the Global South: Towards Increased Online Visibility of Scholarly Research Output in Zambia", 2018

Vehicle Tracking System with Passive Security using Pythagoras theorem (VTSPS)

Matengele Kaira
 School of Graduate Studies
 Copperbelt University
 Box 21692, Jambo Drive, Kitwe, Zambia
 matengele.kaira@mopani.com.zm

Jameson Mbale
 School of Information Communication Technology
 Copperbelt University
 Box 21692, Jambo Drive, Kitwe, Zambia
 jameson.mbale@gmail.com

Abstract— The REST based Vehicle Tracking System with Passive Security using Pythagoras theorem is a low cost vehicle tracking system that is adept for tracking in regions that experience poor network availability (this is the case in Zambia). Most tracking systems require active input from a user to repeatedly send commands to the vehicle unit usually through SMS to get back coordinates which are displayed on a map. The difference with the proposed approach is that the coordinates which will be polled at a specified interval will be cached on local storage and then a RESTful service will be invoked by a C program running on an Arduino microcontroller at a specified interval, this will then log all previously collected coordinates into a database on a server allowing for replay of the movements of the vehicle. Furthermore the proposed system will utilize Pythagoras theorem implemented on the microcontroller in order to detect when a vehicle has travelled beyond an allowable radius.

Index Terms—Internet Of Things, Pythagoras Theorem, Representational State Transfer, Embedded Software, Distributed Systems, Geometry, Latitude, Longitude.

INTRODUCTION

Currently, a large number of the general public own private vehicles, these vehicles are usually at high risk of theft as most of the public rely on traditional safety features like the door lock and the police check points. As thieves become more conversant with how to break into vehicles and evade check points when on the run with stolen vehicles there is a need for the regular person to invest in a vehicle security system that is both affordable and effective given our environment in Zambia. Network connectivity is not everywhere and so existing tracking systems that rely on SMS communication between the vehicle unit and the monitoring unit or have server side or end user initiated security features can be ineffective in our environment. Furthermore all reviewed systems rely on active input by the stolen vehicle's owner to activate additional security features like cutting the engine and locking the doors. In most cases when a car is

stolen the owner is probably unaware, either because he or she is asleep at the time of the incident or it could be the thieves have been profiling the owner and know he or she is out of town or the country, the proposed system takes security a step further by implementing passive security where the owner of the vehicle can limit the radius the vehicle is allowed to travel on a map, for example the airport car park. The moment the vehicle is stolen and the thief attempts to drive outside the allowed radius the engine automatically cuts and the doors lock thereby deterring the efforts of the thieves. Additionally notifications can be sent to the owner and the authorities that the car is being stolen.

REVIEW OF VEHICLE TRACKING SYSTEMS

A. Vehicle Tracking and Locking System Based on GSM and GPS

In the proposed work, a novel method of vehicle tracking and locking system is used to track the stolen vehicle by using GPS and GSM technology [1]. This system goes into sleep mode while the vehicle is being handled by the owner or authorized person otherwise it goes to active mode. The mode of operation can be changed manually in person or remotely. The operation is such that if any interruption occurs on either side of the door, then the Infra-Red sensor senses the signals and sends an SMS to the microcontroller. The controller issues the message about the location of the vehicle to the car owner or authorized person. When the owner sends an SMS to the controller, it issues the control signals to the engine motor. Engine motor speeds are gradually decreased and come to rest. After that, all the doors are locked. To open the door or restart the engine, an authorized person needs to enter the password. Using this method, tracking of vehicle's location is easy and since doors lock automatically, the thief cannot get away from the car without drawing attention to himself by breaking the windows for example.

B. GPS based Advanced Vehicle Tracking and Vehicle Control System

A vehicle tracking system is an electronic device that tracks the vehicle's location. Most of the tracking systems use a GPS module to locate the vehicle's position. Many systems also combine communication components such as satellite transmitters to communicate the vehicle's location to a remote user. Google maps are used to view the vehicle's location [2]. The design of the tracking system is divided into three parts; basic design, intermediate design and an advanced Design. The basic design of the vehicle tracking system consists of a GSM module, a GPS module, a MCU (ATMEL), a Relay circuit and a LCD. The user sends an SMS and the system responds to the user's request by providing the coordinates of a location in accordance to the requirements of mobile phone users through the GPRS network. The intermediate and advance design is an improvement of the basic design. There are five features introduced in the project. SMS codes are specifically assigned to each of these features. For example, if the user sends „555“ to the tracking system. The GSM modem will receive the SMS and transmit to the MCU unit, where the SMS code will be compared against the codes stored in the library. In this project, the code “555” is assigned to find the location of a vehicle. So, the MCU will get the location from the GPS module and reply back to the user with the location coordinates (i.e. Longitude and Latitude). These coordinates can be used to view the location of a vehicle on Google maps. This vehicle tracking system comprises of a cost effective and special tracking technology. It offers advanced tracking and a variety of control features that facilitate the monitoring and clever control of the vehicle. This tracking system is not only bounded to the shipping industry and fleet tracking but can also be used in cars as a theft prevention tool.

C. GPS and GSM based Vehicle Tracking and Employee Security System

In today’s world, security is a very important aspect for the working professionals. It is the company’s utmost responsibility to take care of its employee’s safety. In order to deal with problems of employee safety, an innovative solution of “GPS & GSM Vehicle Tracking and Employee Security System” was designed [4]. This system detects outsiders as well as locate the cars through RFID & GPS mechanisms. An emergency button for employee’s safety was incorporated. Wherever an employee finds himself/herself in trouble, he/she will press the button and an SMS will be sent to the nearest police station and the company’s monitoring unit so that they will take the necessary action. In this project, a GSM and GPS based system for Vehicle Tracking and Employee Security System was proposed. It consists of a car unit, emergency button and company unit. Car unit is placed inside the car. When the car picks up the employee; he/she needs to swap the RF card. The micro controller matches the RF card number with its database records and sends the employee’s id, cab id and the cab position co-ordinates to the company unit via GSM module, an emergency button is a part of the car unit. There are three to four emergency buttons in the car. These buttons are placed at such positions so that employees can access them easily i.e. near the door unlocking handle. If an employee finds himself/herself in trouble, he/she can press the

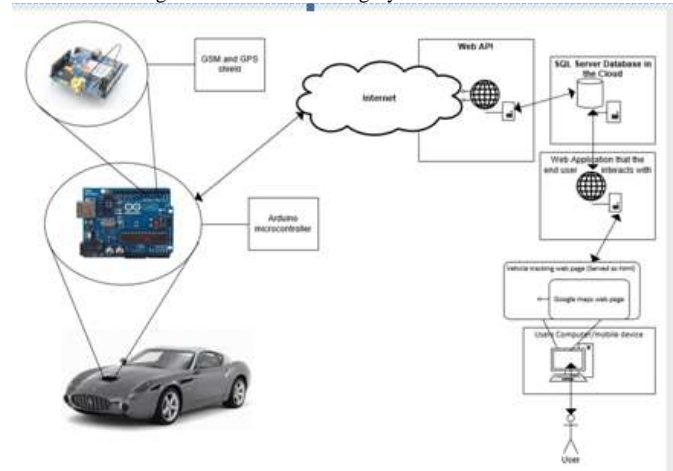
button. The microcontroller will detect the action and send a signal to the GSM which will coordinate with to the company unit and police. The microcontroller will also send a signal to the relay which will turn off the car ignition and stop the car. The company unit consists of GSM modem, RS232 cable and a computer. The GSM Modem will receive the message through GSM. This message will then be transferred to the computer through the serial port. The employee name, employee id and the cars position coordinates (longitude and latitude) get displayed on computer

THEORETICAL AND CONCEPTUAL FRAMEWORK

D. The VTSPS Architecture

The VTSPS is composed of the following functional components: the Vehicle Unit, RESTful Web API, and the Web application and database as demonstrated in Figure 1.

Figure 1. Vehicle Tracking System Architecture



The Vehicle Unit has three functional components and these include the micro-controller, GPS and GSM shields. The micro-controller is the brain of the Vehicle Unit and its function is to process the embedded instructions, to poll for GPS coordinates and initiate a client side call over the internet to send the coordinates to a RESTful Web API. The micro controller used for this research is the Arduino UNO. The GPS shield is responsible for providing the coordinates that indicate the location of the vehicle. The GPS shield used for this work is an Adafruit Fona 808 GPS shield. The third component is the GSM shield, which is used to transmit the coordinates polled by the microcontroller using the GPS shield over the internet. Like the GPS shield, the GSM shield used in this work is an Adafruit Fona 808 GSM shield.

E. Functional Mechanism of the VTSPS

The general functional mechanism of the VTSPS vehicle unit is further explained. First the current coordinates of the car are polled at specified intervals. Second, the coordinates are buffered on local storage in case the server is unavailable. Third, is to add some identification information at the time the coordinates were polled for that particular vehicle and then send the data over the internet to the server for persistence. In addition to sending coordinates, the Vehicle Unit also has the

function of implementing passive security. For instance, the Vehicle Unit has a lightweight efficient radius algorithm utilizing Pythagoras Theorem that calculates the bounds beyond which the vehicle must not exit. If the vehicle attempts to exit this boundary, the unit triggers the passive security features such as locking the doors and cutting the fuel supply to the engine. In view of this, the system has the capacity to notify the authorities and the owner of the ongoing theft by sending messages to them.

The RESTful Web API is responsible for receiving the coordinates from the vehicle unit and persisting them to a database on a server. It is imperative to first understand and define the concept REST. REST, an acronym for Representational State Transfer is a lightweight technology used for interoperability enabling technologies that under normal circumstances would not interoperate to do so. REST allows a client such as a thin client to invoke functionality usually programming logic that is hosted remotely on a server by invoking http links with the URL format mapped onto methods on a remote server. REST is a stateless technology and therefore is very light on resource requirements. The web API also provides an interface off which the vehicle unit can make calls to pick up messages from the server. This allows a distributed architecture of sending messages from the user to the vehicle unit. The user simply persists a message into the database on the server using a web application and waits for a separate operation originating from the vehicle unit to pick up any new messages and respond accordingly. Apart from REST, another critical technology in this work is JSON. JSON is an acronym for JavaScript Object Notation and is a lightweight data-interchange format, most of the web technologies picked for the API are lightweight so as to allow a client to communicate with the server even with very little bandwidth at a very low cost. The data is sent to the server as JSON and the result of the invoked method if it returns anything is returned as JSON. Both REST and JSON are open standards and can be used by almost every programming language or technology out there. Some of the operations that are implemented on the server are; first, within the Values Controller, a get command is used to log coordinates to the database by invoking a “SendCoordinates” method; second within the messages controller, a get command provides a JSON string of different messages for the vehicle units from the server such as the starting coordinates and the radius value used to calculate the allowed coordinates. The web API also logs any exceptions encountered when executing server side code such as any errors when posting coordinates into the database. This helps the client with knowing what to persist locally on its end and what to clear off so it isn’t synced again by the vehicle unit, the REST API responds with a true when coordinates are successfully posted and a false otherwise. The web application and its database is the part of the system that the end user interacts with, this is the GUI app that a user can use to see where his car is and to track its movements. The web application can also be used to take input from the user and persist that as instructions for the vehicle unit in the database.

METHODOLOGY

A. *The Simulation Approach/Software Production*

The approach used in this research is the Simulation/Prototyping approach which ultimately results in software production of an evolutionary prototype that’s further refined with each iteration. This prototype is then tested and results are analyzed. Depending on the results the next course of action in line with the functionality requirements highlighted in the conceptual framework is determined. An evolutionary prototype is one which eventually becomes the finished product. The following components will be built and tested fist in isolation and then together during integration testing;

The GPS module, this module polls coordinates at a specified interval with values such as latitude, longitude, direction, speed, and altitude. The GPS module uses a SIM 808 GPS module with an antenna a LIPOLY battery to be used a power supply. The SD Card module, this module is responsible for persisting the results returned from the GPS module to the SD Card connected to the Arduino microcontroller. The GSM Client, this module reads the data persisted onto the SD Card and sends it over the internet to a web server using the GPRS network, it also erases synced lines of data off the SD Card. The Allowable Radius Algorithm, this module utilizes Pythagoras theorem with a few alterations to calculate distances on the ground from GPS coordinates to determine if a car has travelled further than allowed in any given direction. The tracking web application, this module consists of a web app that utilizes google maps JavaScript API’s to serve mapping capabilities to clients. The RESTFUL web API, this is the module that can receive data from a primitive client and also serve data to a primitive client on request. It can support both the GET and POST HTTP commands.

B. *Pythagoras Theorem Implementation*

Consider a map with Cartesian coordinates in degrees represented by x and y . let’s assume the x coordinates represent longitudes and are the horizontal axis while the y coordinates represent latitudes and are the vertical axis. If the vehicle is at a starting position of x_1, y_1 and then moves to a new position x_2, y_2 , we are tasked to calculate how far the vehicle is from its original starting point [6]. The challenge with this is a vehicle can travel in any given direction from the starting point, visualize x_1, y_1 as the center of a circle of radius 100. The vehicle could now be at any location within the circle any direction from the starting point but within 100 units of the starting point as long as it is within the circle. Another challenging fact to consider is when the vehicle is now at x_2, y_2 we know the new Cartesian coordinate but we need to convert those to meters or kilometers on the ground as that’s what the users of the system will use to restrict the distance the vehicle can travel. For example a user might set his vehicle not to travel more than 100 meters from where it is parked because he/she knows it is at an airport parking lot and it must not exist that parking lot while he/she is away. To solve these problems we use the Pythagoras theorem but we must add something to it so we can convert from Cartesian coordinates of degrees to meters or kilometers. For a car that

has travelled in any direction, the difference between the current position and the starting position in Cartesian coordinates is hypotenuse of the triangle formed calculated by the square root of the square of the difference between the x coordinates added to the square of the difference between the y coordinates illustrated as distance travelled = $\text{Sqrt}((x_2-x_1)^2 + (y_2-y_1)^2)$. Two other challenges that quickly come up when a decision to use Pythagoras theorem to calculate the distance between two points on a map are;

1. The GPS coordinates that are polled by the microcontroller that serve as input coordinates to the formula are in degrees and
2. Due to the earth not being a perfect sphere, as you move away from the intersection of the equator and the prime meridian there is some margin of error that's introduced into the calculations.

To cater for the above two challenges, the coordinates polled are converted to radians which represent a length of an arc and when the arc length and radius of the circle are equal that signifies 1radian. To convert the radians to distances on the ground we multiply the radians by 6,371km which is approximately the radius of the earth. This enables us to apply Pythagoras theorem and arrive at an answer representing a distance in meters. The second challenge or error can be solved by multiplying the cosine of the radian average of the two latitudes by the difference in longitudes. Refer to the code sample below for a testable C++ implementation of the Pythagoras theorem for travelled distance calculation based on GPS coordinates specified in degrees.

```
#include "pch.h"
#include <iostream>
#include <cmath>
using namespace std;

bool CalculateAllowableRadius(float x1, float y1, float x2, float y2, float R, float& x, float& y, float& d)
{
    const float PI = 22 / 7;
    //multiplying by PI/180 converts degrees to radians
    x = ((PI / 180) * (x2 - x1)) * cos((PI / 180) * ((y1 + y2) / 2));
    y = ((PI / 180) * (y2 - y1));
    d = sqrt(pow(x, 2) + pow(y, 2)) * R;

    return true;
}

int main()
{
    float x1, x2, y1, y2, R, x, y, d;
    // where x1, x2 are longitudes, y1, y2 are latitudes,
    // x and y are the differences between the longitudes and latitudes respectively
```

```
// R is the radius of the earth in kilometers and d is the hypotenuse in kilometers

cout << "enter longitude 1 : ";
cin >> x1;
cout << "enter latitude 1 : ";
cin >> y1;
cout << "enter longitude 2 : ";
cin >> x2;
cout << "enter latitude 2 : ";
cin >> y2;
R = 6371;

bool success =
CalculateAllowableRadius(x1,y1,x2,y2,R,x,y,d);
cout << "difference between longitudes in Radians : " << x << endl;
cout << "difference between latitudes in Radians : " << y << endl;
cout << "distance travelled in kilometers : " << d << endl;
}
```

C. Prototype Implementation (Individual Components)

The implementation of the product was tackled in two, iterations, the first was to assemble, configure and write code to enable the individual components to function and the next was to integrate the individual components after they have been tested to ensure they function as needed. In the first iteration the individual components that make up the vehicle unit were assembled. The vehicle unit is made up of an Arduino Uno microcontroller, a SIM808 GPS and GSM shield, Plain headers, slim sticker-type GSM/Cellular Quad-Band Antenna - 3dBi uFL, Passive GPS ANtenna uFL - 15mm * 15mm 1 dBi gain and Lithium Ion polymer battery 3.7v 1200mAh.

The vehicle unit components all come with sample code that can be used to test that they function correctly. The sample code libraries were then installed on the development computer and deployed module by module to the microcontroller and each component tested to determine it works correctly standalone. The modules tested include polling for GPS coordinates, writing to and reading from an SD Card and making an HTTP get request. The C algorithm that was written to calculate distance travelled given a starting point was tested standalone in a visual studio as a C++ console application to make it easy to work with user input for coordinates and output to the console window for feedback. It is this algorithms result that is compared with the allowable radius to detect a vehicle that is being stolen. Further to that the RESTFUL Web API was implemented and the client side code that would be used for tracking vehicles was also implemented. The tracking code instantiated a google maps map and placed some markers in arbitrary positions and shifted the markers by a certain offset repeatedly after a specified number of seconds thereby validating that it can asynchronously update the position of the vehicles being tracked.

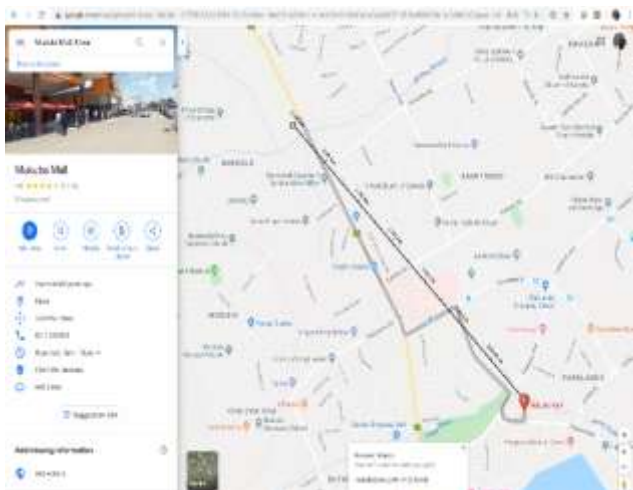
RESULTS

A. Architectural and Design based results

The VTSPS envisages the use of data instead of SMSs and also buffers coordinates locally on the controller and repeatedly retries at a defined interval. The VTSPS, sends the coordinates using only a few bytes of data per transmission and a few hundred bytes per minute. This calculates to a few kilo bytes a day which is almost of negligible cost compared to the SMS system. It further distributes the ability to detect theft to the vehicle units and does not rely on a central server or the user. In case the vehicle has been stolen and cannot be tracked currently, it allows for the history of the vehicles movements to be replayed. The replaying of Global Positioning System (GPS) coordinates is possible because the unit polls and buffers GPS coordinates for sending every 1 minute meaning as soon as network is available all the unsent coordinates are transmitted. The result is a system that accounts for the whereabouts of each vehicle every 1 minute and this data is persisted on a server and can be used days or weeks or months later if needed.

B. Pythagoras Theorem Test Results

The algorithm was put to the test to determine its level of accuracy for both short distance (within the same town) and long distance (across provinces) calculation of distance travelled by a car. The coordinates for both Copper Hill mall and Mukuba mall were searched for on google after which google maps was used to show directions from copper hill to Mukuba mall. Since the directions follow the road network Mukuba mall was right clicked and on the context menu was an option to select the starting point to measure distance from. Copper Hill mall was selected and google maps drew a straight line between the two points with a distance of 3.48km as shown by the figure below.

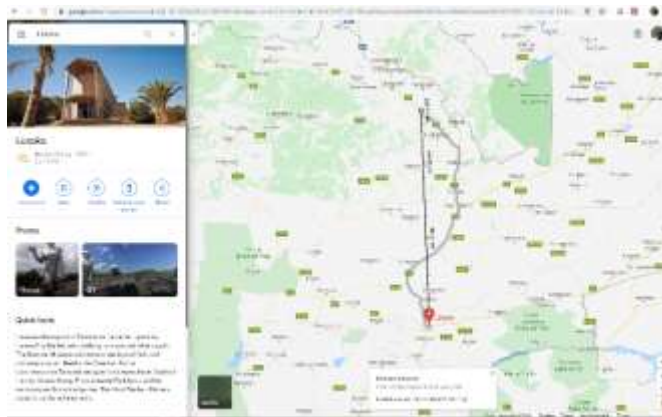


For the short distance tracking test the coordinates of Copper Hill mall and Mukuba mall were used. The coordinates of Copper Hill mall and Mukuba mall respectively are -12.7843, 28.1967 and -12.8070, 28.2204. After entering those coordinates into the pythagoras theorem program the output is as shown below 3.48474km which when rounded off to two

decimal places yields 3.48km.



For the long distance test the google maps coordinates of towns Kitwe and Lusaka were used. The process followed to get the straight line distance between the two towns in google maps was repeated and the resulting distance was 268.87km as shown below.



The coordinates of Kitwe and Lusaka towns respectively are -12.8232, 28.2176 and -15.3875, 28.3228. After entering these coordinates into the pythagoras theorem program, the output is as shown below 272.503km which when rounded off to two decimal places yields 272.50km.



As can be seen from the results above the algorithm's results are satisfactory with almost no variance on short distance tracking and an error margin of error of about 1.35% which is more than acceptable.

In addition, the employed Pythagoras theorem does not require much memory or processing power for a vehicle unit to utilize it as opposed to other systems discussed earlier on that rely on server side processing. The proposed design decouples the ability to decide whether or not the vehicle is being stolen from the server or server side application by distributing it to each vehicle tracking unit thereby reducing the need for a very high spec server as the number of vehicles increases making the VTSPS very scalable.

DISCUSSION AND CONCLUSION

The purpose of this study was to apply Pythagoras theorem to Vehicle Tracking as a solution for distributing geo fencing capabilities to microcontrollers thereby making the proposed system more resilient to network or communication lapses. The following are some key takeaways from this paper.

A. REST Client

A vehicle unit which acts as a rest client was created, REST was chosen because it enables even the most primitive of clients to invoke programming logic on a remote server application. REST further allows for very low data footprint when messages are sent over the network from the client to the server making the cost of data transfer very low allowing for more frequent data transfer. In this paper a REST based client communicating over HTTP was implemented on the vehicle unit as an alternative to communicating with the server when sending coordinates for server persistence.

B. Pythagoras Theorem

One of the most challenging aspects of this research was to figure out how to determine if a vehicle has left its allowable radius when parked at an airport or home or is supposed to operate only within a certain town for example as this is the trigger that sets off all other passive security features. Part of the challenge is that the vehicle can head in any direction so the algorithm must be able to detect distance covered in either direction from the starting point. Another challenge was that the information polled by the GPS shield is in longitudes and latitudes not meters and kilometers, hence converting those positions represented by the GPS coordinates into distances on the ground in comparison to a starting point. While this task in itself poses quite a challenge, there was the added complexity of performing this computation on a device that has low processing speeds and capabilities as that way the computation is independent of connectivity to the server. Pythagoras theorem was chosen for this purpose as the difference in coordinates between the longitudes and the latitudes signifies the two adjacent sides of a triangle and the hypotenuse the current distance from the starting point, to convert this to distances on the ground the degree coordinates were converted to radians and then, a cosine function was applied to the average of the latitudes and multiplied by the radian difference in longitudes. The radius of the earth 6,371km was multiplied by the final answer to convert it to km from radians which are an arbitrary length unit. This allows even a primitive client device to calculate in all directions if the vehicle has left its allowable radius in relation to its starting point.

C. Coordinate Buffering

Another key design consideration for areas of poor connectivity is caching, the system should be able to cache data when unable to communicate and attempt to resend at a later time. This design approach allows for the system to operate seamlessly in areas of poor connectivity and to catch up by sending all missing coordinates to the server. This makes it possible for investigators using the system to carry out coordinate replay and see everywhere the vehicle has been in the last 48 hours for example at a specified interval of every 1 hour for example.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Doctor Josephat Kalezhi for his willingness to assist with information about the format and structure of IEEE papers which influenced the structure of this paper.

REFERENCES

- [1] Ramanil, R. Valarmathy, S. Vanitha, N. S. Selvaraju, S. Thirupathi, M. Thangam, R. (2013). Vehicle Tracking and Locking System Based on GSM and GPS. Tamilnadu: I.J. Intelligent Systems and Applications.
- [2] Mukhtar, M. (2015). GPS based Advanced Vehicle Tracking and Vehicle Control System. London: I.J. Intelligent Systems and Applications.
- [3] ElShafee, E. ElMenshawi, M. Saeed, M. (2013). Integrating Social Network Services with Vehicle Tracking Technologies. Egypt: International Journal of Computer Applications.
- [4] Pethakar, S, S. Srivastava, N. Vanitha, N. S. Suryawanshi, S. D. (2013). GPS and GSM based Vehicle Tracing and Employee Security System. Pune: International Journal of Computer Applications.
- [5] Maliti, B. Taylor, T. (2008). Contemporary Research Handbook. Kitwe: The Copperbelt University.
- [6] Veness, C. (2019). Calculate distance, bearing and more between Latitude/Longitude points. Cambridge: <https://movable-type.co.uk/scripts/latlong.html>
- [7] Wiegers, K, E. (1999). Software Requirements Specification. Retrieved at: https://web.cs.dal.ca/~hawkey/3130/srs_template-ieee.doc

Exploring High Spatial Resolution Sentinel-2 Imagery for Land Use and Land Cover Mapping - A Case of Kopa area in Kanchibiya District, Zambia

Kennedy Kanja¹, Chimanga Kashale²

¹Department of Agriculture and Aquatic Sciences, Copperbelt University-Kapasa Makasa University Campus, Chinsali, Zambia, ² Department of Information and Communications Technology, Copperbelt University-Kapasa Makasa University Campus, Chinsali, Zambia

¹kennedy.kanja@cbu.ac.zm, ²kashalechimanga@gmail.com

Abstract—Accurate land cover mapping is an important component in land use land cover change analysis. Land cover mapping is accomplished using Remote Sensing by way of image classification. With recent developments in Remote Sensing, freely available high resolution images are now available with a resolution of as high as 10m. This study therefore explored the potential of the freely available Sentinel-2 imagery for land cover classification. Sentinel-2 image was downloaded from ESA website and Landsat-8 OLI image was downloaded from USGS website. Random Forest (RF) algorithm was used to classify the two images. The overall classification accuracy was 60% and 72% for landsat-8 OLI and Sentinel-2 imagery respectively. The Kappa indexes were 52% for Landsat-8 OLI and 61% for Sentinel-2 imagery. These results proves that a higher classification accuracy can be achieved by using 10m resolution bands from Sentinel-2 imagery. This implies that land use land cover change analysis can be achieved more accurately and at a finer scale using Sentinel-2 imagery. Land use land cover change analysis is an important component in climate change analysis.

Keywords— Remote Sensing, Land use land cover mapping, Sentinel-2, Random forest, Zambia

I. INTRODUCTION

Land cover refers to the natural and man-made coverings on the earth's surface such as vegetation, soil and water; while land use is considered as the exploitation of the Earth's biophysical attributes by humans for a particular purpose [1]. Land use and Land cover changes (LULCC) are among the most significant human modifications to the earth's terrestrial surface [2]). Although anthropogenic LULCC has been in existence since time immemorial, research indicates that it has been more rampant in the past decades and this can be attributed to the growth in human population coupled with developmental projects [3]). Clearing of natural vegetation may predominantly modify the vegetation cover and in turn have a long-term impact on sustainable food production, forest resources and climate change [4]). Large amounts of natural vegetation (tropical forest/woodlands and savannas/grasses) have been cleared over the years for agricultural purposes in response to increasing population and

the need to produce more food. It is therefore important that LULCC are closely monitored time and again in order to create a balance between agriculture expansion and environmental degradation [5].

Monitoring land use land cover changes using satellite images has been gaining attraction since the past two decades. Remote sensing by way of satellite imagery is advantageous among the surveying methods owing to their periodic acquisition which enables LULCC analysis to be conducted at regular intervals. For many years, data from various optical sensors (Landsat, SPOT, MODIS, etc.) have been analyzed at different scales to produce agricultural land use maps [6]. Reference [7] used both Quickbird imagery and Landsat 7 ETM+ to derive land cover layers for eastern Zambia and obtained better classification accuracies. Reference [8] used Landsat 8 ETM+ to monitor agriculture expansion in Mwekera Forest reserve Zambia. Recent developments in remote sensing has resulted in high spatial resolution and free satellite images data such as Sentinel-2 imagery which has spatial resolutions of as high as 10m for bands in visible light (Red, Green and Blue) and the near-infrared (NIR) band. Sentinel 2A satellite imagery is composed of 13 spectral bands [9], details of these bands are shown in Table 1. The Copernicus Sentinel-2 mission comprises a constellation of two polar-orbiting satellites placed in the same sun-synchronous orbit, phased at 180° to each other. It aims at monitoring variability in land surface conditions, and its wide swath width (290 km) and high revisit time (10 days at the equator with one satellite, and 5 days with 2 satellites under cloud-free conditions which results in 2-3 days at mid-latitudes) will support monitoring of Earth's surface changes [15].

There is need therefore to explore the effectiveness of Sentinel-2 imagery in producing finer land cover maps for purposes of LULCC analysis. This study therefore aims at exploring higher resolution Sentinel-2 imagery for purposes of producing finer land use land/cover maps and compare it with the Landsat-8 operational land imager (OLI) imagery.

II. STUDY AREA

This study was conducted in Chief Kopa’s Chiefdom which is located in the newly created Kanchibiya District which used to be part of Mpika District (Fig. 1). It was selected owing to the fact that a lot of land clearing for agriculture expansion has been happening through the ZamPalm plantation project started by ZAMBEEF and now under industrial development cooperation (IDC). Kopa area is also well known for caterpillar picking (finkubala) where trees are cut down for purposes of accessing these caterpillars. The area caters for both grassland and forests.

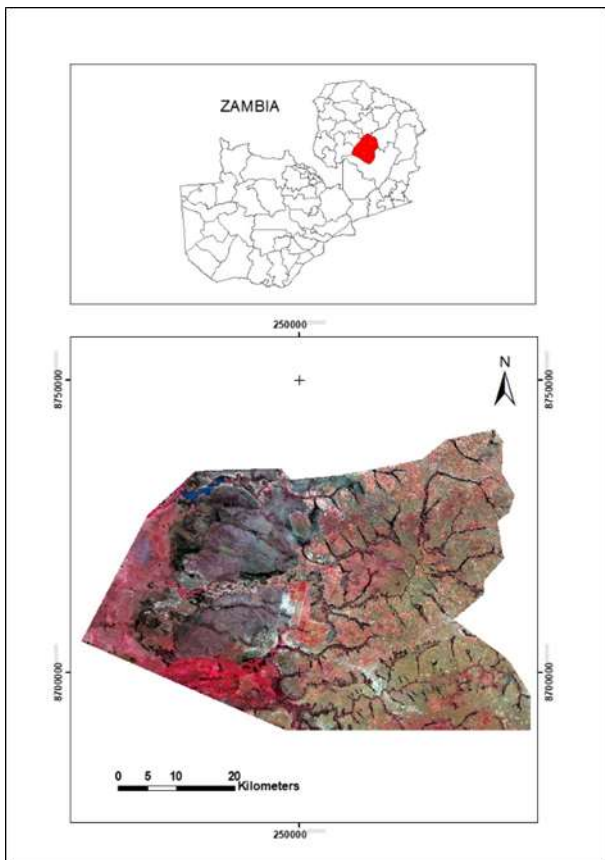


Fig. 1 Colour infrared composite (8 4 3) of study site area

TABLE I
SENTINEL-2 MULTISPECTRAL BANDS

Band Number	Band Name	Spatial Resolution (m)	Central wavelength (nm)
B1	Coastal/aerosols	60	443
B2	Visible Blue	10	490
B3	Visible Green	10	560
B4	Visible Red	10	665
B5	Red edge 1	20	705
B6	Red edge 2	20	740
B7	Red edge 2	20	783
B8	NIR	10	842
B8a	Narrow NIR	20	865
B9	Water vapour	60	945
B10	SWIR Cirrus	60	1375
B11	SWIR 1	20	1610
B12	SWIR 2	20	2190

III. METHODOLOGY

Sentinel 2A satellite imagery was obtained from European space agency (ESA)’s Copernicus Open Access Hub website (<https://scihub.copernicus.eu/dhus/#/home>). Landsat-8 OLI satellite imagery were obtained from the United States Geology Survey (USGS) Earth Explorer website (<https://earthexplorer.usgs.gov/>). Training classes were obtained by visual interpretation of very-high-resolution satellite data of Google Earth. For both sentinel-2 and landsat-8 OLI images, only the near infrared (NIR), Red (R), Green (G) and Blue (B) bands were stacked together to form a multispectral image that was used for classification by forming a false colour composite band combination (colour infrared). Supervised machine learning method random forest (RF) was used to classify the images.

The Random Forest (RF) is a successful ensemble classifiers developed based on the concept of learning strategy. It is an ensemble learning method for classification that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes classification of the individual trees. RF is a non-parametric machine learning algorithm which produces high

classification accuracy and capable of determining the variable importance [9].

Accuracy assessment was performed on the classified images. This is usually performed to determine how well the classification process accomplished the task. This task was accomplished by compiling error matrices. An error matrix is a table of values that compares the value assigned during classification process to the actual value from an aerial photo. Cohen’s kappa coefficient was then calculated from the error matrix. Kappa provides insights into the performed classification scheme and whether or not it was achieved better than would have been achieved strictly by chance [10], [11]. It’s given by the formula:
$$\text{Kappa} = \frac{\text{Observed} - \text{expected}}{1 - \text{Expected}}$$

IV. RESULTS AND DISCUSSION

In the current study, Sentinel-2A and Landsat-8 OLI images were acquired on different dates but for the month of September 2019. The bands for these images were stacked and later used for classification. Random Forest (RF) classifier was implemented using R open software. Open software QGIS and licenced ArcGIS were both used for sample trainings. The accuracy of LULC maps is one of the major indicators that determine the quality of the map produced, the fitness for a specific application and an understanding of error and its implications [12]. The performance evaluation overall accuracy and the Kappa coefficient were computed and used in the current study for the two images.

For the Sentinel-2A imagery, at a resolution of 10m, observed accuracy of 72% and Kappa coefficient of 0.61 or 61% were obtained (Table II). The classified image is shown in Figure 2. It was observed that Open Forests achieved the best accuracy followed by water bodies, agriculture fields and grasslands. Palm plantations were also classified and came out clear although there were areas where availability of these palm plantations could not be confirmed with ground truth data. Landsat-8 OLI imagery with a resolution of 30m produced an overall accuracy of 60 % and a Kappa coefficient of 0.52 or 52% (Table II). The classified image is shown in Figure 3. The best accuracy for this image was observed in Wetlands followed by built up areas and open forest. Palm plantations were classified as well though not accurately as this was observed in far areas where palm fields have not yet been established.

TABLE II
CLASSIFICATION ACCURACIES

	Landsat-8 OLI (%)	Sentinel S2A (%)
Overall	60	72
Expected	16.6	28

d	52	61
Kappa	52	61

Conflict was observed between the classes for grasslands and classes for open forests. This can be attributed to the fact that the images were acquired in September when grasslands were dry making them appear like sprouts. The kappa coefficient and overall accuracy of 0.61 and 72% respectively that was obtained for Sentinel-2 image is similar to that obtained by [13] in his crop type mapping study who obtained Kappa index of 61% and overall accuracy of 70%. Other similar studies are that of [13] and [9] who obtained overall accuracies of 83% and 84.22% respectively using Sentinel 2 imagery and concluded that Sentinel-2 has potential in vegetation mapping domain in remote sensing.

V. CONCLUSION

The current study explored the potential of the freely available high resolution Sentinel-2 imagery in fine scale land use land cover mapping. Landsat-8 Operational Land Imager and Sentinel-2 imagery were compared. The overall classification accuracy was 60% and 72% for Landsat-8 OLI and Sentinel-2 imagery respectively. The Kappa indexes were 52% for Landsat-8 OLI and 61% for Sentinel-2 imagery. These classifications were attained using the random forest (RF) classification algorithm. Land-use and land-cover maps are important when it comes to land use land cover change (LULCC) analysis. They are also necessary in agriculture for monitoring crop field sizes which can help in crop harvest surveys and in crop identification surveys. Land use land cover maps also aid in the continuous monitoring of agriculture expansion vis-à-vis forest ecosystem degradation, something that requires balance so that the effects of climate change can be mitigated. The results of this study have shown that Sentinel-2 imagery at 10m resolution gives finer scale classified land use land cover maps over the 30m resolution Landsat-8 OLI images. However, lack of cloud free Sentinel-2 images during the rainy season might be a hindrance if one wants to perform crop classification.

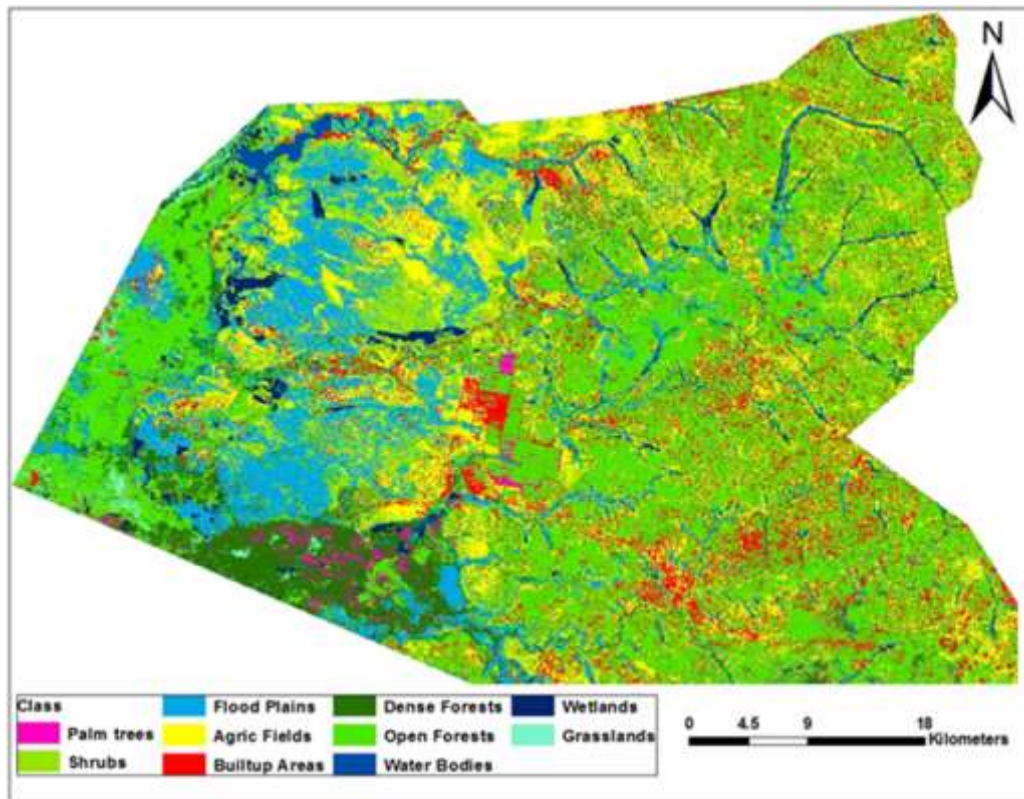


Fig. 2 Land Cover map obtained using Sentinel 2A (10m) imagery

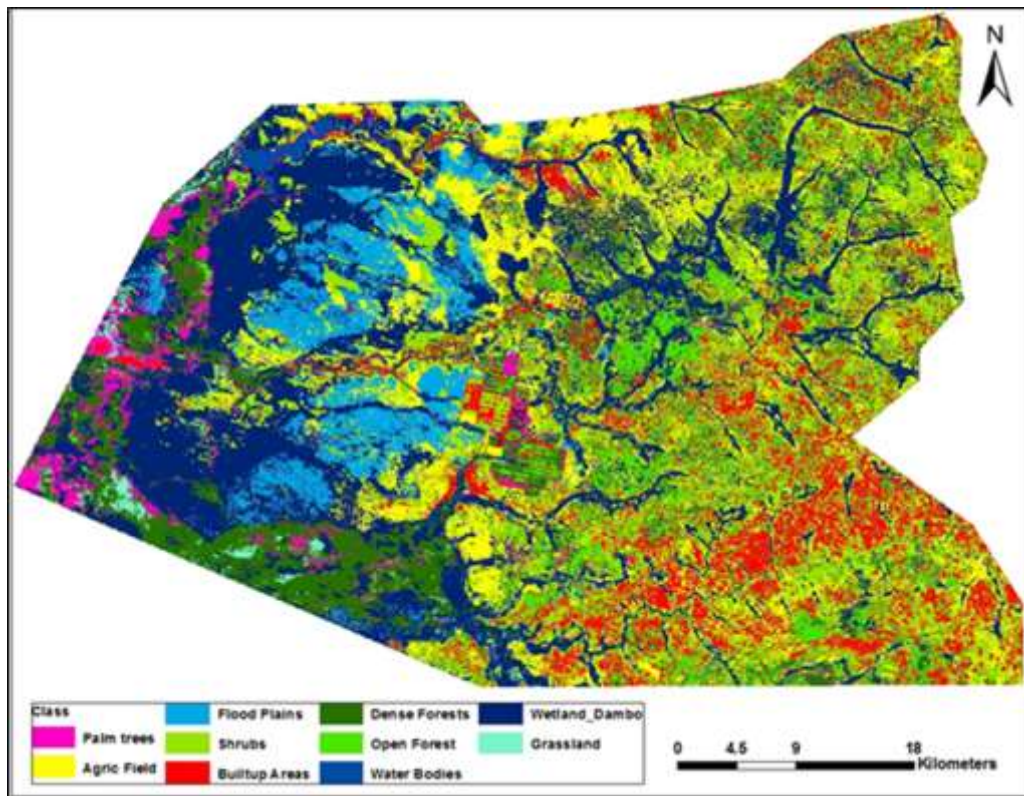


Fig. 3 Land Cover map produced from Landsat-8 OLI (30m) imagery

REFERENCES

- [1] P.G. Chandra, *Remote Sensing of Land Use and Land Cover: Principles and Applications*, CRC Press Published May 2, 2012 Reference - 477 Pages - 65 Color & 136 B/W Illustrations ISBN 9781420070743 - CAT# 70746. 1st Edition
- [2] E.F. Lambin, B.L. Turner, H.J. Geist, et al., The Causes of Land-Use and Land-Cover Change Moving beyond the Myths. *Global Environment Change*, 11, 261-269, 2001. [https://doi.org/10.1016/S0959-3780\(01\)00007-3](https://doi.org/10.1016/S0959-3780(01)00007-3)
- [3] E.F. Lambin, B.L. Turner, H.J. Geist, et al., Global Land Use Change, Economic Globalization, and the Looming Land Scarcity. *Proceedings of the National Academy of Sciences*, 108, 3465-3472, 2011. <http://dx.doi.org/10.1073/pnas.1100480108>
- [4] A.B. Brink, and H. D. Eva, Monitoring 25 years of land cover change dynamics in Africa: a sample based remote sensing approach, *Appl. Geogr.* 29: 501– 512, 2009.
- [5] A. Wood, P. Stedman-Edwards, and J. Mang, *The Root Causes of Biodiversity Loss*. World Wildlife Fund and Earthscan Publications Ltd., London, UK, 2000.
- [6] S. Foerster, K. Kaden, M. Foerster, and S. Itzerott, Crop Type Mapping Using Spectral–Temporal Profiles and Phenological Information, *Computers and Electronics in Agriculture* 89: 30–40. doi:10.1016/j.compag.2012.07.015, 2012.
- [7] Y. Her and C. Heatwole, Land use classification in Zambia using Quickbird and Landsat imagery. ASABE Paper No. 072017. St. Joseph, Mich.: ASABE, 2007.
- [8] K. Kanja, M. Mwemba, and K. Malunga, MONITORING THE RATE OF EXPANSION OF AGRICULTURAL FIELDS IN MWEKERA FOREST RESERVE USING REMOTE SENSING AND GIS, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-3/W6, 377-380, <https://doi.org/10.5194/isprs-archives-XLII-3-W6-377-2019>.
- [9] R. Saini and S.K. Ghosh, CROP CLASSIFICATION ON SINGLE DATE SENTINEL-2 IMAGERY USING RANDOM FOREST AND SUPPORT VECTOR MACHINE, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-5, 683-688, <https://doi.org/10.5194/isprs-archives-XLII-5-683-2018>.
- [10] (2019) The Virginia Tec website. [Online]. Available: <http://www.virginiaview.net/education.html>
- [11] J.B. Campbell and R.H. Wynne, *Introduction to Remote Sensing Fifth Edition*. The Guilford Press, New York, 2011.
- [12] G.M. Foody, Status of Land Cover Classification Accuracy Assessment. *Remote Sensing of Environment*, 80, 185-201, 2002. [http://dx.doi.org/10.1016/S0034-4257\(01\)00295-4](http://dx.doi.org/10.1016/S0034-4257(01)00295-4)
- [13] D. Vijayasekaran, SEN2-AGRI – CROP TYPE MAPPING PILOT STUDY USING SENTINEL-2 SATELLITE IMAGERY IN INDIA, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-3/W6, 175-180, <https://doi.org/10.5194/isprs-archives-XLII-3-W6-175-2019>
- [14] N. Puletti, F. Chianucci and C. Castaldi, Use of Sentinel-2 for forest classification in Mediterranean environments. ISSN: 2284-354X, 2017. <http://dx.doi.org/10.12899/asr-1463>
- [15] (2019) The European Space Agency website. [Online]. Available: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>

An Expert System for the Diagnosis of Poultry Diseases

Macmillan Simfukwe

Department of Computer Science and
Information Technology, School of
Science, Engineering and Technology
Mulungushi University
Kabwe, Zambia
msimfukwe@mu.ac.zm

Kondwani Chanda

Department of Computer Science and
Information Technology, School of
Science, Engineering and Technology
Mulungushi University
Kabwe, Zambia
chandakondawani@gmail.com

Christopher Chembe

Department of Computer Science and
Information Technology, School of
Science, Engineering and Technology
Mulungushi University
Kabwe, Zambia
cchembe@mu.ac.zm

Abstract—An expert system is a computer program that can perform tasks that are normally performed by human beings, using intelligence. Expert systems keep knowledge in their knowledge bases and use this knowledge to make inferences and perform reasoning. Expert systems are widely used in various fields such as diagnosis of diseases, diagnosis and repair of machinery, creation of investment portfolios, analysis of the chemical composition of substances, etc. Poultry farming has become a very popular economic activity in Zambia. Thousands of Zambians are engaged in poultry farming as a business, because it is easy to run a poultry business and it also does not require a lot of capital to start. However, the birds, especially the broilers, are susceptible to a variety of diseases, most of which require special veterinary knowledge to correctly diagnose. Unfortunately, there are very few veterinary clinics in Zambia, especially in rural areas. Hence, in times of epidemics, the poultry farmers have nowhere to run to for professional help and end up incurring losses when their birds die. In order to assist the poultry farmers with disease diagnosis and treatment advice, we have developed an expert system that is capable of diagnosing poultry diseases. The aim is to detect the diseases as early as possible in order to treat them early and avoid epidemics. In this paper, we present the expert system for the diagnosis of poultry diseases that we built.

Keywords— *expert system, poultry disease, Prolog, diagnosis.*

I. INTRODUCTION (HEADING 1)

Poultry farming is the process of raising domesticated birds such as chickens, ducks, turkeys, geese and quails for the purpose of both meat and eggs. Generally chickens that are raised for meat purposes are called broilers, whilst those that are raised for the purpose of producing eggs are called layers.

There are various types of poultry farming, but in our study we focused on the raising of broilers and layers. Poultry farming in Zambia has created huge employment opportunities for many people. Thus, there is a growing demand for chicks, related equipment, vaccines and medicines.

It is very easy for diseases to spread among the chickens, owing to the fact that they are crowded together in small and confined spaces. To compound the problem, there are few veterinary centres and specialists in Zambia, where the farmers can go to and seek help for their chickens, and even when such centres are available, many

small scale farmers are unable to access them and afford paying the associated costs. The most prevalent poultry diseases in Zambia were reported by [1].

An expert system is an intelligent computer program that uses knowledge and inference procedures to solve problems difficult enough to require significant human expertise to solve them [2]. Expert systems have been used various fields such as medical diagnosis [3], diagnosis and repair of machinery, diagnosis and troubleshooting of computer faults[4], chemical analysis of substances, etc. Therefore, expert systems can also be adopted for the diagnosis of poultry diseases in order to assist the farmers. In [5], an expert system was reported to assist farmers with the management of poultry diseases. However, this system is based on diseases that are prevalent in Nigeria, and thus does not contain some of the diseases that are critical for Zambia. The same is true for the expert system reported by [6].

In this paper we present an expert system that diagnoses poultry diseases that are very common in Zambia. This paper is organized as follows: In Section 2, we present a background of poultry diseases and expert system, in Section 3, we present the process we followed to develop our expert system, in Section 4, we present and discuss the results of using our expert system and Section 5, we conclude this paper and highlight the future direction of this study.

II. RELATED WORK

A. Poultry Diseases

A detailed, but incomplete presentation on poultry diseases is given in [7]. Nevertheless, the most common poultry diseases in Zambia are presented in [1]. The following are the diseases which we were considered in our study and can thus be diagnosed using our proposed expert system:

1) *Newcastle Disease* (ND): It is caused by the paramyxovirus. All kind of birds can suffer from PMV-1. Infection is spread directly between birds and indirectly via egg handling equipment and people. It is not transmitted from hen to egg. The virus can also be spread through the air about 70 m. Some birds can carry the virus subclinical and wild birds can spread the infection to domestic birds. Details about the epidemiology, diagnosis and control techniques of this disease are presented in [8]. ND has an incubation period of 5-7 days and its symptoms

include laboured breathing with wheezing, paralysis, neck-twisting, disruption of egg production and high mortality. Diagnosis is done by autopsy, serology and virus isolation (with or without determination of pathogenicity by DNA-sequencing and analysis of amino acid sequence). The following are some of the changes noticeable at autopsy; tracheitis, bronchitis, pneumonia, broken ovarian follicles, peritonitis caused by yolk in the abdomen, bleedings in viscera, lymphatic organs and in the glandular stomach, encephalitis and meningitis. ND is not treatable. Therefore, prevention, via vaccination or good hygiene in the flock, is essential for control of this disease. Sometimes, antibiotic treatment does help to reduce the losses via mortality.

2) *Infectious Bursal Disease (IBD)*: It is also known as Gumboro. IBD is caused by birnavirus. The infection affects chickens, ducks and turkeys, but only chickens get clinical symptoms. The incubation period is 2-3 days, and the chickens are most vulnerable to the infection at 3-6 weeks of age, when the bursa of Fabricius is in its largest size. Chickens older than 15 weeks do not get sick because the bursa is then almost completely degenerated. The disease is spread directly and indirectly, but not vertically to the eggs. The virus is very resistant and chickens excrete it in the faeces for up to two weeks after sickness. The Symptoms include watery white diarrhoea, loss of appetite, ruffled plumage, soiled sewer feathers, trembling and closed eyes. The disease quickly leads to death. Almost all birds in a flock get sick, but mortality varies, usually between 0-20 %, depending on the breed, age, immune status and viral virulence. IBD reduces the capability of the immune system and increases susceptibility to other diseases. Gumboro is diagnosed by autopsy, immunohistochemistry and serology. At post mortem a swollen bursa of Fabricius is seen, which is oedematous and sometimes even bleeding in the mucosa. The chicken also suffers dehydration. There may be a white exudate in the lumen of the bursa; comprising rejected dead cells. For prevention of Gumboro, it is important to have good hygiene, i.e. thorough cleaning and disinfection between batches. Vaccination against Gumboro works effectively. Vaccination of parents provides passive immunity to the chicks in their first weeks by maternal antibodies from the egg.

3) *Swollen Head Syndrome (SHS)*: This virus spreads via direct contact, from one bird to another, and via contaminated personnel, water and equipment. Chickens and turkeys are most susceptible to this disease. The symptoms include sneezing, nasal discharge, foamy conjunctivitis, swollen infraorbital and peri orbital sinus "swollen head". Further, shaking of the head and coughing can occur in older birds. For layers, egg production may drop and the eggs produced tend to have shells of poor quality. SHS is not treatable, thus vaccination is the most effective method for disease control.

4) *Marek's Disease*: This disease is caused by an alpha herpes virus and is a highly contagious disease. It is mainly transmitted in premises where the birds are being

kept. Day-old chicks may become infected via the oral and respiratory routes. Dander from feather follicles of infected chickens can remain infectious for more than one year. As such, it is important to clean and disinfect the premises before a new batch of birds arrives. Young chickens are particularly susceptible to horizontal transmission, and the older the birds, the less susceptible they become. The infected Birds show weight loss or/and some form of paralysis. This paralysis scientifically called (neurolymphomatosis) in its classical form is with leg nerve involvement causes the bird to lie on its side with one leg stretched forward and the other backward. When the gizzard nerve is affected, the birds will have very small gizzards and intestines will waste away. Marek's Disease causes paralysis, mortality and infiltrations/tumours in multiple organs. Marek's disease is also untreatable, and disease control can only be achieved via prevention and good hygiene.

5) *Aspergilliosis*: It is also known as fungal pneumonia is caused by a fungus called Aspergillus. Aspergillus is transmitted via inhalation of fungus spores from contaminated feed. Fungal Pneumonia affects young chickens more than it affects older chickens and its symptoms include depression, thirst, Gaspings and heavy breathing. Further, yellowish-white lesions can be found inside the lungs and air sacs. Sometimes, yellowish-green fungal growth can be found in body cavities. The best treatment is to get rid of the infected birds.

It can be noted that most poultry diseases have same/similar symptoms, thus making it difficult to do correct diagnosis. The type of medication to be administered depends on the disease, thus correct diagnosis is essential for disease control and overall viability of poultry business. Correct diagnosis requires specialist knowledge and skills, which are in short supply in Zambia due to inadequate number of veterinary centers and trained personnel.

B. Expert Systems for Poultry Diseases' Diagnosis

Expert systems have long been used to perform tasks that normally require human intelligence to perform. Fig. 1 below shows the structure of an expert system.

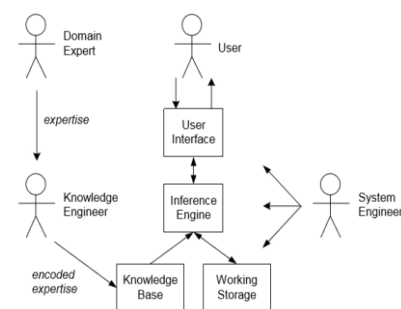


Fig. 1. Expert System's Structure[3]

It consists of the following: a Knowledge base which collection of domain knowledge and expertise, represented in the form of "IF THEN" rules, an Inference engine which acts as the core of the system at it derives recommendations from the knowledge base, based on the user's inputs, a user interface that acts as a means of communicating between the user and the system and the

working storage for storage of intermediary results. The major role of humans is to act as domain experts, end users, knowledge and system engineers. Domain experts are individuals who are trained and specialized in solving the problem of interest. In our study, the domain experts are the veterinary officers and doctors. The knowledge engineers encode the domain experts' knowledge in a declarative form, and this knowledge is to be stored in the knowledge base. The end user is the individual or group of individuals who shall be using the expert system for solving their problems of interest. In our study, the poultry farmers are the end users. The system engineer is the one who maintains the expert system and all other related components to ensure that it is always in working order.

A number of expert systems have proposed in literature for the diagnosis and management of poultry diseases. In [9], an expert system was reported for the management of poultry diseases. However, it only covered diseases prevalent in Nigeria, hence it would be difficult to use by Zambian farmers. Another expert system for was reported by [6], but it also focused on diseases prevalent in Indonesia and only on broiler chickens. In [10], an expert system for managing the final and economic side of the poultry business, called Xlayer, was reported. It helps farmers to make financial decisions, such as the projections of their profits and losses, etc.

To the best of our knowledge, no expert system has been developed to diagnose poultry diseases prevalent in Zambia. This is the motivation behind this study, i.e. to develop computer-based solutions to assist in the management of poultry diseases, and one such solution is the expert system that we have developed.

III. METHODOLOGY

The whole idea is that we would like the farmers to use the expert system on their own, in the comfort of their homes, in order to save them the time and effort required to visit a human veterinary expert. The expert system should be able to give a correct diagnosis and also offer drug prescriptions and related advice on the disease in question.

We developed our expert system using the waterfall model for software development [11]. The waterfall model entails rigidity and consists of the following phases: feasibility study, requirements gathering and analysis, design, construction, testing and deployment.

1) *Feasibility Study*: During this phase, we established why an expert system for poultry disease diagnosis is needed in Zambia. We based our conviction of why such a system is needed on the complaints of the small scale poultry farmers around Zambia's Central Province, regarding the unavailability of veterinary centers, where they seek help. We concluded that an expert system that can be installed on the farmers' computers, tablets and mobile phones would be of great help to them.

2) *Requirements gathering and Analysis*: During the phase, we spoke to various farmers and veterinary specialists to understand what type of functionalities they expect the expert system to possess. We also took this opportunity to gather knowledge about the various

poultry diseases that are prevalent in Zambia, i.e. the symptoms of the diseases, how they are diagnosed, how they should be treated, how they can be prevented and controlled, etc. The following is a summary of the user requirements for our expert system:

- The system shall be in English language.
- The system shall have a user interface".
- The system shall prompt the user to enter symptoms in order to conduct diagnosis.
- The system shall identify the disease based on the symptoms entered by the user
- The system shall recommend treatment and other prevention and control measures, based on the identified disease.

3) *Design*: During this phase, we modeled all the components that shall comprise the expert system. Fig. 2 and 3 below show the data flow diagram (DFD) and the flow chart for our system.

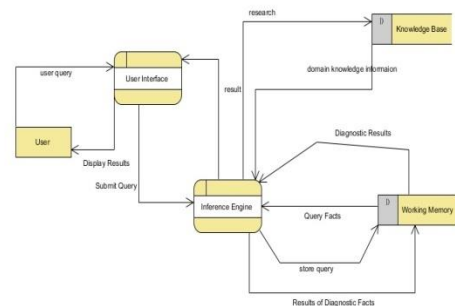


Fig. 2. DFD for Proposed System

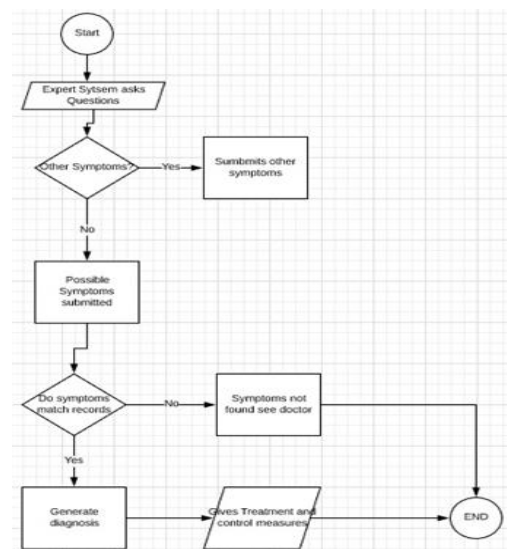


Fig. 3. Flow Chart for Proposed System

4) *Construction and Testing*: During these phases, we built and tested the expert system using Prolog language. We used SWI-Prolog as the IDE, which is an open source software for the implementation of the prolog language. We picked the SWI-prolog editor shell for MS-Windows, because the majority of our end users are familiar the

Microsoft-Windows environment. Fig. 4 below shows the results of testing our expert system.

```

SWI-Prolog -- f:/school docs/4th year/Sem 2/The Build/THE REAL DEAL/esiter.pl
File Edit Settings Run Debug Help
Welcome to SWI-Prolog (threaded, 64 bits, version 7.6.4)
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is free software.
Please run ?- license. for legal details.

For online help and background, visit http://www.swi-prolog.org
For built-in help, use ?- help(Topic). or ?- apropos(Word).

?- go.
Do the birds have the following symptoms:respiratory_distress? yes.
Do the birds have the following symptoms:coughing? |: yes.
Do the birds have the following symptoms:sneezing? |: yes.
Do the birds have the following symptoms:bloody_mucus? |: yes.
Do the birds have the following symptoms:facial_swelling? |: yes.
Do the birds have the following symptoms:pink_eye? |: yes.

prevention_and_control
vaccinate against disease if necessary
Do Not allow visitors from other farm
Do not share equipment/vehicles
clean and disinfect poultry house and equipment every time new birds are added
Please follow the advise strictly because
I believe that the chickens have: infectious_laryngotracheitis
TAKE CARE
true.
?-
    
```

Fig. 4. Testing the Proposed System

The deployment phase has not been done yet, because we are still carrying out the study.

IV. RESULTS AND DISCUSSION

We have managed to gather a lot of information about the common poultry diseases in Zambia. This information has been obtained from both human veterinary specialists and related literature. We are still in the process of converting this information into declarative knowledge that can be used for building an expert system. So far, we have managed to do so for the five diseases discussed in Section 2.A above.

We have built the expert system for poultry diseases' diagnosis using Prolog language. We opted to use prolog because of its ability to express relationships between/among objects, a feature that is cardinal for expressing the connection between a certain disease and a respective set of symptoms.

The accuracy rate of our expert system is 100%, i.e. it is capable of correctly identifying all the diseases contained in the knowledge base, when a set of symptoms is presented.

However, our expert system only contains five diseases, which is by no means an exhaustive list of all the poultry diseases prevalent in Zambia. Our expert system also does not have a graphical user interface. It uses a console-based user interface, which may pose a lot of challenges for the end users, since most of them have low computer literacy skills. The use on English as the only language for the system is also a source of challenges for

the end users, who are mostly small-scale farmers, since some of them may not be conversant with the language.

V. CONCLUSION

In this paper, we have presented some results of our study on the development of computer-based solutions for the management of poultry diseases. The study focuses on the poultry diseases that are common in Zambia. Based on the vast quantity of information about poultry diseases common in Zambia, we have built an expert system for the diagnosis of these diseases. Our proposed expert system has an accuracy rate of 100%.

In future, we would like to add more diseases to the knowledge base. We would also like to add a graphical user interface, instead of the current console-based one. Further, we would like to add local languages to the system, so that they are use alongside English.

REFERENCES

- [1] Andersson, C., Broiler Production in Zambia - Management, Growth, Diseases and Welfare. 2014, Swedish University of Agricultural Sciences: Uppsala.
- [2] Feigenbaum, E.A., Knowledge Engineering for the 1980s. 1982, Computer Science Department, Stanford University: Stanford, California.
- [3] M. Simfukwe, D. Kunda., M. Zulu, Addressing the Shortage of Medical Doctors in Zambia: Medical Diagnosis Expert System as a Solution. International Journal of Innovative Science, Engineering and Technology, 2014.
- [4] S. Mandal, S. Chatterjee, B. Neogi, Diagnosis and Troubleshooting of Computer Faults Based on Expert Systems and Artificial Intelligence. International Journal of Pure and Applied Mathematics 2013: p. 717-729.
- [5] H. Arowolo, B. Amosa., E. Faleye, An Expert System for Management of Poultry Diseases, in International Conference on Computer Technology and Science (ICCTS 2012). 2012, IACSIT Press: Singapore.
- [6] D. P. S. Setyohad, R. A. Octavia, T. D. Puspitasari. An Expert System for Diagnosis of Broiler Diseases using Certainty Factor. in The 2nd International Joint Conference on Science and Technology (IJCST). 2017. IOP.
- [7] M. R. Islam, B. C. Das., K. Hossain, N.S. Lucky, M. G. Mostafa, A Study on the Occurrence of Poultry Diseases in Sylhet Region of Bangladesh. International Journal of Poultry Science 2003.
- [8] N. Yune, N. Abdela, Update on Epidemiology, Diagnosis and Control Technique of Newcastle Disease. urnal of Veterinary Science and Technology, 2017.
- [9] H. Arowolo, B. Amosa., E. Faleye. An Expert System for Management of Poultry Diseases. in International Conference on Computer Technology and Science (ICCTS 2012). 2012. Singapore: IACSIT Press.
- [10] E. Schimisseur, J. Pankratz, Xlayer: An Expert System for Layer Management. Poultry Science, 1989.
- [11] Sommerville, I., Software Engineering, 9th Edition. 2010: Addison-Wesley.

AN INQUIRY INTO THE PUBLIC SECTOR EMPLOYEE RESISTANCE TO INFORMATION AND COMMUNICATION TECHNOLOGY BASED CHANGES: THE CASE OF IFMIS, IN THE MINISTRY OF FINANCE AND GENERAL EDUCATION

Hastings M. Libati ^a, Paul Chima ^b

^a*Copperbelt University, Jambo Drive Riverside, Box 21692, Kitwe, Zambia*

^b*University of Lusaka, Off Alick Nkhata Road, Long Acres, Box 36711, Lusaka, Zambia*

ABSTRACT

Change is inevitable in organizations. It happens continuously and often at such a rapid rate to the extent that it influences the efficiency and effectiveness with which an organization carries out its business mandate. Because change has become an everyday part of organizational dynamics, employees who resist change can actually cripple an organization. One of the major changes taking place today in businesses is technological change. Change in technology can have both positive and negative effects on the performance and attitude of employees. The general objective of this study was to establish the reasons why employees resist technological change in this case the Integrated Financial Management Information Systems (IFMIS) based change in public organizations. The study also sought to identify change management strategies and approaches that can be used to manage or prevent resistance to change. The study employed a descriptive approach to establish factors that influence employees' resistance to Information Communication Technology (ICT) change in the Ministries of Finance and General Education of the Government of Zambia. Major reasons for employee resistance to technological change were found to be uncertainty of the outcome of the technological change, lack of capacity building and skills training, poor and untimely communication about change, lack of management support, lack of employee involvement in the change process and fear of loss of their control. The study also suggests other avenues for further studies on the same topic

Key Terms: Change, Management, Resistance, IFMIS (Integrated Financial Management Information System), GRZ (Government of the Republic of Zambia)

INTRODUCTION AND BACKGROUND

In our economy today, change cannot be avoided in organizations. It happens continuously, and often at such a rapid rate to the extent that it influences the efficiency and effectiveness with which an organization carries out its business mandate. Because change has become an everyday part of organizational dynamics, employees who resist change can actually weaken an organization.

Resistance is an inevitable response to any major change. Individuals naturally rush to defend the status quo if they feel their status or security is threatened. Folger & Skarlicki (1999) [9] claim that "organizational change can generate skepticism and resistance in employees, making it sometimes difficult or impossible to implement organizational improvements".

With the coming of technology world over, governments and organizations have not been left out by these changes. These changes have necessitated enhancement in service provision with regards efficiency and effectiveness; examples include, education departments, finance departments and agriculture departments to name a few.

Globally, governments have resolved to heighten the quality of public financial management with many developed and developing countries making vital and remarkable achievements in strengthening management of finance in their public sector. Developing countries have been in the past few years encouraged to reform their public expenditure management systems and have increasingly embarked on a lot of projects to computerize their government operations.

In response to increasing demands for greater transparency and accountability in the management of the public's finances, the Government of the Republic of Zambia (GRZ) since 2009 has been undertaking a strategic transformation that involves substantial changes to the manner in which it manages its financial information used for managing both public financial resources and decision making in the Integrated Financial Management Information System (IFMIS). IFMIS refers to the computerization of public financial management processes, from budget preparation and execution to accounting and reporting, with the help of an integrated system for the

purpose of financial management (Lianzuala & Khawlhiring, 2008) [11] GRZ is expected to efficiently prepare and implement sound budgets, effectively manage and monitor expenditure, and improve economic and financial management. Above all the IFMIS will assist improve service delivery through the generation of reliable and accurate economic and financial data. After the implementation of IFMIS, GRZ is expected to have the necessary tools and skills to manage economic growth, development and poverty reduction. The system is intended to increase GRZ ability to monitor its expenditure, to access financial information easily and in real time, improve internal control and avoid fraud; support accrual accounting and have a better view of its cash position. The implementation of Systems Application Products (SAP) in the government will have a direct impact on its current ways of working, therefore, there is need to review and improve upon them, by changing processes, ensuring that the new way of doing things is more efficient and is able to provide more meaningful information to the policy managers and civil servants.

Statement of the Problem

The public feels the government is inefficient and ineffective in its operations and service delivery to the public. However, the government has over the past few years implemented a number of business process re-engineering strategies and most of which have the goals to bring about improvement (efficiency and effectiveness) in the service delivery to the public as well as financial benefits to Government Ministries, Provinces and Spending Agencies (MPSA). However, most often than not these strategies either bring on board new problems leading to looking for an alternative solution or they do not provide the expected benefits on time, consistent with intended goals or objectives. For example, the government has been implementing reforms to improve its service delivery. The notable one is Public Service Reform Programme (PSRP) which was initiated in 1993 and whose aim was to improve the quality, efficiency and effectiveness of public service delivery. In 1999 it was acknowledged that the measures taken under PSRP had not led to tangible improvements in the service delivery of the public services. Therefore, Public Service Capacity Building Project (PSCAP) was designed to support the implementation of PSRP. One of the components of PSCAP was the “Financial Management, Accountability and Transparency” and IFMIS was a project within it. A fully functioning IFMIS can improve governance by providing real-time financial information and enhance transparency and accountability, reducing political discretion and acting as a deterrent to corruption and fraud (Diamond and Khemani, 2006) [6]. To ensure proper management of funds, efficiency and effectiveness in its delivery of service to the public, the government resolved to use IFMIS as a measure that will provide internal controls in managing resources in the public sector. There are a number of factors that may affect negatively the successful implementation of IFMIS. Among the many challenges is resistance of employees to work with such new information and communication technology (ICT) driven approaches when brought on board. The problem would be looked at, at different levels including the organization itself, the kind of employees hired and the management approaches to managing the organizational change.

It is in this spirit that the study intended to establish factors influencing the implementation of integrated financial management information systems and recommend strategies that would help in addressing the problem.

Objectives of the Study

The study sought to investigate why people resist ICT based Integrated Financial Management Information Systems changes in public sector organizations. It was guided by the following objectives:

- To explore and establish key factors that influence employees’ resistance to IFMIS driven organizational change.
- To determine the nature and degree of this resistance.
- To map the socio-economic profiles of employees that generally resists change.
- To identify change management strategies and approaches that can motivate employees to embrace organizational change.

Research Questions

Generally, what factors drive employee resistance to ICT based institutional change in the public sector?

1. What are the factors that cause employees to resist IFMIS based organisational change in the public sector?
2. In what ways do employees show this resistance and to what degree is this resistance to change?
3. What category of employees generally resists these organisational changes?
4. What change and risk management strategies are likely to encourage employees to embrace change?

LITERATURE REVIEW

A number of literature was reviewed. Review on the study sought to explore writings on the subject of change which De Jager (cited by Egan and Fjernerstad, 2005) [7] said change “occurs whenever we replace the old with

the new. It is about travelling from the old to the new, leaving yesterday behind in exchange for a new tomorrow.”

New technologies and rising customer demands increase the pressure to make changes. Change helps organisations to navigate successfully and facilitate growth. Subsequently, change can be seen as an important element for development and success that requires careful and well thought planning in order to bring about positive outcomes and a smooth progress to advancement.

It is also important that employees be afforded key opportunities to participate in change initiatives. Zimmerman (1995) [20] argued that employees should always be the key players in the facilitation, implementation and management of effective change because employees are directly involved in the process of change in some form or the other. It is just as important to recognize that employees can also be the main obstacle to implementing change.

Innovation comes from changes in technology and technology change has become an increasingly important subject due to the fact that innovation is vital and of significance for the survival of any organization, whether it is a public sector firm that needs to improve their services to public and attract attention or a private sector firm that needs to increase their market share. Innovation matters because change is required constantly. It has been mentioned by many researchers that companies in any industry look at innovative strategies and strive hard to get success because innovation is imperative and it is necessary to either innovate or move out of the competition, that firms have to decide before taking challenges of innovation (Eveleens, 2010) [8].

IFMIS is largely a new concept or system that is yet to take sufficient roots in government ministries and provincial spending agencies. Needless to say, this system is bound to face significant resistance from the staff expected to implement and use it.

Resistance is defined as behavior intended to prevent the implementation or use of a new system or to prevent the system designers from achieving their objectives (Goldberg et al., 1999; Egan and Fjermestad, 2005) [10,7]. Manuela and Clara (2003) [14], explains that resistance is a phenomenon which affects the change process by delaying its beginning or hindering its implementation which, in turn, increases cost. Resistance is also defined as a conduct which tries to maintain the status quo and avoid changing.

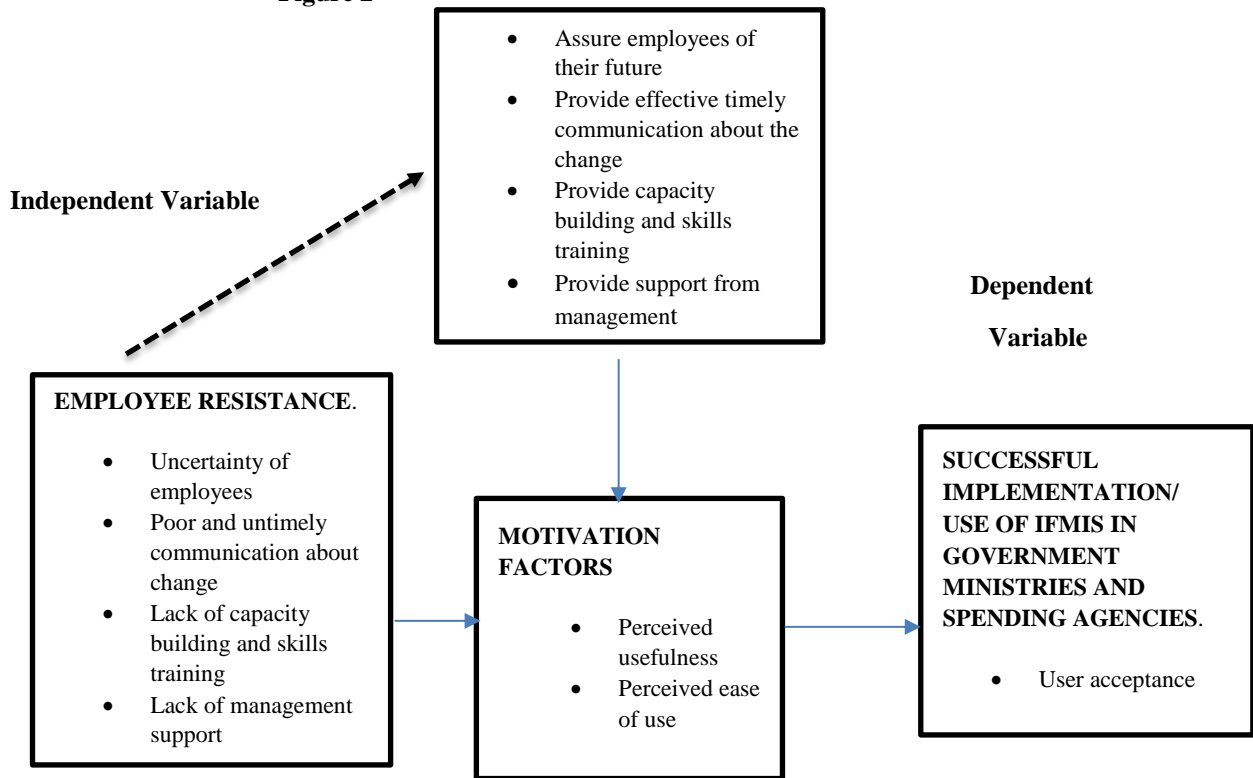
The implementation of an IFMIS is a complex, risky, resource-intensive process that requires major procedural changes and often involves top-level management officials who lack incentives for reform (Chene, 2009) [4]. It demands a commitment to change: change in technology; in processes and procedures; as well as changes in skills, responsibilities and behaviors (Rodin-Brown, 2008) [18]. Therefore, it is essential for all participants to be fully aware of the magnitude of the undertaking. Decision-makers must be convinced that the benefits of an IFMIS exceed the risks, and participating Ministries, Provincial offices and other government Spending Agencies (MPSA) must recognize the need for a new system (Chene 2009) [4].

To implement any system in the organization brings about change in processes, roles, policies and functions. With these changes come resistance from those affected. Hendricks (2012) [11] asserts that resistance to change may come from various stakeholders in the organization such as individuals with vested interests who benefited from previous methods, civil servants who see it as a threat to their jobs and people who resist change for the fear of the unknown. Change management is therefore, the process of creating, maintaining and thoroughly evaluating changes that occur in the organization (Barcan, 2010) [1]. This can be achieved through clear communication, education, training and other means that emphasize on the need and benefits of the change.

CONCEPTUAL FRAMEWORK AND THEORETICAL MODELS USED IN THIS STUDY

Below is a conceptual framework and theoretical models on related studies which this study made use of. The models listed below were coined by different writers. These models were adopted and criticized by other researchers based on new technological era. The models used in this research are: Lewin's Change Model, LCM [12], Technology Acceptance Model, TAM [5].

Figure 2



Based on literature review, a conceptual framework of the study was formulated consisting of independent variables as well as dependent variables as shown in figure 1.

METHODOLOGY

The study employed a descriptive approach to establish factors that influence employees’ resistance to ICT change in the Ministries of General Education and Finance of the Government of the Republic of Zambia. According to Mugenda (2008) [16], the purpose of descriptive research is to determine and report the way things are and help in establishing the existing status of the population under study.

The study only covered two government ministries as these were among the first ministries to start using IFMIS during the first roll out. Since it was not easy to select specific respondents from a large population, care was taken to select those who had key information in regard to IFMIS implementation and use. Management, Finance/ Procurement officers, ICT officers, and key system users formed the population of the study, bringing the total to 200 officers (150 males and 50 females).

To reduce the risks of biasness, purposive sampling procedure was used to select the respondents due to the nature of the study which required individuals who have information on the topic under investigation. This helped in the collection of adequate and sufficient information.

According to Mugenda (1999) [17], a sample ratio of 0.3 should be adequate to obtain a representation of each stratum of respondents from all the four groups as tabulated below in Table 1. The sample size was therefore, 60 respondents derived from the target population.

Table 1. Sampling Procedure

CLUSTERS	NUMBER OF TARGET POPULATION	SAMPLING RATIO	SAMPLE SIZE
Management Officers	20	0.3	6
Accounts/ Procurement officers	120	0.3	36
ICT Officers	20	0.3	6
Key system Users	40	0.3	12
Totals	200		60

Source: *Field Data 2018*

The instruments used to collect primary data for the study were structured questionnaires which had closed and open ended questions and face to face interviews. The basis of employing these instruments is driven by the assertion that they are easy to use and administer to respondents.

The study depended on the primary data which was collected from the four major stakeholders (Management, Accounts/ Procurement officers, ICT officers, and key system users). The two instruments used to collect data were structured questionnaires and face - to - face interviews. According to Sekeran (2009) [19], a questionnaire is a pre-written set of questions to which respondents record their answers. And Creswell (2003) [3], states that “interviews enable participants to discuss their interpretation of the world in which they live and express how they regard situations from their own point of view”. The questionnaire comprised of closed and open ended questions aimed at achieving the study objectives. The questionnaires were self-administered and interviewer - administered especially to senior management who were busy and key informants. The questionnaires were administered using drop and pick later approach so that respondents have ample time to respond to questions. The responses to interviewer - administered questionnaires were recorded by the interviewer on the basis of exact answers provided by the interviewee. The face-to-face interview resulted in gathering of accurate information, good response rate and completion of questionnaires in time.

FINDINGS, ANALYSIS AND PRESENTATIONS

For analysis, thematic analysis was used. According to Miles (1994) [15], thematic analysis is an approach to dealing with data that involves the creation and application of ‘codes’ to data. The ‘data’ being analyzed might take any number of forms e.g. an interview transcript, field notes, policy documents, photograph and video footage. The major themes were drawn from questionnaires and interviews with participants and description of each theme. The statistical package for Social Sciences (SPSS) version 20 was used to analyze the data together with excel.

The sample size was 60 respondents, 30 from each Ministry (Finance and General Education). Fifty-four (54) respondents were targeted for questionnaires and 6 respondents were earmarked for face to face interviews. Out of 54 respondents, 50 successfully completed the questionnaires and out of 6 respondents 4 were interviewed giving a total number of 54 representing a response rate of 90%. A total of 6 respondents did not respond giving a non-response rate of 10%. Of the 54 successful respondents, Ministry of Finance recorded 100% response rate meaning all the 30 respondents targeted answered the questionnaires and were interviewed while for Ministry of General Education only 24 responded; 4 failed to answer the questionnaires and 2 were not interviewed giving a total of 6. Mugenda and Mugenda (1999) [17], asserts that a response rate of 50% is adequate for analysis and reporting; a response rate of 60% is good and over 70% response rate is considered excellent. Based on this assertion, the rate of 90% was considered to be excellent meaning the data collected could be analyzed and the outcome would certainly represent the general population.

By age results indicate that most respondents were in the age range of 36 to 40yrs representing 46%, followed by 31 to 35yrs representing a percentage of 26%, then above 40yrs representing 17% and the least respondents were in the range of 25 to 30yrs years representing 11% respectively.

The majority of the respondents 26 (48%) had basic degree level of education while 6 (11%) had postgraduate degree level and 15 (28%) had advanced diploma level, 6 (11%) had diploma level of education respectively. All respondents had general certificate of education ((GCE) “O” level certificate) as it is a primary requirement.

Reasons why Employees Resist Change

Table 1 shows the frequencies and percentages of reasons why employees resist new system change.

Table 1

Reasons why Employees Resist Change		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total
Uncertainty of the outcome of the technological change	Freq.	0	0	9	11	34	54
	%	0%	0%	17%	20%	63%	100%
Lack of capacity building and skills training	Freq.	0	0	0	7	47	54
	%	0%	0%	0%	13%	87%	100%
Poor and untimely communication about change	Freq.	0	0	21	16	17	54
	%	0%	0%	39%	30%	31%	100%
Lack of management support	Freq.	0	11	21	19	3	54
	%	0%	20%	39%	35%	6%	100%
Lack of employee involvement in the change process	Freq.	10	7	11	7	19	54
	%	19%	13%	20%	13%	35%	100%
Fear of losing their control	Freq.	3	12	9	17	13	54
	%	6%	22%	17%	31%	24%	100%

Source: *Field Data 2018*

Uncertainty of the outcome of the Technological Change: The findings from the study show that 83% (20% + 63%) of the respondents positively agreed that uncertainty has great influence on employees’ resistance to change as employees may not know their fate. Seventeen percent (17%) of the respondents were uncertain. This was one of the major aspect of the study which the researcher sought to investigate. Change brings about loss of employment in most cases. If employees do not know what is at stake or outcome for any new development, they will resist change as they do not know what that change will bring whether good or bad.

Lack of Capacity Building and skills Training: The findings in table 1 indicates that 100% (87% + 13%) of the respondents agreed that lack of capacity building and skills training is a major contributing factor to employee resistance because if they do not know how to use technology or lack skills on how to manage the new system, the only way out is to resist change. For any changes happening in the organization, employees should have the capacity to manage and handle such. If they do not have the necessary skills and capacity, they become irrelevant in the organization and this may result in some employees losing jobs or resigning. So those who do not have the skills and or capacity they resolve to resist change knowing they won’t be able to handle the new systems or cope with change. This shows that lack of skills and training has a huge impact on employees’ acceptance to change. If employees are well trained with adequate skills, they will gladly accept the change.

Poor and Untimely Communication about Change: The study findings show that 61% (31% + 30%) of respondents indicated that poor communication is one of the causes of resistance with 39% respondents being uncertain. If communication is properly done, it clears any concerns and problems surrounding change initiatives which employees may have and any other stakeholders. If communication is properly done, employees will accept the change taking place but if they lack information about the changes taking place and are left in the dark, they won’t buy into the new strategy hence would resist change.

Lack of Management Support: Results table 5 indicate 41% (35% + 6%) of respondents supported the factor that lack of management support causes employee resistance with 39% being neutral. Twenty percent (20%) disagreed.

Lack of Employee Involvement in the Change Process: Results in table 1 indicates that 48% (13% + 35%) of the respondents agreed that lack of employee involvement in the change process contributes to employee resistance while 20% were uncertain and 32% (13% + 19%) disagreed. This shows that it is important to involve users as they are the custodians of the new systems to be implemented or will be the most affected individuals once the change process is completed. The involvement of employees in the change process regarding the present situation and desired future state inculcate trust among the employees. This will also create an environment where change agents or management will be able to receive sound advice from employees that will assist in the change process and will support the change and go along with it.

Fear of Losing their Control towards Work: The findings in table 1 shows that 55% (31% + 24%) of the respondents indicated that fear of employees losing control over their work was one of the factors why employees resisted technology change while 17% were not sure and 28% (6%+ 22%) disagreed. Results showed that employees wanted to have full control of their work as this made it easier to do what they wanted but with the new system it was difficult.

Once such change is effected employees will have little or nothing in terms of control as every transaction will be electronically controlled. With such an environment employees are given little room or none to manipulate transactions as everything is done electronically. This brings about resistance to accept the change. If employees understand how ease and beneficial the new systems are they will definitely buy into the vision and accept to support the idea

Strategies to Mitigate Employee Resistance

Results in Table 2 indicate the respondents’ views on the strategies that must be used to mitigate employee resistance to technological change. Results are indicated both in frequency and percentage.

Table 2. Strategies to mitigate employee resistance

Strategies to Mitigate Employee Resistance		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total
Communication about change should be timely and relevant	Freq	0	0	8	11	35	54
	%	0%	0%	15%	20%	65%	100%
Provide capacity building and skills training continuously	Freq	0	0	1	13	40	54
	%	0%	0%	2%	24%	74%	100%
Management should involve employees in the change process	Freq	4	1	20	15	14	54
	%	7%	2%	37%	28%	26%	100%
Effective rewarding policy to pay for performance	Freq	10	14	20	5	5	54
	%	19%	26%	37%	9%	9%	100%

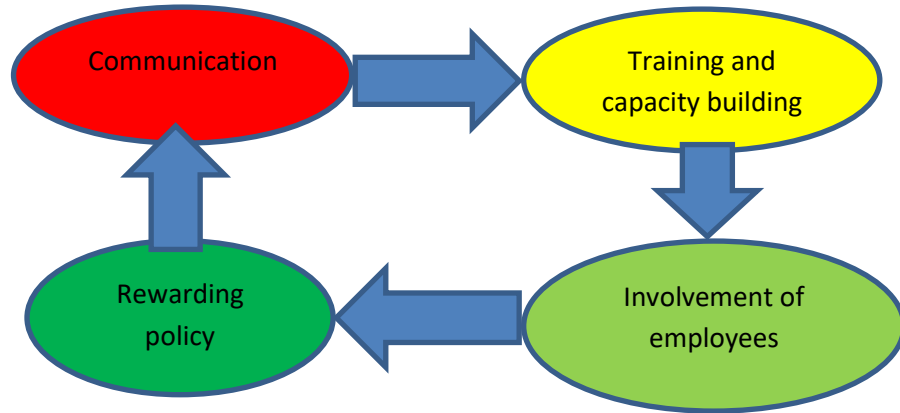
Source: Field Data 2018

Communication about Change should be Timely and Relevant: Results indicate that 85% (20% + 65%) of the respondents agreed that communication is a key strategy to mitigate employee resistance to technology while 15% were uncertain. If employees are communicated to timely and given relevant information about change, chances are that employees would accept change without problems. **Provide Capacity Building and Skills Training continuously:** Provision of training and capacity building was seen as a major strategy by respondents that could be used to mitigate resistance. Results indicated that 98% (24% + 74%) respondents agreed that training and capacity building would play a key role in reducing resistance to change. Only 2% were uncertain. Once employees are trained and given skills, fear would be driven out and would be ready to handle and accept the new technology. **Management should Involve Employees in the Change Process:** On the part of management involving employees in the change process, results showed that 54% (28% + 26%) were in agreement that employees should be involved in the change process as they play a key role in the change itself. 37% were neutral about the strategy and 9 (7+2) disagreed. **Effective Rewarding policy to Pay for Performance:** In table 2 results from the study indicated that 45% (19% + 26%) disagreed with the strategy that it had no major impact to make employees effectively work as they were on a permanent salary. 37% were uncertain and 18% (9% + 9%) agreed that once employees are motivated they tend to work well. According to the results obtained from the study, there is a clear and positive perception about change and resistance. Change is inevitable and significant in an organisation if done the right way and that people will always resist change. Change is vital to both the organisation and employees. Therefore, it can be concluded that employees in both Ministries of Finance and General Education welcome the IFMIS change as long as things are done in the correct manner without negatively impacting their plight.

Strategies to mitigate employee resistance

To manage or prevent harmful effect of resistance to change, the author came up with a simple model with four pillars. This model is based on the findings from the study.

Figure 1: Strategies to mitigate employee resistance model



Source: Author 2018

The study revealed and pointed out on some strategies that can be used to mitigate employees’ resistance to Information Communication Technology based changes as indicated in figure 20. Three main strategies came out, the fourth one was partially noted. The three main ones were; communication, capacity building and training, involvement of employees and the fourth one was rewarding policy which was partially accepted as one of the strategies.

The first strategy to mitigate employee resistance is effective and active communication. Communication must be clear, transparent and honest regarding the need for change and its impact as it is essential to the success of the initiative. It is an important tool that must be established in the organisation in order to mitigate resistance from employees that are fighting against change. Effective communication should segment and target each audience focusing on what employees care about and need to know. Once fear about what will follow with regards to job security and career advancement is reduced and contained, the change will be more easily accepted within the organisation. Leaders should put in place effective channels of communication to ensure everyone is informed of what is about to happen and underline why this change will benefit them. The second strategy is training of employees involved. The objective is to educate employees about the change before it is implemented and to help them understand the process of change. Training and capacity building is the cornerstone for building knowledge about change and required skills. Training is an essential tool that would help employees to learn new things and acquire new skills in order to meet the job requirements. Butterfield (2010) [2], posits that it is essential to implement innovative training sessions so as to put up a strong environment in the organisation as this is one way of motivating employees. Thirdly, management should involve employees in the change process. This will help managers to obtain support and collective efforts required to make the change. Employees tend to fully support initiatives in which they are directly involved and have a say. This makes them feel part of the organisation and the change itself. This involvement of employees has a greater impact on the change initiative, resulting from the fact that employees participate with proposals and implement effective solutions. This will enable a good working environment and employees will fully support the change initiative as they are part of the process. Lastly rewarding policy, employees need to be rewarded positively by providing financial incentives as a motivational factor. This will contribute in increasing employee morale and their personal development and thus creating a good climate in the organisation for high performance. In other cases, managers can offer financial incentives to lure those who may oppose change so that change should not be blocked. This solution is good when dealing with persons who will lose something valuable when change is effected.

CONCLUSION AND RECOMMENDATIONS

The study was conducted at the Ministry of Finance and General Education in Zambia to carry out an inquiry into the public sector employee resistance to information and communication technological based changes. The specific objectives of the study included;

- To explore and establish key factors that influence employees' resistance to IFMIS driven organizational change.
- To determine the nature and degree of this resistance.
- To map the socio-economic profiles of employees that generally resists change.
- To identify change management strategies and approaches that can motivate employees to embrace organizational change;

The researcher used descriptive statistics where frequencies and percentage of responses were obtained. The findings of the study revealed that most of the respondents went for improving organizational communication, capacity building and training, involvement of employees and rewarding policy as the major factors that influence the implementation of IFMIS. The results revealed that motivating the employees and management support on the new technological changes would lead to successful implementation and acceptance of IFMIS.

In terms of resistance to change, it was discovered that most of the employees become more resistant to change when they do not fully understand the system in this case IFMIS. Additionally, future uncertainty results in employee resistant. Thus, to reduce resistance, the organization must clearly state the processes and benefits of the new system that is to be implemented. However, the findings of the study also revealed that employee resistance reduces with an increase in employee perceived usefulness of the new system. Perceived ease of use was also found to be one of the factors that reduce employee resistance to change. Hence, an elaborate benefit of the new system would reduce the employee resistance to change.

Conclusion

The study was carried out to inquire why people resist ICTs based IFMIS changes in public organisations.

Based on the study findings; the researcher makes conclusion that the successful implementation and user acceptance of an IFMIS is hampered by employees' resistance to change.

The study revealed that employees resisted IFMIS change because of:

1. Ineffective and insufficient communication about the changes taking place. Employees were not fully educated/oriented on the benefits that IFMIS would bring to both them and the organization.
2. Employees lack of skills to handle the new technology being implemented. The study discovered that employees were reluctant to accept technological advancement as they lacked confidence to effectively run an IFMIS based system. Skills training and system orientation should have been undertaken before implementation of the system.
3. Employees lack of involvement in the change process despite being key stakeholders. This important aspect was overlooked by implementers of IFMIS hence alienating the program from the end user. It made them feel left out and unmotivated. Motivation is an important psychological need that plays a key role in the life of an employee. If not motivated; employees tend to feel discouraged and resort to resist change.

Employee resistance to IFMIS cut across the socio-economic divide in the organisation. The researcher concludes that employees resisting IFMIS where from different levels in the organisation and not from a particular profile. Individuals in all levels of the organisation displayed resistance due to the aforementioned factors namely ineffective communication; lack of education/ orientation and skills required in order for them to confidently run the system as well as sidelining of employees in the roll-out phase of the program.

Overall, successful implementation of IFMIS would bring more benefit to the organisation as it would improve the way financial information is collected, processed and subsequently aid the nation as a whole in budget preparation and economic forecast.

The researcher recommends that success of IFMIS can be achieved through:

Improving communication; that is, making it effective, clear and transparent. This is essential to the success of any initiative and an important tool that can help mitigate resistance from employees.

Capacity building and training; Employees must be trained or educated about the change before it is implemented in order to help them understand the change process. This will help employees to learn and obtain new skills in order to face the new job challenges.

Employees must be involved in the change process happening in the organisation; Change managers should make sure employees are fully involved as this will help them (managers) obtain full support and effort required to make these changes. This makes employees feel part of the organisation and the change itself.

Rewarding policy should be effective; Employees must be motivated by way of providing incentives. This will increase employees' morale and will in turn make them give their best performance and support to the change process.

Recommendations

Based on the findings of the study, the researcher would like to make the following recommendations;

1. The Ministry of Finance and General Education should ensure that there is a clear explanation of the process of IFMIS to the employees. This can be achieved through having an effective communication channel, workshops and trainings.
2. The management of these public Institutions should fully involve employees in the change process so as to gain support from employees throughout the change process.
3. There is need to enhance employee motivation through incentives, trainings and management support so as to make them feel appreciated and part of the organisation and this will encourage them to be ready for change and accept it as it comes.

REFERENCES

- [1] Barcan, L., (2010). New concepts in the change management within public organizations. *Young Economists Journal Revista Tinerilor Economisti*, 8(14), 93– 97.
- [2] Booth-Butterfield, M., Horan, S. M. (2010). Investing in affection: An investigation of affection exchange theory and relational qualities. *Communication Quarterly*, 58, 394–413. doi:10.1080/01463373.2010.524876.
- [3] Creswell, JW (2003). *Research design: Qualitative, quantitative and mixed methods approaches*, Thousand Oaks, CA: Sage Publications.
- [4] Chene, M. (2009). The implementation of integrated financial information management systems (IFMS).
- [5] Davis, F.D (1989), User acceptance of computer technology: A comparison of two theoretical models, *Management Science*, 35 982-1003.
- [6] Diamond, J. & Khemani, P. (2006). Introducing financial management information systems in developing countries. *OECD Journal on Budgeting*, 5(3), 97-132.
- [7] Egan, R. W., and Fjermestad, J. (2005). Change and Resistance Help for Practitioner of Change. In *Proceedings of the 38th Hawaii International Conference on System Sciences*, pp 219-226.
- [8] Eveleens, C. (2010). Innovation management; a literature review of innovation process models and their implications. *Science*, 800, 900.
- [9] Folger, R. & Skarlicki, D. (1999). Unfairness and resistance to change: hardship as mistreatment, *Journal of Organizational Change Management*, 35-50.
- [10] Lewin, K. (1958). "Group decision and social change," in *Readings in Social Psychology*, eds
- [11] Lianzuala, A. & Khawlhing, E., 2008, Mizoram IFMIS Project, viewed 14 May, 2018, from <http://www.docstoc.com/docs/39661608/Mizoram-IFMIS-Project>
- [12] Goldberg, S. Grusec, J. E., Jenkins, J. M., (1999). Confidence in protection; *Journal of Family Psychology*
- [13] Hendriks, C.J. (2012). Integrated financial management information systems: Guidelines for effective implementation by the public sector of South Africa. *SA Journal of Information Management*, 14(1), 1-9.
- [14] Manuela Pardo del Val and Clara Martinez Fuentes (2003). Resistance to change: a literature review and empirical study. *Management Decisions*. 42 (7), pp.148-155.
- [15] Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis an Expanded Sourcebook*. Thousand Oaks, CA Sage Publications.
- [16] Mugenda, O. M., & Mugenda, A. G. (2008). *Social Science. Applied Research & Training Services*. Nairobi: Kenya.
- [17] Mugenda, O. M., & Mugenda, A. G. (1999), *Research methods: Quantitative and qualitative approaches*: Olive M Publication: Nairobi Acts 1999.
- [18] Rodin-Brown, E. (2008). *Integrated Financial Management Information Systems*. New York: The Louis Berger Group, Inc.
- [19] Sekaran, U. and Bougie, R. (2009) *Research Methods for Business a Skill-Building Approach*. 5th Edition, John Wiley and Sons Inc., Hoboken.
- [20] Zimmerman Marc A., (1995). *Psychological empowerment: Issues and illustrations*.

A Comparative Study of the Performance between Boys and Girls in Computer Studies in three selected Secondary Schools in Lusaka District, Zambia

Kuliwa S. Syakwasia¹, Christopher Chembe²

¹Department of Computer Science and Information Technology, Mulungushi University, Kabwe, Zambia,

²Department of Computer Science and Information Technology, Mulungushi University, Kabwe, Zambia

¹kuliwas@yahoo.com, ²cchembe@mu.ac.zm

Abstract - This study investigated the performance between secondary school boys and girls and their attitudes towards computer studies in Lusaka district, Zambia. The study adopted a descriptive survey method. The instruments used were a questionnaire and past academic performance results in computer studies. Firstly, to determine difference in performance, a sample of 387 grade 9 results for pupils who sat for their computer studies examination in 2018 from three selected schools were collected and analyzed. Secondly, a structured questionnaire on pupils' attitude towards computer studies with a battery of 19 questions on a five point Likert scale was administered on a sample of 131 grade 9 computer studies pupils preparing for their 2019 Junior Secondary examination. The respondents' results and responses were scored and analyzed using t-test and descriptive statistics. The results of the study revealed a significant difference between secondary school boys' and girls' performance in computer studies. However, there was no evidence to show that all boys perform better than girls. Findings on boys' and girls' attitude towards computer studies showed no significant difference. However, the study found boys to have a slightly higher positive attitude than girls. Evidence suggests that agents of socialization play a major part in the observed performance difference. Recommendations were made based on the findings of this study.

Keywords - Computer Studies, Gender, Performance, Attitude, Secondary School, Boys, Girls, Lusaka.

I. INTRODUCTION

A. Background of the Study

Today, a major issue attracting public concern in Zambia is the issue of gender gap in academic performance of pupils in secondary schools. Gender differences in academic achievement contribute to disparities in the allocation of cognitive roles in the world of work and limit access to colleges and universities.

The use of computers in secondary schools in Zambia dates back to the mid-seventies when Mpelembe secondary school introduced computer studies with the help of teachers drawn from Mathematics department. In the mid-eighties more private schools that could afford Information and Communication Technology (ICT) infrastructure introduced the computer studies either as a subject or for purposes of computer literacy. The mid-nineties saw an addition of more private schools including International Community and Chengelo secondary schools. Government aided schools such as Matero Boy's and Ibenga Girls also enlisted.

The late-nineties to early two thousand saw the introduction of computer studies in public schools including Hillcrest and David Kaunda National Technical Secondary Schools. The first national ICT working group meeting held at the Ministry of Education Headquarters in January 1998, led to the establishment of SchoolNet Zambia a vehicle

formed to spearhead the formulation of a National ICT Policy in Education and the introduction of computer studies/science at secondary school and University levels in Zambia Syakwasia [1].

Subsequently in 2015, computer studies was introduced as a compulsory subject at junior secondary school while new topics were introduced into the Grade 12 curriculum. In 2015 when the subject was first introduced, a total of 350,656 candidates sat for their grade 9 computer studies examination with each pupil scoring 30% on average. The 2015 examinations for computer studies was faced with a number of challenges. According to Zambia Institute for Policy Analysis and Research [2] the Examination Council of Zambia (ECZ) attributed the poor performance in computer studies to limited capacity of teachers, inadequate infrastructure and lack of adequate teaching and learning materials.

Several comparative studies have been undertaken to examine gender differences in academic performance resulting in a substantial body of literature, Dania [3]; Agbaje & Alake [4] and Atovigba et al, [5]. Some of these researchers pointed out that there is no significant gender difference in pupils' academic performance in various subjects while others found a significant difference with either the boys or the girls performing better.

It therefore remains unclear whether gender differences in performance and attitude towards computer studies can be generalized across the country as none of the reviewed literature investigated the performance gap in computer studies at secondary school level in Zambia.

B. Purpose of the Study

The purpose of the study was to compare between secondary school boys' and girls' performance and their attitude towards computer studies in Lusaka district.

C. Objectives of the Study

The objectives of this study were:

- i. To compare between secondary school boys' and girls' performance in computer studies.
- ii. To compare between secondary school boys' and girls' attitude towards computer studies.

B. Research Questions

The research questions of the study were:

- i. Is there any significant difference between secondary school boys' and girls' performance in computer studies?
- ii. Is there any significant difference between secondary school boys' and girls' attitude to computer studies?

C. Research Hypotheses

The following null hypotheses were tested at the 0.05 level of significance:

- i. Null Hypothesis 1 (H_0): There is no significant difference between secondary school boys' and girls' performance in computer studies.
- ii. Null Hypothesis 2 (H_0): There is no significant between secondary school boys' and girls' attitude to computer studies.

II. REVIEW OF RELATED LITERATURE

A. Introduction

Review of literature in this study was organized under the variables; performance, parental attitude, teacher attitude, pupil attitude, and school type.

B. Performance

Comparative studies of the performance of pupils in various subjects and the factors that affected their performance have been carried out in Zambia and elsewhere.

Okwo and Otunba [6] reported that, gender influences performance by 13.39% of the total influence factor. Studies by Aremu [22] revealed that boys are better than girls in science subjects and mathematics. Toh [7] reported that girls performed better than boys in some other school subjects, while other researchers found no difference in some cognitive tasks involving both boys and girls.

In the recent past, a study by Chikumbi [8] on the performance of secondary school students in computer studies in Malawi revealed significant differences with male pupils doing better than their female counterparts. However, Adigun et al [9], reported that there was no significant difference between male and female pupils in computer studies achievement.

C. Attitude and Self-concept

Smith [10] defines Attitude towards Computer Studies as "a person's general evaluation or feeling towards computer technology and specific computer related activities." Given the pervasiveness of computers at all levels of the educational system, pupils are likely to develop some attitude toward these machines. In a school environment, attitudes can be formed regarding particular teachers, school type, subjects, facilities, syllabus and other aspects of pupils' daily lives.

Raimi and Adeoye [11] reported a significant difference between male and female students' attitude towards integrated science in favour of male students. This finding corroborates earlier findings of Sáinz and López-Sáez [12] in Spain, who found more positive computer attitudes in secondary school boys than girls. However, a South African study focusing on pupils' attitude towards technology did not reveal gender differences in attitude towards technology van Rensburg and Ankiewicz [13]. Both girls and boys shared similar views about the importance of technology in society and believed that technology should be accessible to all.

D. Pupils, Parents' and Teachers' Attitude

Studies by Sutton [14] revealed that pupils' beliefs and attitudes have the potential to either facilitate or inhibit learning. Goode [15] showed that male pupils have more

favorable attitudes toward computers than their female counterparts. Female pupils appear to be more anxious with computers and regard them as more difficult to deal with than male pupils. In contrast, Sáinz and López-Sáez [12] reported that females and males did not display significant differences in attitudes toward computers such as liking, comfort, confidence and perceived usefulness.

A parents' attitude and gender stereotyped views about computers can have a positive effect on computer attitudes of their sons, but a negative effect on their daughters' attitude Dania [3]. Most parents are unable to help cultivate in their children, a positive attitude towards computers because they themselves never studied computer studies at school or because they do not see its relevance to their daughters' career aspirations.

Sutton [14] contends that teachers tend to have lower expectations of success for girls and make less academic demand of them compared to boys. Girls tend to be valued for neatness, conformity and good behavior, while boys are commended for exuberance, excellence and creativity.

Jones and Clarke [16] found that female pupils from single sex settings had more time with computers and more positive attitudes than those from mixed schools.

E. Theoretical Framework: Social Learning Theory

This study was guided by Bandura's Social Learning Theory. According to Bandura's [17] Social Learning Theory, human behavior is a function of the person plus the environment. In other words, the person, the behavior, and the environmental situations are highly interrelated. Each is capable of influencing the other. The theory assumes that when two persons or groups interact, they become part of each other's environment.

The theory argues that learning is a product of the individual and the environment. This implies that the interaction between a student and the learning environment determines, to a certain extent, the performance of the student in a given school subject. The theory further, holds that performance in school learning is traceable to gender stereotype of learners.

The environments include both the school environment, the home environment and other factors that are within the environment such as stereotypic view, gender, school type, and individual's self-concept. This theory has a link with this study as it highlights a link between academic performance and the students' attitude and gender stereotype. Bandura [17]

III. METHODOLOGY

A. Study Site

The study was delimited to Lusaka district of Lusaka province in Zambia. The district has a surface area of 418 square kilometers at an altitude of 1,312m above sea level and a population density of 6,045/km².

B. Design

The research design that was employed in this study was descriptive survey method. The aim of the researchers was to record, analyze and report the way things are and establish the current status of the population under study. This design also accommodates generalization of findings of the study

population from a representative sample. This made descriptive survey research design suitable for this study. The research is non-experimental and therefore variables were not manipulated.

A quantitative research method was chosen for this study in order to give a detailed description of a comparison between secondary school boys' and girls' performance and their attitude towards computer studies. This method was appropriate for the research objectives of this study as the aim of the study was not to infer cause and effect but to describe the nature of the research topic and objectives by collecting numerical data that was analyzed using statistical tools.

C. Population

The population of the study comprised of all the 2018 and 2019 grade 9 computer studies pupils in Lusaka district of the Republic of Zambia.

D. Sample

Simple stratified random sampling technique was used to select the three schools. The schools were first divided into three strata, which are boys, girls and co-education secondary schools. The group from each stratum was then selected using simple random sampling. Three (n=3) schools offering computer studies were selected. They included; Woodlands "A" secondary school, Lusaka Boys' secondary school and Lusaka Girls secondary schools.

According to Krejcie and Morgan [18] the usage of the Morgan table for sample size determination for Educational and Psychological tests, when population size is 40,000, confidence level of 95% and a margin of error of 5% a sample size of 381 is adequate.

Thus to compare between secondary school boys' and girls' performance in computer studies, the research used a sample of 387 grade 9 computer studies Junior Secondary school Leaving Examination results for 2018 from the three selected secondary schools. The breakdown of the sample consisted of 204 girls and 183 boys. 172 were from Woodlands "A" – 70 boys and 102 girls (co-education school), 102 from Lusaka Girls secondary school, and 113 from Lusaka Boys' secondary school.

To compare attitude of secondary school boys and girls towards computer studies, a sample consisting of 2019 grade 9 computer studies pupils totaling 131, 66 girls and 65 boys aged between 12-17 years with a mean age of 14.3 and 14.03 for boys and girls respectively. Of the 131 respondents, 38 were from Lusaka Boys' secondary school, 39 from Lusaka girls' secondary school and 54 from Woodlands 'A' secondary school. The respondents were believed to be preparing for their 2019 Grade 9 computer studies Junior Secondary Leaving Examination. Simple random sampling was used in selecting a class in the selected schools with more than one stream. One class from each of the three schools was selected.

E. Research Instruments

The researchers developed and administered two instruments.

- i. Secondary data analysis of 387 pupil's results for the 2018 Grade 9 computer studies Junior Secondary School Leaving Examinations was utilized for determining

performance between secondary school boys and girls in computer studies.

- ii. A questionnaire for comparing attitudes of secondary school boys and girls towards computer studies was administered on the 2019 grade 9 computer studies pupils totaling 131.

a) *Computer Studies Performance Instrument (CSPI)*

To compare the performance between secondary school boys and girls in computer studies, a document review of the 2018 Grade 9 computer studies Junior Secondary Leaving Examination results for 387 pupils' were collected from the three selected schools in Lusaka district for analysis. Invariably, secondary data analysis or re-use of qualitative data was employed to afford the researchers an opportunity to investigate research questions using a large-scale data set while saving time and resources.

b) *Computer Studies Attitude Questionnaire (CSAQ)*

To compare attitude of boys and girls towards computer studies a questionnaire was selected because it helped the researchers to be consistent in asking questions and data yielded was easy to analyze. The questionnaire consisted of a battery of 19 Likert Scale type statements formulated in using a simple and clear language. A five-point rating scale ranging from Strongly Agree (SA) = 5, Agree (A) = 4, Neutral (N) = 3, Disagree (D) = 2 and Strongly Disagree (SD) = 1 was adopted. Any response with a mean score of 2.5 and above was considered as "agree" while responses below the mean score of 2.5 were considered as "disagree". This was done to enable consistency such that low scores corresponded to negative attitudes while high scores reflect positive attitudes towards computer studies.

Section "A" of the questionnaire comprised demographic data, which sought to obtain respondents' details such as age, sex, and type of school. Section "B" sought to determine the attitude of pupils towards computer studies. Section "C" was aimed at getting pupils' suggestions and recommendations on how to improve the performance of girls and boys in computer studies.

F. Validity and Reliability of the Research Instruments

The researchers focused on content validity, which refers to the accuracy with which an instrument measures the factors under study. Being a grade 9 Junior Secondary School Leaving Examination used for certification and set using measurement approach theories, the validity and reliability of the instrument used to compare performance between boys and girls was assumed to have been based on Blooms' Taxonomy: knowledge, comprehension, application, analysis, synthesis and evaluation.

Accurate and careful phrasing of each question to avoid ambiguity and leading respondents to a particular answer ensured reliability of the tool. In addition, respondents were informed of the purpose of the study and of the need to respond truthfully in order to establish the reliability of the questionnaire.

G. Piloting of Research Instruments

A pilot study was conducted to test the questionnaire for reliability. Ten respondents from Chelstone Secondary School (n=10) with similar characteristics to the research sample who were not part of the main study were

interviewed. Following the pilot study, some ambiguous questions were rephrased to give greater clarity and some questions were discarded, as they proved irrelevant.

H. Collection of Data

Data was collected between March and April, 2019 using a structured questionnaire administered on the 2019 grade 9 computer studies pupils and a document review of the 2018 grade 9 computer studies results. Respondents in schools were selected and requested to participate in the study. Detailed information about the study was given to the respondents. Verbal consent was obtained before distributing the questionnaire. Secondly, the 2018 grade 9 computer studies Junior Secondary School Leaving Examination results generated by Examinations Council of Zambia on computer sheets were collected from the three selected schools for coding, tabulation and analysis.

I. Analysis of Data

The process of data analysis involved data collection, data cleanup and interpretation. Data cleanup involved editing, coding and tabulation. This was done in order to detect anomalies in the responses and assign specific numerical values to the responses for further analysis. In addition, the questionnaire on attitude towards computer studies was edited for completeness and consistency.

Findings were presented in the form of descriptive and inferential statistics. Significance tests were also carried out. Percentages, means, standard deviations, independent two-tail t-tests, probability tests, frequency table, pie charts, normal distribution curve and bar graphs were presented. To provide meaningful and final results, the data was analyzed and interpreted using Microsoft Excel Able bits and statistical Package for Social Sciences (SPSS).

J. Ethical Considerations

Permission to conduct the study was obtained from respective secondary school Head Teachers explaining the purpose of the research and requesting them to allow grade 9 pupils participate in the study. After a full explanation of the nature of the study, participants were asked to give a verbal consent of their willingness to participate in the study. To ensure anonymity, steps were taken to protect the identity of the individual by neither giving their name nor including identifying details which may reveal their identity.

IV. PRESENTATION OF RESULTS

A. Demographics

The age group of respondents ranged between 12-17 years old. The mean age for boys was 14.30 while that of girls was 14.03. The boys with a higher mean age were therefore expected to be marginally more mature in mental capacity than girls. The respondents where a representative of all social economic groups with the majority having come from the lower to middle income groups.

B. Research Question 1: Is there a difference between secondary school boys' and girls' performance in computer studies?

Table 4.1: Descriptive Statistics Gender based Performance

Gender	N	Mean	Std. Deviation	Std. Error Mean
Boys	183	54.16	15.43	1.14
Girls	204	49.99	15.68	1.10

Table 4.1 reveals that boys did better than girls in the 2018 computer studies examination. The boys mean score was 54.16% while that of girls was 49.99 % or 4.17% less than that of boys. This shows that on average more boys did better than girls.

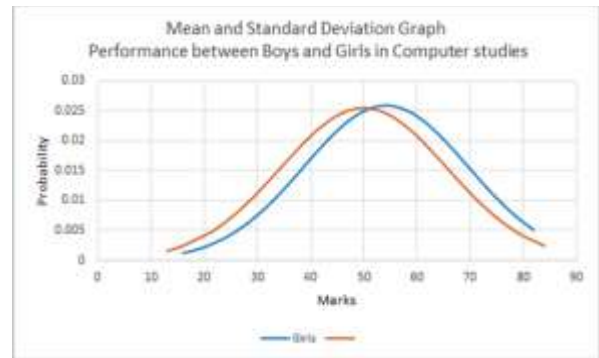


Figure 1: Spread of marks Gender based Performance

Fig 1 shows the spread of marks with a standard deviation for boys and girls of 15.43 and 15.68 respectively. Though the deviation for girls' marks was more spread than boys' they did not perform as well as the boys did.

C. Hypothesis 1 H_0 : There is no significant difference between secondary school boys' and girls' performance in computer studies.

An independent t-test was used to check the difference between boys' and girls' performance in computer studies. Alpha level of significance was set at 0.05 (5%) with a Confidence Level (CL) set at 95%. A confidence level of 95% implied that in the event of dismissing H_0 , we are 95% sure that we made the best choice.

Table 4.2 Independent t-test of the Gender based Performance

Gender	N	Mean Diff.	df	t-value	p-value	Sig. 2 tailed	Decision
Boys	183	4.17	382.63	2.66	0.008	0.05	Significant
Girls	204						Reject H_0

“N” = Number of pupils, “df” = Degree of freedom

The results were significant ($t(387) = 0.008, p < 0.05$), p value of 0.008. The table shows that the 4.17 more marks gotten by the male students is significant since the significance value of the test (0.008) is less than 0.05. The statistics failed to support H_0 and is therefore rejected.

D. Research Question 2: Is there a difference between secondary school boys' and girls' attitude towards computer studies?

Table 4.3 Analysis of Responses to Selected Statements

	Statement 5	Statement 6	Statement 13
Gender	I would rather be taught computer studies by a teacher of the same sex	In general boys do better than girls in computer studies	I feel anxious and fearful, when am asked questions in computer studies
Boys Mean	2.59	2.97	2.15
Girls Mean	2.32	1.80	2.46

Table 4.3 revealed that both boys and girls exhibited a positive attitude with mean scores of 2.89 for boys and 2.95 for girls. In response to statement 6 boys exhibited a positive attitude with a mean score of 2.97 compares to a mean score of 1.50 for girls. Both boys and girls exhibited a negative attitude with mean scores of 2.15 for boys and 2.46 for girls. 18(27.69%) girls agreed that they feel anxious and fearful compared to 12(18.18%) boys.

E. Null Hypothesis 2 (H₀): There is no significant difference between secondary school boys' and girls' attitude towards computer studies.

Table 4.4 Independent t-test of the Gender based Attitude

Gender	N	Mean Diff.	df	t-value	p-value	Sig. 2 tailed	Decision
Boys	66	0.06	128.77	0.03	0.77	0.05	Not Significant
Girls	65						Do not Reject H ₀

The results were not significant ($t(131) = 0.30, p > 0.05$), p-value of 0.77). The table shows a mean difference of 0.06 in attitude between boys and girls. The difference is not significant since the significance value of the student test (0.30) is more than 0.05. Hence the stated hypothesis is accepted.

Further, the results presented in Table 4.4 revealed that the mean difference attitude towards computer studies between boys and girls was 0.06 with boys scoring a mean of 2.36 compared to girls' mean response of 2.30. In general, both boys and girls seemed to be quite negative about computers.

V. DISCUSSION OF RESULTS

A. Research Question 1: Is there a difference of between secondary school boys' and girls' performance in computer studies?

The results of the study revealed that there is a significant difference between boys' and girls' performance in computer studies. However, while the study showed that boys performed better than girls, there was no evidence to show that all boys perform better than girls. There were many girls with excellent potential for computer studies and many boys who had great difficulties.

Results of this study uphold the findings of Chikumba [8] who showed significant differences between male and female pupils in computer studies achievement. The results negates Adigun et al [9] who reported that there was no significant difference between male and female pupils in computer studies achievement.

The higher deviation around the mean for girls (15.68) revealed that the performance of the girls is not as uniform as

that of boys (15.43), that is, male pupils have similar performance as opposed to the female pupils. However, the mean mark for boys (54.16) was found to be higher than that of girls who had a mean mark of 49.99. The mean difference between the two groups was 4.17. The difference in performance is most likely due to environmental variables which are academic treatments pupils are exposed to.

B. Research Question 2: Is there a difference between secondary school boys' and girls' attitude towards computer studies?

The findings of this study revealed that there was no significant difference in attitude between secondary school boys and girls towards computer studies. However, the study, found boys to have a slightly higher positive attitude than girls with a mean difference of 0.06.

This confirms previous findings by Sáinz & López-Sáez [12] who demonstrated that secondary school pupils' attitude scores did not vary with gender. Çelik and Ceylan [19] also came to the same conclusion that gender did not have an influence on attitudes towards computer studies.

This contradicts Goode [15] who reported significant difference between male and female pupils in attitude to computer studies. Adesoji [20] reported that there is significant difference between male and female pupils' attitude to computer studies but the male had a little more positive attitude than the female pupils.

The study revealed that girls tended to exhibit less confidence than boys did. A lot more girls are held back in computer studies by lack of confidence and misplaced modesty to an extent that were they do succeed they do not believe their eyes and continue to belittle their achievements. The study also revealed that male pupils have higher mean attitude than their female counterparts. This is in line with the finding of Steele et al. [21].

The large number of boys and girls who were in neutral category are not really neutral but unconcerned and lacking focus. Ways of capturing this group to be more focused and interested to excel in computer studies must be found and implemented by schools and parents. Doing so would shift the performance positively thus raising the number of pupils passing in the subject.

Further, the study revealed that, teachers do seem to have different expectations about the behavior and performance of girls and boys in computer studies. They tend to have low expectations of success for girls than boys. Teachers also tends to favor children who seem to understand the task set and are able to do it correctly.

In addition, most parents tended to think of computer studies as a more appropriate career for boys than girls. Many parents take pride in professing ignorance of computer studies while only a few would admit having been poor history pupils. It is in this atmosphere that the teacher must work.

VI. CONCLUSION

A. Performance between Boys and Girls in Computer Studies

The study revealed that there is a significant difference between secondary school boys' and girls' performance in

computer studies. However, there was no evidence to show that all boys perform better than girls.

B. Attitude of Boys and Girls towards Computer Studies

The study revealed that there is no significant difference between secondary school boys' and girls' attitude towards computer studies. However, the study found boys to have a slightly higher positive attitude than girls.

VII. RECOMMENDATIONS

Based on the findings of this study, the following recommendations were made:

- i. Parents should help their children cultivate a more positive attitude towards computer studies.
- ii. Teachers should provide selective help to improve attitudes of girls towards computer studies and guard against gender stereotyping.
- iii. Schools need to strengthen counselling departments to tackle negative attitudes formed.
- iv. The Ministry of Education should provide schools with electricity, adequate computers, teaching/learning aids and trained computer studies teachers.

VIII. REFERENCES

[1] Syakwasia K. S. (1998). "Information and Communication Technology: Opportunities and Challenges", Ministry of Education, Lusaka.

[2] The Zambia Institute for Policy Analysis and Research: (2018): "The Teaching of Computer Studies and a Review of the Quality of Higher Education in Zambia", Submitted to the Committee of Education, Science and Technology. Lusaka, Zambia.

[3] Dania Peter O. (2014). "Effect of Gender on Students Academic Achievement in Secondary School Social Studies". Journal of Education and Practice. <http://www.iiste.org> (Online) Vol.5, No.21, 2014.

[4] Agbaje, Rashidat O, & Alake, Ese M (2014). "Students' Variables as Predictor of Secondary School Students' Academic Achievement in Science Subjects". *International Journal of Scientific and Research Publications*, Volume 4, Issue 9, September 2014.

[5] Atovigba M.V., Vershima A. Michael, & O'Kwu E.I. and Ijenkeli Emmanuel (2012). "Gender Trends in Nigerian Secondary School Students' Performance in Algebra". *Research Journal of Mathematics and Statistics* 4(2):42-44, 2012.

[6] Okwo, F. A. and Otunba, S (2007). "Influence of Gender and Cognitive Styles in Science Achievement in Physics Essay Test". Journal of Science Teachers Association of Nigeria.

[7] Toh (1993). "Gender and Practical Tasks". *Science Education Research* 35, 225-265.

[8] Chikumba, P. (2012). "Challenges in computer education computer studies ICTs in education Malawi education system". Springer DOI: 10.1007/978-3-642-23828-4_11.

[9] Adigun J. ; Onihunwa, J. Irunokhai, E., Sada Y, Adesina O., "Effect of Gender on Students' Academic

Performance in Computer Studies in Secondary Schools in New Bussa, Borgu Local Government of Niger State", *Journal of Education and Practice*, V6 n33 p1-7 2015

[10] Smith B, Caputi P, Rawstorne P (2000a). "Differentiating Computer Experience and attitudes towards Computers". *An Empirical Investigation. Comput. Hum. Behav.* 16: 59-81.

[11] Raimi, S. A. & Adeoye, F. A. (2002). "Gender Difference among College Students as Determinant of Performance in Integrated Science". *E-ISSN 2039-2117 ISSN 20399340 Mediterranean Journal of Social Sciences MCSEER Publishing, Rome-Italy Vol 5 No 2 January 2014*

[12] Sáinz, M. & López-Sáez, M. (2010). "Gender differences in computer attitudes and the choice of technology-related occupations in a sample of secondary students in Spain". *Computers & Education*, 54, 578-587. doi:10.1016/j.compedu.2009.09.007.

[13] Van Rensburg, S., & Ankiewicz, P. (1999). "Assessing South Africa learners' attitudes towards technology by using the PATT (Pupils Attitudes towards Technology) questionnaire". *International Journal of Technology and Design Education*, 9, - 151.

[14] Sutton RE (1991). "Equity and Computers in the schools": A decade of research. *Rev. Educ. Res.* 61:475

[15] Goode, J. (2010). "Mind the gap: The digital dimension of college access". *The Journal of Higher Education*, 81, 583-618.

[16] Jones T., & Clarke V. A. (1995). "Diversity as a determinant of Attitudes: A possible explanation of the apparent advantage of single-sex settings". *Journal of Educational Computing Research*, 12 51-64.

[17] Bandura, A. (1977). *Social Learning Theory*. New York: General Learning Press.

[18] Krejcie, R.V., & Morgan, D.W. (1970). Determining Sample Size for Research Activities. *Educational and Psychological Measurement*, 30, 607-610

[19] Çelik HC, Ceylan H (2009). "The comparison of high school students' mathematics and computer attitudes according to various variables", *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*. 26:92-101.

[20] Adesoji Francis A. & Olatunbosun Segun M. (2008). "Student, teacher and school environment factors as determinants of achievement in senior secondary school chemistry in Oyo state, Nigeria". *The Journal of International Social Research Volume 1/2 winter 2008*.

[21] Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual ability and performance. *American Psychologist*, 52, 613-629.

[22] Aremu, S. A. (1999). "Games and Primary School Pupils' Achievement in Some Aspect of Geometry". *Unpublished Ph.D. Thesis*, Department of Teacher Education, University of Ibadan, Ibadan.

AUTHORS

First Author – Kuliwa Sibajene Syakwasia is a Microsoft Systems Engineer. He is currently the CEO for AfriCount Limited. His research interests include Mobile Based Elections Results Collation and Transmission Systems and Computer Science Education.

Second Author – Christopher Chembe received the B.Sc., M.Sc., University of Zambia, Ph.D. degrees, Computer Science, University of Malaya. He is currently the Head of Department – Computer Science and Information Technology at Mulungushi University, Kabwe, Zambia. His research interests include Cognitive Vehicular Networks, Optimizing Wireless Channels, Naive Bayes and Artificial Neural Networks and Big data.

Geospatial, Cloud and Web Based Model for Evidence Based Decision Making in Tuberculosis Prevention

Innocent Mwila
 Department of Electronics and Electrical Engineering
 University of Zambia
 Lusaka, Zambia
 innocentmwila@yahoo.com

Jackson Phiri
 Department of Computer Science
 University of Zambia
 Lusaka, Zambia
 jackson.phiri@cs.unza.zm

Stakeholders in the fight against Tuberculosis (TB) face a dilemma of the quick spreading disease that must be fought with minimal resources. The scarcity of the resources available for the fight against TB calls for creative and strategic ways if this fight is to be won. Information and Communication Technologies (ICTs), particularly cloud computing, geospatial data analysis and web technologies presents an opportunity towards the creative fight against TB. The ability of these technologies to capture and present spatial data in real-time on mapping applications makes them a good candidate for exploration in the fight against TB. The aim of this paper is to develop a decision support model that would help in the prevention of the spread of TB through improved monitoring and tracking using cloud computing, geospatial data and web technologies. This system will display on a mapping application the geographical distribution of TB cases and other reports which can be used to make decisions in the prevention of TB. Using the information provided by this system, TB program managers will be able to put the scarce resources where it matters more in the fight against the disease - where the cases are.

Keywords – Tuberculosis Prevention, Cloud Computing, Geospatial Data Analysis, Web Technologies, Model, Decision Support

I. INTRODUCTION

Zambia is among the 30 high TB burden countries in the world [1]. Each year, more TB cases are noted partly resulting from weak surveillance among other causes, as a result TB keeps spreading. In Zambia TB caused 17,000 deaths in 2015 [2]. This has a negative impact on development of the nation more especially that TB is more rampant in the productive age groups [3]. TB is an infectious disease which spreads from an infected person to one not infected and one way of preventing the spread that people who infected must be identified and then put in measures to curtail the spread to protect the ones not infected and ICTs can be utilized for this purpose, to improve surveillance. In Zambia ICTs, particularly cloud computing, geospatial data and web technologies have been less utilized in TB monitoring and tracking. Little is known of geospatial distribution of TB cases and therefore it is unclear as to which areas and communities TB program managers should urgently direct the resources in the fight against TB. This study looks at how cloud computing, geo spatial data analysis and web technologies can be used to equip TB program managers with knowledge needed to direct and apportion scarce resources into places and interventions which will result in the prevention of the spread of TB.

The lack of evidence based information for decision making in the fight against TB resulting from the lack of or

underutilization of ICTs such as cloud computing, geospatial data and web technologies, has led to little being known about the geospatial distribution of TB cases and other information needed for decision making in the fight against the spread of TB. This information is not readily available when required and as such it is unclear as to which areas and communities TB program managers should urgently direct the resources in the fight against TB. The aim of this paper was to develop a decision support model that would help in the prevention of the spread of TB through improved monitoring and tracking using cloud computing, geospatial data and web technologies. This paper sought to establish factors which escalate the spread of TB in Zambia, to design a model to be used to monitor and track the spread of TB based on cloud computing, geospatial and web technologies, and based on the designed model develop a prototype system to help with decision making in the fight against TB.

II. LITERATURE REVIEW AND RELATED WORKS

There have been efforts to use ICTs to help in the fight against the spread of Tuberculosis (TB). The ability of ICTs such as real time reporting coupled with the ability of Geographic Information Systems (GIS) to identify and display the geographical position of objects have been leveraged to improve disease surveillance and this has resulted in improvements in disease prevention activities. This section looks at the TB disease, factors which promote the spread of the disease and how cloud computing, geospatial data and web technologies have been and are being used in countries such as India, United States and Brazil to help with the prevention of spread TB through improved monitoring and surveillance. This section also looks at how cloud computing is being used to improve data capturing and reporting efficiencies in the agriculture sector in Zambia.

A. Tuberculosis

TB is an airborne disease which mainly affects the lungs and can spread to other parts of the body [4]. It is caused by the Mycobacterium tuberculosis bacterium and spreads from one person to the other by the exhaling of the bacteria from the infected person's lungs and the subsequent inhaling of these bacteria by a non-infected individual [4]. TB is a highly infectious airborne disease which is curable [5]. The disease is preventable and being aware of the factors which lead to its spread is key in stopping the further spread. Below are some of the factors that leads to the spread of TB as listed by [5][6];

Susceptibility – this is the immune strength of the person exposed to the TB bacteria. The weaker the immune system, the higher the likelihood for one to get infected with TB.

Infectiousness – this is the amount TB bacteria that an infected person can expel into the air. The more bacteria in air, the higher the chances for the next person to get TB.

Environmental factors which includes; (i) Concentration of infectious droplet nuclei in the environment. The more the TB bacteria in the air, the higher the likelihood of transmission of the disease. Exposure in small, enclosed spaces is likely to result in transmission of TB. (ii) Inadequate ventilation that results in insufficient dilution or removal of infectious TB bacteria from the environment supports the transmission of TB. Poor air circulation causes the recirculation of air containing the TB bacteria in the environment and this supports the spread of the disease.

Proximity and length of exposure; (i) Longer duration of exposure to a person with infectious TB accelerates the risk of transmission of the disease to the people around that person. (ii) The more frequent the exposure to infectious person, the higher the risk for transmission of TB. (iii) Physical proximity to infectious person is likely to support the transmission of TB.

Influencing factors (ii) and (iii) – which we have control over is important in the fight against the spread of TB and much of the emphasis in the fight must be focused on altering or being aware to these factors.

B. Information and Communication Technologies (ICTs) In Disease Surveillance and Spread Prevention

Information and Communication Technology (ICT) is the capturing, processing, storing and communicating information electronically within a digital medium. It includes the storage, communication and transmission of information using a variety of communication methods and devices such as satellite networks, mobile phones, wired and wireless networks as well as other digital forms [7]. The prevention of the spread of TB can be enhanced through better disease surveillance and ICTs can be harnessed for this purpose. ICTs have been used in different ways to improve disease surveillance [8][9] and thereby preventing further spreading or at least reducing the rate at which some diseases are spread. Disease surveillance is the ongoing systematic monitoring, tracking, collection and analysis of data and the provision of information which leads to action being taken to prevent and control a disease, usually one of an infectious nature [10]. Using ICTs, in particular, the leveraging of cloud computing, improved data management (through database technologies), geospatial data management and web technologies in TB disease surveillance can help to prevent the spread of new cases or reduce the rate at which the

disease is spread as is the case in India, United States and Ireland [11][12].

Spatial data indicates where things are within a given coordinate system while geo is the absolute or relative position of an object on the surface of the planet, Earth in this case [13]. Geospatial data, therefore, represents the positioning of an object on earth's surface and is associated with a coordinate system and so can be transformed onto a map. Vector data is used to depict a given point on a map and is in form of latitude and longitude pairs and these points can show the location of objects on a map can be analyzed [14] and used to action decisions. The scrutinizing and using of spatial data to interpret the relationship of objects and other associated attributes on the earth's surface is known as Geospatial Data Analysis [14] and Geographic Information Systems (GIS) are software systems used to capture, store, retrieve, analyze and display spatial data [15]. Advantages of GIS includes cost effectiveness, handling large volumes of spatial data and the ability for real time reporting have made GIS to be widely applied for disease surveillance [16].

Cloud Computing is the use of remotely hosted servers for data management instead local computers and servers [17]. The servers and other computer resources, hardware and software are hosted offsite and made available via the internet by service providers. Cloud Computing offers several advantages such as the pulling of resources to utilize economies of scale and scalability, cost efficiency, device independency, high computing power, location independency, redundancy, improved availability while allowing users to use low specification computers to take advantage of maximum computing capabilities [17]. Cloud Computing has presented itself as tool that can be harnessed for disease surveillance and help with the prevention of the spread of disease. Cloud Computing services are now being used in disease surveillance systems such as early warning systems of some infectious diseases so that timely steps can be taken to avoid epidemic breakout [18]. Whereas traditional systems of disease monitoring are time consuming, slow and use less efficient data collection and monitoring systems ICTs, Cloud Computing in particular provides speed, accuracy and efficiency in the epidemic outbreak detection using intelligent systems, databases, improved analytics, modeling, visualization and ontology mapping offering real time options for decision making [18].

Web Technologies collectively describe the types of technology used in communication over the Internet. The Internet is a global system or world-wide interconnected computer networks [19][20]. This web of connections enables people to communicate and access files and documents located on other computers (and other electronic devices) in this network using web browsers. A web browser is a program that runs on the client computers and requests

for information, files, documents or services from other computers [19][20]. Examples of web browsers includes Microsoft Internet Explorer, Mac OS X Safari, Mozilla Firefox, Netscape Navigator, Google Chrome and Opera.

C. Related Works

A research study conducted by [21] in the United States entitled Using GIS Technology to Identify Areas of Tuberculosis Transmission and Incidence illustrates how GIS and spatial data analysis can be leveraged to contribute to the fight against the spread of TB. With the limitation in the resources allocated towards TB screening and treatment programs, innovations through which high risk populations are targeted for testing and treatment were encouraged. Apart from identifying and clustering risk factors involved in the spread of TB in the areas being researched, residential addresses of TB patients were geocoded and mapped according to the disease strain. Going by these findings, the TB program managers started planning and implementing geographically informed TB prevention activities such as screening and treatment.

Spatial data analysis was once again put to task in this study done in Brazil to establish the effect of socioeconomic status on the transmission and spread of TB. The objective of this research study conducted by [22] in Vitoria, Espirito Santo in Brazil was to investigate the spatial patterns of the incidence of Pulmonary TB (PTB) and to establish if socioeconomic status contributes to the transmission and spread of the disease. The results from this research study revealed a strong spatial autocorrelation among incidence rates and that TB incidence and socioeconomic status had a significant curvilinear relationship. The study established a strong correlation between TB incidence and decreasing socioeconomic index as neighborhoods with lower socioeconomic index had increased TB incidence and severe rates. This finding confirmed the hypothesis that neighborhood poverty is strongly associated with increased TB incidence with individuals with lower socio-economic status being at greater risk of becoming infected with TB. Lessons learned from the geospatial analysis of the TB cases distribution helped TB control programs in Vitoria to allocate resources and target control efforts to the areas more prone to TB transmission and spread via a structured territory-based surveillance system. The results of the study provided an opportunity to plan interventions directed towards specific areas highlighted in the cluster analysis as TB hotspots. GIS is a powerful tool and [22] recommends its use for TB control and prevention in developing countries as it offers real-time surveillance and supports decision making.

The use of the cloud infrastructure to speed up the efficiency of data capturing and reporting and the display of information on a mapping application to help with the quick

understanding of activities and how they relate to their geographic position cannot be over emphasized. Researchers [23] and [24] used cloud computing and mapping technologies respectively to improve efficiencies in data capturing, reporting and mapping in the Zambia Agricultural sector. In their study whose objective was to support the Zambia Food Reserve Agency and Farmer Input Support Program (FISP) to improve efficiency in data capturing and to support the decision making process, [23] used the cloud computing architecture to achieve the set objective. In this study a cloud based system was developed to assist with the registration and verification of farmers and to also help the officers at the Ministry of Agriculture monitor the FISP activities closely. This system runs on a dedicated database of a web server on a networked infrastructure. Agriculture Extension Officers access the database via a mobile application on a smartphone or tablet using the available connection type such as 3G, 4G or WiFi. The system provides the officers with statistics and other reports needed to run the FISP efficiently.

III. METHODOLOGY

The aim of this study was to develop a decision support model that would help in the prevention of the spread of TB through improved monitoring and tracking using cloud computing, geospatial data and web technologies. A baseline study was used to collect data and descriptive statistics, correlation and regression analysis were used to analyze the results. To help the TB program managers with decision making to counter the identified risk factors which support the spread of TB, the a model was designed that would be used to monitor and track TB cases based on cloud computing, geospatial and web technologies. Finally, a prototype based on the model was developed to validate the results. The prototype development is based on the Cloud Computing Architecture, KoboToolBox running on Android Operating System, Google Mapping API and PHP Programming Language. The development of the prototype is in progress.

A. Data Collection and Population Study Sample

The proposed research study was carried out in Lusaka district. TB patients on TB treatment formed the sample study population. Participants of the study were identified from clinical records at health facilities from which they were obtaining their TB drugs and randomly picked. The questionnaire for interviewing participants for baseline study data collection was built on the elements in the *Identifying TB Transmission Risk Factors* section below. 183 participants were surveyed as at the time of writing. Data collection was done using KoboToolBox (KoboCollect) Application running on Android and the interview only took place once participants consented to take part in the study.

B. Identifying TB Transmission Risk Factors

The figure below shows the factors which promote the transmission and spread of TB as a summary of findings presented by [5] and [6] and upon which the questionnaire for answering research question 1 was developed;

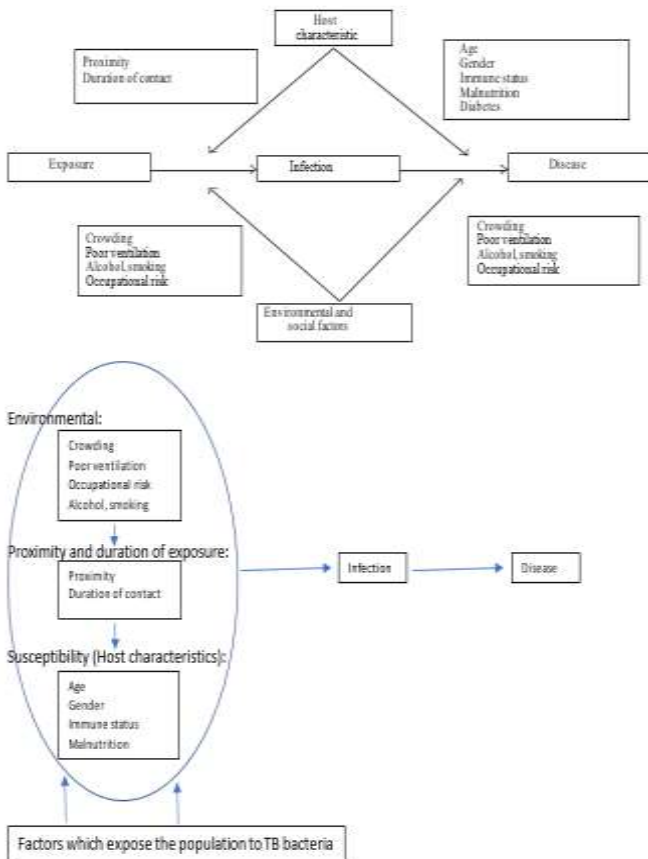


Figure 1 – TB transmission risk factors [5][6]

The disease is preventable and being aware of the factors which lead to its spread is key in stopping the further spread. Influencing factors such as proximity and length of exposure as well as environmental factors – which we have control over is important in the fight against the spread of TB and much of the emphasis in the fight must be focused on altering or being aware to these factors and spatial data analysis of already existing TB cases using GIS can help identify the geographical patterns of the transmission and spread of the disease and therefore inform strategic decisions in the fight against TB.

C. Model Design Methodology

The model under consideration in this paper is a data analysis decision support model. The model works by synthesizing the raw data into meaningful reports and map displayed information which TB program managers can use to make quick but effective decisions regarding TB prevention activities on which to focus on. The design model of the TB tracking and monitoring system consists of the data capturing application which sits on the mobile device. The application on the mobile device sends the captured data onto the database in the cloud. The web application on user’s computer embedded with a GIS application is used to display geo locations of TB cases and some summary reports which can be used for strategic decision making by TB program managers to help with TB prevention. The figure below shows the summary of the design of the model;

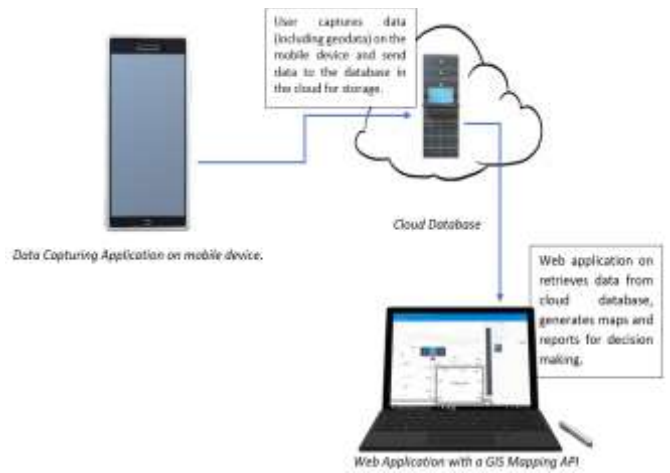


Figure 2 – proposed system application design

The user captures data (including geodata) on the mobile device and the data is sent to the database in the cloud for storage. The web application on the TB program manager’s computer then on retrieves data from the cloud database, generates maps and reports for decision making to help with resource allocation for prevention activities.

D. System Modelling

The model has two actors, the data capturer, who does the capturing of the data in the communities and the TB program manager, who retrieves the information, views maps and runs summary reports for decision making.

The figures below show the use cases for the two actors;

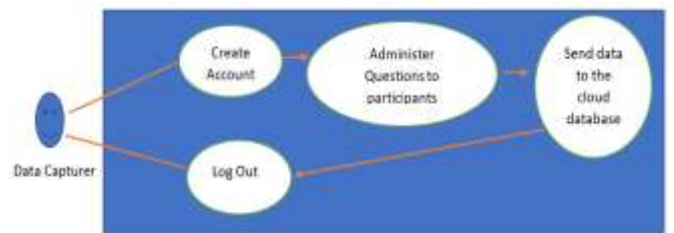


Figure 3 – Data capturer use case diagram



Figure 4 – TB program manager use case

IV. RESULTS AND DISCUSSION

The preliminary results of the data which has been collected so far show that crowding is the major driver of the TB disease spread with 76.9% (141 out of 183) of the surveyed TB patients living in homes which are considered to be crowded as shown by the graph below.

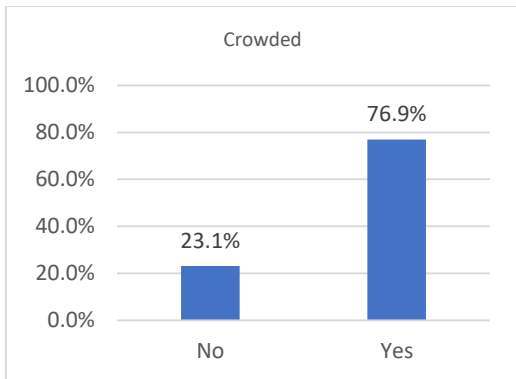


Figure 5 – Graph depicting crowding is a factor in TB spread

While crowding is eminent, 69.2% (127 out of 183) of the patients who were interviewed indicated that they did not consistently open windows of the rooms which they shared with the people who presented with likely symptoms of the TB disease such as coughing and unexplained weight loss as shown by the graph below.

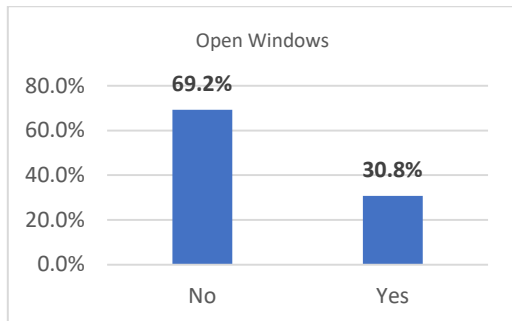


Figure 6 – Graph showing participants who never opened windows

70 out of 183 which translates into 38.5% of participants confirmed having being with a person who was on TB treatment or who was coughing, complained of having fever and chest pains or was unintentionally losing weight. Crowding being common and unimpressive window opening rates being observed, the rate of exposure to the TB disease causing bacteria is increased there by promoting the spread of the disease.

While smoking has been strongly associated with causing TB by the literature reviewed, only 37.5% (69 out of 183) participants indicated they were smoking prior to being diagnosed with TB as shown by the graph below.

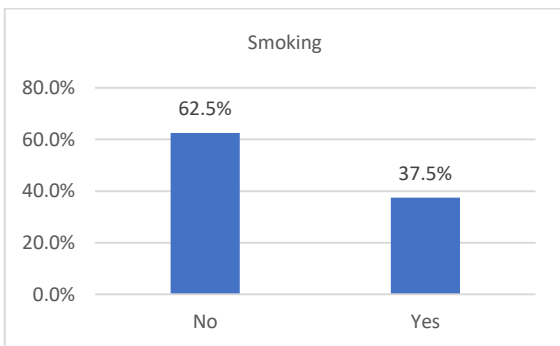


Figure 7 – Graph showing participants who smoked prior to TB diagnosis

Further analysis of the preliminary data has shown that 77.9% (143 out of 183) of the participants are males while 22.1% representing 40 are females as shown in the figure 7.

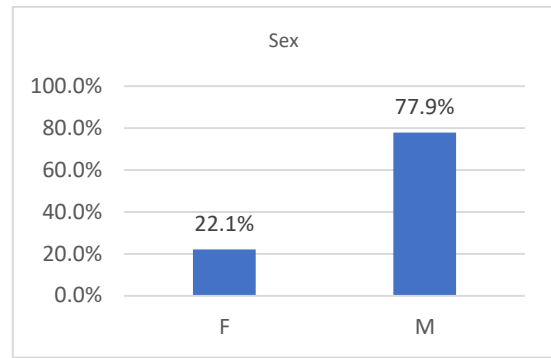


Figure 8 – Study participants by sex

From these findings it can be established that crowding and poor ventilation resulted into prolonged exposure to the TB causing bacteria for the individuals who contracted the disease. While smoking has been strongly associated with TB, the preliminary results have shown that only 37.5% of the participants were smoking prior to being diagnosed with TB. On the hand being male raises the likelihood of contracting TB to 77.9% compared to the 22.1% chances of the females. This can be loosely linked to the idea that males are mostly found in poor ventilated places like Bars and this is evident by the 70% of the participants who indicated that they were drinking before being diagnosed with TB.

There's a strong negative correlation, r between not opening window and getting the TB disease while there's slightly strong positive correlation between crowding and the TB disease. Below are some results of the correlation analysis of some selected variables.

Table 1 – Sample correlation

Variable	Response	r
Crowded	Yes	0.335047
Open Windows	No	-0.87242
Sex	F	-0.66489

Though there could be other factors which promote the spread of TB, TB program managers can put in intervention which can help to reduce the rate at which the disease is spread if they have this information and other reports this system will readily avail. Coupling this information with the geospatial display of TB cases on a mapping application will help TB program managers to put resources where it matters the most in order to fight TB.

V. CONCLUSIONS AND RECOMMENDATIONS

This paper presents a way of making well informed decisions in the face of constrained resources to help with the fight against TB. It utilizes the flexibility of cloud computing, the accuracy and reliability of geo spatial analysis and display of information on a mapping application made feasible by GIS and the ability to create reports and help to make decisions via the web application. It is recommended that the Ministry of Health, National TB Program adopts this model to help in the fight against TB.

A complete system prototype will be ready by the end of November, 2019 and by the same date data collection would have been completed.

VI. ACKNOWLEDGEMENT

I would like to acknowledge the National TB Program Management to facilitating the data collection used in this study.

REFERENCES

[1] UNAIDS, 2018. UNAIDS DATA 2018. [unaids.org](http://www.unaids.org) [Online] <http://www.unaids.org>. Available at: http://www.unaids.org/sites/default/files/media_asset/unaids-data-2018_en.pdf [Accessed 12th March, 2019].

[2] Government of the Republic of Zambia - Ministry of Health, 2016. National Strategic Plan (NSP) for Tuberculosis Prevention, Care and Control (2017-2021) - "Towards Elimination". Lusaka. Ministry of Health.

[3] Government of the Republic of Zambia - Ministry of Health, 2015. Zambia National Tuberculosis Prevalence Survey 2013 – 2014 Technical Report. Lusaka. Ministry of Health.

[4] The Global Fund to Fight AIDS, Tuberculosis and Malaria, 2017. Stop TB and Global Fund Deepen Cooperation to Find Missing Cases of TB. [Online] <https://www.theglobalfund.org>. Available at: <https://www.theglobalfund.org/en/news/2017-12-18-stop-tb-and-global-fund-deepen-cooperation-to-find-missing-cases-of-tb/> [Accessed 11th March, 2019].

[5] Narasimhan et al, 2013. Risk Factors for Tuberculosis. Volume 2013, Article ID 828939, 11 pages. <http://dx.doi.org/10.1155/2013/828939>.

[6] CDC, Transmission and Pathogenesis of Tuberculosis. Available Online at: <https://www.cdc.gov/tb/education/corecurr/pdf/chapter2.pdf> [Accessed on 29th May, 2019].

[7] Salvation Army. Poverty, and Information and Communication Technology. Salvation Army: Available Online at: https://www.salvationist.org/poverty_nsf/bda26fd1f7ee9db680256af00049eed1/9dd3057f1dea1d9080256b230032109a?OpenDocument [Accessed 23rd February, 2017].

[8] Doosje J, et al, 2004. Improving the Use of information and Communication Technology in Infectious Disease Control. Published In: EUROPEAN JOURNAL OF PUBLIC HEALTH VOL. 14 2004 NO. 4.

[9] Rukshan A and Miroshan A. AN ICT-BASED REAL-TIME SURVEILLANCE SYSTEM FOR CONTROLLING DENGUE EPIDEMIC IN SRI LANKA.

[10] William C. Shiel Jr. Medical Definition of Disease Surveillance. [Online] <https://www.medicinenet.com>. Available at: <https://www.medicinenet.com/script/main/art.asp?articlekey=26702> [Accessed 29th March, 2019 – 20:37].

[11] United States Centre for Disease Control and Prevention (CDC). 2015. KNOW YOUR EPIDEMIC: DISEASE SURVEILLANCE FOR TARGETING EFFORTS TO END TB.

[12] Health Protection Surveillance Centre, 2019. What is disease surveillance? [Online] <http://www.hpsc.ie>. Available at: <http://www.hpsc.ie/about/whatisdiseasesurveillance/>. [Accessed 7th April 2019 - 10:07].

[13] Mathworks, 2019. What Is Geospatial Data? Online. <https://www.mathworks.com>. Available at: <https://www.mathworks.com/help/map/what-is-geospatial-data.html> [Accessed 23rd May, 2019]

[14] Mathworks, 2019. Vector Geodata. Online: <https://www.mathworks.com>. Available at: <https://www.mathworks.com/help/map/vector-geodata.html> [Accessed 27rd May, 2019].

[15] Rezaeian, etal, 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. Published In: J Epidemiol Community Health 2007; 61:98–102. doi: 10.1136/jech.2005.043117.

[16] R, ZERBE et al, 2012. AN ANALYSIS OF BENEFITS FROM USE OF GEOGRAPHIC INFORMATION SYSTEMS BY KING COUNTY, WASHINGTON. RICHARD ZERBE AND ASSOCIATES, 939 21ST AVE EAST, SEATTLE WA, 98112

[17] Soniya K, Dr. A. Senthil Kumar. A Basic study on Cloud Computing. Published In: IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727 PP 43-46. www.iosrjournals.org.

[18] Sunil Chowdhary and Abhishek Srivastava, 2013. Cloud Computing: A Key to Effective & Efficient Disease Surveillance System. Conference Paper: Proc. of Int. Conf. on Advances in Computer Science and Application 2013

[19] Veer Surendra Sai University of Technology. INTERNET & WEB TECHNOLOGY-II LECTURE NOTES. DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, IT

[20] G. PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY. Lecture Notes on WEB AND INTERNET TECHNOLOGIES. Department of Computer Science and Engineering.

[21] Patrick K Moonan, et al, 2004. Using GIS technology to identify areas of tuberculosis transmission and incidence. Published In: International Journal of Health Geographics 2004, 3:23 doi:10.1186/1476-072X-3-23. <http://www.ij-healthgeographics.com/content/3/1/23>.

[22] E. L. N. Maciel, et al, 2010. Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil. The Union. INT J TUBERC LUNG DIS 14(11):1395–1402.

[23] A. Simukanga, J. Phiri, M. Nyirenda and M. M. Kalumbilo-Kabemba, 2018. E-governance systems: A case study of the development of a small-scale Farmer database. Published In: Zambia Information Technology (ICT)Journal, vol. 2, no. 1, pp. 7-15, 2018.

[24] S. Chihana, J. Phiri and D. Kunda, 2019. An IoT based Warehouse Intrusion Detection (E-Perimeter) and Grain Tracking Model for Food Reserve Agency. Published In: (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 9, 2018

Model Development of a Recommender System for Cognitive Domain Assessment Based on Bloom's Taxonomies

Raphael Banda
Department of Computer Science
The University of Zambia
Lusaka, Zambia
raphael.banda@yahoo.co.uk

Jackson Phiri
Department of Computer Science
The University of Zambia
Lusaka, Zambia
jackson.phiri@cs.unza.zm

Mayumbo Nyirenda
Department of Computer Science
The University of Zambia
Lusaka, Zambia
mayumbo@gmail.com

Abstract - Assessment is used a lot in the schools and universities and is synonymous to evaluation. It is used to measure learners' competences and proficiency after undertaking some course or a set of instructions. Online or hard copies Tests used to assess learners are considered as assessment tools. Using outmoded tools can be a big challenge in assessing the learners. Some of the challenges are failure to interpret and systematically award marks according to the questions asked. We introduce a system that uses the question ranking system based on action verbs of Blooms Taxonomies (BT). This technology automatically assigns each question a term weight according to the level of the verb on BT. Most methods used in designing questions do not take care of verb weights used as a result the output or interpretation of such results is exceedingly flawed and its credibility questionable. Developing an advanced model to decode the Bloom's Taxonomy of the Cognitive Domain in analysing marks to questions can be useful. The system can be used to assist educators to find appropriate action verbs to use when assessing candidates.

Keywords – test item, information coding, ranking system, Bloom's Taxonomy, cognitive domain, e-assessment, term weighting

I. INTRODUCTION

When we use sentences like: "I like the book", "I don't like the way the book was written" we are technically assessing the book by making a judgement maybe after reading the book. We assess things almost all the time and without even knowing it [1]. We say sentences such as "I like it", "This looks good", "Well done" or the opposite, "I should have prepared better", "I did not do so well" or "I hate it" [1]. In education assessment is done to test the competences that the learners have acquired. this competence test is used to award marks to the learners or for promotion purposes to another grade. However most of the testing or assessment tools that are used lack scientific backing. They are done in a hit-or-miss manner such that most people question their credibility as they lack scientific backing.

The marks awarded to the test items are normally done at random and may not represent the true mark to be awarded to the question. It is also clear that in some cases some questions may have been over rated and others underrated as it is the case with human errors. Common assessment schemes pose a lot of errors ranging from wrong entries for the learners by the teacher

to bulkiness of the entries to be entered on the sheet. Other challenges may include erroneous awarding of the marks to questions. It's also difficulty for the teacher to construe the results of such an assessment scheme. As a result, in wide range information searchers are not able to access the whole site causing incapability to obtain specific information [2] [3] [4].

In trying to make the results of similar tests justifiable some scholars have used the Jaccard coefficient for keywords. This technique uses key words from text to award the weight to such words. This technique can also help in awarding the necessary or justified marks to the questions depending on the frequency of such words. Technically, [2] developed a measure of similarity Jaccard with Prolog programming language to compare similarity between sets of data. This technique when employed can justify the marks that are given to each question in Figure 1 below.

II. LITERATURE REVIEW

A. E-Assessment

There are a lot of advantages associated with e-assessment. As stated above huge classes can undermine the way we assess learners as this may also result in inaccurate awarding of marks or appraisals to the learners due to human errors. Universities adopted e-assessment, in order to obtain develop, accurate and faster method to assess students, rather than traditional measure (paper-test) [3]. Challenges of e-assessment can also adversely affect the implementation of e-assessment. At lower grades the children need to be trained or assisted for them to accurately use computers or other related gadgets to be assessed like tablets. Although this may not be a big challenge there is also a problem of lack of such facilities that can be used to electronically assess the learners. Students need a training at the beginning to be familiar with E-assessment [3].

Data mining or word mining use similar architecture to execute the work. It starts with raw data that is processed so as to target the required data. This required data is transformed to create patterns that can be recognised by the research. It is these patterns that the research is interested to interpret to get some knowledge about the data collected. [5] have summarised this diagrammatically as shown in Figure 1 below.

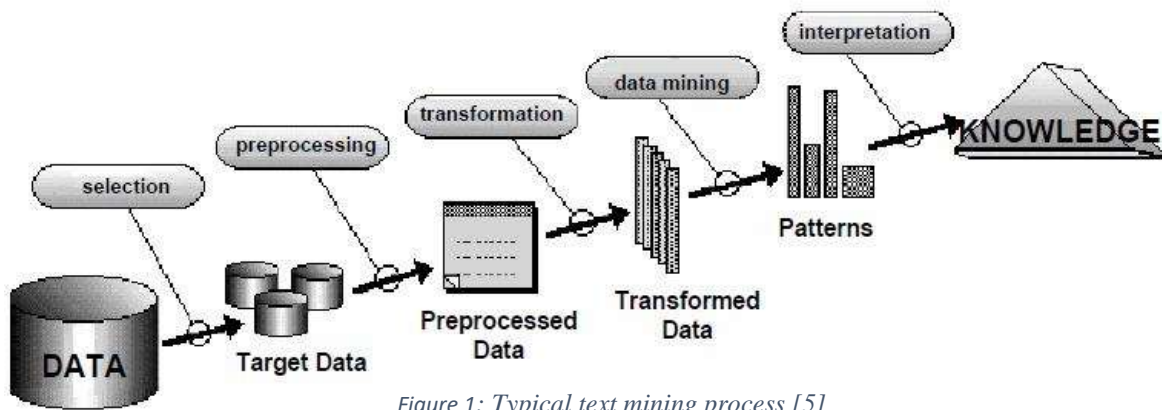


Figure 1: Typical text mining process [5]

There are five basic text mining steps as under [7]:

- a) Collecting information from unstructured data.
- b) Convert this information received into structured data
- c) Identify the pattern from structured data
- d) Analyse the pattern
- e) Extract the valuable information and store in the database.

B. Challenges of Natural Language

Natural language and natural writing are complicated for a computer to decode. It is this complexity that makes coding or programming the computer a great challenge. Complexity of natural language is main challenging issue in text mining [4]. The natural language is not free from the ambiguity problem [4]. One word may mean different things depending on the context under which it is used. This ambiguity leads to noise in extracted information and it cannot be entirely eliminated from the natural language as it gives flexibility and usability [4]. However there are problems that can be outlined as follows [4] [7]:

- i) The names of different entities and relationship between them can easily be found from the corpus of documents set using the technique such as information extraction.
- ii) The challenging problem of managing great amount of unstructured information for extracting patterns is solved by text mining. This also gives rise to some of the challenges or demerits of Text mining [4] [8]:
 - The information which is initially needed is nowhere written.
 - To mine the text for information or knowledge no programs can be made to analyse the unstructured text directly.

Before text is processed to have meaning to the translation there are some procedures that have to be taken by the system:

- Colons, semi-colons, opening parenthesis, closing parenthesis, hyphens, double hyphens, full stops, exclamation marks, question marks, single quotes and double quotes must be removed from the words or sentences.
- We removed the stop words from the translation, i.e., those words which are quite commonly used in English and have no direct relation to the questions

will have to be removed from the rest of the questions.

C. Technologies For E-Assessment

We have selected a few technologies relevant to developing of the comprehensive e-assessment model that will address the current problems of assessing learners

a) Term weighting

Term weighting is used to convert documents as vectors in the term space. Term weighting scheme, which has been used to convert the documents as vectors in the term space, is a vital step in automatic text categorization [5]. Term weight a technique in text mining and entropy from Shannon’s information theory are both used to quantify information [11].

There are two types of term weighting methodologies that are commonly used. These are; Traditional Term Weighting Methods and

i) Traditional Term Weighting Methods

The traditional term weighting methods for text categorization are usually borrowed from information retrieval field and belong to the unsupervised term weighting methods [5]. For example, the simplest binary representation is adopted by text classifiers which only involve binary operation rather than floating number operation, such as Decision Tree [5]. The term occurrences alone could be used as term weighting methods without other factors, such as raw term frequency (tf), $\log(1 + tf)$ and ITF [12][6]. [6] have adopted the most popular $tf.idf$ method as a baseline in their study.

ii) Supervised Term Weighting Methods

“Text categorization” is a form of supervised learning [5]. It makes use of prior information on the membership of training documents in predefined categories. This training of data set is effective and has been widely used in the step of feature selection and the construction of text classifier to improve the performance [6]. It is important to note that a natural generalization consists in giving each term a separate weight, according to its importance [7].

To correctly assign a mark to each question asked on the question sheet we must measure its constituent terms used in the question and assign a weight to each term used. [7] have proposed a new measure for the importance of

terms. In their analysis they have made consideration on which terms can be of great importance in the sentence. For example, the sentence “apple iPad. The term “iPad” carries most of the information, as it, alone, can tell us what kind of product is expected; while the term “apple” can be associated to a broader range of products [7]. In the context of our questioning techniques in the sentence “count objects” the term “count” carries a lot of information as it can tell us what action we need to do to the “objects”; while the term “objects” can be associated to a broader range of items

[7] have advocated for a method that when the same term is used in a large variety of purchases, it is less important than another term which is systematically associated to the very same purchases. The implementation of the idea is based on Shannon Entropy of a discrete probability distribution.

iii) *Shannon's Entropy Metric For Information [11].*

The entropy of a variable is the "amount of information" contained in the variable. Informally, the amount of information in an email is proportional to the amount of “surprise” its reading causes [8]. For example, if an email is simply a repeat of an earlier email, then it is not informative at all [8]. Shannon’s entropy quantifies the amount of information in a variable. This provides the foundation for a theory around the notion of information.

iv) *Mathematical Framework of Entropy-Based Term Weighting and how it works*

Shannon Entropy of a discrete probability distribution states that for a random variable X with values in a finite set X

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \geq 0 \quad [1]$$

The Shannon Entropy of a discrete probability distribution quantifies the *unevenness* of the probability distribution p . The minimum $H(X) = 0$ is reached for a constant random variable, i.e. a variable with a determined outcome, which reflects in a fully localized probability distribution $p(x_0) = 1$ and $p(x) = 0$ for $x \neq x_0$ [9]. At the opposite, $H(X)$ is maximal, equal to $\log_2(|X|)$, for a uniform distribution. $H(X)$ is also denoted [9]:

$$S(p) = - \sum_{i=1}^{|X|} p(x_i) \log_2 p(x_i) \quad [2]$$

which underlines the fact that entropy is a feature of the probability distribution p [9]. Entropy does not depend on the graph $x \rightarrow p(x)$, i.e. it is not a feature of the random variable itself but only of the set of its probability values [9]. This property reflects in a permutation invariance of $H(X)$: let the variable $\sigma.X$ obtained by a permutation of the states, namely $\text{Prob}(\sigma.X = x\sigma(i)) = p(x_i)$, then $H(X) = H(\sigma.X)$ [9]. Entropy trivially increases with the number of possible states: for an unbiased coin, $H = \log_2 2 = 1$ while for an unbiased dice, $H = \log_2 6 > 1$ [9]. Note that $\chi = \{1, 2, 3, \dots, 10\}$ or \mathbb{N}

Alternatively; Given a probability distribution π on a finite set I, the Shannon Entropy is defined as:

$$H(\pi) \stackrel{def}{=} - \sum_{i \in I} \pi_i \log \pi_i \quad [3]$$

Equation 3 log is the base 2 logarithm. Although it does not really matter what base we use as we can convert the figures according to the log we are working in. To each term t , associate the following probability distribution, referred to as *term purchase distribution*:

$$\pi(t) = \frac{1}{Z_t} \sum_{\{q,p\} \in \mathcal{D}} \mathbb{I}\{t \in q\} \delta_p \quad [4]$$

where δ_p denotes the probability distribution with all its mass on product p and Z_t , corresponding to the number of purchases associated to t is a normalization term such that π_t be a probability distribution over P . This is denoted as $H_{(t)} = H(\pi_{(t)})$ for the sake of simplicity [7].

Considering the above method we can simplify our questions on “Count fruits”

Table 1: Math illustration

Query	Result
Count fruits	P1
Count fruits	P2
Count oranges	P3
Count banana	P4

The entropy of terms in the previous sample can be calculated as follows:

$$\begin{aligned} H(\text{Count}) &= -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - 2 \times \frac{1}{4} \log_2 \left(\frac{1}{4}\right) = \frac{3}{2} \times \log_2 2 \\ H(\text{Count}) &= -2 \times \frac{1}{2} \log_2 \frac{1}{2} = \log_2 2 \\ H(\text{oranges}) &= -\log_2 1 = 0 \\ H(\text{pc}) &= -\log_2 1 = 0 \end{aligned}$$

Table 2: Term-products

Query	P1	P2	P3
Count fruits	0.5	0.25	0.25
Count fruits	0.5	0.5	0
Count oranges	1	0	0
Count bananas	0	0	0

This implies that there are 4 values altogether and we will need the following number of bits to represent them [8]:

- o $\log_2(4) = \log_2\left(\frac{1}{p}\right)$ bits = 2 bits

So, ‘c’ will need 2 bits to be represented as distinct from those other values.

Similarly, to represent the one ‘b’ we need

- o $\log_2(4) = 2$ bits.

To find the amount of storage required for the variable, different individual storage values of the variables are added according to their proportions and frequencies of occurrence [8]:

- o ‘a’ occurs half of the time and needs 1 bit: $0.5*1$
- o ‘b’ occurs a quarter of the time and needs 2 bits: $0.25*2$
- o ‘c’ occurs a quarter of the time and needs 2 bits: $0.25*2$

The table below is a summary of the calculations done above:

Table 3: Summary of Huffman coding

Value	Bits	Average	Total Bits
A	0	$0.5*1$	0.5
B	10	$0.25*2$	0.5
C	11	$0.25*2$	0.5
Total		1.5	1.5

The table of summary above determines the overall amount of storage required for the variable, we simply added up the above storage requirements in proportion to their frequencies of occurrence:

$$0.5*1 + 0.25*2 + 0.25*2 = 1.5 \text{ bits}$$

This is Shannon's formula for Entropy:

$$E = -\sum_i(p(i) \times \log_2(p(i)))$$

It is a direct measure of the "amount of information" in a variable [8].

D. Related Works

Most of the e-assessment technologies use one or most of the above-mentioned technologies to develop the assessments. Some of the common e-assessment platforms are outlined in the following paragraphs.

a. TAO

This platform was jointly developed by the Public Research Centre Henri Tudor (Centre for IT Innovation – CITI) and the University of Luxembourg (Educational Measurement and Applied Cognitive Science – EMACS). Platform consists in a series of interconnected modules dedicated to the management of subjects, groups, items, tests, planning and results in a peer-to-peer (P2P) network [10].

b. Hosted Test, from HostedTest.com

The platform allows users to create and edit questions, combine them into tests, and administer students using a Web browser [10].

c. Maple TA (MTA) from Maple soft Incorporated

Maple TA (MTA) is an online assessment system capable of handling most assessment tasks, but particularly designed for the assessment of mathematics, where it can call on the power of its Maple engine to probe deeper knowledge and understanding [10].

d. Exam pro plus developed by Cyber Space Ltd.

Exam pro plus is a customized e-assessment platform currently being used to administer electronic examination to students. It is a web application that uses the server-client architecture with the server being connected into an existing LAN [10]. In the platform, there are four integrated but independent modules comprising the global admin, the Lecturers, the technical staff and the students' module [10] [15].

III. METHODOLOGY

A. System Development

We used the well-known system of Blooms Taxonomy (BT) in determining the suitable questions for assessment papers. Blooms Taxonomy is commonly used in mathematics education. BT is hierarchical. This means that each level is a subset of by the higher levels. A learner who has mastered lower levels will find it easy to apply such knowledge at higher levels [10]. We have studied e-platforms and the technologies behind them to develop our model. The system shall concentrate on cognitive domains of *knowledge, comprehension and application* since higher level domains of analysis, synthesis an evaluation are rarely applied at lower grades. Figure 2 shows the levels and key action verbs used by educators and examiners.

The figure also depicts the revised Cognitive Domain of Bloom's Taxonomies.

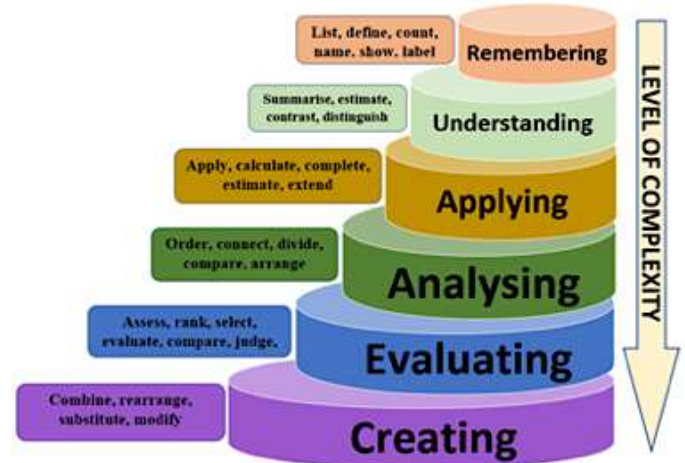


Figure 2: Six levels in Bloom's Taxonomy (BT) of cognitive domain

The Blooms taxonomy categorises levels of abstractions of questions that commonly occur in educational settings [10]. BT provides a useful structure in which to categorise test or examination questions.

IV. RESULTS

There are a lot of verbs that are associated with the Cognitive Domain of the Blooms Taxonomies six stages of *Remembering Understanding Applying, analysing, evaluating, creating*. Revised Bloom's Taxonomies were revised from the original concepts, from 1995-2000 and were made public in 2001 by *Anderson and Krathwohl*. We created the databases to store action verbs of Taxonomies of the Cognitive Domain. These verbs are used to create questions at different levels. *Table 4* shows the number of action verbs that have been collected from different sources and authorities in this field:

Table 4: Example questions according to Bloom's levels

Blooms Level	Example action Verbs	Number of verbs	Example test item
Remembering	Know Identify Relate Define List Recall	126	List three (3) known colours that are not primary.
Understanding	Restate Locate Report Recognize Explain Express	169	Estimate the length of line (b) by comparing it with line (a).
Applying	Apply Relate Develop Translate Use Operate solve	191	Apply quadratic formula to calculate the roots of the given quadratic equation.
Analysing	Analyse Compare Probe Inquire Examine Contrast	184	Examine the results of your experiment and record your conclusions.
Evaluating	Judge Assess Compare Evaluate Conclude Measure	136	Use appropriate method to solve simultaneous equations. Justify your answer.
Creating	Compose Produce Design Assemble Create	208	Develop an idea for a science fair project that focuses on the effects of pollution on plant life.

B. Analysis and justification of system

We wanted 90% confidence about our result. We also wanted the standard deviation for the population to be 2.9. The following steps were used to find the sample for the survey:

- 1: Find z-score by dividing the confidence interval by two, and looking that area up in the z-table:
 $0.90/2 = 0.45$. The closest z-score for 0.45 is 1.65.
- 2: Multiply step 1 by the standard deviation.
 $1.65 * 2.9 = 4.785$
- 3: Dividing Step 2 by the margin of error of 0.5.
 $4.785/0.5 = 9.57$
- 4: Square Step 3.
 $9.57 * 9.57 = 91.58$

We interviewed 91 teachers, lecturers and other education officials through self-administering questionnaires and the results were as shown below. In this survey we wanted to know how many of them took into consideration the awarding of marks to questions using Blooms Taxonomies. The responses to the question:” Do you use the Blooms Taxonomies when awarding marks to the questions that you set for your class? The responses ranged from *No, No idea Not sure, Sometimes, Yes*. The table below shows the responses:

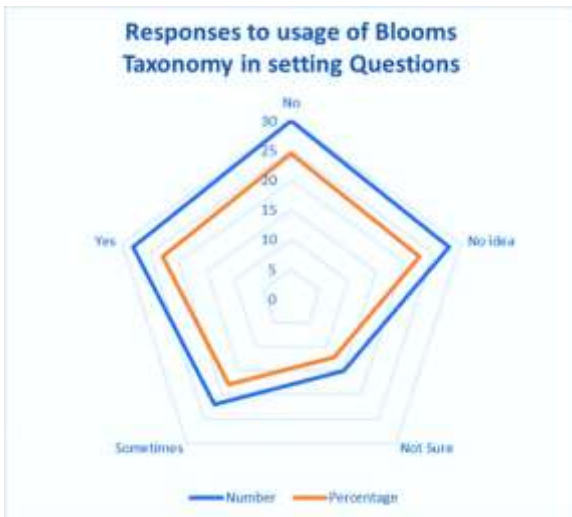


Figure 3: Responses to usage of Bloom's Taxonomies in setting questions

We look at the levels and the suggested marks for each question. Since there are 6 levels each level will have a 1 more than the previous lower level. The marks in this respect will be hypothetically be awarded 1, 2, 3, 4, 5, 6 marks to **Remembering, Understanding, Applying, Analysing, Evaluating, and Creating** verbs in a question that their respective verbs appear. respectively. By hypothesis we can award a mark for the level and another mark for the verb that appears in the question that describes it marks as shown in the table below:

Table 5: Examples of awarding of marks to questions based Bloom's levels

Example test item	Level of verb in the question	Total marks
Count the objects in the diagram below	count(1)	1
Estimate the length of line B by comparing it with that	estimate(2) compare (4)	2 + 4 = 6

Example test item	Level of verb in the question	Total marks
of line A.		
Apply the quadratic formula to calculate the roots of the given quadratic equation.	apply(3) calculate (2)	3 + 2 = 5

Analysis of questions as in Figure 6 can be easy if there are very few questions to be analysed. However, such a method can be tedious if it's done with a lot of question items. To analyse examinations items we have created an algorithms that incorporate:

- 1) a user interface for constructing queries, displaying and synopsising results shall be created.
- 2) a computational API for selecting, mark allocation and integrating content;
- 4) a general text search for retrieving relevant verbs from heterogeneous databases;
- 5) a collection of internal verbs from taxonomy databases

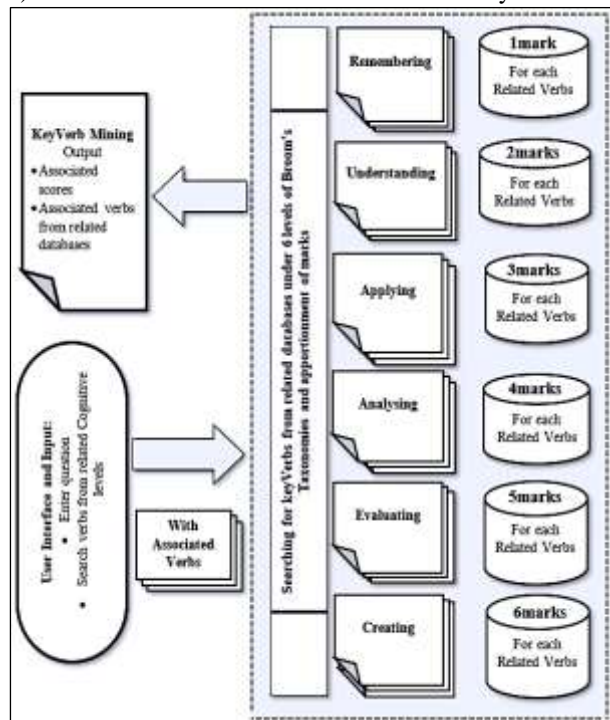


Figure 4: Model for awarding of marks according to Blooms Cognitive Domain

Figure 4 was be used to award marks to questions. This justifies the marks awarded to questions as outlined in Table 6.

C. The Technology Used



Figure 5: The WAMPSEVER logo

Databases as suggested in Figure 4 were created using MYSQL. This is part of the WAMPSEVER Version 3.1.3 Package. WAMP stands for "Windows, Apache, MySQL, and PHP." WAMP is normally installed as a software bundle comprising Apache, MySQL, and PHP. PHP stands for Hypertext Pre-processor. WAMP is often

used for web development and internal testing, but may also be used to serve live websites.

The coding of the programs was done in PHP using *Visual Studio Code*. PHP is a language interpreter that is freely available and used primarily on Linux Web servers. Below is the proposed user interface developed using PHP.



Figure 6: Screen shot of Home GUI of Model

We entered the sample question as shown in table 6. The system was able to count the frequency of action verbs used. Figure 5 shows the input and Figure 6 shows its output.



Figure 6: Screen shot of Sample Output of test items given in table 6

The proposed system will be used to analyse question papers to verify if the total number of mark given to the paper is appropriate. Figure 5 shows the proposed user interface to output appropriate results once the question paper items have been entered in the text box.

V. DISUSSION AND CONCLUSION

The results obtained from the model can be used by the teachers and assessors to prepare assessment questions. Blooms Taxonomy cannot be overlooked when preparing such questions since they measure the correct skills based on the hierarchy and order. System can be used for allocation of marks and standardizing assessment questions. Some systems for e-assessment use text mining to assess what has been learnt and key competences. However, text mining comes with its own challenges bordering on natural language. Computers cannot easily decode natural languages and such programmes can be complex to program. It is this complexity that makes coding or programming the computer a great challenge. Complexity of natural language is main challenging issue in text mining [4]. Our model is recommended to be used by teachers, and lecturers justifiably award marks to questions. With the implementation of text mining, the questions will be assessed according to the level of difficulty and the marks awarded accordingly [11].

ACKNOWLEDGEMENT

We wish to express our sincere gratitude to our survey participants for their contributions towards this paper.

REFERENCES

- [1] D. Wiesnerová, "Benefits of Self-assessment in English classes at Elementary School," *Bachelor Thesis*, pp. 5-62, 2012.
- [2] N. Suphakit, S. Jatsada, N. Ekkachai and W. Supachanun, "Using of Jaccard Coefficient for Keywords Similarity," *Proceedings of the International MultiConference of Engineers and Computer Scientists 2013*, vol. 1, pp. 1-5, 2013.
- [3] G. Salton, *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer.*, Addison-Wesley Publishing, 1989.
- [4] S. Tungwongsarn, *System for storage and retrieval of information by computer*, Bangkok: Pithaki Printing, 2010.
- [5] N. Alruwais, W. Wills and M. Wald, "Advantages and Challenges of Using e-Assessment," *International Journal of Information and Education Technology*, Vol. 8, No. 1, January 2018, vol. 8, no. 1, pp. 34-36, 2018.
- [6] M. Lan, C.-L. Tan, H.-B. Low and S.-Y. Sung, "A Comprehensive Comparative Study on Term Weighting Schemes for Text Categorization with Support Vector Machines," *Department of Computer Science, National University of Singapore.*, pp. 1332-1333, 2005.
- [7] S. Dang and A. P. Peerzada Hamid, "Text Mining: Techniques and its Application," *IJETI International Journal of Engineering and Technology Innovations*, vol. 1, no. 14, pp. 22-25, 2014.
- [8] V. G. Sonali, A. Chaugule and P. Patil, "Text Mining Methods and Techniques," *International Journal of Computer Applications (0975 - 8887)*, vol. 85, no. 17, pp. 42-45, 2014.
- [9] K. Sumathy and M. Chidambaram, "Text mining: Concepts, applications, tools and issues-an overview," *International Journal of Computer Applications*, vol. 80, no. 4, 2013.
- [10] F. Chen, K. Han and G. Chen, "An approach to sentence selection based text summarization," in *IEEE TENCON02*, 2008.
- [11] J. Phiri, T. J. Zhao and J. Mbale, "Identity Attributes Mining, Metrics Composition and Information Fusion Implementation Using Fuzzy Inference System," *Journal of Software*, vol. 6, no. 6, pp. 1025-1032, 2011.
- [12] J. Phiri and T. Zhao, "Identity Attributes Quantitative Analysis and the Development of a Metrics Model using Text Mining Techniques and Information Theory," *International Conference in Information Technology*, pp. 390-393, 2010.
- [13] L. Man, C.-L. Tan and H.-B. Low, "Proposing a New Term Weighting Scheme for Text Categorization," *American Association for Artificial Intelligence*, pp. 764-768, 2006.
- [14] Y. Jiao, M. Cornec and J. Jakubowicz, "An entropy-based term weighting scheme and its application in e-commerce search engines," *1st International Symposium on Web Algorithms*, no. 1, pp. 1-6, 2015.
- [15] S. Vajapeyam, "Understanding Shannon's Entropy metric for Information - A Primer for the Novice," *Research Gate*, pp. 1-6, March 2014.
- [16] A. Lesne, "Shannon entropy: a rigorous mathematical notion at the crossroads between probability, information theory, dynamical systems and statistical physics," *Math. Struct. in Comp. Science*, p. 3, 2011.
- [17] O. F. Bukie, "Understanding Technologies for E-Assessment: A Systematic Review Approach," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 5, no. 12, pp. 936-947, 2014.
- [18] J. Eva-Heinrich and B. Granshaw, "Pathways for improving support for the electronic management and marking of assignments," *Australasian Journal of Educational technology*, vol. 28, no. 2, pp. 279-294, 2012.
- [19] H. S. Tanalola, S. Fattahb, S. R. Sulong and M. Mamat, "Mining Exam Question based on Bloom's Taxonomy," *School of Engineering and Information Technology, University Malaysia*, pp. 424-427, 2017.
- [20] "Mining the Quran," [Online]. Available: <https://sites.google.com/site/miningthequoran/text-mining-of-qur-an>.
- [21] S. MeiLing, "MeiLing's BI E-Portfolio," [Online]. Available: <https://meilingbieportfolio.wordpress.com/2010/01/17/topic-8-text-mining-and-web-mining/>. [Accessed 10 March 2019].

GIS Management Tool for Managing of Underground Electrical and Telecommunication Networks at the University of Zambia

Twataizya Minango¹, Jackson Phiri² and Faustin S. Banda³

*Department of Geography¹, Department of Computer Science², Department of Geomatic Engineering³
The University of Zambia*

The university of Zambia's electricity and telecommunication utilities are supplied by underground cables. However, the current management system lacks spatial information about these cables. This due to the fact that the current system is based on out dated maps and drawings that lack locational data. This problem has resulted in the cutting of these utility cables during day to day construction of new infrastructure within campus. To efficiently manage the underground electrical and telecommunication cables, an improved system is required. In order to develop this system, the cables have to be mapped and digitally stored. Digital mapping of these cables will play a vital role towards their management. This would provide for accurate location of these cables as well as easily make better decisions when installing new network connections. Therefore, this paper focuses on the development of the management system of the existing underground electrical and telecommunication cables at the university of Zambia. To achieve this, the use of free and open source tools such as GPS, PostgreSQL/PostGIS and QGIS desktop software were used to establish a digital management system to easily update, store and analyze the utility network lines. A desktop and web based GIS prototypes were developed in this study.

Keywords: Electrical, Telecommunication, Mapping, QGIS, PostgreSQL, Leaflet JS, Turf JS

I. INTRODUCTION

Electrical and telecommunication networks are essential for the smooth running of a university. Most of the information on these utilities have become largely inaccurate because they have not been updated in a very long time. GIS can be used to improve operations, update network information and provide easy access to spatial data.

The current electrical and telecommunication networks utility management system is based on outdated scanned drawings and old maps (from 1966). These drawings lack spatial information on the network lines and makes it very difficult to precisely locate the exact path of the underground cables. With the day to day construction works within the campus, underground electric and telecommunication cables are vulnerable to being cut due to the lack of spatial information.

Therefore, the aim of this research is to develop a spatial model for managing the underground electrical and telecommunication networks at the University of Zambia using Free and Open Source Software (FOSS). The

University of Zambia will greatly benefit from this study as it would overcome the challenges and shortcomings of the current management system of electrical and telecommunication networks. The proposed will provide digital storage of spatial information which would allow for easy access, querying and analysis of spatial data as well as provide easy way of updating new data into the system. It will also help in making better decisions when planning for new or modification of existing connections.

GIS is a special case of information system in which "information is derived from the interpretation of data which are symbolic representations of features" [1]. In developing countries however, GIS is not used to address pressing needs in ways that are sustainable. This can be due to numerous challenges existing when implementing GIS, which are mostly organizational [2]. The emergence of FOSS based GIS has opened doors towards the implementation of GIS in organizations [3]. FOSS software has become a reliable alternative for many users (Mohammed WE, 2014). According to [4]. GIS can be used to improve operations, update network information and provide easy access to spatial data. [5] also state that by using GIS technology, complexities of utility network can be simplified and maintenance cost can be reduced to some extent.

II. RELATED WORKS

[6] conducted a study at the Copperbelt University. The aim of the study was to develop a WebGIS for effectively and efficiently managing the water utilities at the Copperbelt University using PostgreSQL/PostGIS, QGIS and GeoServer. This was to address the University's dependency on paper-based water utility maps which had not been updated in a long time. The framework for the WebGIS was structured to provide a centralized system with web-based access to accurate and updated information on utility spatial information throughout the University. The results of the study showed successful integration of PostgreSQL/PostGIS database with QGIS for desktop mapping and GeoServer for web mapping. The web application had functions that provided an interactive interface to the spatial data without location restrictions. However, the electrical and telecommunication networks were not mapped.

[7] looked towards improving the management and maintenance of underground electricity network infrastructure that was previously using data on papers,

spreadsheets tabular format, maps, etc. they discovered that this type of data was insufficient to locate the exact positions, maintain and update the data. The aim of the study was to address these issues, they suggested using GIS technology. The methodology used involved database development, map creation and publishing. GPS technology on a smart phone was used to obtain coordinates of the electrical components. A web-based GIS prototype was developed using QGIS. However, a less precise method of mapping was used. Mapping was conducted using a smart phone GPS which gives the approximate location of a mapped point.

[8] described the design and implementation of a GIS for the City of Calhoun was established in order to efficiently manage their utility distribution systems and replace the existing CAD system. The CAD system was not spatially referenced. To overcome this problem a GIS system was built using ArcGIS Server Enterprise Advance. The utility system comprised of 3 layers; water, sewer and electric layer. The electric layer helped the crews quickly identify the location of outages and the Electric Utility locate requests were being done mostly in the office using GIS instead of in the field. However, the GIS system was based on ArcMap software which is a commercial software and expensive.

The reviewed literature showed that the evolution of GIS is a powerful tool that can be used to solve problems and aid in decision making. Some of the tools, methods and methodologies used by previous researchers were adopted in this study. This study aims at extending GIS functionalities to mobile smart phone devices.

III. METHODOLOGY

The methodology of this research study was conducted as indicated in Fig 1.

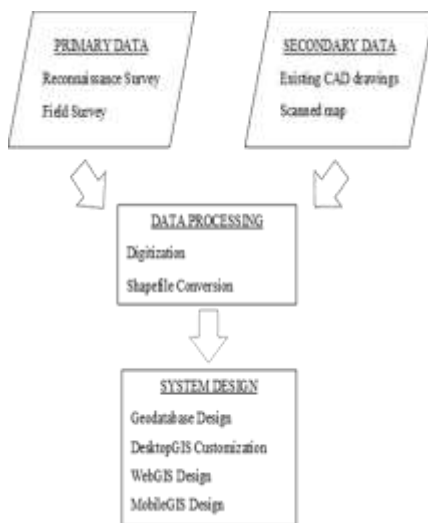


Fig. 1: Methodology Flow Chart

A. Data Collection

Both primary and secondary data were collected. Primary data collection involved firstly by conducting a reconnaissance survey in the field with guidance of the Resident Engineer and the path of the existing underground electrical and telecommunication network was noted using the Measurer mobile application. Precise mapping of underground electrical lines involved using an RTK GPS based on the pinned points obtained from Measurer mobile application. The GPS was used to capture the coordinates of the location of the network lines. Buildings were mapped through on-screen digitization from Google based. Existing CAD drawings and scanned maps showing the campus electricity networks were collected as secondary data. CAD drawings had shown the distribution and connection of the 11KV electrical network lines in form of a circuit diagram. While the scanned maps indicated the distribution of the electrical substations within University of Zambia area. Unfortunately, drawings and maps for telecommunication networks were nonexistent.

B. Data Processing

The csv file containing GPS Coordinates collected from the field survey were loaded into QGIS software as points. These points were saved into shapefile format (.shp). Lines were generated by joining each point was to connected to other corresponding points. The resulting lines represent the spatial arrangement of the existing underground electric and telecommunication network lines. Both electric and telecommunication lines were saved as individual shapefiles. Google satellite imagery basemap layer was uploaded into QGIS. Buildings were mapped through on-screen digitization on the uploaded satellite image using QGIS digitization tool and saved in shapefile format. Attributes of each individual shapefile were entered and stored in their respective tables.

C. SYSTEM DESIGN

The System design has 3 stages, database layer, server-side layer and client-side layer. Fig.2 shows the relationship between the system design and the methodological tools used:

i. Database Layer

PostgreSQL is a general purpose and an advanced open source object-relational database management system. PostgreSQL supports geographic objects so it can be used as a geospatial data store for location-based services and geographic information systems [9]. PostgreSQL allows the addition of custom functions developed using different programming languages. PostgreSQL version 9.5.3 was used in this study. PostGIS is a spatial database extender for PostgreSQL object-relational database. Security parameters were defined to grant administrator privileges [10].

Therefore, access to this geodatabase required the administrator to enter their login details. The substation points, joint boxes, buildings, electrical and telecommunication shapefiles were imported and stored in the geodatabase.

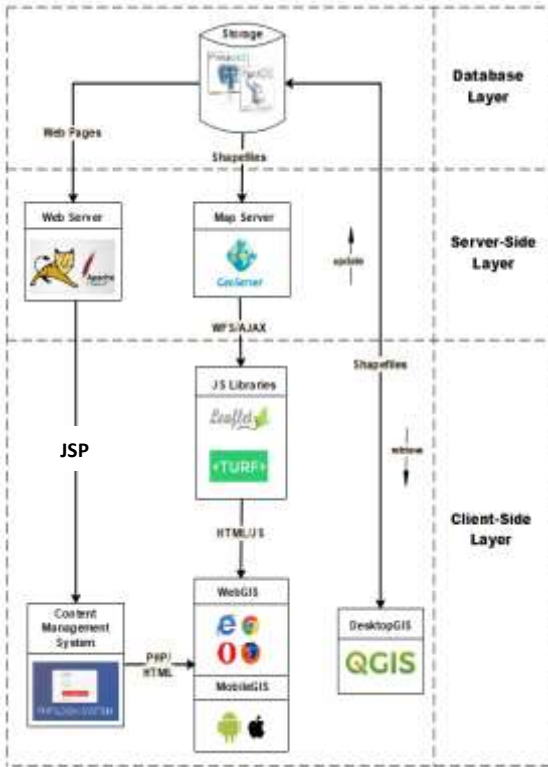


Fig.2: System Design Flow diagram

ii. Server-side Layer

This layer has 2 components, the web server and the map server. The web server was managed by Apache Tomcat. Apache Tomcat was used to implement Java Servlets and JavaServer Pages (JSP) to promote an effective Java server environment. Apache is an HTTP Server, serving HTTP and provides web services. Tomcat is a Servlet and JSP Server serving Java technologies. The map server was managed by Geoserver version 2.15.1. GeoServer is an open source software server written in Java that allows users to share and edit geospatial data (Geoserver, 2019). Geoserver was used to publish the shapefiles over the internet. These shapefiles are accessed via Web Feature Service (WFS), a protocol for transfer the Geographical information, such as maps, spatial attributes and Feature data sets.

iii. Client-side Layer

Also known as the application layer, is the layer where the end user interacts with the system. The end user can interact with the system via the desktop, web and mobile technologies.

iv. DesktopGIS Customization

QGIS toolbar was customized to display the appropriate tools suitable for the end user. The Customize Toolbars plugin was used to select the desired tools that would be available to the user. The toolbar provides the user with quick access to the tools. These tools included distance measuring, area calculation, spatial buffering, zooming and panning capabilities. The PostGIS geodatabase was connected to QGIS to access the shapefiles. For QGIS to access the shapefiles, the administrator’s user name and password were entered. Having successfully logged in all 5 shapefiles were added as layers in QGIS. Google map and Google satellite imagery were added as a basemap raster tiles upon which the shapefile layers were overlaid.

v. WebGIS

The web based GIS application was designed using HTML (Hyper Text Markup Language), CSS (Cascaded Style Sheets), Leaflet and Turf JavaScript libraries. The webpage was created using HTML 5 and styling was done using CSS. Leaflet JS was used to design the web page mapping capabilities. Leaflet JS enabled the display of the online maps. Via Leaflet plugins, the web based GIS application provided zooming, panning, searching, distance and area measurement functions. Buffering function was designed using Turf JS which allows for spatial data analysis functions.

vi. MobileGIS

Just like the WebGIS application, HTML, CSS and Leaflet JS were used to design the mobile based GIS application. However, a GPS navigation tool was designed using Leaflet Routing Machine plugin. It provides the shortest path between two points on a map during navigation. This functionality takes advantage of mobile phone Location Based Services (LBS) to provide the exact location of the network cables.

IV. RESULT AND DISCUSSION

The first objective was to map the existing electrical (11KV cable) and telecommunication networks at the University of Zambia. Fig.3 shows the mapped underground network lines. These network lines were mapped using the RTK GPS equipment. In Fig.3 the red lines represent the 11KV electrical cable and the blue lines represent the telecommunication cable. The red polygons represent the electrical substations and the blue polygons represent the telecommunication joint-boxes.



Fig.3: Mapped underground Electrical and Telecommunication networks

Second objective was to design a desktop based GIS application for managing the electrical and telecommunication networks. A centralized system based on QGIS desktop software was designed to digitally manage the mapped cables.



Fig.4: Desktop based GIS management system for underground Electrical and Telecommunication networks

Fig.4 shows the layers of the mapped network lines, points and digitized buildings in the developed DesktopGIS. These layers were overlaid on a Google map base-map layer. it is the main core of the system as it allows updates and modifications of the respective networks as a whole.

The third objective involved extending web and mobile technologies to the GIS management system. Therefore, WebGIS and MobileGIS applications were developed. The webGIS application allowed the user to visualize and analyze the utility networks via the web. MobileGIS provided navigation services to locate the underground network cables.

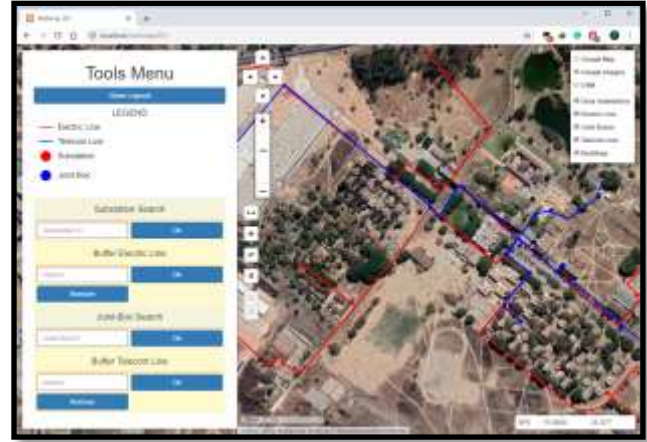


Fig.5: Developed web based GIS management system for underground Electrical and Telecommunication networks

Fig.5 shows the underground electrical and telecommunication networks being accessed via the web.



Fig.6: Developed mobile based GIS management system for underground Electrical and Telecommunication networks

Fig.6 shows the navigation path between 2 points, the first point being the location of the user (starting point) and the second being the destination point.

V. CONCLUSION

This study proposed the development of the underground electrical and telecommunication utility management system using GIS technology. The underground cables were mapped and stored digitally. The desktop, web and mobile based GIS management systems were also developed. The developed system provided tools that could help in decision

making concerning modifications and updates in future. Recommend that in future, the system should integrate with other campus utilities such as water and sewer networks.

REFERENCES

- [1] D. J. Maguire, M. F. Goodchild and D. Rhind, "An overview and definition of GIS, in Geographic Information Systems: Principles and Applications" New York: Wiley & Sons, Inc, 1991.
- [2] P. A. Chikumba, "Integrative Approach Implementing Geographic Information System for Health Management: Lessons Learnt from Malawi" *American Journal of Information Systems*, vol. 6, pp. 5-12, 2018.
- [3] S. Steiniger and E. Bocher, "An Overview on Current Free and Open Source Desktop GIS Development" *International Journal of Geographical Information Science*, vol. 1, pp. 1-24, 2008.
- [4] S. Grise, E. Idolyantes and M. Zeiler, *Water Utilities* New York: ESRI Press, 2012.
- [5] A. J. Rajeshkumar, P. Shashikant and S. Shashikant, "Web-GIS based application for utility management system," *Journal of Geomatics*, vol. 1 no. 1, pp. 1-12, 2017.
- [6] N. Okello, F. Banda and E. Tembo, "WebGIS for Water Utility Management at the Copperbelt University," *Imperial Journal of Interdisciplinary Research (IJIR)*, vol. 3, no. 8, pp. 601-606, 2017.
- [7] S. Bajaj and A. Adhave, "Mapping of Underground Electricity Network using GIS Software for Aurangabad City," *International Journal of Innovative Research in Science, Engineering and Technology* vol. 5, no. 7, pp. 12653-12658, 2016.
- [8] D. Crawford and M. Hung, *Implementing A Utility Geographic Information System For Water, Sewer And Electric*, Marryville: Northwest Missouri State University, 2012.
- [9] PostgreSQL, "PostgreSQL," 2019. [Online]. Available: <https://www.postgresql.org/about> [Accessed 16 July 2019].
- [10] PostGIS, "PostGIS," 2019. [Online]. Available: <http://postgis.net/features/>. [Accessed 16th July 2019]

A Framework For Digital Identity Management

Felix Kabwe
 The University of Zambia
 School of Natural Sciences
 Lusaka, Zambia
 fmlkabwe@yahoo.co.uk

Jackson Phiri
 The University of Zambia
 School of Natural Sciences
 Lusaka, Zambia
 jackson.phiri@cs.unza.zm

Abstract—The growth in the use of services on the World Wide Web has resulted in the proliferation of online cyber mischief, misrepresentation, and cybercrime. Cyber fraudsters and criminals hide their online identities to steal services and other valuables or harm innocent online services’ users. We will reflect on the major sources of identity attributes currently being used in the application and registration forms for the various services offered both in the cyber and real space. The study shall also consider an information systems model that would explain factors that influence users’ decision of who uses these tokens, when the tokens are used, and why identity tokens are used. This study ruminates identity tokens paradigms and their levels of importance. The study further explores the extraction of key identity attributes from identity tokens like identity documents, application and registration forms for the various services offered both in the cyber and real space. The techniques that have been used include data collection techniques, data mining techniques and statistical analysis. The primary data obtained from questionnaires was analysed using Statistical Package for Social Science (SPSS) and Excel. The respondents were drawn from Banks (14%), Churches (12%), Government of the Republic of Zambia (6%), Hospitals (16%), Insurance (10.7%), Mobile Phone companies (2%), and less than 1% from Pensions. Others were Schools (21%), Universities (16%), and Utility companies (1.3%). The paradigms under study include Usefulness, Trust, Ease of use, Image, and User satisfaction. The results helped in establishing a framework that was guiding the research.

Keywords— *Identity, Attributes, Metrics, Model, Digital, Real space, Authentication, Online, Services, framework, constructs*

I. INTRODUCTION

Many computing devices are also deployed in the environments where the users evolve—for example, intelligent home appliances or RFID-enabled fabrics [1]. In this ambient intelligent world the Internet is most likely going to generate more complicated privacy problems [2]. Digital identification has become a challenge in the cyber space. A stolen digital identity is a serious risk to assets, fraud, safety, and privacy; this reduces confidence in electronic business, cyber social networking, and communication in general. Privacy is a real human need which needs to be protected. People have busy lives and should not spend their time administering their digital identities. There is need to build technology that would help users to enable and secure usability. This part of the paper reviews the literature that focuses on similar research work, how the issue of identity management issues have been addressed as well as how the proposed solutions to the challenges have been implemented from previous works. The literature looks at Identity Management Mining, Metrics Composition, Digital Identity Modelling, Identity Quantitative Analysis, Development of a Metrics Model,

Multifactor Authentication System, Fuser Block Technologies Performance, and Design and Implementation of Multimodal Digital Identity Management System. The other literature review covers Using Artificial Neural Networks to Implement Information Fusion in Digital Identity Management Systems. The papers that have been reviewed have been listed at the Reference section of the paper [3][4][5].

With the introduction of internet banking, mobile sending of money through Mobile phone service providers like Airtel, MTN, and Zamtel, Zambia has identified theft of digital identities. Customers would send money electronically, using mobile phones, but their digital identification has been stolen, and sometimes people have been misrepresented or asked to reveal their identities so that their money could not reach the intended beneficiaries but get into wrong hands. Developing techniques of identification of the real owners of the digital identification would help resolve this problem. Zambia has also been rocked with identity management issues when it comes to distribution of farmers’ inputs by the government. People pose to be the true clients of the government to receive government support to receive assistance. Government has spent billions of Kwacha thinking the recipients are the intended beneficiaries when in fact not. When it comes to a lot of Government Ministries, it is a big problem to identify the registered needy for them to receive government support. This has also led to huge government expenditure. Identifying officers who qualify to have fuel from the fuel stations is another area that gobbles huge expenditures, officers use the card to fill in the tanks of their vehicles with very poor identification using electronic smart cards. A project like this would contribute to knowledge towards understanding of digital identity token users and their perception and their choices of use of identity tokens.

The rest of the paper is arranged as follows: In Section II - we have Literature Review, Section III – Methodology, Section IV – Results and Discussion, and V – Conclusion.

II. LITERATURE REVIEW

Elements of Identification In a major identity management initiative digital identity is defined as “the distinguishing character or personality of an individual. An identity consists of traits; digital identity management is a key issue in online service, security and privacy” [6]. For an entity to log in to a service, verified, and authorized, attributes of a digital identity have to be considered. Attributes of an identity token, verification, authorization, and play a major role in digital identity management as indicated in Figure 1 below..



Figure 1: [6] Identity management elements

Identity management elements include the user who has to request to sign in for issuance of a service, or for authentication for access to an electronic or non-asset. The identification request has to be verified by a single or multi-mode verifier. The authorisation of the identification permits the requester access to the service or asset. Authentication is an assumption of trust. Therefore, it is at this point, in this process, where fraud can either prevented or allowed.

The previous works have looked at identification, techniques that could be combined to come up with a multimode identification, methods of mining the attributes that could be used in metric model quantification. In this case, an entity (which could be a person or thing) can be identified using specific attributes of identification like user name, objects like electronic cards, biometrics like voice etc. The frequency of the attributes or identifiers which could be used in the identification would help in establishing metric models quantifications which could further help in uniquely identifying the entity.

A. Entity Identifiers

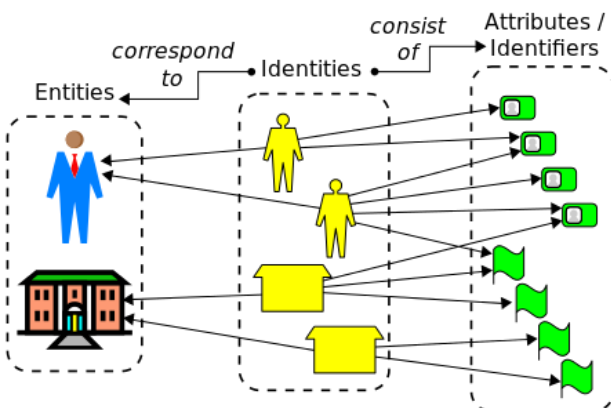


Figure 2: [6] Relationship among entities, identities, and attributes/identifiers

Identity encompasses all the essential characteristics that make each human unique but also all the characteristics that enable membership to a particular group or culture as well as established status within the group [7]. The building blocks of identity of an entity include identifiers which we could term as attributes. Each identifier is different from the other identifier; in that respect we would say identifiers are individually unique. The identity of a person comprises a large number of personal properties. All subsets of the

properties represent partial identities of the person and may relate to roles the person plays. Depending on the context, the person may have multiple different partial identities [8].

Figure 2 above, shows that an entity has attributes which build an identity or identities of an entity.

The relationship among entities, identities, and identifiers [6]:

- Entity: A thing or user that wants to have access to a service
- Identity Provider: the issuer of user identity
- Service Provider, the relay party imposing an identity check
- Identity, a set of user attributes
- Personal Authentication Device, which holds various identifiers and credentials and could be used for mobility

The process on identity management includes authentication. This is a process of verifying claims about holding specific identities. A failure at this stage will threaten the validity of the entire system. Technology is constantly finding stronger authentication using claims based on [9]:

- Something you know (password, PIN)
- Something you have (one-time-password)
- Reachability. The management of reachability allows a user to handle their contacts to prevent misuse of their email address (spam) or unsolicited phone calls.
- Authenticity. Ensuring authenticity with authentication, integrity, and nonrepudiation mechanisms can prevent identity theft.
- Anonymity and pseudonymity. Providing anonymity prevents tracking or identifying the users of a service.
- Organization personal data management. A quick method to create modifies, or deletes an accounts is needed, especially in big organizations [10].

B. Digital Identification Process

In most cases, digital identification requires that the requester for identification must be on the database for identification. The entity must have made an application to be enrolled for recognition. The application could be done by filling in a form where details of personal identification are captured on a database. The applicant becomes subscribed and a token of identification is issued. The subscriber becomes a claimant until specific identification can be made by the applicant that is when authentication has to be made. A session of authentication takes place and verification of details that were submitted by the subscriber are thoroughly checked. The identification of the subscriber is done as verification of the personal details of the subscriber; the verifier has to be a secure mechanism. The attributes of the subscriber are the subject of authentication. It is therefore, important that a lot of attention is paid to the attributes of identification as they are the pillar of the security of identification of a subscriber. In our work, we will consider the application forms and the attributes identified which we shall apply in our work to strengthen further the process of identification.

When identification is done, the subscriber can be authenticated.

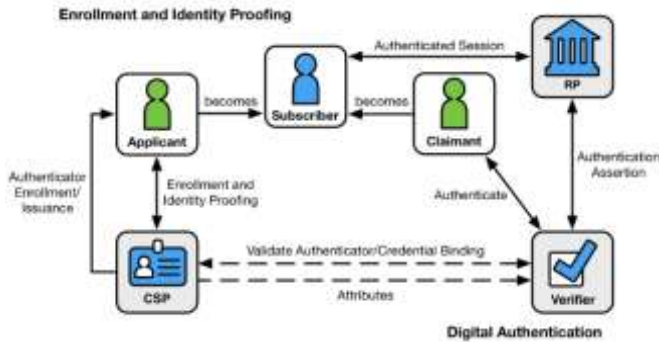


Figure 3: Identification and Authentication process

C. Trust Frameworks

Some Trust Frameworks are being developed by different establishments. The Higgins Trust Framework (HTF or Higgins, for short) is an open source project under development by the Eclipse Foundation that seeks to make sharing of identity information easier and more secure. IBM, Novell and Parity Communications are among the organizations contributing to the project. According to Dale Olds, an engineer at Novell, the purpose of the project is to give users more control over their online identity information [11]. The Higgins framework enables users to securely store identity information and related data and to integrate that data across multiple systems and applications. Stored data can be shared anonymously among Web applications, online vendors and service providers in a controlled manner [12]. HTF API (application program interface) can be thought of as a repository for cookie-like data that makes it convenient for users to conduct ecommerce and interact with Web sites, without the security problems inherent in conventional cookies [13].

D. Framework influencing Research

The works of P. N. Soneka and J. Phiri (2019) indicate there is a model which reflects that factors that influence the level of e-Tax systems adoption in Zambia using Technology Adoption Model (TAM) are useful, easy to use, and secure [14]. The model was based on the tested and well used framework of Technology Acceptance Model.

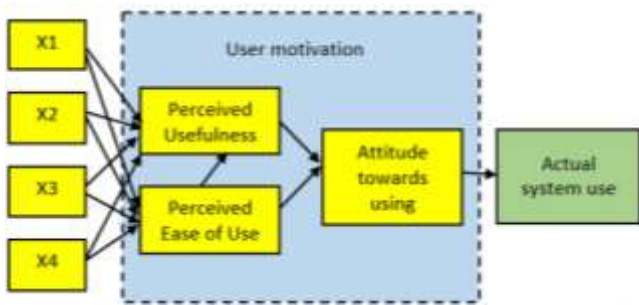


Figure 4: Model influencing this research

From this research, it emerged that the Technology Acceptance model (TAM) was significant in this work. The

perceptions of the respondent in the industries of these study converged on the three constructs to have been influencing the research. In this research, we use TAM to come up with a framework in Digital Identity Management. We identified some constructs to test a variables which were relevant in this research. These included Perceived Usefulness, Perceived Ease of Use, and Attitude towards using technology. These variables showed to be very influential in this study. TAM has been widely applied to a diverse set of technologies and users [15]. TAM provides a basis for tracing how external variables influence belief, attitude and intention to use new technologies [16]. TAM reflects the fact that the actual use of a new technology depends upon the user's attitude towards that technology the perceived ease of use of the technology, the perceived ease of use of the technology, and the perceived benefits that can be derived from using it [16]. From the model, it is deduced that whenever an individual perceives that using a certain technology will bring some benefits to the company, the individual will be eager to adopt it. At the same time, if the new technology is perceived to be easy to implement, people will not be hesitant to adopt it [16].

III. METHODOLOGY

A population of 180 size consisting Banks, Insurance companies, Government Ministries, Departments, organization, Universities, Schools, Hospitals, Mobile phone companies, Utility companies, and Churches was surveyed. Questionnaires were distributed to these organisations. Documents as source of identifications in these respective industries were considered to assess the perception of document users on these documents regarding the following: Usefulness, Trust (secure), Ease of use, Image (status of users), and User satisfaction. The other area of consideration was the identity attributes that these identity documents were interested in regarding the importance each of the attributes ranging from least important to the most important. The questionnaire also considered the demographics of the respondents and frequency of use of the identity documents. From 160 questionnaires that were distributed, there were 150 successful responses from the respondents. Using SPSS software, statistics of the data was analysed in different areas of interest.

We have considered our observations in the model that had influenced our study. We also observe that we have identified some paradigms that came out to be strongly related based on our regression on our statistical evidence. We therefore, propose that the applicable model of our study is the one presented in the image below [17].

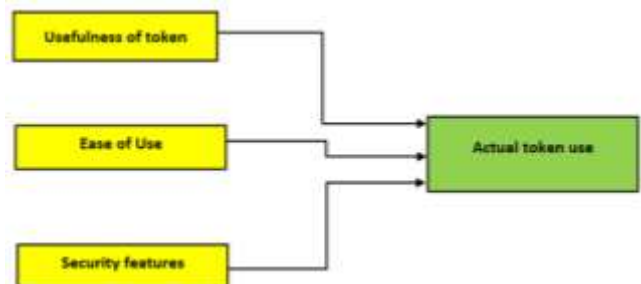


Figure 5: Proposed research framework

IV. RESULTS AND DISCUSSION

1.) *Mean Score of Constructs:* Mean score of the tokens of identity on five given constructs:

TABLE 1: MEAN SCORE OF TOKENS OF IDENTITY

Institution and tokens	Constructs				
	Usefulness	Trust	Ease of use	Image	User satisfaction
Banks	4.59	4.41	4.18	3.80	4.09
Insurance	4.06	3.88	3.88	3.49	4.11
Churches	4.91	4.97	4.97	4.69	4.69
Government	4.94	4.82	4.80	4.50	4.71
Hospital	4.64	4.40	4.53	4.25	4.59
Mobile Phone Companies	4.83	3.58	3.50	3.50	4.83
Schools	4.80	4.66	4.80	4.56	4.49
Universities	4.60	4.42	4.32	4.16	4.49
Utility Bills	5.00	4.00	3.50	4.00	4.50

For the banking industry usefulness was ranked highest (means score 4.59), followed by Trust (4.41)& Ease of use (4.18), user satisfaction (4.09) and the least was image or status (3.80).

The scoring of participants on each each token was out of 5 on average. Considering the outcomes of the research on this table, we have the constructs that were picked to be very strong influencing respondents. It must be observed that “usefulness” of identity tokens has been found to be very important. This was followed by “Trust”. Trust is an element of security. Which implies that security in using identity tokens of great importance. Third in hierarchy comes “Ease of Use” of an identity token. The easier an identity token it is to use, the more favourable it is to the users. “User satisfaction” came forth, in terms of importance. Users of identity tokens respond well to tokens that satisfy them in achieving their objectives. It was observed that “image/status” of using an identity token is equally important as if was above average in rating. However, it came last in the series of importance in this research.

The scores were out of 5 and the mean scores were very high on the paradigms Usefulness, Trust, Ease of use, and satisfaction. However, even if the paradigm of image was the lowest it was above average which was at 3.8 i.e. at 76%. The results show us that for the bank, all these constructs are important.

2.) *Ranking of Constructs for Selected industries:* We had to consider the industries that we sampled on and recognised that Government, Hospitals, Financial institutions, Insurance companies, Schools, and Utility companies were quite sensitive in the use of identity tokens.



Figure 7: Means scores for each industry across the 5 dimensions

3.) *Mean Score of all Attributes from Identity Tokens*

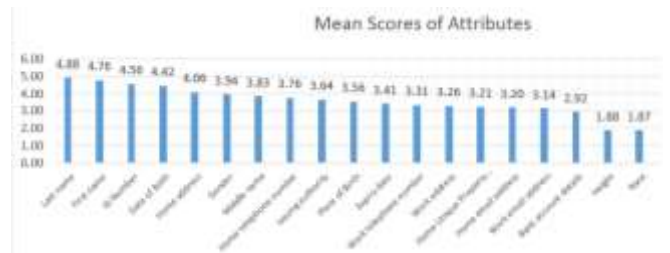


Figure 6: Ranking of attributes in terms of importance

The respondents sampled were main male (53%) and around 40% were single. Close to 50% were aged below 30 years. In terms of education level, only 10% had either grade 12 or less in qualifications – with 37% having a college diploma, another 35% with university degree with the rest have a masters or PhD. Less than 2% were not employed, around 5% being pensioners or self-employed respectively. Almost 88% were salaried employees.

4.) *Top Ten Important Attributes:* The research’s ten most important attributes from primary data from identity tokens

TABLE 2: TOP TEN MOST IMPORTANT ATTRIBUTES FROM DOCUMENTS

Sn.	Attributes	Mean Score on the importance of Attributes
1	Last name	4.88
2	First name	4.76
3	ID Number	4.56
4	Date of Birth	4.42
5	Home address	4.06
6	Gender	3.94
7	Middle name	3.83
8	Home telephone number	3.76
9	issuing authority	3.64
10	Place of Birth	3.56

It was observed that the key attributes that scored highest as the top ten include: Last Name, First Name, ID Number, Date of Birth, Home Address, Gender, Middle Name, Home Telephone Number, Issuing Authority, and Place of Birth.

5.) *Token of Identity Mean Scores:* Mean score of the tokens of identity on five given constructs for the banking industry:

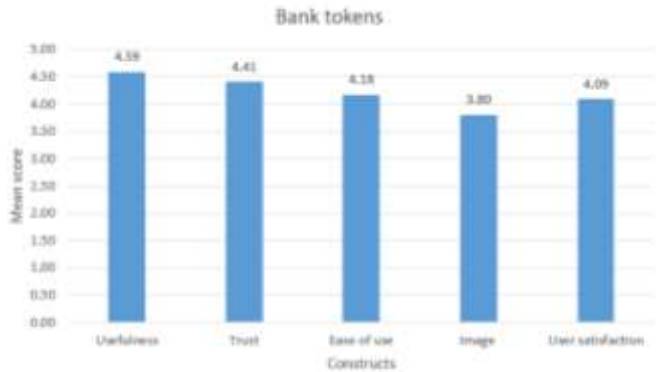


Figure 8: performance of constructs in banking industry

The graph shows the perceptions of the respondents on five different paradigms on identity tokens in the banking industry. Our focus in this article was on the banking industry to consider the perceptions of respondents on five different paradigms.

6.) *Model summary across different industries:* Different industries in our studies included: Banks, Insurance, Churches, Government, Hospitals, Mobile Phone companies, Schools, Universities, and Utility Bill companies. Regression on the statistics from this study was conducted on the five constructs (Usefulness, Trust, Ease of Use, Image, and User Satisfaction). Results of this regression showed that the three (3) constructs showed to be very strong close to each other across these industries. These constructs included: Ease of Use, Usefulness, and Trust.

TABLE 3: RESEARCH MODEL SUMMARY

Model Summary									
Model	R	R Square	Adjusted R Square	Std Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.274 ^a	.075	.054	.39392	.075	3.469	3	128	.016

a. Predictors: (Constant), Ease of use , Usefulness , Trust
 b. Dependent Variable: Offent

7.) *Ranking of constructs of identity tokens:* In general terms, our study observed that the sequence of importance of the constructs in the selected industries were as follows: Usefulness, Trust, Ease of Use, User satisfaction, and then Image or status (in that order).

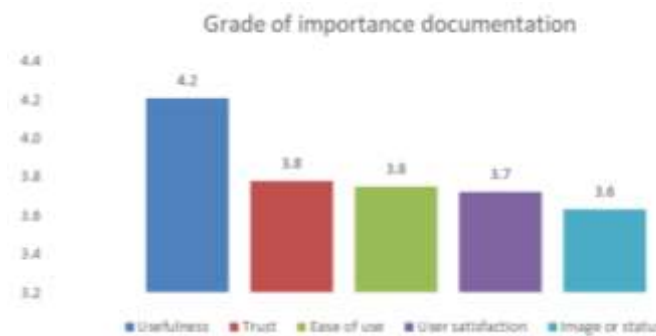


Figure 9: Ranking of constructs

V. CONCLUSION

There are different paradigms that affect the use of identity tokens which affect the use of identity tokens. These include Usefulness, Trust, Ease of Use. These paradigms can be ranked in the way they influence the users of identity tokens. Identity attributes that affect digital identity can also be ranked according to their levels of importance. Identity tokens differ in sensitivity in terms of application from one industry to the other. In choosing identity tokens for the entities,

To add value to identity management, it worth considering the influence the paradigms have on the identity

VI. REFERENCES

- [1] J. Agbinya, N. Mastali, R. Islam, and J. Phiri. "Design and Implementation of Multimodal Digital Identity Management System Using Fingerprint Matching and Face Recognition", 7th International Conference on Broadband Communications and Biomedical , pp. 278 - 277, 2011.
- [2] J. Phiri and T Zhao. "Identity Attributes Quantitative Analysis and the Development of a Metrics Model Using Text Mining Techniques and Information Theory" IEEE International Conference on Information Theory and Information Security, pp. 390 - 393, 2010.
- [3] W. Inambao and J. Phiri. 2018. "Digital Identity Modelling for Digital Financial Services in Zambia". ICTACT Journal on Communication Technology, , Vol.9, No.3, pp. 1829 - 1837, 2018.
- [4] J. Phiri, T. Zhao, C. Zhu. "Using Artificial Intelligence Techniques to Implement a Multifactor Authentication System" International Journal of Computational Intelligence Systems, pp. 420 - 430, Vol. 4, No. 4, 2011
- [5] J.-M. Seigneur and T. E. Maliki, "Identity Management," in Computer and Information Security Handbook, Burlington, Morgan Kaufmann, 2009, pp. 269-297.
- [6] R. Bhasker and B. Kapoor, "Computer and Information Security Hand Book," in Computer and Information and System Security, Burlington, Morgan Kaufmann Publishers is an imprint of Elsevier, 2009, pp. 259-257.
- [7] G. Roussos, D. Peterson, and U. Patel, "Mobile Identity Management: An Enacted View", International Journal of Electronic Commerce, Vol. 8, No1, pp. 81-100, 2003
- [8] S. Clauss and M. Koehntopp , "Identity Management and its support of multilateral security, Computer Networks", Vol. 37, pp. 205-219, 2001
- [9] K. Kathirvel. "Credit Card Frauds and Measures to Detect and Prevent Them," International Journal of Marketing, Financial Services & Management Research, vol. 2, no. 3, pp. 1-8, 2013.
- [10] M. S. Gaigole and M. A. Kalyankar. "The Study of Network Security with Its Penetrating Attacks and Possible Security Mechanisms," International Journal of Computer Science and Mobile Computing, vol. 4, No. 5, p. 729, 2015.
- [11] K. I-Lung. "Securing mobile devices in the business environment," IBM Global Technology Services; Thought Leadership White Paper, pp. 2-10, 2011.
- [12] K. Christian, B. Katja, T. Markus, H. Stephan and R. Kai. "How to Enhance Privacy and Identity Management for Mobile Communities: Approach and User Driven Concepts of the PICOS Project," Mobile Business & Multilateral Security, 2010.
- [13] R. Bhasker and B. Kapoor. "Information Technology Security Management," in Computer and Information Security Handbook, Burlington, Morgan Kaufmann Publishers is an imprint of Elsevier, 2009, pp. 259 -267.
- [14] P. N. Soneka and J. Phiri, "A model for improving E-Tax Systems Adoption in Rural Zambia based on the TAM Model". Open Journal of Business and Management, Vol. No. 7, pp.908-918., 2019
- [15] V. Venkatesh et. al (2003). "User Acceptance of Information Technology: Toward a Unified View", MIS Quarterly Vol. 27 No. 3 pp 428/September 2003

- [16] R. Matikiti, M. Mpinganjira, and M. Roberts-Lombard (2018). "Application of the Technology Acceptance Model and the Technology-Organisation-Environment Model to examine social media marketing in the South African tourism industry" South African Journal of Information Management, ISBN: (Online) 1560-683X, (Print) 2078-1865
- [17] R. J. Holden and B. Karsh. "The Technology Acceptance Model: Its past and its future in health care" Journal of Biomedical Informatics, University of Wisconsin-Madison, Madison, WI, USA; July 2009.

Passenger and Luggage Tracking System Using Sensor Networks for Public Transport

George Kasanga¹ Jackson Phiri²

¹*Department of Electronics and Electrical Engineering, University of Zambia, Lusaka, Zambia,*

²*Department of Computer Science, University of Zambia, Lusaka, Zambia*

¹*ygkasanga@gmail.com,* ²*jackson.phiri@cs.unza.zm*

Abstract— Technology has revolutionized communication and how services can be delivered to the intended end-users. Today many areas such as supply chain, logistics, health, aviation, and education have been greatly influenced. Despite all the technological advancements, bus operators in Zambia still face challenges in computerizing and automating operational processes, to improve service delivery. The problem of misplaced and missing luggage is among the problems the operators are currently facing. This paper proposed and developed methods that can be used to track passenger luggage. The first objective was to identify the challenges faced by bus operators in managing passenger luggage based on Technology Adoption Models (TAM), the results show a link between the use of technology in managing passengers' language, ease of use and usefulness of the technology. To address the second objective, we developed the system models and business processes. This will assist in the development of a prototype to aid bus operators in managing passenger luggage will improve accountability and reduce lost customer luggage incidences will be based on an RFID reader, Arduino, GPS module and a web-based GUI.

Keywords—TAM, Arduino, Transport, GPS, Luggage Tracking

I. INTRODUCTION

Zambia, like other developing countries, in its quest to improve public service and efficient utilization of national resources, has also embarked on infrastructure development, most notably in the road infrastructure, we have had the link Zambia 8000[25], and many road expansions countrywide, this has attracted many bus operators in the provision of the service, to meet the growing demand for more bus schedules. Lusaka is the central hub for communication and trade [26], our capital city receives many travelers from across the country and the region, the city is connected to the four major roads leading North, South, East, and West. The most common mode of transportation is by road. InterCity Bus Terminus was constructed to meet the growing number of travelers, it is centrally located making it the biggest bus station in the country. Livingstone is the tourist capital, home of the Mighty Victoria Falls and many other tourist attractions, there is an influx of tourists during the peak season around April these are both local and international [27]. Many of these tourists opt to visit other parts of the country, hence the need to make use of buses for inter-town commuting. Each bus operator has its own parking slot, ticketing system, and luggage handling mechanism.

According to the RTSA annual report[28], the number of bus operators has increased to 300 this has increased the pressure on the existing infrastructure, it is also worth noting that

among all these operators only 5 are using a computerized system provided by Afrobus which is a bus ticketing solution. When further investigated the available solution was not been used for online bus reservation and did not have a module for luggage tagging and tracking. Passenger Luggage can be classified into a light and heavy luggage, light luggage been handheld luggage that can be handbags, bag packs, and any small packages that can be fitted inside the bus. Heavy luggage is therefore considered as any package that can be placed in the luggage compartment outside the buses. Most travelers carry with them sentimental items, and expensive merchandise on journeys, however, bags get misplaced, lost and worst-case stolen at bus stations. For this reason, it is very necessary to track down the bags in case of loss and theft. There has been an increase in the number of bus operators, which has led to many options for users when traveling [28] however there has been no computerized mechanism to manage, monitor and trace luggage information and mapping it to the rightful owner at bus stations. Another noticeable development is the advent of courier services been provided as a value-added service where customers can entrust their merchandise to a bus operator to deliver to the right owner [29].

The aim of the study is to design and develop passenger luggage system models for the public transport system to help operators in managing passenger luggage, improve accountability and reduce lost customer luggage incidences. The rest of the paper is as follows. In Section II, the literature review and related works are highlighted while in Section II, the methodology used is given. The results are in Section IV and finally, the discussion and conclusions are in Section V.

II. LITERATURE REVIEW

Sensor systems are utilized in ecological monitoring, for example, forest identification, animal tracking, flood location, estimating and climate expectation, and furthermore, in business applications like seismic exercises forecast and observing. [1]. In Military applications, sensors networks are used to detect identify and classify threats based on the count, number, type weather it is armored vehicles or soldiers on foot, and the number of weapons they carry, etc., can be detected in advance. This system provides reliable real-time war pictures and better situational awareness. The sensor nodes from sensor networks are dropped to the field of interest and are remotely controlled by a user. Enemy tracking, security location are likewise performed by utilizing these systems. [1] Health applications, for example, Tracking and checking of patient's response to medication specialists utilize these systems. [1] The most often used wireless sensing element networks applications is within the field of transport systems, for example, checking of traffic, dynamic steering

the board and observing of parking garages, and so on., utilize these networks.[1][2]

A. Characteristics of WSN

These gadgets work under the Transmit, Receive, Idle and Sleep modes having different levels of energy consumption. A detecting subsystem: Low power segments can help to fundamentally reduce power consumption. Since this subsystem (sensors and actuators) is answerable for the sharing of data between the sensor network and the outside world.

A power supply subsystem: It comprises of a battery which supplies power to the node [2]

Some of the unique characteristics of a WSN include

- Limited power they can harvest or store.
- Ability to withstand harsh environmental conditions
- Ability to cope with node failures
- Mobility of nodes
- Dynamic network topology
- Communication failures
- Heterogeneity of nodes Large scale of deployment
- Unattended operation.

RFID is a non-contact, programmed automatic identification technology that utilizes radio signal to distinguish, track, sort and identify an assortment of articles including individuals, vehicles, merchandise and resources without the requirement for direct contact (as found in attractive stripe innovation) or observable pathway contact (as found in scanner tag innovation). RFID innovation can follow the developments of items through a system of radio-empowered examining gadgets over the separation of a few meters. A gadget called an RFID tag is a key segment of the innovation. An RFID tag normally has at any time two components an integrated circuit and antenna for getting signals [5].



Figure.1. shows the components that constitute an RFID system [11].

RFID tags that don't have any incorporated circuit are called chipless RFID labels, RFID tags can be inserted in an article or can be put inside a package [8]. A regular RFID framework establishes four essential segments including RFID Tags, readers, antenna and a central node. A computer system that may house the database server and management software (middleware) [6] as shown in Figure 1.

RFID operates at different frequency bands which are categorized as low, high and ultra-high [9].

- Low frequency: provides a frequency band 30 kHz to 300 kHz (read range between 0 cm and 100 cm), reading is too slow but the data rate of piracy is moderate compared to other frequencies during playback.
- High frequency: frequency band 3 MHz to 30 MHz (read range between 10 cm and 1m), read speed is fast but the piracy rate is higher compared to the low frequency.
- Ultra-high frequency: frequency band 300 MHz to 3 GHz (the reading distance can exceed 15 meters), the playback speed is faster, but the piracy rate is still higher compared with the high frequency, this frequency is used nowadays in different logistics firms, particularly in the area of storage, inventory and stock management.

A powerful RFID tag can prevent RFID readers without special permission (access control or right frequency) to read the content. [10] There are 2 main categories of RFID tags, they can be either passive or active [12]: Passive tag: uses waves in order to transmit information through the energy transmitted by the reader, which supplies the onboard electronic circuits. Active Tag: usually embeds a source of internal energy (battery with up to 10 years autonomy), it sends the various information stored on the electronic circuit to the RFID server. According to R. Yadav and S. Nainan, a hybrid tag can be considered a comparison of features among active, semi-passive and passive tags is made.

TABLE 1. ACTIVE VS. PASSIVE VS. SEMI-PASSIVE TAGS [13]

Feature	Passive	Active	SemiPassive
Read Range	Short (Up to 10cm)	Long (Up to 100m)	Long (Up to 100m)
Battery	No	Yes	Yes
Life Validity	Up to 20 years	Between 5- 10years	Up to 10 years
Storage	128 bytes read/write	128 bytes read/write	128 bytes read/write
Cost	Cheap	Very Expensive	Expensive
Application	Attendance Management System	Monitor the condition of fresh produce	Measurement of temperature periodically

B. Related Works

According to [23] whose point was to propose a viable and practical administration platform to realize real-time tracking and tracing for prepackaged nourishment inventory network dependent on the Internet of Things (IoT) technologies, to guarantee a sheltered nourishment utilization condition. To decrease the production cost while realizing fine-grained tracking and tracing, an integrated solution of utilizing both the QR code and radio-Frequency ID (RFID) tag was proposed. [24] In the proposed work, the authors used an IoT device that retrieves the location Coordinates obtained from the GPS module connected to it and transfers them to Amazon Web Services for further processing and storage. This data is further processed.[7] The proposed IoT model that made use of GPS, RFID, PIR, Wireless radio communication module and GSM technologies. The GPS will be used to monitor and track the location of the vehicle in transit using mobile and web applications. RFID combined with GPRS and Arduino microcontroller was to be responsible for grain bags tallying.

C. Technology Adoption Model (TAM)

There are several models existing that have been utilized to explore Technology adoption. A few investigations concentrating on the adoption of mobile services have their foundations in Technology Acceptance Model (TAM) initially proposed by Davies in 1986. The model is initially intended to predict the user’s acceptance of Information Technology and usage in an organizational context. TAM focuses on the attitude explanations of intention to use a specific technology or service [3].figure 2 shows the TAM model proposed by [4].

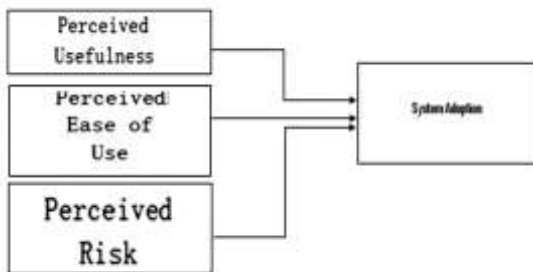


Figure. 2. TAM Model [4]

Perceived usefulness (PU)

Is defined as “the degree to which a person believes that using a particular system would enhance his or her performance”

Perceived Ease of Use (PEU)

Is defined as “the degree to which a person believes that using a particular system would be free of effort,” [20]

Perceived Risk (PR)

Perceived risk is defined as a consumer’s perceptions of the uncertainty and the possible undesirable consequences of buying a product or service [21]. Indeed, people may feel a certain degree of risk when using an Internet-based channel because perceived risk has the characteristics to increase vulnerabilities and generate inhibiting aspects to consumers’.

III. METHODOLOGY

Existing literature on tracking related technologies were studied and reviewed. All the data in the literature review was from published secondary sources like journals, websites, and books.

A. Baseline Study

The purpose of the baseline study was to identify challenges faced by bus operators and passengers in the adoption, of technology to streamline operations. To archive this, different materials were used and methods were followed;

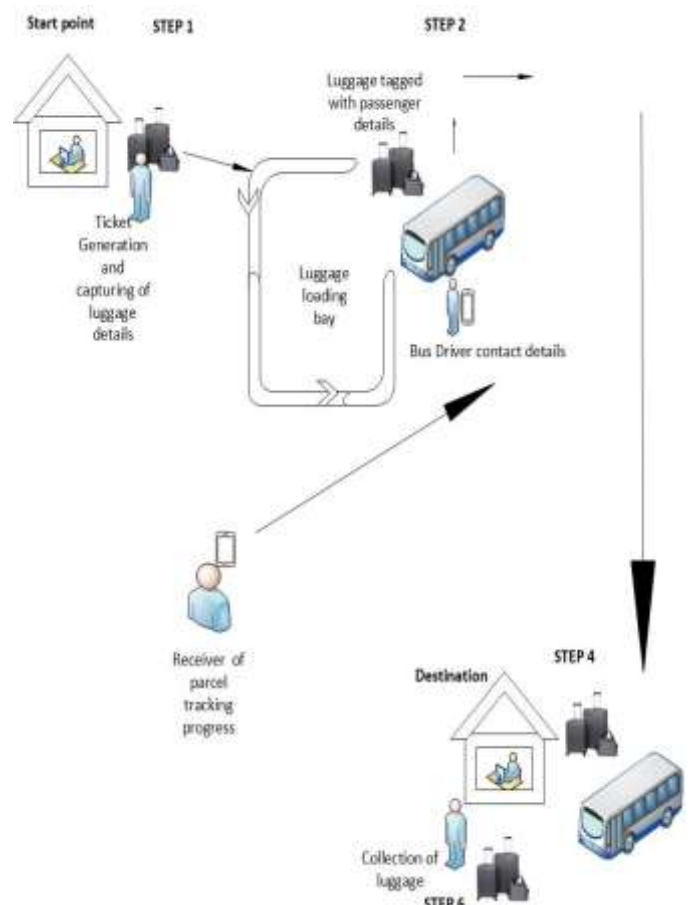
Data Collection: Primary data was collected from both bus operators and passengers based in Lusaka this was done by interviews & structured questionnaires.

Population & Sample: The population in this study included the bus operators, drivers, conductors and passengers on the Livingstone and Lusaka Bus Terminals. Purposive sampling method was used to select the population sample across Lusaka District. Further, a combination of random and convenient sampling method was used to select members of the public to participate in the survey.

Data Analysis & Presentation: Data entry, analysis, and presentation were done using the SPSS statistical software package.

B. Model Design Methodology

In order to formulate a new business model, we had to understand the current business process as shown in figure 3.



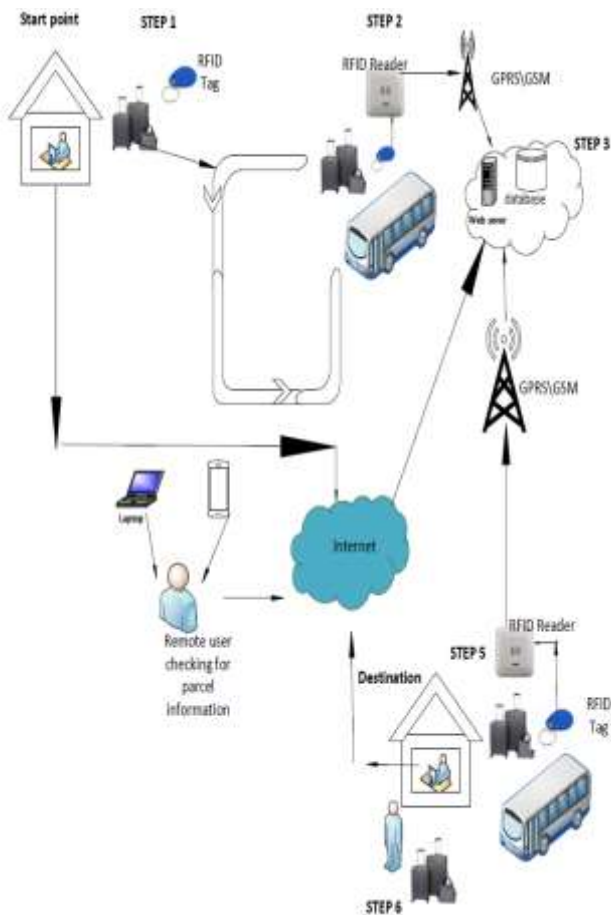


Figure.4. Proposed luggage handling business process

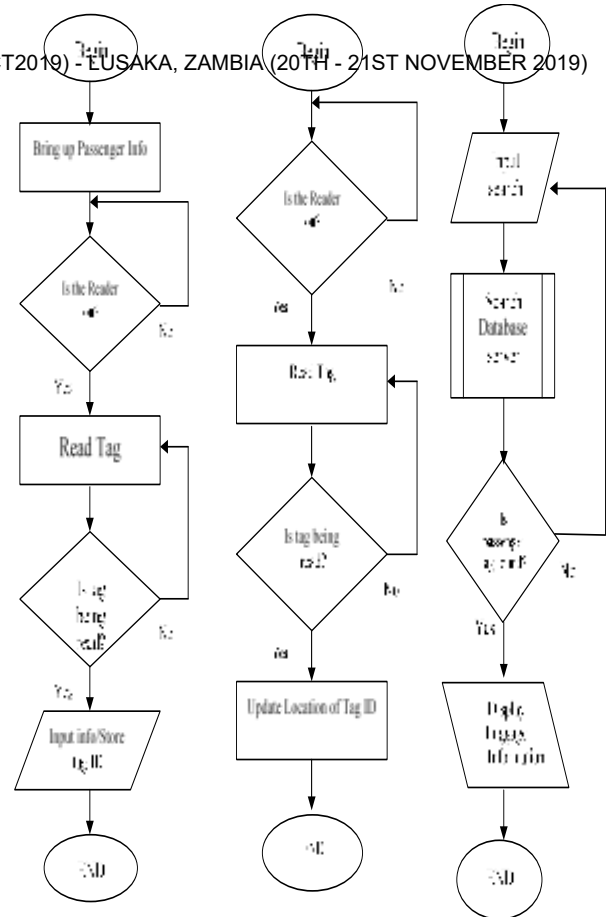


Figure.5. Flow chart of the proposed system Main processes

The current business process requires that a passenger makes a payment for their travel and luggage. A ticket is generated that has the passenger information, such as first, name last name, seat number and drop point, the generated ticket is used when the luggage has to be loaded onto the bus, other bus operators label the luggage with stickers that provide a verification mechanism of the information that is on the luggage owners ticket. Once the bus arrives at its destination passengers are required to produce the ticket in order to claim ownership of the luggage. The proposed luggage handling business process in figure 4 will be used to build a prototype passenger luggage tracking system that will make use of RFID tags, RFID readers, cloud servers and GPRS\GSM technology.as the passenger purchases the ticket their information is captured in the tag, this tag is scanned as luggage is loaded on the bus, capturing the location, user details and stored in a database this information can be retrieved at the destination.

IV. SYSTEM MODELING

This project has at least three processes as shown in figure 5. The first process is for reading the tag and sending information to the database. The second process is when the tag is located at the sender or receiver readers. The third process is a GUI allowing a passenger to search their name or tag ID.

V. DATA COLLECTION

Data was collected from a sample of respondents who are bus operators and passengers within Lusaka city. A descriptive research design was used in describing the current state of affairs, the research will employ a mixed-methods approach, which is a combination of both qualitative and quantitative methods. The study used Purposive sampling and Random sampling. Data Collection for the study was done through, Questionnaires, interviews

We distributed 315 questionnaires and a 63% response rate.

VI. RESULTS AND DISCUSSION

TABLE 2. ATTRIBUTES OF PARTICIPATING PASSENGERS

Demographic characteristics of respondents			
Group	Variable	Frequency	Percentage
Gender	Male	90	58.33
	Female	61	41.67
Age	20 and Below	7	4.64
	21-30	96	63.58
	31-40	37	24.50
	41-50	6	3.97
	51-60	3	1.99
	61 and Above	3	1.99

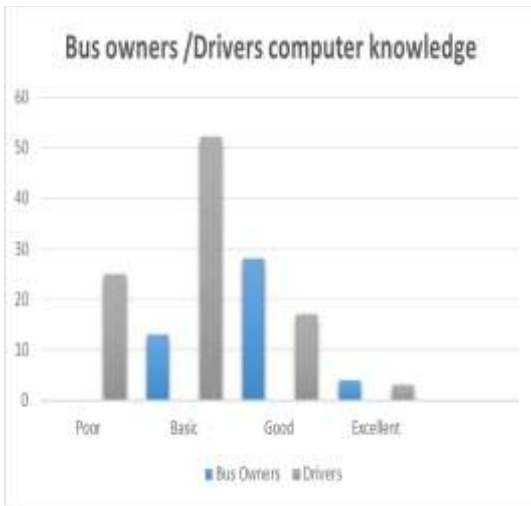


Figure.6. Bus Owners/Drivers Computer knowledge

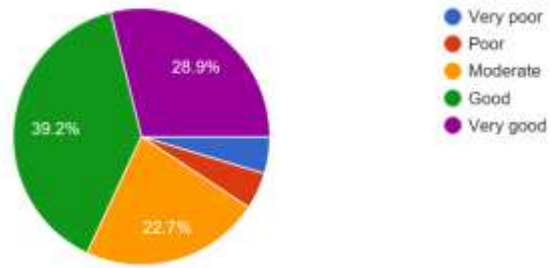


Figure.7. Passenger Computer Knowledge



Figure.8.Usage of bus transportation for inter-town commuting

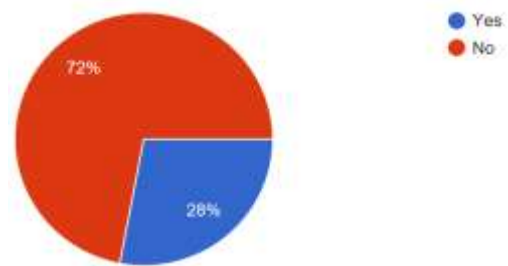


Figure.9. Number of passengers that have lost or misplaced Luggage on a journey.

TABLE 3 – CORRELATION ANALYSIS BETWEEN PE AND PEU

			PUF1	PUF2	PUF3	PUF4	PEUF1	PEUF2	PEUF3	PEUF4
Spearman's rho	PUF1	Correlation Coefficient	1.000	.306**	.425**	.479**	.386**	.277**	.252**	.300**
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000
		N	197	197	197	197	197	197	197	197
PUF2	PUF2	Correlation Coefficient	.306**	1.000	.284**	.297**	.231**	.130	.317**	.223**
		Sig. (2-tailed)	.000	.	.000	.000	.001	.068	.000	.002
		N	197	197	197	197	197	197	197	197
PUF3	PUF3	Correlation Coefficient	.425**	.284**	1.000	.419**	.348**	.249**	.266**	.311**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000
		N	197	197	197	197	197	197	197	197
PUF4	PUF4	Correlation Coefficient	.479**	.297**	.419**	1.000	.429**	.368**	.280**	.334**
		Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000	.000
		N	197	197	197	197	197	197	197	197
PEUF1	PEUF1	Correlation Coefficient	.386**	.231**	.348**	.429**	1.000	.509**	.480**	.547**
		Sig. (2-tailed)	.000	.001	.000	.000	.	.000	.000	.000
		N	197	197	197	197	197	197	197	197
PEUF2	PEUF2	Correlation Coefficient	.277**	.130	.249**	.368**	.509**	1.000	.449**	.455**
		Sig. (2-tailed)	.000	.068	.000	.000	.000	.	.000	.000
		N	197	197	197	197	197	197	197	197
PEUF3	PEUF3	Correlation Coefficient	.252**	.317**	.266**	.280**	.480**	.449**	1.000	.511**
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.	.000
		N	197	197	197	197	197	197	197	197
PEUF4	PEUF4	Correlation Coefficient	.300**	.223**	.311**	.334**	.547**	.455**	.511**	1.000
		Sig. (2-tailed)	.000	.002	.000	.000	.000	.000	.000	.
		N	197	197	197	197	197	197	197	197

** Correlation is significant at the 0.01 level (2-tailed).

The correlation coefficient between perceived ease of use factors of the tracking system and perceived usefulness of

the tracking system factors range from 0.133- 0.547 at (p<0.01). This suggests that there is a substantial positive relationship As depicted in Table.3 The correlation

coefficient between perceived ease of use factors of the tracking system and perceived usefulness of the tracking system factors range from 0.133- 0.547 at ($p < 0.01$). This suggests that there is a substantial positive relationship between respondents' perceived ease of use and perceived usefulness of the tracking system.

VII. CONCLUSION AND FUTURE WORKS

in this paper, we have addressed the two project objectives, of which a baseline study was conducted to understand the current business processes used by bus operators, the study further took time to understand the challenges as to why technology has not been incorporated in the daily operations and understanding the factors affecting adoption of new technology. we have proposed a model based on RFID, tags, GPS and GSM this IoT model will have modules for luggage tracking, fetching customer information and location capturing.

As part of future works and proof of concept, we will develop a prototype for all the modules.

ACKNOWLEDGMENT

We want to thank all participants, bus owners, drivers and members of staff from the Lusaka city council for the support and also for providing us the data we needed for this project. Special thanks to my supervisor Dr.J.Phiri for his guidance and assistance, this also extends to the UNZA Department of Electrical and Electronics Engineering and Computer Science for the Support rendered.

REFERENCES

[1] Stewart, J., Tittel, E. and Chapple, M. (2008). *CISSP*. Hoboken: John Wiley & Sons, Inc.

[2] M. U.Farooq, M. Waseem, A. Khairi and S. Mazhar, "A Critical Analysis on the Security Concerns of Internet of Things (IoT)", *International Journal of Computer Applications*, vol. 111, no. 7, pp. 1-6, 2015. Available: 10.5120/19547-1280.

[3] NIST.(2019).Cybersecurity Framework. [Online] Available at: <https://www.nist.gov/cyberframework> [Accessed 9 Mar. 2019].

[4]"RFID BASED ATTENDANCE SYSTEM", *International Journal of Recent Trends in Engineering and Research*, vol. 3, no. 4, pp. 94-97, 2017. Available: 10.23883/ijrter.2017.3114.wjkh4.

[5]C. Sun, "Application of RFID Technology for Logistics on Internet of Things", *AASRI Procedia*, vol. 1, pp. 106-111, 2012. Available: 10.1016/j.aasri.2012.06.019.

[6] C.L. Muyunda and J. Phiri, "A Wireless Sensor Network Based Grain Inventory Management System for Zambia's Food Reserve Agency, Vol. 5, no.3, pp.1-6,2016 Available: DOI:10.15680/IJRSET.2016.0503144

[7]S. Chihana, J. Phiri and D. Kunda, "An IoT based Warehouse Intrusion Detection (E-Perimeter) and Grain Tracking Model for Food Reserve Agency", *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 9, 2018. Available: 10.14569/ijacsa.2018.090929.

[8]L. Ruiz-Garcia and L. Lunadei, "The role of RFID in agriculture: Applications, limitations and challenges," *Computers and Electronics in Agriculture*, vol. 75, no. 1, pp. 42-50, January 2011

[9] G. Swagarya, L. Boaz, and M. Kisangiri, "International Journal of Scientific Engineering and Research (IJSER) Designing of UHF-Radio Frequency Identification (RFID) Antenna," 2013

[10] M. Abid, b. Nsiri, y. Serhane and h. Aghari, "Tracking Items Through Rfid and Solving Heterogeneity Problems During a Collaboration Between Port Companies", *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 2, 2016. Available: 10.14569/ijacsa.2016.070230.

[11] W. SYSTEM, "WEIGHBRIDGE WITH RFID SYSTEM", *Virtualwarehousestks.blogspot.com*, 2019. [Online]. Available:

<http://virtualwarehousestks.blogspot.com/p/weighbridge-with-rfid-system.html>. [Accessed: 02- May- 2019].

[12] H. M.Quraishi, F.Farheen "Curtain Raiser on Gen-Next of RFID Technology", *IOSR Journal of Electrical and Electronics Engineering*, 2014.

[13] R. Yadav and S. Nainan, "Design of RFID Based Student Attendance System with Notification to Parents Using GSM," *International Journal of Engineering Research & Technology (IJERT)*, vol. 3, no. 2, February 2014.

[14] U. Mamdapur and I. Rajgoli, "Implementing radio frequency identification technology in libraries: Advantages and disadvantages," *International Journal of Library and Information Science*, vol. 3, no. 3, pp. 46-57, March 2011.

[15] K. N. Singh and P. Mahajan, "Application of RFID Technology in Libraries," *International Journal of Library and Information Studies*, vol. 4, no. 2, pp. 1-9, April-June 2014.

[16] O. T. Arulogun, A. Olatunbosun, O. A. Fakolujo and O. M. Olaniyi, "RFID-Based Students Attendance Management System," *International Journal of Scientific & Engineering Research*, vol. 4, no. 2, p. 1, February 2013.

[17] M. Bertrand and S. Bouchard, "The Technology Acceptance Model: A Potentially Useful Tool to Understand Why Therapists Intend to Use or Not Virtual Reality," *PsycEXTRA Dataset*, 2006.

[18] L.Isaiah, O.Tonny and M.Tim,"Application of Technology Acceptance Model (TAM) in M-Banking Adoption in Kenya," *International Journal of Computing and ICT Research*, Vol. 6, Issue 1, June 2012.

[19]R. Boateng, "Mobile phones and micro-trading activities – conceptualizing the link", *info*, vol. 13, no. 5, pp. 48-62, 2011. Available: 10.1108/14636691111160635.

[20] N.Sana , H.Nael and A.Musbah. "A secure mobile payment system using QR code." 111-114. 10.1109/CSIT.2013.6588767. 2013.

[21] QR Code: A New Opportunity for Effective Mobile Marketing

[22]P. Soneka and J. Phiri, "A Model for Improving E-Tax Systems Adoption in Rural Zambia Based on the TAM Model", *Open Journal of Business and Management*, vol. 07, no. 02, pp. 908-918, 2019. Available: 10.4236/ojbm.2019.72062.

[23] H. Kuruva and G. S. P. R. S., "IoT-based anti-theft and remote controlling system for automobiles," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 5, no. 12, pp. 3385-3389, 2016.

[24] H. Alshamisi and V. Képuska, "Real Time GPS Vehicle Tracking System," *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, vol. 6, no. 3, pp. 179-182, march 2017.

[25]"Seventh National Development Plan 2017-2021 | Embassy of the Republic of Zambia in Washington, D.C.", *Zambiaembassy.org*, 2019. [Online]. Available: <http://www.zambiaembassy.org/document/seventh-national-development-plan-2017-2021>. [Accessed: 10- Aug- 2019].

[26]D. Katongo, *Zamreal.com*, 2019. [Online]. Available: <http://www.zamreal.com/2017/wp-content/uploads/2017/10/Daniel-Katongo.pdf>. [Accessed: 12- Aug- 2019].

[27]*Pdfs.semanticscholar.org*, 2019. [Online]. Available: <https://pdfs.semanticscholar.org/fea0/6f3ed9f8dfdf199c1162f4b536594078e471.pdf>. [Accessed: 12- Oct- 2019].

[28]*Rtsa.org.zm*, 2019. [Online]. Available: <https://www.rtsa.org.zm/wp-content/uploads/2019/09/RTSA-ANNUAL-REPORT-2018.pdf>. [Accessed: 12- Nov- 2019].

[29]*Zicta.zm*, 2019. [Online]. Available: <https://www.zicta.zm/Downloads/publications/List%20of%20License%20Postal%20and%20Courier%20Operators-%20August%202019.pdf>. [Accessed: 12- Nov- 2019]

[30]A.Kavianpour,*Peer.asee.org*,2019.[Online]. Available:<https://peer.asee.org/automated-luggage-tracking-system.pdf>. [Accessed: 20- Nov- 2019]

Demystifying Cryptocurrency Mining Attacks: A Semi-supervised Learning Approach Based on Digital Forensics and Dynamic Network Characteristics

Aaron Zimba¹, Mumbi Chishimba², Christabel Ngongola-Reinke³, Tozгани Fainess Mbale⁴

¹Department of Computer Science & IT, Mulungushi University, Kabwe, Zambia, ²Department of Information Technology National Institute of Public Administration (NIPA), Lusaka, Zambia, ³Department of Economics, Mulungushi University, Kabwe, Zambia, ⁴Department of Electrical & Electronic Engineering, University of Zambia, Lusaka, Zambia

¹gvsfj@gmail.com, ²chishimba.mumbi@gmail.com, ³christabel.ngongola@gmail.com, ⁴tozganimbale@gmail.com

Abstract—Cryptocurrencies have emerged as a new form of digital money that has not escaped the eyes of cyber-attackers. Traditionally, they have been maliciously used as a medium of exchange for proceeds of crime in the cyber dark-market by cyber-criminals. However, cyber-criminals have devised an exploitative technique of directly acquiring cryptocurrencies from benign users' CPUs without their knowledge through a process called crypto mining. The presence of crypto mining activities in a network is often an indicator of compromise of illegal usage of network resources for crypto mining purposes. Crypto mining has had a financial toll on victims such as corporate networks and individual home users. This paper addresses the detection of crypto mining attacks in a generic network environment using dynamic network characteristics. It tackles an in-depth overview of crypto mining operational details and proposes a semi-supervised machine learning approach to detection using various crypto mining features derived from complex network characteristics. The results demonstrate that the integration of semi-supervised learning with complex network theory modeling is effective at detecting crypto mining activities in a network environment. Such an approach is helpful during security mitigation by network security administrators and law enforcement agencies.

Keywords—*bitcoin, cryptocurrency, cyber-attack, crypto mining, semi-supervised learning, complex networks*

I. INTRODUCTION

The general aim of conventional cyberattacks has generally been to obtain monetary proceeds of the associated cybercrime. Attackers have had the challenge of acquiring these monetary proceeds with little or no monetary trail since conventional payments leave a trail of traceable financial activities [1]. Cryptocurrencies have alleviated this challenge as they provide for privacy and anonymity [2]. The strong privacy provided in cryptocurrencies makes it almost impossible to trace financial payments [3]. As such, cryptocurrencies have become a de facto method of payments in most finance-related cyber-attacks [4], a trend not uncommon in crypto-ransomware attacks.

However, since victims of cybercrime have had the ability to make payments in cryptocurrencies such as Bitcoin (implying users store cryptocurrencies on computers), attackers have now moved on to attack the very user cryptocurrencies from digital wallets as was evidenced in various attacks [5]. Furthermore, it is not uncommon to find financial malware that seeks to steal cryptocurrencies from targeted users as an extra functionality

[6]. Since not all targeted users harbor cryptocurrencies, attackers have devised a technique of directly generating cryptocurrencies from the victims' CPU (crypto mining¹) by enlisting them to a mining pool. The cryptocurrencies are generated by installable malware or via browser-based crypto mining. Victims are enlisted in a crypto mining pool since solo mining is not efficient [7]. As such, corporate or enterprise networks are attractive to crypto mining attackers because they provide a pool of devices for crypto mining. It is thus not uncommon to find illegal crypto mining cloud computing and IoT environments [8] as well as critical infrastructure systems such as SCADA [9]. Illegal crypto-mining has since been on the rise and costed victims millions of dollars [10]. Consequently, the year 2018 saw the growth of crypto mining malware by 4,000% [11]. As such, crypto mining attacks have proven to be a force to reckon with which can longer be avoided even as attackers have been eschewing the infamous ransomware attacks [12]. The diagram in Figure 1 shows the decline in ransomware attacks versus the rise in crypto mining attacks according to the IBM-X-Force Threat Intelligence Index 2019 [13].

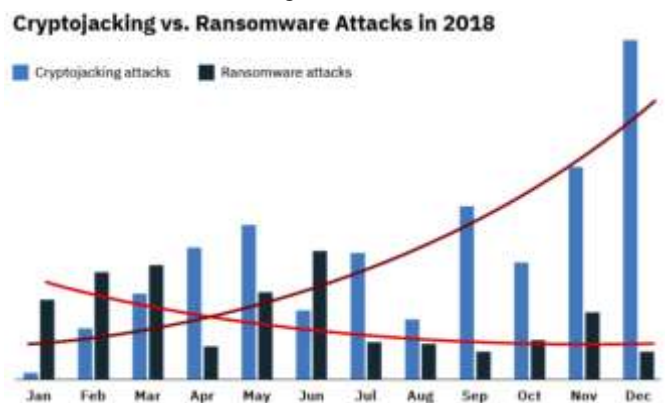


Fig. 1. Crypto mining vs Ransomware attacks in 2018 [13]

Crypto mining is taking over ransomware owing to its ease of administration; easily proliferated by phishing emails, no user input required and the difficulty associated with tracing the perpetrators. These advantages have seen an increase in the stealthier attacks, i.e. crypto mining, by 450% even as cybercriminals pivot from the common ransomware attacks

¹ Crypto mining is a process of using the resources of a computer system to mine cryptocurrencies.

[14]. Crypto mining attacks present a million-dollar industry [15].

In Africa, crypto mining is particularly prevalent in Ethiopia, Tanzania, and Zambia which account for 3 of the top-5 countries largely impacted by crypto-mining attacks, according to a Microsoft report [16]. The most impacted victims are SMEs as the security therein is not as robust as in larger corporations.

Like all malware activities and cyberattacks, crypto mining activities generate noise in the form of network traffic. However, the types of network characteristics associated with these types of attacks are peculiar to crypto mining in that victims enlisted to a mining pool or botnet needs to communicate with the associated C2 servers and mining servers. As such, the detection of crypto mining activities in a network environment calls for an approach that takes into consideration these network characteristics. This paper addresses the detection of crypto mining attacks in a generic network environment using dynamic network characteristics. It tackles an in-depth overview of crypto mining operational details and proposes a machine learning approach to detection using various crypto mining features derived from the network characteristics. The Small-World network models [17] of complex network evolution theory are adopted for attack modeling and we use a semi-supervised approach to machine learning for detection.

The rest of the paper is organized as follows; Section II presents the related works while the methodology and proposed detection framework are brought forth in Section III. The results and the analyses thereof real-world in Section IV and the conclusion is drawn in Section V.

II. RELATED WORKS

Even though crypto-mining attacks are a fairly new phenomenon, they have attracted significant attention in the security landscape. Some research works have concentrated on crypto mining in general computer systems [18] whilst others have narrowed the scope to critical infrastructure and IoT [19]. Authors in [20] propose an end-to-end analysis of browser-based crypto mining by statically and dynamically examining the rise of crypto mining in the real world cases. The proposed approach inspects the traversing traffic between web-sockets without blacklisting of IP addresses. They achieve a detection accuracy of 96.4% using code analysis.

Authors in [21] propose a host-based approach to crypto mining called BitcoinTrap. Modeling via dynamic analysis of executable binary crypto mining files to detect Bitcoin-mining botnets is adopted. The advantage of this approach is that it can detect Bitcoin mining botnets at the lowest level of execution. The Bitcoin block header is centrally used as the pivotal piece of information in this detection methodology. The drawback of this approach is that it specifically applies only to the detection of Bitcoin miners, whereas the crypto mining landscape has seen the emerging of competing and easy-to-mine cryptocurrencies such a Monero and Ethereum.

In [22], the authors examine recent trends towards in-browser mining of cryptocurrencies. They concentrate their efforts on the mining of Monero cryptocurrency via CoinHive and those of similar code- bases. In their model, a web user visits a vulnerable site infected with JavaScript code that executes on the client-side browser, thus mining a cryptocurrency without the user's consent. They further survey the crypto mining landscape in order to conduct measurements to establish the prevalence and profitability thereof. They outline the ethical framework for classifying the attack as an inherent attack or business opportunity. They delineate the various stages

involved in the process crypto mining process and thereafter brief the various terms associated with crypto mining. However, their approach does not address the systematic detection of crypto mining.

In [23], the authors approach crypto mining detection using dynamic opcode analysis on non-executable files. They use a specified dataset to achieve high detection rates of browser-based crypto mining using Random forest (RF) as the preferred classification algorithm. Their model distinguishes between crypto mining websites, weaponized benign crypto mining websites, de-weaponized crypto mining websites, and real-world benign crypto mining websites. As such, their technique offers an opportunity not only to detect but to prevent as well as mitigate crypto-mining attacks.

Authors in [24] present an in-depth analysis of the crypto mining operation. They designed and implemented a passive-active flow monitoring and catalog to detect crypto-mining activities from compromised devices in a network. They tested the feasibility of their approaches to real-life data where passive-active detection is capable of discovering emerging or deliberately hidden crypto mining pools.

III. METHODOLOGY AND PROPOSED FRAMEWORK

The enlisting of vulnerable and exploitable devices to a crypto mining pool is a dynamic process that can be viewed as an evolution of node-addition or deletion in an attack graph. Since the nodes in the mining pool interact one with the other and with the central server, the system can thus be characterized by vertex degrees and clustering coefficients. It is on this premise that we employ the use of Small-World network models of complex network theory to depict the behaviour of the attack network and deduce the corresponding features for purposes of detection. The diagram in Figure 2 shows a time-slice crypto-mining depicting the initialization and growth of a crypto mining pool in a target network.

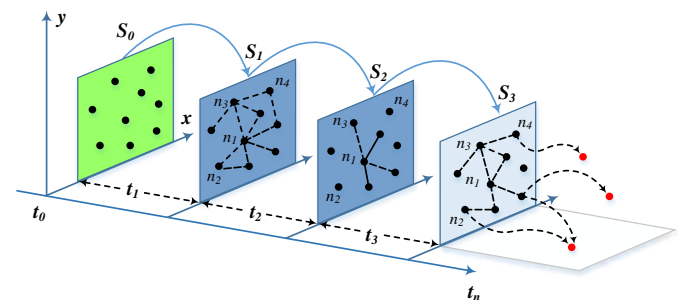


Fig. 2. Dynamic growth of crypto mining pool at various stages of the crypto mining process

The first phase S_0 represents the state of the targeted network system before any device is listed to a crypto mining pool. At this phase, the devices in the network are only susceptible to crypto mining but not yet compromised. As such, the vertex degree and clustering are equal to zero. In phase S_1 , vulnerable nodes are identified and consequently added to the crypto mining pool in phase S_2 . It is worth noting that at this stage, the vertex degree and clustering coefficients are now greater than zero.

$$S_n(t) \rightarrow S_{n+1}(t+1) := \begin{cases} (N_t \cup n_{t+1}^+) - (n_{t+1}^-) \\ (E_t \cup e_{t+1}^+) - (e_{t+1}^-) \end{cases} \quad (1)$$

Phase S_3 represents active crypto mining where the enlisted victims in the crypto mining pool are coordinating and working

together towards the associated proof-of-work². Equation (1) depicts the dynamic transitions of a victim device enlisted to a mining pool at a point in time as echoed in Figure 2.

As the state of the enlisted victim devices transitions from state S_0 to S_n , a series of network traffic is generated which we use to derived features for the detection process.

Members enlisted in a crypto mining pool used dedicated protocols to coordinate the distributed mining process. The 3 common TCP-based crypto mining protocols are GetWork, GetBlockTemplate, and Stratum protocol [25]. Other traffic details found in crypto mining pools include registration and authentication traffic, recurrent assignment of work packages provided by the crypto mining server. It is from these dynamic traffic details that we draw features to devise a detection methodology. In light of this, we present a semi-supervised learning approach to crypto mining detection that takes advantage of the huge amount of unclassified dataset [26] to perform classification of suspicious hosts participating in crypto mining activities and using few labeled instances from the labeled data. The proposed detection framework is shown in Figure 3.

This suits its applicability in complex networks since the clustering coefficient and vertex degrees are dictated by neighbor relations. As such, we adopt the SNN algorithm which apart from considering direct associations between nodes also considers indirect connections. This provides for an ability to detect similarities between nodes that are not necessarily adjacent. Additionally, SNN has the ability to handle clusters of varying sizes, densities and shapes. As such, two nodes that are relatively close but belong to different clusters are handled effectively.

Algorithm 1 illustrates the enhanced SNN algorithm. As shown from Figure 3, our semi-supervised approach consists of two phases: 1) an unsupervised phase that produces complex network characteristics features based on vertex degrees and clustering coefficients. 2) a supervised phase that learns and trains the model. This phase uses the KNN classifier and the labeled data. In short, our semi-supervised learning approach uses the unsupervised learning method to extract features from the unlabeled dataset and the supervised model classifies this data instances of crypto mining using complex network characteristics features. The unsupervised phase utilizes the shared nearest neighbor clustering whilst the supervised phase utilizes the KNN.

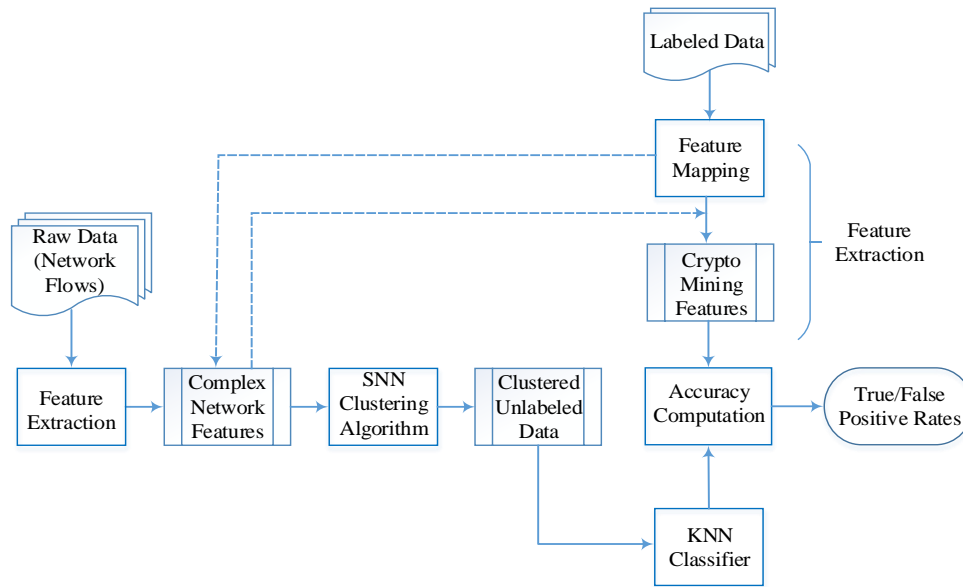


Fig. 3. The semi-supervised approach to crypto mining detection

Our semi-supervised approach shown in Figure 3 shows that in addition to unlabeled raw data (network flow traffic), we have a set of labeled data with features depicted in Table 1. Our semi-supervised approach uses complex network characteristics features of unlabeled data (clustering coefficients and vertex degrees) to create a supervised model. The feature extraction step in the supervised section of our approach uses a mapping scheme to extract hosts from the unlabeled dataset.

We propose a semi-supervised learning approach where we first derive different clusters mainly based on the clustering coefficient and vertex degree. To analyze the normalized data and detect crypto mining, we employ an enhanced semi-supervised algorithm based on the Shared Nearest Neighbour (SNN) clustering algorithm [27]. The SNN clustering defines similarity or proximity between two nodes in terms of the number of directly connected neighbors they have in common.

² Proof-of-work refers to the cryptographic computational puzzle that miners have to solve in order to be issued a crypto currency unit.

Algorithm 1: Enhanced SNN for detecting crypto mining

Input: G - undirected graph, k number of shared nearest neighbors
Output: L^* - list of suspicious hosts participating in crypto mining.

```

1 Initialize  $G^*$  with  $|V(G)|$  vertices, no edges
2 foreach  $i = 1$  to  $V(G)$  do
3   foreach  $j = i + 1$  to  $V(G)$  do
4      $counter = 0$ 
5     foreach  $m = 1$  to  $V(G)$  do
6       if vertex  $i$  and vertex  $j$  both have an edge with vertex
7          $m$  then
8          $counter = counter + 1$ 
9       end
10      if  $counter \geq k$  then
11        Connect an edge between vertex  $i$  and vertex  $j$  in  $G^*$ 
12        for  $S \leftarrow 0$ 
13          if  $\Delta K_i > 1$  at time  $t^*$  for external communications
14            then  $(H_{int-src}) \in S_1$ 
15          else if  $\Delta K_i'' > \Delta K_i \ \&\& \ \Delta C_i > 1 \ \&\& \ \Delta C_i > [\Delta C_{i-1, \dots, 0}]$ 
16            then  $(H_{int-src}) \in S_2$ 
17          else if  $\Delta K_i'' > 1 \ \&\& \ M_v > (X_{threshold}^*)$  for time window  $\Delta t$ 
18            then  $(H_{int-src}) \in S_3$ 
19          end if
20        end
21      end if
22    end
23  end
24  end
25  Return  $L^*$ 

```

The semi-supervised learning approach is summarized in Algorithm 2.

Algorithm 2: Semi-supervised learning for crypto mining detection

Input: $X_{u-d}^i = \{x_{u-d}^1, x_{u-d}^2, x_{u-d}^3, \dots, x_{u-d}^i\}$, unlabeled network flows
 where $x_{u-d}^i \in R^n, i = 1, 2, 3, \dots, n$
 $X_{l-d}^i = \{x_{u-d}^1, x_{u-d}^2, x_{u-d}^3, \dots, x_{u-d}^i\}$, labeled data
 where $x_{u-d}^i \in R^n, i = 1, 2, 3, \dots, n$

Output: TP && FP rates – cryptocurrency mining detection accuracy

1. Read the unlabeled & labeled network flow dataset
2. Normalize original data X_{l-d}^i, X_{u-d}^i , to get the data $\overline{X_{l-d}^i}, \overline{X_{u-d}^i}$
3. Extract k and c , complex network features from X_{u-d}^i
4. From X_{l-d}^i , map corresponding k and c in X_{u-d}^i , generate features
5. Cluster $X_{u-d}^i \rightarrow SNN \Rightarrow C_0, C_1, C_2, \dots, i$
6. Generate cluster states $C_0, C_1, C_2, \dots, i \rightarrow FSM(S_n) \Rightarrow C_i^{S_n}$
7. Train supervised KNN with $\overline{X_{l-d}^i}$ and $\overline{X_{u-d}^i}$
8. Classify $C_i^{S_n}$ with the KNN
9. Compute detection accuracy TP & FP rates based on SNN & KNN
10. Return TP && FP rates for $C_i^{S_n}$ clusters

The unlabeled and labeled data $\{X_{u-d}^i, X_{l-d}^i\}$ from network flows and the labeled data respectively are initialized and read in step 1. In step 2, we normalize the data by converting the input values to a common scale $[\overline{X_{u-d}^i}, \overline{X_{l-d}^i}]$. This enables us to make an effective comparison of the variations of the clustering coefficient and vertex degree. The complex network characteristics feature k and c are extracted from the unlabeled data $\{X_{u-d}^i\}$ in step 3. The labeled data is used in a mapping scheme in step 4 to locate a host and generate crypto mining features from the dataset. The SNN unsupervised clustering algorithm is used to create clusters in step 5 via the process $\{X_{u-d}^i\} \rightarrow SNN \Rightarrow C_0, C_1, C_2, \dots, i$. Step 6 assigns states to the clusters C_0, C_1, C_2, \dots, i generated in step 5. The KNN supervised algorithm is applied to the datasets $\{\overline{X_{l-d}^i}\}$ and $\{\overline{X_{u-d}^i}\}$ in step 7. The clustered data with states $\{C_i^{S_n}\}$ is classified by the KNN algorithm in step 8. The True Positive and False Positive rates (TP && FP) are computed in step 9 and the corresponding TP && FP rates for the clustered states are returned in step 10.

IV. RESULTS ANALYSIS AND DISCUSSIONS

To apply the aforementioned framework and algorithms, we first start by analyzing the unlabeled traffic content with a protocol analyzer for crypto mining and non-crypto mining TCP and UDP traffic. The results are shown in Figure 4 below.

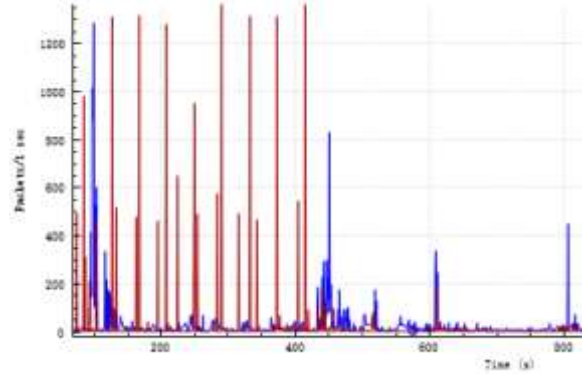


Figure 4. TCP (blue) and UDP (red) traffic from the dataset

We analyze traffic with crypto mining activities for the cryptocurrencies Ethereum, Monero, and Zcash for their corresponding mining pools. The diagram in Figure 5 shows traffic for the Ethereum mining pool captured via Wireshark.

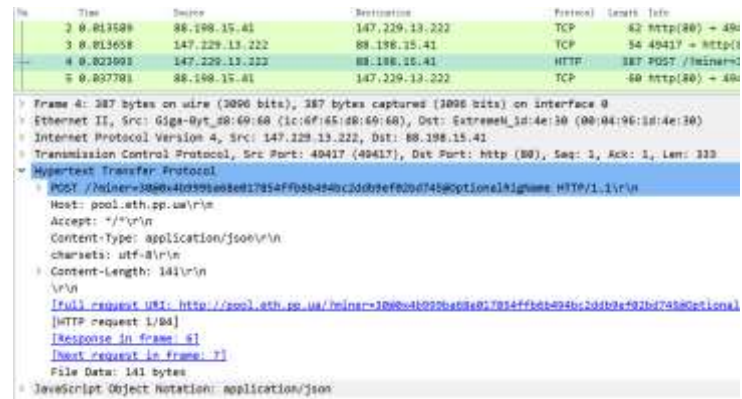


Fig. 5. Traffic for the Ethereum mining pool

It was noted that the mining protocols leverage TCP as the transport layer protocol. In comparison to official crypto p2p clients, the mining protocols do not necessarily use "well-known" port numbers. This is all dependent on the configuration of the administrator. As such, it is not uncommon to encounter port numbers for http=80, https/TLS=443, and SMTP=25 being used to bypass firewalls between the mining victim and the corresponding mining pool server.

The feature vector specified in this dataset [26] which we later use for the supervised learning stage is outlined in Table I.

TABLE I. DATASET FEATURE VECTOR

SN	Feature	Description
1	<i>bpp</i>	Bytes per packet per flow per all flows
2	<i>ppm</i>	Packets per minute
3	<i>ppf</i>	Packets per flow per all flows
4	<i>Ackpush_all</i>	Number of flows with ACK+PUSH flags to all flows
5	<i>Req_all</i>	Request flows to all flows
6	<i>Syn_all</i>	Number of flows with SYN flag to all flows
7	<i>Rst_all</i>	Number of flows with RST flag to all flows
8	<i>Fin_all</i>	Number of flows with FIN flag to all flows
9	<i>class</i>	class - miner or not-miner

We apply different clustering algorithms in Weka [28] and later compare them with the results of SNN clustering. The table below Table II shows the results of different clustering algorithms.

TABLE II. COMPARISON RESULTS OF UNSUPERVISED LEARNING

SN	Clustering Model	Clustered Instances	Distribution (%)
1	Simple-K-Means	2 [C0, C1]	55% : 45%
2	Canopy	3 [C0, C1, C2]	55% : 4% : 41%
3	MakeDensityBased Clusterer	2 [C0, C1]	55% : 45%
4	HierarchicalClusterer	2 [C0, C1]	0% : 100%
5	FilteredClusterer	2 [C0, C1]	55% : 45%
6	FarthestFirst	2 [C0, C1]	100% : 0%
7	SNN	5 [C0, C1, C2, C3, C4]	22%:19%:18%:15%:26%

Application of the SNN clustering algorithm produces clusters of different properties. Table II shows 5 clusters with IDs C0, C1, C2, C3, and C4. As can be seen from Table II, the SNN algorithm performs better clustering with not only the highest numbers of clusters but even a better distribution.

Cluster C0 has a high *bpp* (97.3%) and a high *ppm* (79.2%). It also has a high *ppf* (65.8%) compared to *Ackpush_all* (47.1%). This implies that hosts in this cluster have a higher vertex degree and clustering coefficient with regards to external communications.

On the contrary, cluster C2 has *bpp* (98.5%) and *ppm* (83.6%) but the *Ackpush_all* (90.6%) is greater than *ppf* (75.3%). This implies that hosts in this cluster have a higher clustering coefficient and vertex degree with regards to internal communications. The number of flows with FIN flags to all flows for activities in this cluster is relatively higher than C0.

A lower *bpp* (1.6%) and a high *ppm* (99.6%) corresponding to *ppf* (33.1%) instead of *Ackpush_all* (1.8%) for a smaller time window in cluster C4 entail that hosts in this cluster communicate more with external hosts. Furthermore, hosts in this cluster have a high *Syn_all* (97.8%) value implying a high number of synchronization connections requests to the mining pool.

TABLE III. CLUSTERING RESULTS

Attribute	Cluster-ID				
	C0	C1	C2	C3	C4
<i>bpp</i>	0.973	0.763	0.985	0.861	0.016
<i>ppm</i>	0.792	0.762	0.836	0.582	0.996
<i>ppf</i>	0.658	0.371	0.753	0.864	0.331
<i>Ackpush_all</i>	0.471	0.937	0.906	0.743	0.018
<i>Req_all</i>	0.984	0.735	0.969	0.791	0.092
<i>Syn_all</i>	0.548	0.524	0.81	0.577	0.978
<i>Rst_all</i>	0.471	0.832	0.988	0.72	0.511
<i>Fin_all</i>	0.35	0.526	0.871	0.936	0.302

The clusters C1 and C3 have relatively average network statistics that depict the behaviour of benign hosts. The high *Ackpush_all* (93.7%) in C1 corresponds to a high *Rst_all* (83.2%) which is a correlation expected of normal network traffic. Equally in cluster C3, *bpp* (86.1%) corresponding to *ppf* (86.4%) which is supplemented by average values of other characteristics in the same range. The variations in the clustering coefficient and vertex degree in these network traffic statistics in the respective clusters depict the overall movement of the movement vector from the feature centroid.

After generating the clusters and associating them with crypto mining instances, we use the labeled dataset for classification and evaluate the effectiveness of our proposed

approach. This is because the hosts in the labeled data are technically labeled as malicious for generating crypto mining traffic. However, we do not evaluate which stage of the crypto mining process the traffic belongs to. The detailed characteristics of the model for the hosts classified using the clusters, which are the results of the classification are shown in Table IV.

TABLE IV. MODEL CHARACTERISTICS CRYPTO MINING DETECTION

Class	TP Rate	FP Rate	Precision	Recall	F Measure	MCC	ROC Area	PRC Area
Not Miner	0.998	0.462	1	0.998	0.999	0.276	0.974	1
Miner	0.538	0.002	0.143	0.538	0.226	0.276	0.974	0.504
Avg.	0.997	0.461	0.999	0.997	0.998	0.276	0.974	1

Correctly classified instances represent 99.72% while incorrectly classified instances represent 0.28%. The diagram in Figure 6 shows the confusion matrix of the correctly and wrongly classified instances.

		Was actually	
		Not Miner	Miner
Classified as	Not Miner	355692	882
	Miner	26	147

Figure 6. Confusion matrix for the model

The model has good performance because the weighted average of the ROC Area is near 1 and way above the non-discriminative characteristic (N.D) which represents equal TP and FP rates. The ROC curves for detection of not miner instances and miner instances are shown in Figure 6 and Figure 7 respectively.

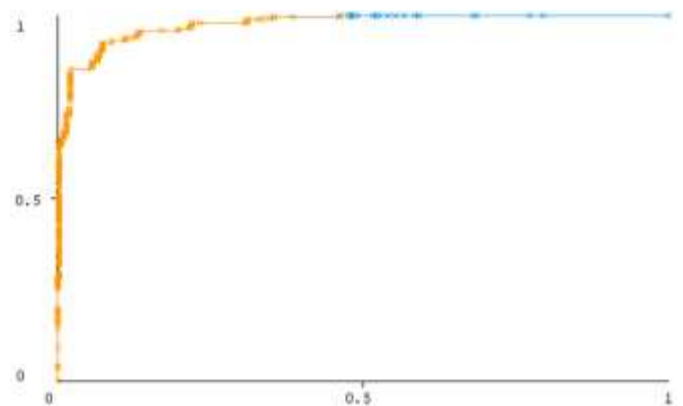


Figure 7. ROC curve for the "Not Miner" class

The ROC Area entails the predictive characteristics of the model to distinguish between the true positives and the true negatives. As such, the model does not only predict a positive value as a positive but as well as a negative value as a negative. The TP Rate represents the instances that are correctly classified as a given class which essentially is the rate of true positives. The FP Rate represents which of the instances falsely classified as a given class which essentially is the rate of false positives.

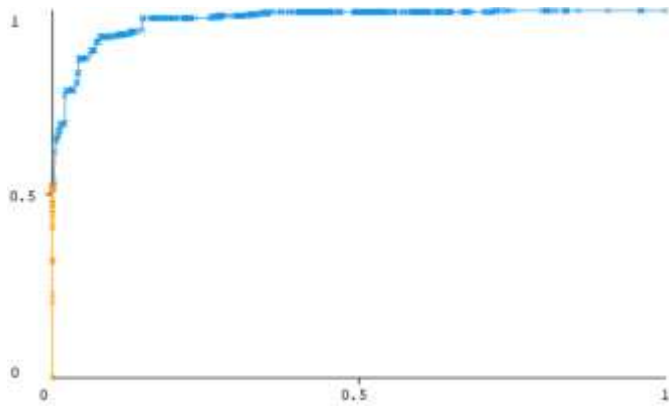


Figure 8. ROC curve for the “Miner” class

The PRC, as opposed to the ROC Area, represents the behavioral characteristics of Precision Vs Recall. The Precision value denotes the ratio of instances that are true of a given class divided by the sum of instances classified as that given class. The Recall value denotes the ratio of instances classified as a class divided by the actual sum in that given class. As such, this is equivalent to the TP rate. The F-Measure is a combined measure that depicts the ratio of double the product of Precision and Recall divided by the sum thereof, i.e. $\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\Sigma(\text{Precision} + \text{Recall})}$.

The MCC is the measure of the quality of binary classifications taking into account true and false positives and negatives. It is a balanced measure that has a range [-1, 1], with -1 denoting a completely wrong classifier and 1 indicating the opposite. The variations in the clustering coefficient and vertex degree in these network traffic statistics in the respective clusters depict the overall movement of the movement vector from the feature centroid. Table V summarizes the differences between our proposed model and existing approaches.

TABLE V. COMPARISON WITH OTHER WORKS

Attribute/Model	Flexibility to Large-Scale Networks	Dynamic Complex Network modeling	Attack Network Formulation	Evaluation of Detection Model	Not Dependent on Attack Vector
Saad et. al [20]	✓	✗	✓	✓	✗
Zareh et. al [21]	✗	✗	✗	✓	✓
Eskandari et. al [22]	✓	✗	✗	✗	✗
Carlin et. al [23]	✓	✗	✗	✓	✗
Vesely et. al [24]	✓	✗	✗	✓	✓
Musch et. al [29]	✓	✗	✗	✓	✗
Proposed Model	✓	✓	✓	✓	✓

As can be seen in Table V, our modeling and detection approach has several advantages not limited to dependency on the prevailing attack vector (i.e. browser-based or installable binary-based) and incorporation of complex network modeling for effective detection.

V. CONCLUSIONS

The results presented in this paper demonstrate that the integration of semi-supervised learning with complex network theory modeling is effective at detecting crypto mining activities in a network environment. Our model's efficiency was enhanced by first clustering unlabeled data based on dynamic complex network characteristics and classifying the resultant clusters using a proximity-based classification algorithm, hence semi-supervised learning. The dynamic network characteristics exhibited in the network traffic generated by crypto mining activities serve as the modeling basis for detection. The presence of such crypto mining traffic in a corporate network is a high indicator of compromise. Our proposed detection methodology is advantageous in that it's independent of the nature of the victim device nor the underlying operating system since it's solely based on dynamic network statistics. Such an approach finds wide application in heterogeneous networks with varied devices such as IoT, SCADA/ICS systems, critical infrastructure, cloud computing, and so forth.

REFERENCES

- [1] N. Kshetri and J. Voas, "Do Crypto-Currencies Fuel Ransomware?," *IT Prof.*, 2017.
- [2] M. C. Kus Khalilov and A. Levi, "A survey on anonymity and privacy in bitcoin-like digital cash systems," *IEEE Commun. Surv. Tutorials*, 2018.
- [3] M. Möser et al., "An Empirical Analysis of Traceability in the Monero Blockchain," *Proc. Priv. Enhancing Technol.*, 2018.
- [4] Branche Patrick O, "Ransomware: An Analysis of the Current and Future Threat Ransomware Presents," Utica College, 2017.
- [5] N. Hampton and Z. A. Baig, "Ransomware: Emergence of the cyber-extortion menace," *Proc. the 13th Aust. Inf. Secur. Manag.*, 2015.
- [6] T. Bamert, C. Decker, R. Wattenhofer, and S. Welten, "BlueWallet: The secure Bitcoin wallet," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2014.
- [7] M. Salimitari, M. Chatterjee, M. Yuksel, and E. Pasilio, "Profit Maximization for Bitcoin Pool Mining: A Prospect Theoretic Approach," in *Proceedings - 2017 IEEE 3rd International Conference on Collaboration and Internet Computing, CIC 2017*, 2017.
- [8] M. A. Dave McMillen, "Mirai IoT Botnet: Mining for Bitcoins?," *Security Intelligence*, 2017. [Online]. Available: <https://securityintelligence.com/mirai-iot-botnet-mining-for-bitcoins/>. [Accessed: 20-Oct-2019].
- [9] "Detection of a Crypto-Mining Malware Attack at a Water Utility," *Radiflow*. [Online] <https://radiflow.com/case-studies/detection-of-a-crypto-mining-malware-attack-at-a-water-utility/>. [Accessed: 13-Oct-2019].
- [10] J. Vijayan, "Crypto-Mining Attacks Emerge as the New Big Threat to Enterprises," *Dark Read.*, vol. 1, no. ATTACKS/BREACHES, p. 1, 2018.
- [11] C. B. Raj Samani, "McAfee Labs Threats Report," Santa Clara, 2018.
- [12] A. Zimba "Recent Advances in Cryptovirology: State-of-the-Art Crypto Mining and Crypto Ransomware Attacks," *KSII Trans. Internet Inf. Syst.*, 2019.
- [13] M. A. Chenta Lee, Andrey Iesiev, Mark Usher, Dirk Harz, Martin Steigemann, Marc Noske, Abby Ross, Scott Moore, Limor Kessem, Tomer Agayev, Claire Zaboeva, Camille Singleton, Dave McMillen, Dave Bales, Joshua Chung, "IBM X-Force Threat Intelligence Index 2019," United States of America, 2019.
- [14] J. Zorabedian, "Cryptojacking Rises 450 Percent as Cybercriminals Pivot From Ransomware to Stealthier Attacks," *Secur. Intell.*, vol. 1, no. 2, 2019.
- [15] J. Bloomberg, "Top Cyberthreat Of 2018: Illicit Cryptomining," *Forbes*, vol. 1, no. 3, p. 1, 2018.
- [16] C. Haan, "African Countries Most At Risk of Ransomware and Cryptomining Attacks," *Microsoft*, 2019. [Online]. Available: <https://www.crowdfundinsider.com/2019/04/146716-african-countries-most-at-risk-of-ransomware-and-cryptomining-attacks/>. [Accessed: 23-Sep-2019].
- [17] X. F. Wang and G. Chen, "Complex networks: Small-world, scale-free and beyond," *IEEE Circuits and Systems Magazine*. 2003.
- [18] A. Zimba, Z. Wang, and M. Mulenga, "Cryptojacking injection: A paradigm shift to cryptocurrency-based web-centric internet attacks," *J. Organ. Comput. Electron. Commer.*, vol. 29, no. 1, 2019.
- [19] P. R. K. S. B. Kumar, C. U. Om, "Detecting and confronting fash attacks from IoT botnets," *J. Supercomput.*, vol. 1, no. 1, pp. 1–27, 2019.
- [20] A. M. Muhammad Saad, Aminollah Khormali, "End-to-end analysis of in-browser cryptojacking," *ArXiv*, pp. 1–15, 2018.

- [21]A. Zareh and H. R. Shahriari, "BotcoinTrap: Detection of Bitcoin Miner Botnet Using Host Based Approach," in *2018 15th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology, ISCISC 2018*, 2018.
- [22]S. Eskandari, A. Leoutsarakos, T. Mursch, and J. Clark, "A First Look at Browser-Based Cryptojacking," in *Proc - 3rd IEEE European Symposium on Security and Privacy Workshops, EURO S and PW 2018*, 2018.
- [23]D. Carlin, P. Orkane, S. Sezer, and J. Burgess, "Detecting Cryptomining Using Dynamic Analysis," in *2018 16th Annual Conference on Privacy, Security and Trust, PST 2018*, 2018.
- [24]V. Veselý and M. Žádník, "How to detect cryptocurrency miners? By traffic forensics!," *Digit. Investig.*, 2019.
- [25]S. Delgado-Segura, C. Pérez-Solà, J. Herrera-Joancomartí, G. Navarro-Arribas, and J. Borrell, "Cryptocurrency Networks: A New P2P Paradigm," *Mobile Information Systems*. 2018.
- [26]M. Veselý V. & Žádník, "Dataset: How to Detect Cryptocurrency Miners? By Traffic Forensics!," *GitHub*, 2018. [Online]. Available: <https://github.com/nesfit/DI-cryptominingdetection/blob/master/README.md>.
- [27]H. B. Bhavsar and A. G. Jivani, "The Shared Nearest Neighbor Algorithm with Enclosures (SNNAE)," in *2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009*, 2009.
- [28]"Weka 3: Machine Learning Software in Java." <https://www.cs.waikato.ac.nz/ml/weka/>.
- [29]K. R. Marius Musch, Christian Wressnegger, Martin Johns, "Thieves in the Browser: Web-based Cryptojacking in the Wild," in *Proc. of the 14th Int. Conference on Availability, Reliability & Security. ACM*, 2018.

A Conceptual Secure Blockchain Based Settlement and Clearinghouse for Mobile Financial Services in Zambia

Fickson Mvula¹, Jackson Phiri², Simon Tembo³

^{1,3}*Department of Electronics and Electrical Engineering, School of Engineering, UNZA, Lusaka, Zambia,* ²*Department of Computer Science, School of Natural Sciences, UNZA, Lusaka, Zambia*
¹*fickson.mvula@gmail.com,* ²*jackson.phiri@cs.unza.zm,* ³*simon.tembo@unza.zm*

Abstract— Developing Countries in Africa in general and Zambia in particular, have seen a rapid rise in use of mobile payment platforms. This has not only revolutionized access to finance for the poor but also allowed them access to other financial products such as savings or insurance. Mobile financial wallets are being used by different mobile network operators to extend their product offering beyond the traditional voice services. Equally other non-mobile network providers have joined the race in providing mobile money wallets. As a result of these different mobile wallet providers, subscribers are presented with a problem of interoperability between them where transfer of value from one provider wallet to another on a different network is not possible. In this study we first review the technical landscape and features of mobile payment systems in Zambia and then assess the feasibility of using blockchain technology in proposing a settlement and clearing system that will allow mobile money interoperability. A decision model is used to test what form the proposed prototype system design would take. A prototype is then designed in which amounts being interchanged between providers are managed as assets on a permissioned blockchain. The system runs a distributed shared ledger which provides non-repudiation, data privacy and data origin authentication, by leveraging the consistency features of the blockchain. Development of the prototype is being undertaken as the third objective of the study.

Keywords— *Blockchain, Mobile Money Interoperability, Clearing and Settlement, Blockchain Security*

I. INTRODUCTION

There is a growing number of mobile money wallet services providers in Zambia which has led to the creation different autonomous financial ecosystems with little to no interoperability between them. We define interoperability as an ability of one mobile money subscriber on one network being able to transfer value to another on a different network. Attempts have been made to close this gap through provision of bilateral arrangements between mobile money providers which has proved problematic as there are delays in settlement due to ledger trust issues. Integration of wallet providers' systems through a central clearing house for purposes of clearing and settlements is thus necessary to achieve interoperability. Blockchain technology presents a perfect opportunity as a potential technology to disrupt payment, clearing and settlement because of its ability to introduce a set of synchronized ledgers managed by one or more entities rather than individual non communicating ledgers. This would lead to

a reduction in the reliance on traditional central ledger managed by a trusted entity for holding and transferring funds.

II. MOBILE MONEY INTEROPERABILITY IN ZAMBIA

A. Current Status

As at now, there is currently no live implemented system that allows interoperability between the different mobile financial services wallet providers in Zambia. The proposed Zambia National Switch project [1] being undertaken by the Zambia Electronic Clearing House Limited (ZECHL) will among others enable participants in the mobile financial ecosystem to interchange money by providing a clearing and settlement platform. The system implementation will be phased and the first phase expected to cater for interoperability of commercial banks and expected to be launched at the end of 2018. The second phase will cater for integration of other financial services such as mobile money and telegraphic money transfers [2].

B. Challenges With The Proposed Approach

The National Financial Switch system however, being a traditional database based central system will have a number of shortfalls in as far of effective provision of the desired features identified for clearing and settlement of account to account (A2A) interoperability transactions. Firstly, there will be integration complexity as every participant will be required to connect to a central node. This central node of processing will hinder efficiencies in end-to-end processing speed and thus availability of funds may be hampered. Further, there will be no network resilience offered by distributed data management system such as one provided by a distributed ledger system. And furthermore, there may be operational and financial risks as a result of a single central node rather than a distributed one.

III. CLEARING AND SETTLEMENT ON A BLOCKCHAIN

This paper proposes the design of a secure and trusted blockchain based clearing and settlement architecture for mobile financial services in Zambia.

A. Blockchain Defined

A blockchain can be defined as an immutable ledger for recording transactions, maintained within a distributed network of mutually untrusting peers. Every peer maintains a copy of the ledger. The peers execute a consensus protocol to validate transactions, group them into blocks, and build a hash chain

over the blocks. This process forms the ledger by ordering the transactions, as is necessary for consistency. Blockchains have emerged with Bitcoin and are widely regarded as a promising technology to run trusted exchanges in the digital world [3].

The two main categories of blockchains are public and private blockchains. In a public or permission-less blockchain anyone can participate without a specific identity. Public blockchains typically involve a native cryptocurrency and often use consensus based on “proof of work” (PoW) and economic incentives. Permissioned blockchains, on the other hand, run a blockchain among a set of known, identified participants. A permissioned blockchain provides a way to secure the interactions among a group of entities that have a common goal but which do not fully trust each other, such as businesses that exchange funds, goods, or information. By relying on the identities of the peers, a permissioned blockchain can use traditional Byzantine-fault tolerant (BFT) consensus.

B. The Case For Blockchain In Clearing And Settlement

The proposed solution is directed primarily at arrangements that involve restricted ledgers (access to which is for approved users only) or permissioned blockchains, reflecting the main types of arrangement currently being developed in the financial sector, such as one required for a mobile money account to account interoperability among a number of disparate network providers.

Clearing and settlement of a financial transaction, regardless of the asset type, requires a network of participants, an asset or set of assets that are transferred among those participants, and a transfer process that defines the procedures and obligations associated with the transaction. Typically, the set of direct participants are financial institutions such as banks or brokers and indeed mobile wallet providers in the case of mobile financial services. Indirect participants include end users such as subscribers in this case. An asset can be any financial instrument, such as a monetary instrument, security, commodity, or derivative. Again in a mobile financial services ecosystem the asset type of interest is virtual money (or e-money) being transferred from one wallet to another across the network of participants. Communications among the participants in a network involve sending electronic messages, acknowledgements, statements, and other information between computer systems typically maintained by a network operator and its participants.

It is worth noting at this stage that the current implementation of such networks is such that each participant maintains and is responsible for their own financial ledger which acts as their single source of truth on the status of their data. To achieve interoperability, a common central authority may be necessary which would be entrusted by their participants with updating and preserving the integrity of a central ledger and, in some cases, managing certain risks on behalf of participants.

The case for Distributed ledger technology (DLT) as a potential technology to disrupt payment, clearing and settlement implementations is because of the technology’s ability to introduce a set of synchronized ledgers managed by one or more entities rather than individual non communicating

ledgers. This would lead to a reduction in the reliance on traditional central ledger managed by a trusted entity for holding and transferring funds and other financial assets.

DLT may radically change how assets are maintained and stored, obligations are discharged, contracts are enforced, and risks are managed. Proponents of the technology highlight its ability to transform financial services and markets by [4]:

- Reducing complexity
- Improving end-to-end processing speed and thus availability of assets and funds.
- Decreasing the need for reconciliation across multiple record-keeping infrastructures
- Increasing transparency and immutability in transaction record keeping
- Improving network resilience through distributed data management
- Reducing operational and financial risks

IV. LITERATURE REVIEW

In their basic sense Mobile payments platforms allow their users to pay and transfer funds in mobile money, but also offer access to other financial products, such as savings and bill payments. [5] reviewed the economic features of mobile payment systems in developing countries, and studied the cooperation models that can emerge between the different firms potentially involved in a mobile payment transaction. Focus was drawn on the main competition concerns that public authorities should be concerned about, and which regulatory tools could be considered as a remedy.

Key among some of the key challenges in mobile money schemes was the issue of interoperability. Different concepts of interoperability are relevant and need to be distinguished according to their implications for regulation and business models differ.

Different approaches have been undertaken by different countries in an attempt to implement interoperability for their mobile money financial systems. This section reviews a number of such proposed architectures for mobile payments that support interoperability. These have been drawn from well-developed mobile money markets and they include India, Kenya, Rwanda, and Tanzania. Next, a number of blockchain based use cases were reviewed and presented to support the case for use of blockchain in a system model proposed.

A. Interoperability Approaches

The In 2008 the Reserve Bank of India (RBI) provided an interoperability platform called UPI [6]. This is however, a central integrating node which suffers integration complexities.

Alternative architecture approaches proposed [7] with hierarchical lookup. Kumar et al. also proposed architectural choices [8]. However, their model is specific to highly

regulated financial environment in India, where every transaction is processed by a bank.

Interoperability is not mandated under the Kenyan National Payments System (NPS) regulations but instead payment service use bilateral arrangements [11], [12], [13] rather than through a common central switch system. But as has been observed by [14] a common switch, with its own set of rules for participation, technical and operational issues, improves coordination and customer experience, and allows for a much faster implementation of interoperability, as compared to private switches or bilateral agreements. Other options in the Indian landscape include, the Mobile Payment Foundation of India [8] which is also developing a model for interoperability. Further, Kumar et al. have proposed architectural choices for interoperability [8]. However, their model is specific to highly regulated financial environment in India, where every transaction is processed by a bank.

There are four different mobile network operators all providing mobile money services to their subscribers in Tanzania [16]. Tanzania is one of the most successful mobile money markets in the world with more than 25% of the population being active mobile money users (with almost 11 million in December 2013) and transacting an estimated USD 2 billion in transactions per month in 2014 [16]. According to a study [17] by the GSMA on account to account (A2A) interoperability models in Tanzania and Pakistan, A2A interoperability was launched in Tanzania in 2014, and in Pakistan in 2015. The study found that in both Pakistan and Tanzania, the regulatory environments were enabling for A2A interoperability and that providers freely choose the technical model that best suited their commercial interests rather than being restricted to a pre-determined or preferred model defined by regulation. This has led to Tanzanian mobile money providers opting for bilateral point to point integrations as a preferred model for interoperability.

As been pointed out [18], bilateral models may seem easy to deploy where there are limited parties involved but later suffer several disadvantages including the increase in complexity with number of parties, duplication of efforts and an increases in complexity of maintenance over time.

Literature studied showed a number of different approaches to interoperability employed in different countries. One such an approach is the use of a Central Bank led national switching system for clearing and settlement. Mobile money services in Zambia are regulated by the Central bank and therefore, this makes the use of a central switch an ideal and suitable enough approach to interoperability. So far, the technological setup used in such an approach has been with a central database system. A number of problems with this approach have pointed out including, complexity of integration, introduction of a single point of failure and lack of trust. This paper therefore proposed a blockchain based solution approach to address these shortfalls. A number of blockchain use cases are presented in the following section to highlight some of the properties of blockchain that make it a suitable technology to address these problems.

B. Blockchain Use Cases

A number of blockchain based solutions have been proposed by various researchers across multiple industries over the last few years that the technology has matured. This section highlights some of these solutions.

Firstly, [4] examined the use of Distributed Ledger Technologies (DLT) in the area of payments, clearing and settlement and identified both a number of opportunities and challenges facing its long-term implementation and adoption. Further calls for tamper-resistant data stores solutions are made in [19] by proposing the use of a write once and read multiple times data storage solution. Similar calls are echoed in an attempts to solve problems in the management of clinical records [20]. It is argued that a blockchain technology has the potential to solve the records management problems by providing a single, secure, decentralized storehouse of clinical data for all patients.

A solution for parking slot management in a trust less network is proposed here [22] which seeks to provide a platform capable of being used without a third trusted party.

In [25] a decentralized traceability system based on IoT and blockchain is proposed for the food industry. In [26] proposes a hybrid architecture for supply chain management based on a set of private distributed ledgers for storing sensitive customer information and a public ledger where a hash of each private event is stored along with the monitoring events.

Like many such similar use cases proposed, the goal to implement a secure and trusted system that takes advantage of the blockchain properties of transparency, immutability and shared consensus [29].

V. RESEARCH PROCESS

The study was guided by three (3) objectives. Firstly, a targeted survey and interviews were conducted to establish how mobile financial services are currently implemented in Zambia. Further, literature and documentation on mobile money system and service implementation was consulted to understand how they are setup. The goal was to try to establish and highlight short falls and inefficiencies in implementation that prevent interoperability and thereby identify opportunities for improvements.

Secondly, an analysis as to whether a conceptual model for inter operator mobile financial transactions payments, clearing and settlement in a secure, transparent and trusted manner could be proposed and designed. The goal was to establish if blockchain technology would be an ideal technology to achieve the proposed design.

Finally, carry out implementation of a prototype that demonstrates Blockchain security services in a permissioned and regulated environment. The designed system was a prototype system in which amounts being interchanged between mobile money providers are managed as assets on a permissioned blockchain. The system runs a distributed shared ledger which prevents amount theft as well as fraud such as transferring invalid amounts, or transferring multiple copies of

an amount, by leveraging the consistency features of the blockchain.

A. Baseline Study

A list of interview questions were designed into a survey and administered to a target audience of respondents, deliberately selected according to set criteria. Further, walk in interviews were conducted with subject matter experts to validate and verify researched literature and documentation on mobile money systems and service implementation. The goal was to try to establish and highlight short falls and inefficiencies in implementation that prevent interoperability and thereby identify opportunities for improvements in the solution design.

B. Target Group

The research participants were purposively selected basing on their expertise, experience and skills relating to the subject under study in order to get rich and relevant information. Survey participation was drawn from employees of Zambia’s mobile money operators and employees from Zambia’s mobile money regulatory and supervisory authority, Bank of Zambia. The operators included the major Mobile Network Operators (MNOs), Airtel (Airtel Money), MTN (MTN Mobile Money) and Zamtel (Zamtel Kwacha). Participation was further extended to non-MNO providers who have been running money transfer services on mobile and have since extended their product offerings to include the mobile wallet feature on their services, which allows customers to hold value and transact off those accounts. These included Zoona (who run the Zoona Plus wallet), Broadpay (who run the Broadpay wallet) and cGrate (who run the Konse Konse wallet).

C. Sampling Rationale

Due to the specialized nature of the data that the research required, survey respondents had to be conveniently sampled. The Bank of Zambia, for example, is the regulatory authority that supervises and regulatory financial services providers in Zambia. They do this through among others registration and designation of payment systems and institution as well as oversight of both systemic and non-systemic payment systems. The central bank is also responsible for the clearing and settlement infrastructure and processes in the country. It was felt strongly therefore, that they would be well positioned to provide information on payment system interoperability from regulatory and standards perspectives. Participation therefore, was also drawn from a number of Bank of Zambia staff with varying specializations. These included Payments Systems specialists, Financial Institutions Supervision specialists, Information Systems specialists and Information Systems Security specialists.

D. Decision Model

For the second part of the study, we looked at whether and how a blockchain based solution would be ideal for this use case. A flow chart based decision model was used to determine the suitability of the technology to be adopted as proposed by Wüst and Gervais [30]. Other such similar models have been proposed [31]. This model was found more suitable as it provides a detailed description of the decisions leaving less

room for misinterpretation. The model consist of a decision tree based on the following scenario properties:

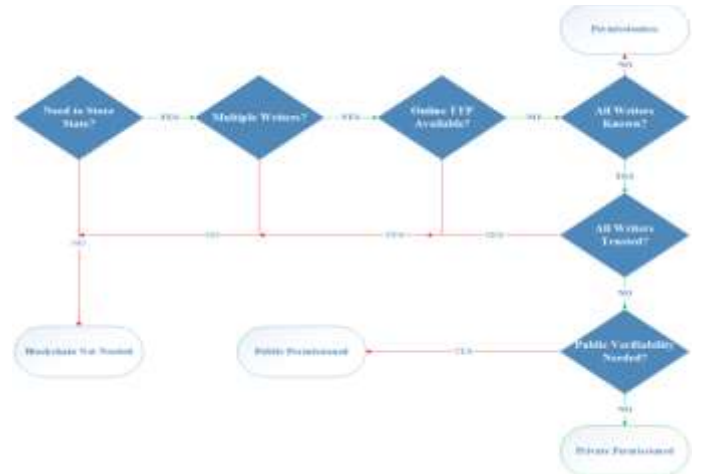


Fig. 1. Decision model adopted on blockchain use case (Source: Wüst and Gervais [30])

- a) *Storing state.* Refers to the need of storing data that may change both in volume and in content over time.
- b) *Existence of writers* that have a common interest in agreeing on the validity of the stored state.
- c) *Is there a Trusted Third Party?* A Trusted Third Party (TTP) is a centralized entity that could manage changes and updates the state. A TTP, if present, may also control who can read the state stored.
- d) *Are all writers known?* This refers to knowing the identity of all writers.
- e) *Are all writers trusted?* When writers are trusted, they are expected not to behave maliciously. When writers are not trusted, they may behave maliciously.
- f) *Public verifiability of state.* This property determines who may read the state stored on the blockchain, and verify the integrity of the ledger.

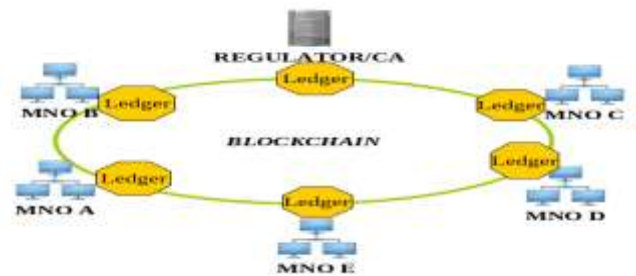


Fig. 2. Proposed Network Architecture

E. System Design

A formal software development methodology was followed in the design and implementation of the solution prototype proposed. Object Oriented Analysis and Design methodology using the Object Modeling Techniques (OMT) phases to model the different aspects of the prototype was used. The proposed

framework consists of a common replicated ledger in which transferred amounts are managed as assets on a permissioned blockchain based on Hyperledger Fabric (Fig 2). Hyperledger Fabric is an open source permissioned distributed ledger technology (DLT) platform, designed for use in enterprise contexts. Fabric was chosen because of its highly modular and configurable architecture that makes it adaptable to a number of use cases. Fabric also supports the use of general purpose programming languages such as Java in the development of smart contracts and therefore, was an ideal choice for this prototype. Blockchain approach was used to provide key security requirements of confidentiality, origin authentication, non-repudiation and availability.

i. Network Architecture

The network layout is depicted in Fig 2 as a shared, replicated, permissioned distributed ledger where all participants have a copy of the ledger alongside their data. The blockchain architecture gives participants the ability to share a ledger that is updated every time a transaction occurs through peer-to-peer replication.

Fig 3 shows the main nodes and components that make the proposed solution. Each participants (labelled as MNO) maintains their own mobile money systems. As part of the Fabric network, each participant also runs nodes called Peers which allows them to connect to the rest of the blockchain network. These peers receive transaction requests from participant systems through an Application Programming Interface (API) provided by the Software Development Kit (SDK). Each pair of participants (MNO A and MNO B for example) connect through a separate channel interface that allows them to maintain data privacy between the two. The Orderer node is responsible for ordering and writing transaction requests to the ledger before replication.

ii. Use Case Model

The main asset that is transacted on the proposed network is a **transfer** and this represents a request made by one subscriber through a participant to transfer an amount to another subscriber on a different participant's network. Fig 4 shows the main use cases in the system.

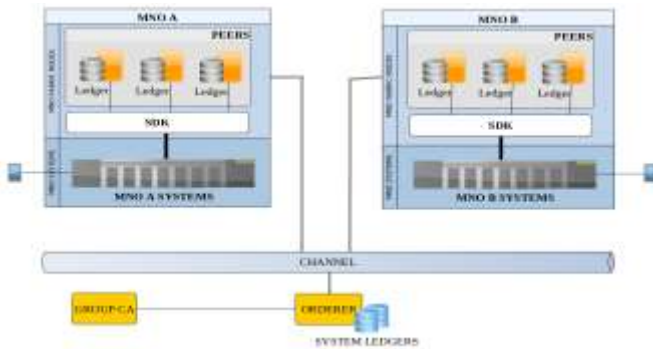


Fig. 3. Blockchain Solution Architecture

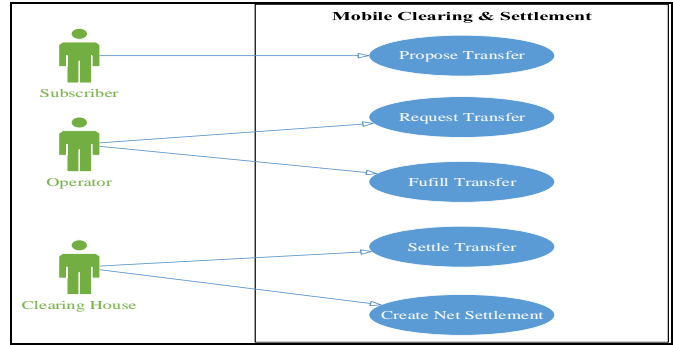


Fig. 4. Use cases

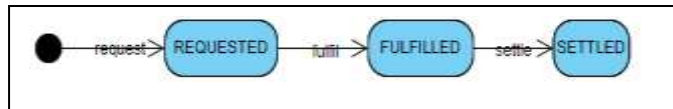


Fig. 5. State Transition Diagram

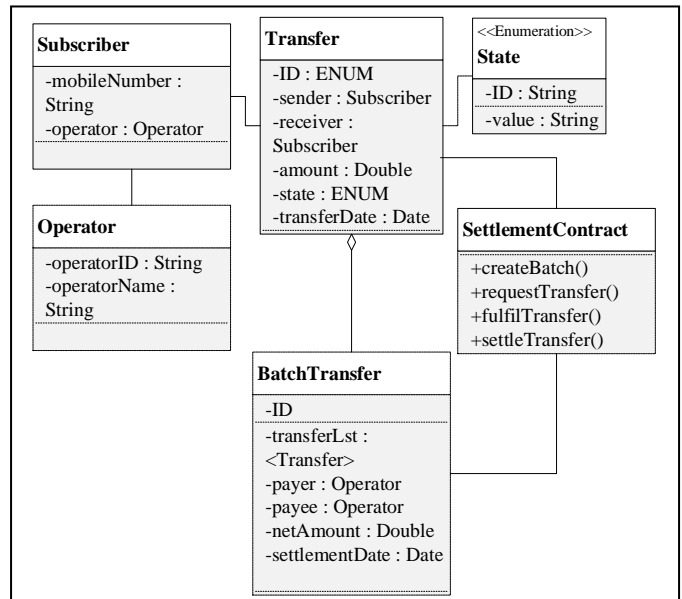


Fig. 6. Class diagram

iii. Smart Contract Design

On the Blockchain, the Transfer represents the conceptual object of value and is modeled as states, whose lifecycle transitions are described by transactions. Fig 5 shows the life cycle of a transfer. The transfer transitions between requested, fulfilled and settled states by means of the request, fulfill and settle transactions. A special program called a smart contracts was implemented that models this transaction logic that transitions the transfer between their different states. Smart contracts allowed us to define the key business processes and data that are shared across the different organizations collaborating in the network. Fig 6 shows the class model that

captures the smart contract and depicts the main objects that make up the smart contract.

VI. RESULTS AND ANALYSIS

A. Baseline Study Results

Participation was drawn from providers with varied subscriber bases (Fig 7) and each using different platforms and reported facing integration challenges (Fig 8) Overall, on inter operator integration, the general feeling was that it was manageable and could be eased with the use of a central integrator rather than having every operator to integrate individually with every other provider.

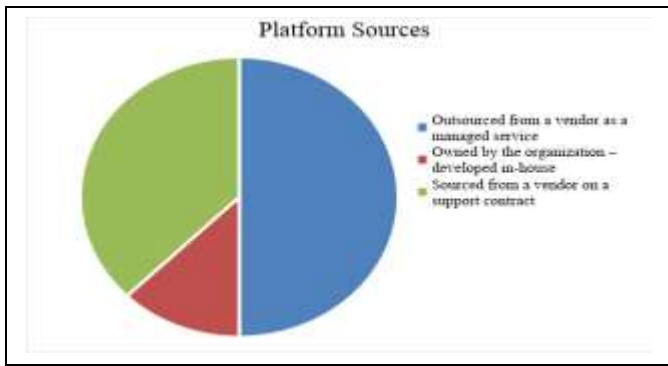


Fig. 7. Platform Sources

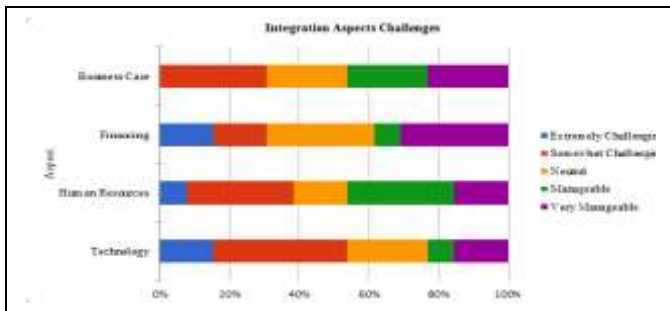


Fig. 8. Integration Challenges

B. Decision Model Analysis

Table 1 summarizes key survey findings. The Decision model in Fig 1 was used with these findings to determine blockchain suitability for this use case and showed that the use case could use a permissioned blockchain. A number of important aspects such as the need to store state, existence of multiple writers were used to arrive at the decision of the solution. The next section highlights the main implementation aspects of the proposed solution.

TABLE I. KEY SURVEY FINDINGS

Decision Model Analysis		
Decision State	Finding Description	Result
Storing state	Existence of different independent mobile money operators	YES
Existence of writers	Existence of technological platforms or systems on which these operators run their services	YES
Trusted Online Third Party	Controlled access to the network with permissioning.	NO
Are all writers known	The need for integration among these systems to provide interoperability	YES
Are all writers trusted	Security and privacy of transactions	NO
Public verifiability of state	Security and privacy of transactions	NO

^a Decision tree based on [30]



Fig. 9. Object Class Members

Fig. 10. Settlement Class Methods

C. Implementation

The smart contract was developed from the design using java thanks for high level language support on the Hyperledger Fabric. The main classes are the Transfer and BatchTransfer which represents the main objects (Fig 9). The SettlementContract (Fig 10) is the main logic class which models the transaction definitions. This class implements the following methods that control application lifecycle:

requestTransfer() – Creates a new transfer context object between two participants (sender and receiver) which is saved on the ledger as an asset.



fulfilTransfer() – Gets a transfer object in *REQUESTED* state and sets it *FULFILLED* (after the receiver has fulfilled the transaction).

settleTransfer() – Sets fulfilled transactions to *SETTLED* after net settlement.

createBatch() – At end of day, all transactions between pairs of transacting parties are batched into a BatchTransfer object which is used for settlement.

CONCLUSION

The study proposed the use of blockchain technology to solve the problem of mobile money interoperability in Zambia. A structured approach was used to confirm the gap and then decide a technological solution through the use of a structured decision model for careful determination. We further designed a prototype system on the Hyperledger Fabric network which could developed in an Object Oriented language such as Java for deployment. We conclude that mobile money interoperability settlement is a valid use case for a permissioned blockchain technology and would be an ideal solution approach rather than the traditional central processing database systems.

ACKNOWLEDGMENT

Acknowledgement of special thank you go to the University of Zambia (UNZA) Department Of Computer Science and Department of Electronics and Electrical Engineering for the support and guidance during the study.

REFERENCES

[1] Bank of Zambia, “Payment Systems Vision and Strategy 2018-2022,” Bank of Zambia, Lusaka.

[2] T. TEMBO, “National financial switch to start this year,” *Daily Mail*, 2017. [Online]. Available: <http://www.daily-mail.co.zm/national-financial-switch-to-start-this-year/>. [Accessed: 21-Jan-2018].

[3] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” pp. 1–9, 2013.

[4] D. C. Mills *et al.*, “Distributed Ledger Technology in Payments, Clearing, and Settlement,” 2016.

[5] M. Bourreau and T. Valletti, “Competition and Interoperability in Mobile Money Platform Markets: What Works and What Doesn’t? (*),” *Commun. Strateg.*, 2015.

[6] R. B. of I. Department of Payment and Settlement Systems, “Prepaid Payment Instruments (PPIs) – Operational Guidelines for Interoperability,” no. 808. Mumbai, India, pp. 2–4, 2018.

[7] D. Kumar, T. A. Gonsalves, A. Jhunjhunwala, and G. Raina, “Mobile payment architectures for India,” *Proc. 16th Natl. Conf. Commun. NCC 2010*, no. 1, pp. 4–8, 2010.

[8] K. Kumar and M. Tarazi, “Interoperability in Branchless Banking and Mobile Money. Consultative Group to Assist the Poor (CGAP).,” *CGAP*, 2012. [Online]. Available: <http://www.cgap.org/blog/interoperability-branchless-banking-and-mobile-money-0>.

[9] B. Andiva, “Mobile Financial Services and Regulation in Kenya,” in *1st Annual Competition and Economic Regulation (ACER) Conference*, 2015.

[10] B. Muthiora, “Enabling mobile money policies in Kenya - Fostering a digital financial revolution,” *GSM Assoc.*, no. January, p. 30, 2015.

[11] M. Mamabolo, “Kenya’s Central Bank gives mobile money interoperability thumbs up,” *IT WebAfrica*, 2018. [Online]. Available: <http://www.itwebafrica.com/fintech/842-kenya/243988-kenyas-central-bank-gives-mobile-money-interoperability-thumbs-up>. [Accessed: 20-Dec-2018].

[12] Central Bank of Kenya, “Mobile Money Interoperability,” no. January, p. 2018, 2018.

[13] M. Mamabolo, “Kenya pilots mobile money interoperability,” *IT WebAfrica*, 2018. [Online]. Available: <http://www.itwebafrica.com/mobilex/309-kenya/242443-kenya-launches-mobile-money-interoperability-pilot-today>. [Accessed: 20-Dec-2018].

[14] M. Bourreau and S. Hoernig, “Interoperability of Mobile Money: International Experience and Recommendations for Mozambique,” no. December, 2016.

[15] J. Argent and J. A. Hanson, “The Regulation of Mobile Money in Rwanda,” *J. ICT*, 2013.

[16] Bank of Tanzania, “National Payment System Directorate Statistics:” [Online]. Available: <http://www.bot-tz.org/PaymentSystem/statistics.asp>. [Accessed: 18-Dec-2018].

[17] GSMA, “Choosing a Technical Model for A2A Interoperability: Lessons from Tanzania and Pakistan,” no. December, 2015.

[18] D. Clark and G. Camner, “A2A Interoperability: Making mobile money schemes interoperate,” no. February, 2014.

[19] B. Smith and K. Christidis, “IBM Blockchain: An Enterprise Deployment of a Distributed Consensus-based Transaction Log,” *Proc. Fourth Int. IBM Cloud Acad. Conf.*, pp. 1–4, 2016.

[20] D. Ivan, “Moving Toward a Blockchain-based Method for the Secure Storage of Patient Records,” 2016.

[21] R. J. Reisman, “Air Traffic Management Blockchain Infrastructure for Security , Authentication , and Privacy,” pp. 1–14, 2019.

[22] L. García-Bañuelos and F. Milani, “Sven Mitt Blockchain Application-Case Study on Hyperledger Fabric Blockchain Application-Case Study on Hyperledger Fabric.”

[23] M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman, “Applied Innovation Review,” *Appl. Innov. Rev.*, no. 2, pp. 5–20, 2016.

[24] K. Korpela, J. Hallikas, and T. Dahlberg, “Digital Supply Chain Transformation toward Blockchain Integration Kari,” *Hawaii Int. Conf. Syst. Sci.* /, pp. 4182–4191, 2017.

[25] F. Tian, “A Supply Chain Traceability System for Food Safety Based on HACCP, Blockchain & Internet of Things,” *14th Int. Conf. Serv. Syst. Serv. Manag.*, vol. 323, no. 2, pp. 511–519, 2015.

[26] H. Wu, Z. Li, B. King, Z. Ben Miled, J. Wassick, and J. Tazelaar, “A distributed ledger for supply chain physical distribution visibility,” *Inf.*, vol. 8, no. 4, pp. 1–18, 2017.

[27] I. Haq and O. Muselemu, “Blockchain Technology in Pharmaceutical Industry to Prevent Counterfeit Drugs,” *Int. J. Comput. Appl.*, vol. 180, no. 25, pp. 8–12, 2018.

[28] Y. Omran, M. Henke, R. Heines, and E. Hofmann, “Blockchain-driven supply chain finance: Towards a conceptual framework from a buyer perspective Yaghoob Omran,” 2009.

[29] A. Baliga, “The Blockchain Landscape Office of the CTO,” *Persistent Syst. Ltd.*, p. 21, 2016.

[30] K. Wüst and A. Gervais, “Do you need a Blockchain?,” *IACR Cryptol. ePrint Arch.*, no. i, p. 375, 2017.

[31] F. Lamberti, V. Gatteschi, C. Demartini, C. Pranteda, and V. Santamaria, “Blockchain or not blockchain, that is the question of the insurance and other sectors,” *IT Prof.*, 2017.

CRIME MAPPING MODEL BASED ON CLOUD AND SPATIAL DATA

A CASE STUDY OF ZAMBIA POLICE SERVICE

Jonathan Phiri ¹, Jackson Phiri ²

¹ Department of Electrical and Electronic Engineering, University of Zambia, Lusaka, Zambia

² Department of Computer Science, University of Zambia, Lusaka, Zambia

¹jonathanphiri82@yahoo.com, ²jackson.phiri@cs.unza.zm

Abstract—Crime mapping is a strategy used to detect and prevent crime in the police service. The technique involves the use of geographical maps to help crime analysts identify and profile crimes committed in different residential areas, as well as crafting best methods of responding. The development of geographic information system (GIS) technologies and spatial analysis applications coupled with cloud computing have significantly improved the ability of crime analysts to perform this crime mapping function. The aim of this research is to automate the processes involved in crime mapping using spatial data. A baseline study was conducted to identify the challenges in the current crime mapping system used by the Zambia Police Service. The results show that 85.2% of the stations conduct crime mapping using physical geographical maps and pins placed on the map while 14.8% indicated that they don't use any form of crime mapping technique. In addition, the study revealed that all stations that participated in the study collect and process the crime reports and statistics manually and keep the results in books and papers. To address the second objective, the results of the baseline study were used to develop the business processes and a crime mapping model, this was implemented successfully. The proposed model includes a spatial data visualization of crime data based on Google map. The proposed model is based on the Cloud Architecture, Android Mobile Application, Web Application, Google Map API and Java programming language. A prototype was also developed and the test results of the proposed system shows improved data visualization and reporting of crime data with reduced dependency on manual transactions.

Keywords:Zambia Police, web application, Mobile application, cloud model, crime mapping, Spatial Data.

1. INTRODUCTION

1.1 Background

Challenges in preventing and reducing crimes are what most governments around the world are struggling to deal with, every family and business have been directly affected by different kinds of crimes like robberies, vandalism, burglaries, sexual and other crimes [1]. Crimes affect the quality of life, economic growth, and reputation of a nation. There is need for the law enforcements to take tough preventive measures to reduce crimes in communities [2]. In Zambia, the Zambia Police is considered as the main law enforcement agency in Zambia mandated to enforce law on Zambian citizens, combat crime thereby playing a critical role in the Zambian criminal justice system [3]. A criminal justice system comprises of government institutions mandated to detect and mitigate crime, it focuses on how criminal cases flow from the time they are reported and investigated up to when they are disposed off. A well

established and effective criminal justice system is the key to the reduction of crime in a nation [4]. The Zambian criminal justice system comprises of the Zambia police service as the main law enforcement agency, prosecution services, judicially and Zambia prison services now changed to Correction services. The prosecution together with the judicially services ensure that accused criminal offender are tried in the courts of law and fined or sentenced to prison if found guilty. The Zambia Prisons now known as the Correction services are there to provide human custodian and correction services to inmates with the view to reform and rehabilitate all sentenced offenders [5]. Ordinary citizens and communities in a Government expect the criminal justice system and its general capacity to not only protect the communities and deal with criminal offenders but also interact with different various parties including victims, witnesses, accused as well as criminal justice professionals [6], therefore the Zambia police being the first to have contact with these people can be regarded as the gate keeper of the justice system. The Zambia Police was established in 1891 under British South African Company known as Northern Rhodesia police force, and later in 1964 upon attainment of independence was established then under Article 103 (3) of the constitution and now under Article 193 (2) of the 2016 amended constitution of Zambia and also under the Zambia police amendment act number 30 of 2016 of the laws of Zambia, the name was changed from Northern Rhodesia to Zambia Police force which later in 1994 changed to Zambia police service. Article 193 (2) of the 2016 amended constitution clearly outlines the roles and functions of the Zambia police service, it mandates the agency to ensure protection of life and property, preservation of peace, maintenance of law and order, upholding bill of rights and most importantly detect and prevent crime [7].

One of the key strategies used to detect and prevent crime is crime mapping. The technique involves the use of geographical maps to help crime analysts identify and profile crimes committed in different residential areas, as well as crafting best methods of responding [8]. It facilitates visual and statistical analysis of spatial crime data for a specific area by linking it with geographical variables like bars, schools, streets and others. Crime does not spread across the space evenly or equally but rather clumps on some specific areas while absent in other areas. Crime mapping is devoted at identifying high crime areas or neighborhoods also known as hotspots, hotspots are areas with high criminal activities [9]. The development of geographic information system (GIS) technologies and spatial analysis applications coupled with cloud computing have significantly improved the ability of crime analysts to perform this crime mapping function [10]. However, the Zambia Police Service is still using the manual and traditional way of mapping crimes. In order to

address the challenges of manual data management in the Zambia Police, this paper highlights a proposed computerized crime mapping model based on cloud and spatial data.

2. LITERATURE REVIEW

Harries [11] describes a crime in four dimensions; Legal – a law must be broken, Victim – someone or something has to be targeted, Offender – someone has to do the crime and Spatial – a crime has to happen at a place somewhere in space and time. Therefore, a crime is defined as the breach of criminal law that govern a particular geographical area, the criminal law that aims at protecting the lives, property and rights of citizens within a particular jurisdiction. Daglar & Argun further add that a place of a crime and any other geographical information connected with a criminal incident can give a lot of information about characteristics of possible criminals [12], this type of information is referred to as crime data. In criminology it is called spatial crime data because it contains geographical referenced attributes like geographical coordinates (longitude and Latitude) that can be used to establish the exact location of an object or incident on the map [13]. Crime does not spread across the map evenly or equally but rather clumps on some specific areas while absent in other areas. Crime mapping is devoted at identifying high crime areas or neighborhoods also known as hotspots, hotspots are areas with high criminal activities. Proactive policing pushes police officers to identify areas with high concentration of crimes, determine what causes these concentrations and find methods of reducing these concentrations [14]. Crime mapping helps police management in decision making on the allocation of resources basing on which areas have high demand for policing and areas which have low demand for policing. This chapter will highlight the history of crime mapping crime mapping technologies, cloud computing and related works.

2.1 History of Crime Mapping

The use of traditional crime pin maps for data visualization dates back in early 1830s. The crime map was a representation of pins stuck on it, the pins on the map were useful for showing where crimes occurred [11]. There are three schools that were recognized and pointed out in the study by Philips [13], the first was called **cartographic/geographic school** – it dominated between 1830s and 1880s originated from France and later spread to England. In this work governments begun to collect social data, the intended purpose was to center on the influence of variables such as wealth and population density on levels of crime. The second was called **Typological school** – The work dominated between 1880s and 1900s, it focused on the relationship between the mental and physical characteristics of people and crime. The third was called **Social ecological school** – which has dominated from 1900s up to date, it concentrated on the geographical variations on social conditions assuming that they were related to crime patterns. It focused primarily on recognizing and classifying areas in the cities with similar social characteristics [11].

2.2 Crime Mapping Technologies

The manual and traditional pin maps had a number of limitations among them were loss of data as they were updated and also the maps were static, they could not be manipulated or queried. Mapping of crimes has become a

primary function in law enforcement agencies, the advancements in computing have facilitated the development of geographical system applications. There are three main categories of technologies that can be used for crime mapping [15]; the first one is **open source maps** – these are maps that available over the internet for free, examples of common open source maps are google maps and open-street maps. Crime analysts are able to perform basic crime mapping duties at free cost. The disadvantage of open source maps is that crime analysts cannot perform complex queries. The second one is **GIS-Software** – these tools are commercial, they provide the necessary geospatial analytical functions needed for basic, intermediate and advanced mapping queries. The main disadvantage of GIS-software is the high cost of acquiring a license, it also requires special training. The third one is **Online Dashboard** – in this category, law enforcement agencies outsource the crime mapping duties from agencies outside the law enforcement. Agencies like Bair Analytics can provide crime mapping services at a nominal fee. The main advantage of outsourcing is that it reduces the workload of geocoding and data organization. It also provides better data storage and security as data is store via cloud-storage.

2.3 Cloud Computing

The other technology that has emerged, developed so fast and contributed to the spatial data technology is cloud computing. The evolution of GIS technology and of spatial information acquisition technology have led to more and more collection of spatial data through various approaches for different services like emergence services, crime mapping and other reasons, this has caused the demand for high information processes and computing environment [16]. Cloud computing has emerged as a technology primarily focusing on large scale resource sharing and low cost for big data storage technology. According to [17] in their paper they proposed an approach which employs cloud-based service to solve the big spatial data technology in emergence management for better spatial analysis. They further described cloud services as the form of centralized web-based applications, where applications and files are hosted on a “cloud” consisting of thousands of computers and servers, all linked together and accessible over the Internet. Mwansa and Phiri in their paper [18] a model for the inventory system based on quick response and cloud computing integrated with mobile application for real-time capture of grain bags brought in by farmers at the setline depot. The results of the system showed an improved and acceleration of grain stock statistics in real-time.

2.4 Related Works

Most Police stations and other law enforcement agencies in developed countries like USA and UK already migrated from traditional pin maps to computerized crime mapping systems [19]. Tong in her paper [20] also proposed a crowdsourcing based crime mapping system, it's a system based on cloud architecture integrated with iPhone mobile application. The system composed of a server running in the cloud and a client application that includes a website and iPhone mobile application to enable users to interact with crime contents. The system gathered both crime reports from the crowd together with crime contents and displayed them on a crime map. Users of the system are able to review and comment on crime incident on a digital crime map.

Singh et al. [21] in their paper proposed and developed a cloud GIS crime mapping blog that could be used by the police for crime mapping. The system generates daily, weekly and monthly crime maps which would help to identify crime patterns and clusters. The crimes in the blog are categorized into theft, murder, snatching and other classes, the system was developed on a cloud architecture using google cloud and google map as source of spatial data. The system uses RSS (Really Simple Syndication) feeds from various news websites as a source of crime data, meaning that crime related news is manually extracted from the collection of news and formatted into desired format with attributes like location, type of crime, details and link to the news web page. The location data is further transformed into georeferenced data, the process commonly known as geocoding. Geocoding is described as the process converting or transforming a description of a location such as pair of coordinates, an address into a location on the earth's surface and used to point a location in the GIS digital map [22]. The limitation of this system is that the location of the crime specified in the news is generalized therefore the generated point location is not very accurate.

Some of the solutions provided by applications in the related works would be of great benefit if adopted in the crime mapping model for the Zambia Police. The source of crime data for the proposed system is the live crime reports from the general public, to achieve this a crime reporting component would be added to the system consisting of a web application and mobile application adopted from Mwangala et al [23]. In order to enhance accuracy, users would specify crime location by selecting a name of the location from the google map, then the system saves the georeferenced data into the firebase cloud database. The benefit of this approach is that it does not require crime data to be geocoded to produce crime maps as everything would be done by the system.

3. METHODOLOGY

We reviewed existing literature on crime mapping related technologies used for mapping crimes. All the data in the literature review was from published secondary sources like journals.

3.1 Baseline study

The purpose of the baseline study was to identify challenges in the current crime mapping system used by the Zambia Police. To archive this, different materials & methods were used;

- A. *Data Collection:* Primary data was collected from police officers based in Lusaka working in the CID and VSU departments through the use of interviews & structured questionnaires. Further, general members of public were also interviewed through self-administered questionnaires. Open-ended and closed-ended questions were included in the questionnaires to capture both the qualitative and quantitative responses.
- B. *Population & Sample:* The population considered for this study was the Lusaka based police officers. Purposive sampling method was used to select 88 police officers from ten police stations across Lusaka District. Further, a combination of random and

convenient sampling method was used to select members of public who were visiting the police stations for various reasons.

- C. *Data Analysis & Presentation:* Data was entered, analyzed and presented using the Statistical Package for Social Scientists (SPSS) version 16. The analysis was mainly descriptive in nature.

3.2 Model Design Methodology

The proposed model was designed using Cloud Architecture, Android Mobile Application, Web Application, Google Map API and java Programming Language.

3.3 Mapping Business Processes

Fig.1 shows the crime mapping business processes for the current crime mapping system



Fig.1 Current crime mapping business processes.

As shown in fig.1 above, crime mapping process begins at the time when a crime is officially reported by a member of public. The case including all the details of the complainant and suspect if any are recorded into an occurrence book by an officer on duty at the front desk known as inquiries. Thereafter the case is forwarded to the CIO for review, approval and assigning of case to an investigator. The investigating officer investigates & analyze a crime by placing a pin on a geographical map depicting its location. The investigating officer forward the case details to the Records/Statistics officer who enters the case into the crime register for record purposes. The records/Statistics officer generates crime statistics and share the information with an investigation officer and others.

The Proposed automated crime mapping business processes are derived from the current business processes presented in fig. 1. The proposed model is designed in two parts, the mobile application to be used by the general public to report crimes and also the Web application to be used by the police to not only capture and view crime reports but also generate crime statistics and crime maps. Fig.2 shows the proposed web application.

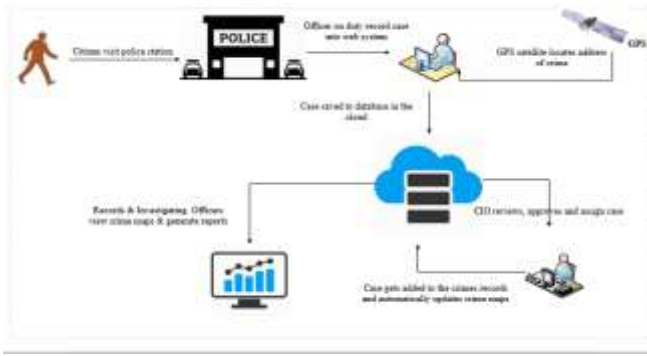


Fig.2 – Proposed Business processes - Web application

A crime is directly reported by a member of the public, the officer on duty records the crime case into the system, the GPS satellite through google maps captures not only the actual location of crime but also residential address of both the complainant and suspect, the case details including geo-referenced data will be saved into the cloud database. The case is automatically forwarded to the CIO for assigning, upon assigning the case to the investigator, the case is added to the crime register and automatically updates the crime map. Both the crime investigator & statistics/records officers will be able to view and generate crime maps & reports. Fig.3 below shows the proposed mobile application business processes.



Fig.3 - Proposed Business processes – Mobile application

The mobile application allows the citizen/user to report crime case in using a mobile device. The police will use the web platform to view the reported crimes and map them. The citizens are also able to view the status of the case they reported.

3.4 Proposed System Architecture

The proposed system architecture shown in fig.4 below utilizes the private cloud infrastructure where ZAMTEL the largest telecommunications company in Zambia provides the MPLS network while Zambia Police provides the private cloud services. The MPLS backbone comprises of fiber and microwave. The Zambia Police exclusively operates computing resources in the cloud in which different servers are installed such as Email server, Database server, Application server, Web server, Real-time communication server and many others. The rationale herein is to make sure the system is accessed by different police stations located across the country thus include the general public who are the mobile application users with less convenience of configuring hardware, software and security of information. The proposed architecture will allow Zambia Police to have complete control of the system including how data is managed and what security measures are to be put in place.

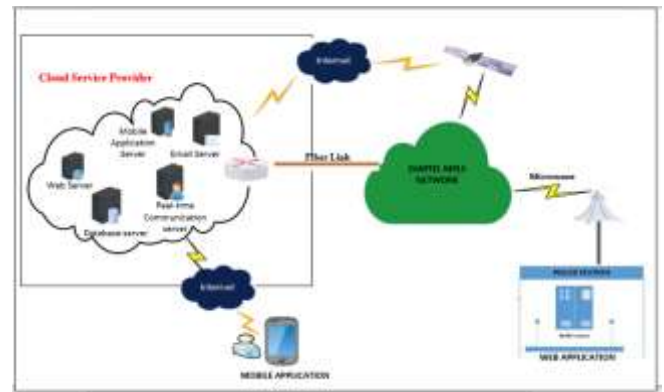


Fig.4 – Proposed System Architecture

3.5 System Modelling

The use case for web application is shown in fig.4 below.

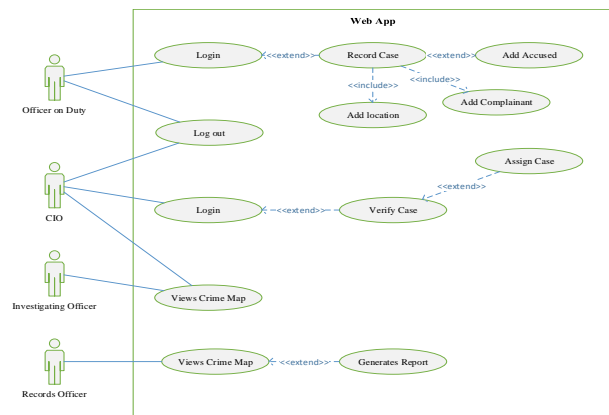


Figure 5– Use case for web app

As shown in fig.5 above, the main actors in the web system are; officer on duty, CIO, investigating officer & records/statistics officer.

The entity relationship diagram (ERD) is shown in fig.5 below.

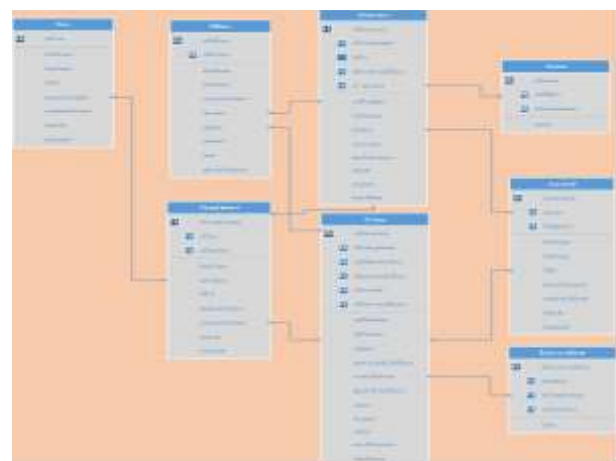


Fig.6 - ERD

The ER model diagram in fig.6 above shows the relationships between entities and attributes in the proposed system.

Fig 7 bshows the sequence diagram (SD) for recording a case.

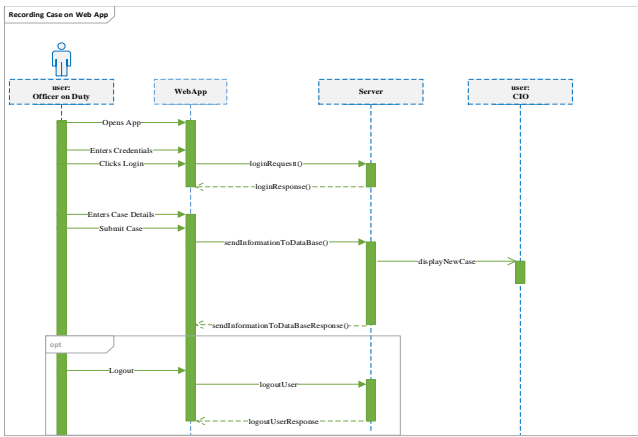


Fig.7 – SD -case recording

Fig.8 below shows a sequence (SD) for assigning of cases by the CIO.

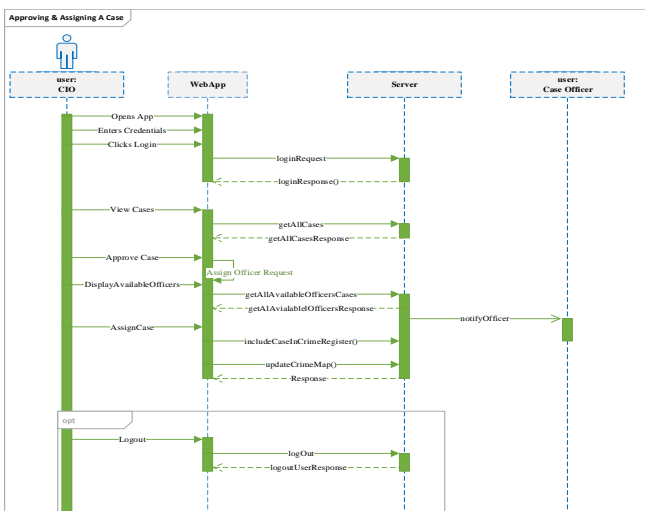


Fig.8 – SD -case assigning

4. RESULTS AND DISCUSSIONS

The results in Fig.8 show the levels of knowledge in computers among the police officers.

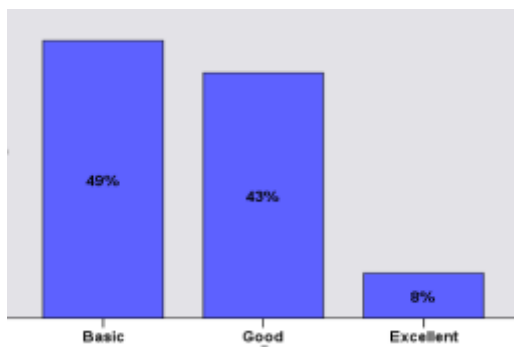


Fig.9 – Computer Knowledge

As shown in fig.9 above, all the officers interviewed indicated that they had at least some basic knowledge of computers, with more than 50% of the respondents indicating that their level of knowledge in computers was either good or excellent.

Fig. 9 below shows the usage of crime mapping technique in police stations.

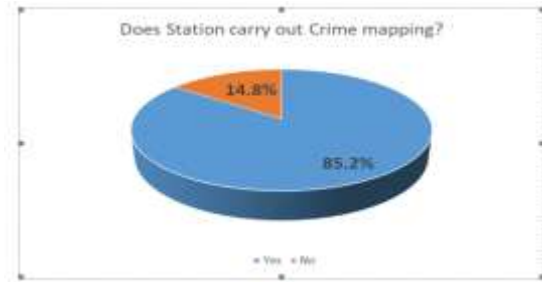


Fig.10 – Crime mapping usage

The results in fig 10 above shows that 85.2% of the stations conduct crime mapping using physical geographical maps and pins placed on the map while 14.8% indicated that they don't use any form of crime mapping technique. Some challenges indicated by respondents include ;non instant availability of statistics on areas prone to crimes and also it's not easy to analyze & manage crime data with manual maps.

Table.1 below shows the crime reporting methods.

	Responses		Cumulative frequency
	Frequency	Percent	
Does the public walk in to the police station when making crime reports?	88	75.9%	100.0%
Does the public make a call to the police when making crime reports?	28	24.1%	31.8%
Total	116	100.0%	131.8%

Table.1 – Crime reporting methods

The results in fig.9 above indicate that 75.9% of people walk to the station to report crimes. Only 24.9% make phone calls. In addition, all crime cases and statistics are manually recorded, processed and stored in books & papers.

The study further revealed in fig.11 that 86% of the general public own phones or other mobile devices that have access to internet. 72% of those have android operating system,14% have Microsoft windows while only 8% have apple as shown in fig.11.

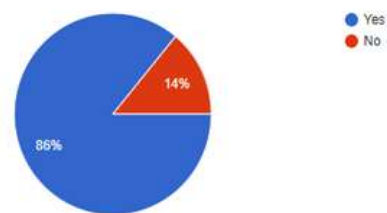


Fig.11 – Mobile devices

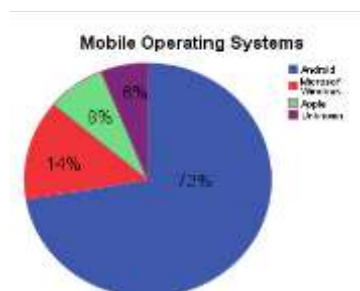


Fig.12 – Mobile operating system

Fig.13 below shows a screen shot window displaying sample spatial location of crime spots.



Fig.13 – Spatial crime spots

As shown in fig.13, the system is able to display spatial crime spots as they are being reported by the general public. Each crime spot shows name, location, date and time the crime was reported.

5. CONCLUSION AND RECOMENDATION

The study proposed a computerized crime mapping model based on cloud and spatial data to address the challenges in the current system used by the Zambia Police. Test results of the proposed system shows improved data visualization and reporting of crime data with reduced dependency on manual transactions. With police personnel that exhibited high levels of knowledge in computers, and a general population that is drifting towards the use of ICT, we recommend the adoption of the proposed model as it will improve the work efficiency within the Police Service.

ACKNOWLEDGMENT

We would like to thank the Zambia Police Service for the support and also for providing us with all relevant information towards this project. Special thanks goes to UNZA Department of Computer Science and Department of Electronics and Electrical Engineering for the advice and expertise rendered to this project.

REFERENCES

[1] A. Sighn and S. Baghel, "A Predictive Model for Mapping Crime using big data Analytics," *International Journal of Research in Engineering and Technology*, vol. 4, no. 4, pp. 343-348, 2015.

[2] H. Kumar, B.Sania and G.Mahujana, "Crime Prediction & Monitoring Framework Based on Spatial Analysis," in *International Conference on Computational Intelligence and Data Science*, Jaipur, 2018.

[3] V. Kappeler, "What is the Place of Police within the Criminal Justice System?," Eastern Kentucky University, October 2012. [Online]. Available: <http://plsonline.eku.edu/insidelook/what-place-police-within-criminal-justice-system>. [Accessed 5 November 2016].

[4] [correctionalofficer.org](http://www.correctionalofficer.org), "U.S Criminal Justice System," [correctionalofficer.org](http://www.correctionalofficer.org), [Online]. Available: <http://www.correctionalofficer.org/us-criminal-justice-system>. [Accessed 13 November 2016].

[5] Policy Brief, "The Criminal Justice System in Zambia," African Human Security Initiative, Pretoria, 2009.

[6] Department of Justice & Equality, "White paper on crime-The Community and the Criminal Justice System," Department of Justice and Equality, Dublin, 2011.

[7] Zambia Police, "Zambia Police History," Zambia Police, 2008. [Online]. Available: https://www.google.co.zm/?gws_rd=cr&ei=Pb0aWL3JJciUsgG3ka34Ag#q=history+of+zambia+police+service. [Accessed 29 October 2016].

[8] C. Chandrasekar and M. Kumar, "GIS Technologies in Crime Analysis and Crime mapping," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 1, no. 5, pp. 115-122, 2011.

[9] D. Wang and D. Wei, "Crime hotspot mapping using the crime related factors—a spatial data mining approach," Springer Science+Business Media, New York, 2012.

[10] D. Breetzke, "Geographical Information System (GIS) And Policing In South Africa- A Reveiw," *Policing: An International Journal of Policing Strategies and Management*, vol. 29, no. 4, pp. 723-740, 2007.

[11] K. Harries, *Mapping Crime: Principle and Practice*, Washington: National Institute of Justice, 1999.

[12] U. Argun and M. Dağlar, "Crime Mapping and Geographical Information Systems in Crime Analysis," *International Journal of Human Sciences*, vol. 13, no. 1, pp. 1-14, 2016.

[13] K. Hansgen, "Crime Mapping In Law Enforcement: Identifying Analytical Tools, Methods, And Outputs," *Culminating Projects in Criminal Justice*, vol. 1, no. 1, pp. 1-248, 20 October 2016.

[14] P. Kedia, "Crime Mapping and Analysis using GIS," *International Institute of Information Technology*, vol. 1, no. 1, pp. 1-15, 2016.

[15] J. Western and A. Murray, "Exploratory Spatial Data Analysis Technologies For Examining Urban Crime.," *Center for Crime and Justice Studies*, vol. 1, no. 1, pp. 1-21, 2001.

[16] Z. Liu and X. Li, "A Spatial Data Security Model Under the Cloud Environment," *Proceedings of the 2nd International Conference On Systems Engineering and Modeling*, vol. 1, no. 1, pp. 1-4, 2013.

[17] H. Wang and G. Xiaosan, "Cloud-based Service for Big Spatial Data Technology in Emergence Management," Henan Polytechnic University, jiaozuo, 2001.

[18] N. Mwansa and J. Phiri, "Real-Time Data Capture odel for Accelerated Payment of Small-Scale Farmers," *American Journal of Engineering and Applied Sciences*, vol. 11, no. 3, pp. 1164-1177, 2018.

[19] C. Mamalian and Nancy. N. LaVigne, "The Use of Computerized Crime Mapping by Law Enforcement: Survey Results," USA National Institute of Justice, Washington, 1999.

[20] H. Tong, "A Crowdsourcing Based Crime Mapping System," Waseda University, Tokyo, 2014.

[21] H. Singh, R.Kumar, A.Singh and P.Litoria, "Cloud GIS for Crime Mapping," *International Journal of Research in Computer Science*, vol. 2, no. 3, pp. 57-60, 2012.

[22] D. Mazeika and D. Summerton, "The impact of geocoding method on the positional accuracy of residential burglaries reported to police," *Policing: An International Journal of Police Strategies & Management*, vol. 40, no. 2, pp. 2-15, 2017.

[23] M. Mwiya, J. Phiri and G. Lyoko, "Public Crime Reporting and Monitoring System Model Using GSM and GIS Technologies: A Case of Zambia Police Service," *International Journal of Computer Science and Mobile Computing*, vol. 4, no. 11, pp. 207-226, 2015.

A Spatial Framework for Managing Sewer and Water Networks Using Sensor Networks: A Case of the University of Zambia

Gabriel Chibuye
*Department of Geography, School of
 Natural Sciences
 The University of Zambia
 Lusaka, Zambia
 gab.chibuye@gmail.com*

Jackson Phiri
*Department of Computer Science,
 School of Natural Sciences
 The University of Zambia
 Lusaka, Zambia
 jackson.phiri@cs.unza.zm*

Faustin Banda
*Department of Geomatic Engineering,
 School of Engineering
 The University of Zambia
 Lusaka, Zambia
 faustin.banda@unza.zm*

Abstract— The University of Zambia lacks accurate up to date locations of manholes, sewer lines and water lines which makes maintenance, management and planning of these utilities very difficult. Therefore this study attempts to build a framework for Managing Sewer and Water Networks Using Sensor Networks. A Web-based GIS or WebGIS application was used to develop the spatial framework. WEBGIS is a powerful mapping and analytical functionality expressed within a web browser. Using internet of things (IOT) technology, water level sensors were placed on some critical manholes to help with the monitoring of the Sewerage network. Utility coding and tagging was done using QR code which hugely helped in the identification of all the infrastructure. The QR codes could be scanned using any QR code reader and could provide attribute data including the location for easy identification for any field personnel.

Keywords— GIS, WEBGIS, Sensors, Internet of Things, Coding, Tagging, QR Codes

I. INTRODUCTION

The University of Zambia is rapidly expanding as observed from the new developments all around campus [1]. The growth naturally necessitates the acquisition of new properties and development of University infrastructure. One of the key areas of focus for managing the University facilities is optimal use of utility spatial information. In particular, water sewerage utilities are essential for the smooth running of a university. Water provides inevitable sustenance to life and sewerage systems provide a means of discarding waste water. As such, it is essential for water utilities to be managed well.

Most of the water utility spatial information used in managing the University facilities have been primarily paper based maps. These hardcopy maps and documents have become largely inaccurate because they have not been updated in a very long time. The maps have become old and inaccessible over time as most maps are torn and are now lost. These hardcopy maps are no longer appropriate for real time decision making because they are unable to act in response to changing circumstances. The field technicians locate most lines and other facilities using memory and experience. Consequently, there is need to introduce better systems that are more efficient in spatial data organization, manipulation and visualization.

The aim of the study was therefore to build a spatial framework for managing the existing sewer and water networks at the University of Zambia. There first objective was to conduct a baseline study of the existing sewer and water networks. The second objective was to design a framework for the GIS sewer and water networks integrated with sensors based on the findings in the first objective. The final objective was to develop a prototype based on the framework developed in the second objective using the web, mobile and sensing technologies.

The Research Questions formulated were in line with each objective. For objective one the research question was: Where are the water and sewer networks located and how are they currently identified around the University? In objective two the research question was: Can a framework be designed for the findings in the first objective to monitor, manage and integrate the utility system with sensors? Finally the third objective question was: How can a prototype of the framework designed in the second objective be developed?

A Web-based GIS or WebGIS application was used to develop a spatial framework. WEBGIS is a powerful mapping and analytical functionality expressed within a web browser. The web application had some tools imbedded in it including; query, measure, add point/vertex, zoom, view, search, identify and update. Using internet of things (IOT) technology, water level sensors were placed on some locations to monitor some manholes of the Sewerage network. Utility coding and tagging was done using QR code. The code was used to uniquely identify all the structures.

II. LITERATURE REVIEW

The literature reviewed involved looking at the various geospatial and other relevant technologies that could be of use in water and sewer networks. These included aspects of Land Surveying, Traditional GIS and Mapping, Web Mapping, Spatial Database, QR Code technology and Internet of things technologies.

A. GIS AND OTHER TECHNOLOGIES

Geographical Information Systems (GIS) technology offers combined power of both geography and information systems an ideal solution for effective management of water and sewer utility infrastructure. The effective management of water utility network can be possible by proper

representation and analysis of network data [2]. The most important applications of GIS for water utility management are mapping, monitoring, modelling, infrastructure planning, maintenance, water conservation and response to emergencies [3].

Much recent attention in GIS has been focused on developing GIS functionality in the Internet, Worldwide Web, or a private intranet, which is sometimes termed WebGIS. It hosts traditional GIS functionalities on the internet web, coupled with the powerful ability to integrate information and tools from multiple sources.

Web Application development uses both off the shelf software packages (ESRI DesktopGIS) and Free Open Source Software (FOSS) i.e geospatial platform for building WebGIS applications across web browsers, desktops, and mobile devices. The softwares that may be adopted include database server, a front-end comprehensive database design and management system [4], a map server to link the database to the client and acquires and processes requests coming from the user and return output results. An app for displaying dynamic map data in web browsers from multiple sources [5]. In addition an open source sever implementation of the Java Servlet, JavaServer Pages, Java Expression Language and Java Websocket technologies [6].

Internet of Things is the system of physical objects or things hooked up with hardware, software, sensors, and system connectivity which empowers these objects to gather and alternate information. IoT makes use of different kinds of protocols to work with exclusive objects [19].

Wireless Sensor Networks are part of the perception layer found in the Internet of things [7]. Wireless Sensor Network requires multiple nodes to form a network [8]. The sensor nodes continuously or at set intervals sense data from the environment and send the data to the sink node [9]. The sink node collects the data from sensor nodes for viewing locally and sending for storage. The data from the sensor nodes can be accessed remotely over the Internet or other means by users [10]. The wireless sensor networks provide better methods of monitoring environmental conditions than manual methods [11]. Wireless Sensor Networks are now a reality with applications in Smart Grids, Smart Environments and machine to machine communication.

B. Related Works

Most of the related works that haven done combine various technologies to develop frameworks and prototypes. However not much has been done to combine utility mapping with sensors and addition of the QR Codes to the web applications.

Nickolas Okello [12] developed a GIS based decision support tool for effectively and efficiently managing the water utility networks at the Copperbelt University using PostgreSQL/PostGIS, QGIS, GeoServer and GXP template built on GeoExt and OpenLayers. The goal was to have a centralized system with easy access to accurate and updated information on water utility network spatial information

throughout the university by means of the web. In his findings he sufficiently demonstrated the application's ability to act as a decision support tool for the end-users and decision makers in the management of water utility networks at the university.

Zambia's agricultural sector through Food Reserve Agency (FRA) while still underdeveloped faced many challenges that ranged from marketing, spoilage, infestations, and theft at site, spillage and storage among others. The methods used by FRA in their business processes were largely manual as there were no systems in place. In order to help curb these problems, Chihana [13] proposed and developed novel methods that could be used to sense real-time warehouse intrusion and grain tracking within the FRA circulation. The IoT based prototype model made use of the APC220 transceiver, GSM, GPRS, RFID, PIR and cloud storage. To curb theft of grain at storage points, the system used motion sensing through the use of PIR sensors, wireless radio communication module and the GSM/GPRS technologies such that when anyone comes in the range of PIR sensor, then the sensor will send a logic signal to the microcontroller. Lastly, the RFID combined with GSM and Arduino microcontroller responsible for grain tracking. From the results obtained in the experiment conducted it was believed that once this technology was adopted, thefts would be reduced and grain management in the FRA satellite Depots dotted around the country would improve.

Chilela [14] developed a WebGIS at University of Houston in United States of America. It was capable of saving, organizing, and geo-spatializing information from all over the campus and facilities. The WebGIS could process data and manipulate devices connected to the internet using concepts and tools of Internet of Things and provide the number of people inside of buildings. The information about the number of people inside of buildings was a new technology known as smart door. In case of a public business building (bank or shopping for instance), the WebGIS was capable, for example, of showing the average time that a client would spend in that building to be attended. In addition, the study provided a mobile application to allow users to interact with smart campus and campus facilities and to facilitate access to basic information about public transportation and data sensors which had been placed in various locations around campus.

Mwansa and Phiri [15] proposed a model for the inventory system based on Quick Response (QR) and cloud computing for real-time capture of grain bags brought in by farmers at the satellite depot for the Food Reserve Agency (FRA). The government supplied farming input to local farmers and bought the grain back from the farmers. The study looked at part of the buying process which required the movement of grain from the local farmer to the government. The proposed system first required tagging the grain bags then capturing the details of the farmer and attached this to the grain bags. Their proposed model based on cloud technologies was integrated with the mobile application used to read the QR code attached to the grain bags. These details were then linked to the details of the farmer in the database. These

captured data regarding the farmer and grain bags supplied at the satellite depot were made available to the decision makers in real-time. The results of the study showed that the proposed model would help to address a number of challenges that the current system had been facing. These included accelerated process of paying the local farmers supplying grain to the government which used to take months. It would also help to give the grain stock statistics in real time per region and the country at large. This model would be very useful for most developing countries in managing their grain.

In order to introduce modern warehousing, improve upon the storage of grain and grain marketing business processes for the Food Reserve Agency in Zambia, Chibuye [16] developed a prototype of a remote sensor network and built as a proof of concept for a much wider deployment using cloud computing and the internet of things concept. It was determined that a wireless sensor network would aid the Food Reserve Agency in analytics, timely action and real-time reporting from all its food depots spread-out throughout Zambia. Google's Android Things Platform was used in order to achieve the objectives. Advantages of Android Things over traditional platforms that had been used to develop wireless sensor networks were looked into and presented in this paper.

From rainfall data, patterns and topography Fujitsu Technologies [17] were able to come up with the best locations for placing sensors in a sewer network. They developed a technology that uses ICT for lowcost detection of early signs of sewer system overflows in order to mitigate damage in cities stemming from torrential downpours. The technology for determining the locations and number of manholes where sensors should be installed, based on an analysis of the time required for water to flow from upstream pipes to downstream locations in accordance with land topography and the shape and length of sewer pipes, which makes it possible to track and predict the overall flow through a sewer system using only about one-fifth as many sensors.

Mutale [18] proposed a document archiving system which would integrate barcoding, time stamping and mobile technologies to improve on the achieving and retrieval processes. The study began with the baseline study in an effort to establish the challenges faced by institutions of learning in document archiving. The study was based on the University of Zambia. The results from the study indicated that 70% of the offices used the manual box file system in Document achieving. They also lacked storage space for box files with 80% having difficulties in retrieving old documents. Based on this study, Mutale designed and implemented a document archiving system. The system was developed in C#. It had time stamp integrated with short messaging system (SMS) for reminders and barcode which was electronically stapled on the soft copies before being archived to improve on the identity of the document.

The study used some of the methodologies from the various works outlined and the framework developed involved combining some of these methodologies.

III. METHODOLOGY

The methodology involved the understanding of the spatial distribution of the utilities installed cross campus. This involved collection of data from the Resident Engineer's Office and interviews of key personnel. The designing of the framework was done based on the literature that was reviewed. The prototype has been designed and developed based on the framework and it will be used by the Resident Engineer's office for monitoring, management and planning.

Equipment and Materials used; GPS GNSS RTK Set (High Target V30), Total Station, Tablet PC, Laptop, arduino kit, Water level Sensors and accessories, tags.

Software; GPS Essentials, Ht Target Software, Microsoft Office (Word, Excel, Access, Powerpoint), Mapping Tools (ArcGIS, QGIS, Bing Maps, SASplanet), Automation Tools (Web and Web Mapping Development Apps, Database Development tools, Sensor integration tools)

A. User Needs and Data Collection

Interviews of key personnel of the RE's Department was done to ascertain the user needs.

Data Collection tools included the following, Records and Secondary Data, Field Observations, and field Surveys and Maps. Figure 1 shows the summary of the workflow.

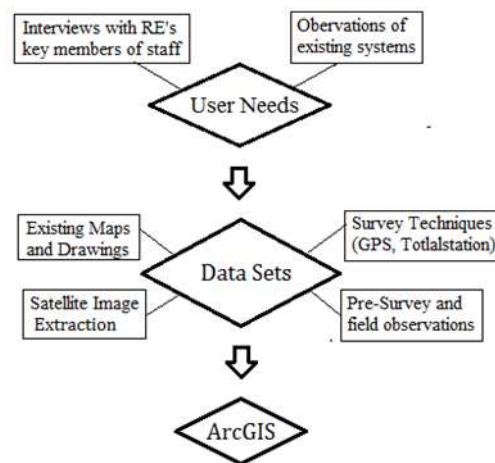


Figure 1: Data Collection Workflow

The collected data in figure 1 was input into ArcMap. Field Survey data was processed prior to its entry in arcMap and was combined with the rest of the data. Images were georeferenced using the surveyed field data, Digitizing was further done to create new shapefiles of roads, buildings, water lines, Sewer lines, water tanks, fire hydrants and all other existing infrastructure.

B. Development of the Framework

Designing of the framework began with the conceptual Design of the system, a summary is shown in figure 2. From the data collected in objective 1, point, line and polygon features were created. A 15 digit code was formulated and each feature was given a unique serial number which was scanned to generate QR Codes.

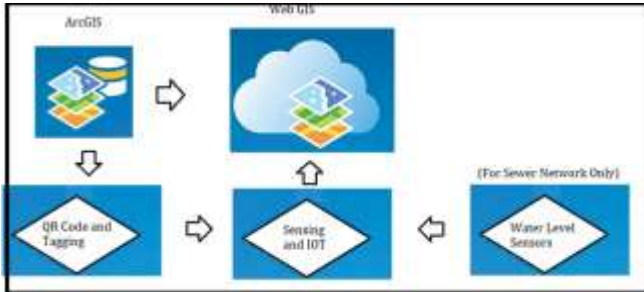


Figure 2: Proposed Framework

The created shapefiles were loaded into a spatial database PostgreSQL and using PostGIS the dataset were connected to geoserver for uploading on to the web app. The layers were customized and functionalities were added using open layers and JavaScript library tools. The figure 3 shows the flow chart of the framework leading to the proposed prototype;

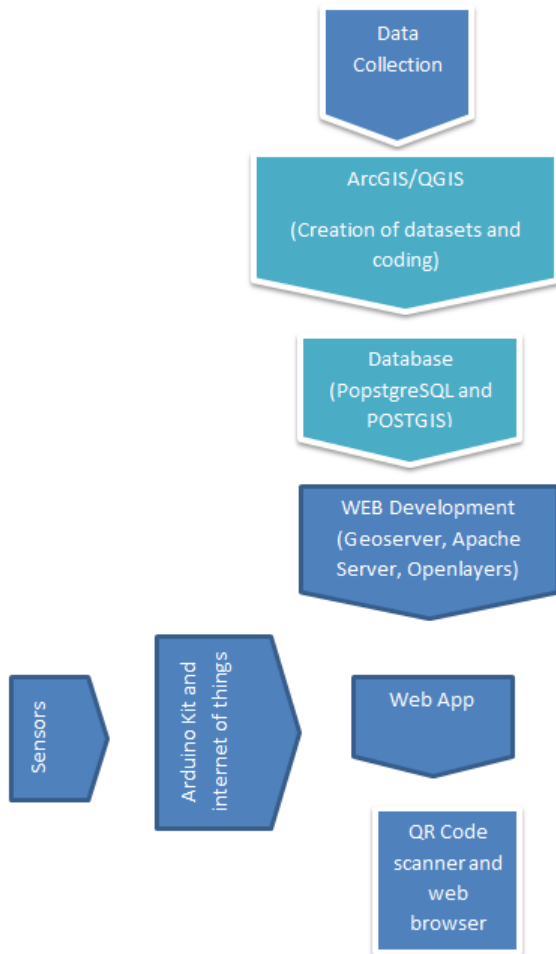


Figure 3: Flow Chart of the Methodology

C. Development of the Web GIS Prototype



Figure 4: Water and Sewer Network Systems WebGIS Prototype

In the prototype development, the structures will all be tagged with the generated QR Codes. A user with a mobile device and QR Code reader can scan the code on any tag to know which structure is being looked at by clicking the link which would open a web map app where more details can be obtained. For the Web app the user can view the whole network immediately the page is opened. In both mobile and web browsers, there will be login requirements before any user can proceed. For some manholes, water level sensors will be installed to monitor the water level. The sensor will send data to the arduino kit which will send data to the web app as an alert. Figure 4 shows a pictorial summary of the prototype.

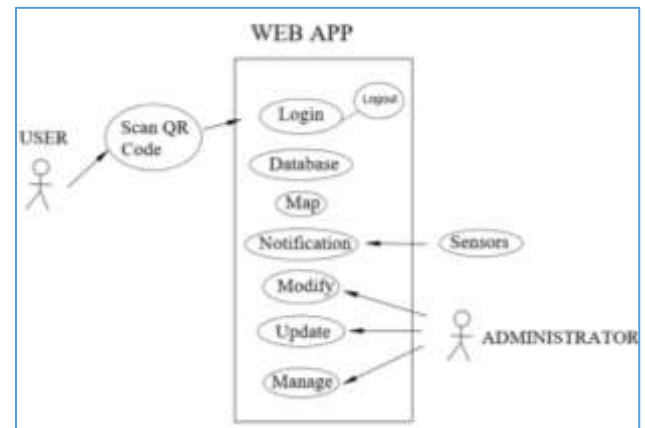


Figure 5: System Use Case Diagram

Figure 5 shows the use case diagram of the process. The user scans the tag with the QR code using a phone, logs into the system (mobile or web) is able to view the map and details of the structures. Sensors send notifications for any blocked manhole. Administrator has additional privileges of modifying and editing any structure.

IV. RESULTS AND DISCUSSION

The data collection began after the research problem was defined and research design/ Plan had been formulated. The various data collected were processed in arcMap and more datasets were created for different features. The field survey data was first processed in order to be used in ArcMap. The images were georeferenced and new shapefiles were created with the layers shown in figure 6 and 7 in relation to the updated water and sewer networks.

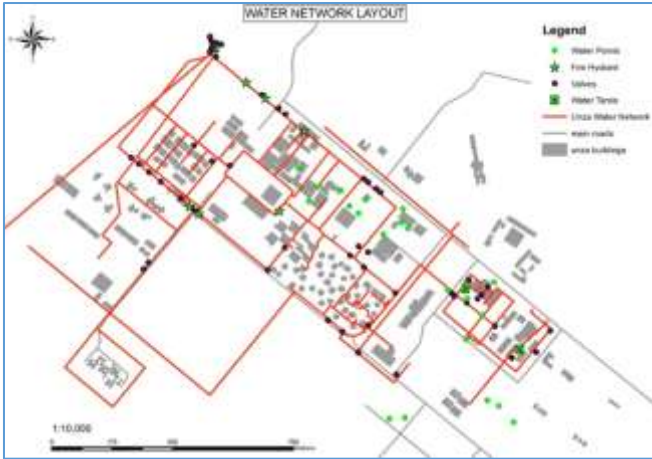


Figure 6: Existing Water Network

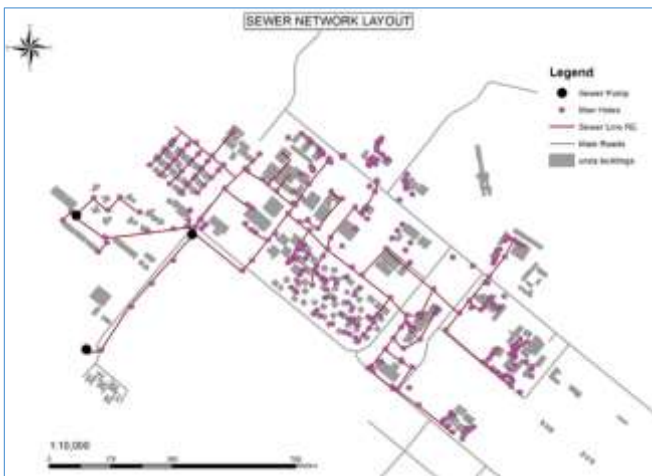


Figure 7: Existing Sewer Network

More attribute data was added for each layer obtained in objective 1 for both water and sewer networks. There were basically two types of features i.e. nodes and lines. Nodes included point data like manholes, water tanks, water valves; fire hydrants. Lines included line features which were sewer lines and water lines. Each feature was given a unique code, line segments were also given a tag including their starting point and finishing point.



Figure 8: Attribute table and layout of connected network

The tagging system that was used had 15 digit code e.g. for a manhole RE01MHS10000001; a brief description of the coding system is as shown in figure 9:

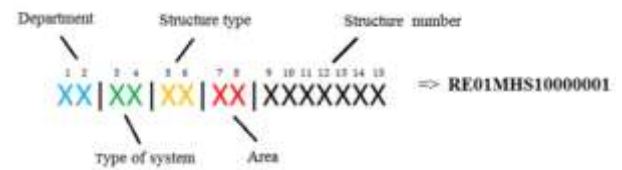


Figure 9: 15 digit serial number

Table 1 is a summary of the possible codes of the various features (manholes, water tanks, water valves, fire hydrants, water pipes, sewer lines (pipes);

Table 1: Codes for Features of the Water and Sewer Network:

Digit	1 and 2	3 and 4	5 and 6	7 and 8	9 to 15
Name	Department	Type of System	Structure	Area code from figure 13	Actual number of the structure
Possible Codes	RE – Resident Engineer	01 – Sewer System, 02 – Water System, 03 – Storm Water	SL – Sewer Line, MH – Man Hole, WL – Water Line, WT – Water Tank, FH – Fire Hydrant, BH – Bore Hole, SP – Septic Tank, WV – Water Valve	S1, S2, S3 and S4	000001 to 999999

QR Code Generation - The codes were input into an online QR code Generator which produced png images of each QR code; The QR codes stored the unique ids for each node that in turn would have a link to the web map once a user clicks it. The figure 10 shows an example of three manholes' generated QR Code.

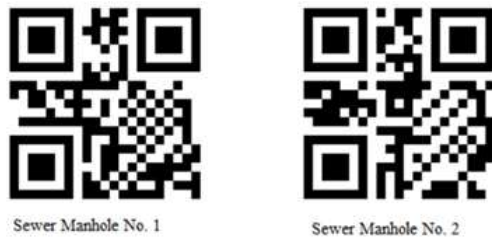


Figure 10: Examples of QR Codes for some Manholes

The shapefiles created from ESRI’s ArcGIS were exported to the PostgreSQL/PostGIS database. The open source QGIS was installed and used to create semi-interactive digital thematic maps using spatial data from the geo-spatial database created using PostgreSQL/PostGIS. The web application used PostgreSQL, PostGIS, Geoserver, OpenLayers and Apache server. A WebGIS portal was created which was the main access for users to the university GIS utility data. Access to data and services has been designed to require a user name and user password. The password has to be alphanumeric with a minimum of six characters. Access can also be obtained from scanning the QR code which gives the name of the structure as a link and it takes the user to the portal which in turn displays the map.

The water level sensors were installed on a few manholes and using Arduino Development Kit and internet of things, the sensors were connected to the web app. The series of sensor network would be connected to a personal area network and an aggregator that has a GSM shield. This GSM shield will support the common mobile network technologies such as GSM, 3G, LTE. The GSM shield is connected to an Arduino aggregator and an operating system will be used in order to communicate with the shield and upload data to a remote server. A cloud based service was used as a broker to gather all the information from the remote sensors, the most popular open source broker being the Mosquitto server was used. The sensor was able to send notification to the web app and the user could tell which manhole is affected prior to going to the field. The users could log in and view any information required for each structure. Administrator rights were also included to allow top management to be able to edit, update and manipulate the data as required.

V. CONCLUSION

The WebGIS tools were used to build a working prototype of the spatial framework for the University of Zambia. Data was collected from various sources, analysed and processed and finally used to build a spatial Database using PostGIS and PostgreSQL and a Web Map. Using internet of things sensors were integrated into the web based framework after coding and tagging was done for each feature using QR coding system. It is envisioned the developed prototype will ease operations for the Resident Engineer in terms of maintenance, management and overall workflows.

VI. RECOMMENDATIONS

The system can be greatly improved if accurate elevations of the water and sewer lines are known. At the moment the accurate locations are in the horizontal i.e. X and Y. A

topographic survey of the entire corridor of the sewer and water lines could be done and an accurate digital terrain model can be generated and uploaded to the system hence improving the surface accuracy. In addition a scan of the lines of the water and sewer lines can be done using Ground Penetrating Radar (GPR). This would give precise elevations of the top of the pipes which can also be included into the system. Accurate information of both the round elevation the pipe elevations can help when making new installations and can prevent damage of pipes as a result of any excavations within the University.

VII. ACKNOWLEDGEMENTS

We would like to thank the University of Zambia Resident Engineer’s office for their assistance, Department of Computer Science, School of Geography and School of Engineering for the advice and expertise rendered in this study.

REFERENCES

- [1] UNZA PRESS - The University of Zambia, “Strategic Plan 2018-2022” The Strategic Planning Office of the University of Zambia. P.O. Box 32379, LUSAKA 10101, ZAMBIA, 2018.
- [2] GIS Landmark, “Utility Mapping” Accessed [online] <http://www.gislandmark.com/utility-mapping>. Retrieved on 5th April, 2019.
- [3] CPHEEO, Government of India, Ministry of Housing and Urban Affairs. “Manual on water supply and treatment”. Accessed [online] <http://mohua.gov.in/publication/manual-on-water-supply-and-treatment-systems-cpheeo-1999.php>. Retrieved on 22nd March, 2019.
- [4] PostgreSQL. (2018). About: PostgreSQL. Retrieved June 10, 2018, from PostgreSQL: <https://www.postgresql.org/about/>
- [5] OpenLayers. (2018). Home: OpenLayers. Retrieved July 20, 2018, from OpenLayers: <https://openlayers.org/>
- [6] Tomcat-Apache. (2018). Home: Apache Tomcat. Retrieved July 28, 2018, from Apache Tomcat: <http://tomcat.apache.org/>
- [7] Tzounis, N. Katsoulas, T. Bartzanas and C. Kittas, “Internet of Things in agriculture, recent advances and future challenges,” *Biosystems Engineering*, vol. 164, pp. 31-48, 2017. <https://doi.org/10.1016/j.biosystemseng.2017.09.007>
- [8] M. Hossain, M. Fotouhi and R. Hasan, “Towards an Analysis of Security Issues, Challenges, and Open Problems in the Internet of Things,” *IEEE World Congress on Services*, 2015.
- [9] M. A. Mahmood, W. K. G. Seah and I. Welch, “Reliability in Wireless Sensor Networks: A Survey and Challenges Ahead,” *Computer Networks*, 2014.
- [10] O. A. Chase, J. F. Sousa de Almeida, J. R. Brito de Souza and C. T. da Costa Junior, “Sensory platform architecture for IN SITU monitoring the thermal comfort in rural environments – The case study at Federal Rural University of Amazonian, Brazil,” *Measurement*, Elsevier, vol. 58, pp. 294-300, <https://doi.org/10.1016/j.measurement.2014>.
- [11] M. F. Othman and K. Shazali, “Wireless Sensor Network Applications: A Study in Environment Monitoring System,” *International Symposium on Robotics and Intelligent Sensors*, Procedia Engineering, Elsevier, vol. 41, pp. 1204-1210 <https://doi.org/10.1016/j.proeng.2012>
- [12] Okello, N., Banda, F. & Tembo, E., “WebGIS for Water Utility Management”. *Imperial Journal of Interdisciplinary Research (IJIR)*, 3(8), pp. 1-6, 2017
- [13] Sipiwe Chihana, Jackson Phiri and Douglas Kunda, “An IoT based Warehouse Intrusion Detection (E-Perimeter) and Grain Tracking Model for Food Reserve Agency” (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 9, No. 9, 2018
- [14] Júlio Gabriel Chilela, “Web Geographic Information Systems (Webgis) For Smart Campus and Facility Management” *Mathematics of Faculty of Science and Technology at University of Coimbra, Portugal*, 2016.

- [15] Natasha Mwansa Kachingwe and Jackson Phiri, “Real-Time Data Capture Model for Accelerated Payment of Small-Scale Farmers” American Journal of Engineering and Applied Sciences 2018, 11 (3): 1164.1177 DOI: 10.3844/ajeassp.2018.1164.1177
- [16] Mulima Chibuye and Jackson Phiri , “A Remote Sensor Network using Android Things and Cloud Computing for the Food Reserve Agency in Zambia” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 11, 2017
- [17] Fujitsu Laboratories Ltd, “Fujitsu Develops Technology for Low-Cost Detection of Potential Sewer System Overflows” Accessed [online], <http://www.fujitsu.com/global/about/resources/news/press-releases/2015/0210-03.html>, Retrieved on 15th June, 2018.
- [18] Brian M. Mutale and Jackson Phiri (2016). “Web Based Document Archiving Using Time Stamp and Barcode Technologies – A Case of the University of Zambia” International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol. 5, Issue 4, April 2016.
- [19] Hazael Phiri, Douglas Kunda and Jackson Phiri. “An IoT Smart Broiler Farming Model for Low Income Farmers” International Journal of Recent Contributions from Engineering, Science & IT (iJES) – Vol. 6, No. 3, 2018, <http://www.i-jes.org>, 2018.

Enhancing Security of Examination Question Papers Through a Tracking System Based on Spatial and Cloud Technologies

Banji Milumbe
 Department of Computer Science
 School of Natural Sciences
 The University of Zambia
 Lusaka, Zambia
 banjimilumbe@gmail.com

Jackson Phiri
 Department of Computer Science
 School of Natural Sciences
 The University of Zambia
 Lusaka, Zambia
 jackson.phiri@unza.cs.zm

Mayumbo Nyirenda
 Department of Computer Science
 School of Natural Sciences
 The University of Zambia
 Lusaka, Zambia
 mayumbo@gmail.com

Abstract— *This study first investigated the transportation of examination materials in relation to examination malpractices and leakages. Based on the findings, the study explored the use of GPS, GPRS/GSM and cloud technologies to enhance the distribution process of examination question papers and answer scripts. We proposed a GPS tracking system which would ensure real-time tracking and monitoring of the examination distribution process. The main advantage of this system is that the transportation vehicles will always be tracked and remain monitored such that if there is any holdup, GPS tracking allows immediate action to be taken. Also since GPS allows constant updates in real-time, any sort of uneven event happening, the main station will be informed immediately. This helps timely action to be taken and also eliminates human interventions or rather depending on humans to report any uneven event. The baseline study indicated that the challenges faced in the distribution process was mainly lack of reliable transportation vehicles which accounted for 67% of those who did not have any designated vehicles.*

Keywords—*GPS, GSM, GPRS, Google Earth, Tracking, cloud technology, style, examination malpractices, leakages*

I. INTRODUCTION

Efficient and effective management of the examination process has become a very important part of the service delivery due to the increasing demand in the qualification awarded by the Examinations Council of Zambia (ECZ). In the amended Zambian Constitution of 2016, a new requirement was added where all those vying for political positions should have a minimum of a grade 12 or school certificate [1]. This has raised the demand for the grade 12 certificate and hence the more need to ensure that the distribution process of examination papers is tightened through (remote monitoring using GPS tracker) vehicle tracking to help know in real-time or near real-time the location of the truck delivering examination papers. This would help in detecting and reacting quickly to any uneven situations and probably resolve the problems or at least minimise the damage [2].

II. BACKGROUND OF THE STUDY

Security is key to all types of organisations be it a profit making, non-profit making or governmental as having a secure environment would increase efficiency and reduce on losses and liabilities.

Ensuring that examination papers are securely transported from the source to the various destinations which are schools or examination centres needs to be addressed. In Zambia, the school examinations question papers and answer scripts are delivered to various district education offices as stipulated in [3]. The main concern is the security of these

examination materials during transportation to various destinations throughout the country i.e. in schools or districts as there is no way of tracking and monitoring in the process.

The ECZ, created by an Act of Parliament Number 15 of 1983 chapter 137 [4] [5] is responsible for conducting examinations and awarding certificates and diplomas to deserving candidates. The whole examination process includes preparing and distributing the question papers and answer scripts to the district education offices who in turn distribute to schools in their respective districts [3] using their own transport arrangements. The National Policy on Education, Educating Our Future, emphasizes on improving educational quality, access to good quality education, fair assessments of all learners, and upholding the education standards [6]. It is important that all learners should have a fair national examination at the end of their schooling thus making it more important to safeguard the examination materials. In addition, with the education system having been liberalized and decentralized in accordance with the democratic principles of local government, the focal point of educational administration and management for schools are districts education boards and centres [6]. The increase in the number of candidates, examination centres, districts and also examinations offered by the ECZ makes it difficult to physically monitor all examination centres and the manner in which they keep and transport the examination materials. Therefore, in order to improve on delivery of services in government, the National Policy on Information and Communication Technology gives a commitment to support the implementation of the decentralisation process by extensive integration and utilisation of ICTs at all levels [7]. This research is riding on this policy statement to implement an automated tracking and monitoring system for managing examination question papers and answer scripts during transportation.

A. Distribution Process

The distribution of examination materials to district education offices is done by the examining body under the auspices of the Ministry of General Education. According to [1] which states that ‘The Examinations Council of Zambia shall deliver question papers directly to the District Education Board Secretary’s (DEBS) Office under police escort’. This process of distribution also ties well with what Malawi National Examinations Board (MANEB) also does as indicated in the Malawi National Examinations Board Act (CAP 30:04) that ‘All national examination papers shall be delivered and kept securely at a distribution centre designated by the District Education Managers prior to delivery to examination administration centres. [8]. This kind of system that is dependent on humans only is bound to

abuse as the same people entrusted to secure the materials can abuse their authority to gain undue advantage for themselves or others. Just as [9] said ‘All security products are only as secure as the people who configure and maintain them.’ From this statement by [9], it means that these examination materials transported in this manner are only as secure as the people who maintain or keep them. Beyond that it becomes very difficult to know what happens as there are no systems in place to remotely monitor the distribution process.

The current distribution process shown in Figure 1 is such that ECZ delivers the examination materials to the DEBS’ office after receiving the same from the printing company. The Ministry of General Education takes charge of the security of the materials once they have been delivered by the ECZ and received by the District Education Board Secretary and of their further distribution to zones and schools, and transportation of scripts to marking centres.

Current Distribution Model

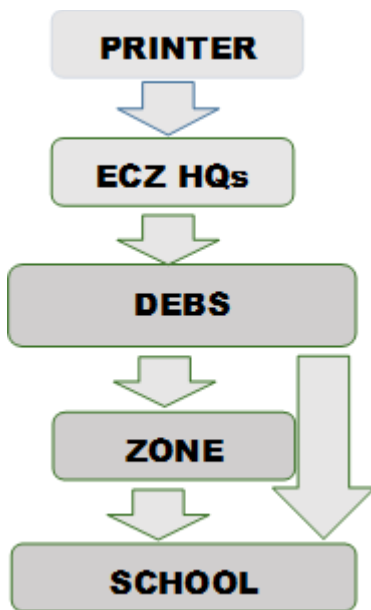


Fig. 1. Current Examinations Materials Distribution Process

Technology has advanced so much that it is not limited on where it can be applied thereby reducing on most of the manual interventions. In logistics management, use of GPS tracking and GPRS has been there for sometime now. In this research, we attempted to borrow some concepts and technologies used in logistics management to come up with a framework for managing, securing and monitoring examination materials in transit.

Why then are we talking about securing examination materials and not any other items? As already alluded [1] the desire to gain the academic qualification at school certificate or Grade 12 level has tremendously increased with the Law passed in the Amended Zambian Constitution of 2016, cap 70. (1) which states that ‘Subject to clause (2), a person is eligible to be elected as a Member of Parliament, if that person— (d) has obtained, as a minimum academic qualification, a grade twelve certificate or its equivalent’ [1]. This condition in the Law also applies to any other person seeking to be elected to political office as stated in cap. 82, 100 and 53 of [1]. This mandatory requirement has put the examination process in Zambia at a higher risk than before as

more and more people are trying to gain this qualification as they seek political office. The need to safeguard this qualification and ensure its credibility and international recognition has become even more important than before. As such ensuring the security of examination materials at all stages in the examination process is very critical. This underscores the need to have a system that will help monitor the transportation of these examination materials to respective destinations throughout the country. Hence the need to put stringent measures to secure the examination materials to prevent unauthorised access before and after writing the examination is important.

Reports on thefts of examination question papers and answer scripts [10] [11] [12], different forms of examination malpractices, among others are some of the challenges that have continued to occur in the examination process [13] in Zambia. Despite efforts to safeguard the examination materials put in place, these have proved to be unreliable as new ways of unauthorized access to examination materials have continued to emerge. Apart from the existing traditional way of securing examination materials through using security personnel and multiple locking system, there has not been a reliable system that would help monitor, track and ensure the security of examination materials throughout the examination process. The study evaluated and identified the current security challenges in the examination management process during transportation of examination papers. Based on the evaluation results, we introduced Global Positioning System (GPS) tracking, GPRS/GSM and cloud technologies to propose a framework for secure transportation, and tracking of examination materials in order to improve on the security of the examination materials during transportation in Zambia.

III. LITERATURE REVIEW

Various literature was reviewed to the proposed technologies used in this research. Such literature as the GPS technology, GPRS Technology, Google Earth, cloud technologies as well as a review of some of the reported incidences theft and loss of examinations while in transit.

A. GPS Technology

The GPS tracking has become a very essential part of any logistic business because of the advantages and facilities that it offers. GPS can provide extremely accurate location information for mobile objects and people which is far superior to earlier tracking techniques. As more devices become GPS enabled, accuracy will increase and the system's scale and global reach will benefit everyone [14] GPS consists of a network of 24 satellites in six different 12-hour orbital paths spaced so that at least five are in view from every point on the globe. [15]

B. GSM/GPRS Technology

GSM has been widely adopted and offers data services already but they have been constrained by the use of circuit switched data channels over the air interface. The General Packet Radio Service (GPRS), a data extension of the mobile telephony standard GSM, stands out as one major development in the GSM standard that benefits from packet switched techniques to provide mobile subscribers with the much needed high bit rates for bursty data transmissions. [16]

C. Google Earth

Google Earth is a three-dimensional software model of the earth. It is a commonly used tool to explore the geography of the world and displays satellite images of varying resolution of the earth's surface. Although it has limitations, it is a useful tool that will be used in this research together with google maps. [17] In our proposed system, we will employ Google Earth software and Google Map to show locations of transportation trucks.

D. Cloud Technology

Cloud computing a web-based technology where quality services are provided to users including data and software, on remote servers. [18] It provides clients with numerous capabilities like getting to an extensive number of uses without the requirement for having a permit, buying, introducing or downloading any of these applications. Clients can access data anyplace; all they require is to interface with a system (usually the Internet).

Cloud computing encompasses a whole range of services that can be hosted in a variety of manners depending on the nature of the service involved and the data / security needs of the contracting organisation. A cloud can be public, private or hybrid. A public cloud is owned and maintained by a cloud service provider while a private cloud is operated solely for a specific organization. It may be managed by the organization or a third party and may exist on the premise or off premise. [19]

E. Related works

Real-time and tracing systems were considered by [1] in tracking and tracing shipments. A sophisticated overview of the technology based methodology or approach for solving the complete tracking and tracing systems in the logistics and supply chain network was presented. It was noted that through the implementation of tracking and tracing, it is possible to detect and react to any uneven situations and probably resolve the problems or at least the damage can be minimised.

To improve on the distribution of examination question papers, [20] implemented a geographic information system (GIS) that was to perform all the tasks of the current manual system and in addition, provide functionality to aid in the efficient management of the Kenya National Examinations Council (KNEC) data. The study emphasised on efficiency in determining the shortest distances analysis and efficient distribution route determination were performed using spatial analysis and network analysis tools.

[21] developed a system that remotely monitors the movement and storage of examination materials. The Secure Exam Management system employs the use of uniquely designed electronic seals and remotely located Smart keys, as well as a centralised management software application, to not only manage the opening of examination scripts at the appropriate time before commencement of the examination, but also the time window within which examinations are to be conducted. It further geographically locates the examination venue and ensures that all opening of scripts and sealing of the answer sheets, can only be effected at these geographic locations which are predefined on the system.

A lot of work has been done that relate to tracking and remote sensing in different fields which this particular research will borrow from. Many researchers have come up

with different ways of remote tracking which can also be applied to tracking examination question papers and answer scripts in Zambia as we have seen how important these materials are and the security of such.

F. Security Breach during transportation of Examination Materials

A number of examination bodies have suffered loss of examination question papers or answerscripts during the transportation process. Such incidents on security breaches concerning examination question papers and answer scripts have been reported and are as many as those that cared to report about them Some of those reviewed indicate that in Brazil, [22] reported in an article entitled 'Truck Full of CFA Exams Hijacked in Rio Crime Wave' that a truck carrying CFA Institute exams was hijacked and all the cargo was stolen, the Brazilian branch of the CFA Institute said in a document obtained by Bloomberg, adding that chances of recovering the material were slim.

A theft was reported by [23] in Britain where Singapore Chemistry A-Level test papers were stolen in transit to British examiner. The theft affected more than 200 students in Singapore who had part of their British-administered school leaving exam paper stolen when they were sent to be marked in Britain, the Singapore exam board said. Again in Britain, a parcel van that was carrying AQA examination papers to the schools was stolen forcing them to reset the A-level and GCSE exam papers at the last minute after a van delivering the papers to schools was stolen. In 2008 AQA said it reset 40 papers in 15 subjects after a Parcelforce van in Hertfordshire was stolen. In 2005 the same board had to replace half a million exam papers after another Parcelforce van was stolen [24].

This phenomenon is not only in the outside countries, in Zambia a report 'Stolen exam papers shock NGO - Daily Nation' and 'Robbers steal G9 exam papers', it was reported that unknown criminals stole 2016 grade nine(Junior Secondary School Leaving Examination) examination papers on a containerised truck destined for the North Western Province of Zambia. It was reported that the incident happened in the night when criminals broke into the containerised truck that was ferrying the papers to Solwezi while in motion at Mushishima area in Chingola [25]. This is only for reported cases as there may be other unreported cases which may have been experienced both outside and inside Zambia. It is for this reason that this research embarked on using the existing technologies and apply to the examination management process. These technologies such as GPS tracking, GSM/GPRS and cloud technologies have been used in logistic management, vehicle tracking for fleet management to mention a few but have not been used in the examination management process in Zambia.

IV. SCOPE OF THE STUDY

The baseline study considered issues pertaining to the transportation of the examination materials and the security related aspects during transportation to various examination centres across the country. It was important to know whether schools or district offices had specific vehicles for transporting examination materials because that was going to help in determining a viable solution for secure transportation of examinations. This research utilised both qualitative and quantitative research types.

The baseline study was conducted in all the ten provinces of Zambia and 99 out of 115 gazetted districts at the time of data collection in June, 2019.

The questionnaire was administered to the respondents with the help of research assistants who visited the provinces. The researcher also visited two provinces and administered the questionnaire there. The returns were received from 296 respondents, out of which 17 were not properly filled in and were not included in the analysis. The remaining 279 that were fully completed were used for the study which made up 275 teachers, two (2) Education standards Officers and two (2) others who were Statisticians from the District Education Board Secretary’s office. Of the 279 participants, 199 were from ordinary standalone centres and 76 were Zonal Centres which are used as zonal centres for storage and distribution of examination materials and four (4) DEBS office.

Data obtained from the questionnaire were analysed using descriptive statistics of frequency counts and percentages. The data were analysed for statistical significance to establish the relationship found in the sample using the Chi-square tests and Cramer’s V (coefficient of association) to determine the strength of association between variables. The convention that we used in our research for the level of significance is $p < 0.05$ which is the maximum level of statistical significance that is acceptable and is widely used by most social researchers [26].

The respondents comprised 20.1 % females and 79.9 % males. The highest level of education ranged from Certificate to Master’s Degree of whom were Teachers (98%), Education Standards Officer (1%) and other profession (1%) drawn from ordinary school, Zonal school and DEBS office.

The results of the baseline study contributed towards the development of the proposed model that is using GPS tracking, GSM/GPRS and cloud technologies in the examination distribution process depicted in Fig. 3

A. Current process

The current distribution process begins at ECZ and examination papers are delivered to various districts accompanied by a police officer. ECZ delivers the examination materials to the DEBS’ office. The Ministry of General Education takes charge of the security of the materials once they have been delivered by the ECZ and received by the District Education Board Secretary and of their further distribution to zones and schools. The diagram in Figure 2 represents the existing distribution model for examination papers.

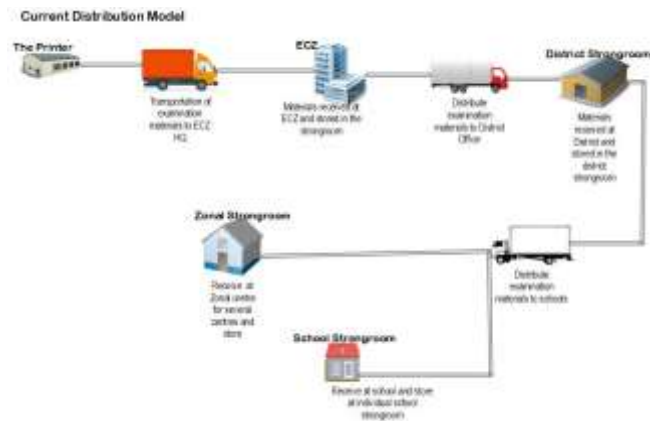


Fig. 2. Current Examination Papers Distribution Model

B. Proposed Distribution model

The proposed business process integrates technologies such as use of GPS for tracking the movements of the vehicle and determine its location, GSM/GPRS for transmitting data to the web server/ cloud. In this proposed model, delivery of examination papers should be direct from the printing company to the provincial strongroom. With the decentralization of ECZ where service centres are being opened in all the provinces, this could be used as receiving and distribution points of examinations in the provinces. The boxes / containers where examination papers are packaged should be tagged with RFID tags so that data can be captured about the cartons loaded onto the transportation vehicle. Data read from the RFID tags are transmitted to the server via GSM/GPRS. The distribution vehicles should be installed with GPS Tracker to enable monitoring and tracking during the whole distribution process. Putting RFID tags will ease the process of identifying the examination papers for each centre. The information about the trucks and examination papers will be sent to the webserver via GSM/GPRS. The proposed model is shown in Figure 3. These technologies have been used in logistics management as in [27]. We propose use a private cloud due to the sensitivity of the data. While the cloud server is prone to attacks and manipulation even if a private cloud is used, there is need to secure it using smart and efficient techniques to avoid attacks [28].

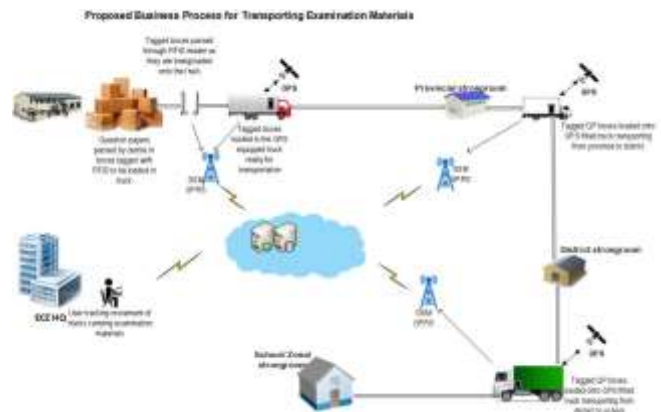


Fig. 3. Proposed model

System Architecture

The system architecture in Figure 4 shows the truck with GPS device to enable know its location in the distribution process. The user at the district monitors and tracks transportation vehicles for their district, the user at the province monitors and tracks movement of the transportation vehicles in their province while the user at ECZ-HQ tracks and monitors for the whole transportation process in the country. Any device with internet connection is expected to work.

V. RESULTS AND DISCUSSION

The focus of the study was to establish the causes of examination malpractices and leakages as well as understand the security systems that were existing in the distribution of examination materials. The results of the baseline study identified that among the causes of examination malpractices and leakages was due to poor security during delivery of examination materials (13.3%) which led to the incidence in

[25]. The research results confirmed that most of the schools did not have vehicles for use to transport examination question papers and answer scripts to and from the examination centres. This is because a bigger percentage of 67% said they did not have vehicles specifically designated for transporting examinations while 33.0 % said they had. Even those who had rely on vehicles provided by the DEBS office and those centres are not involved in transporting examination materials.

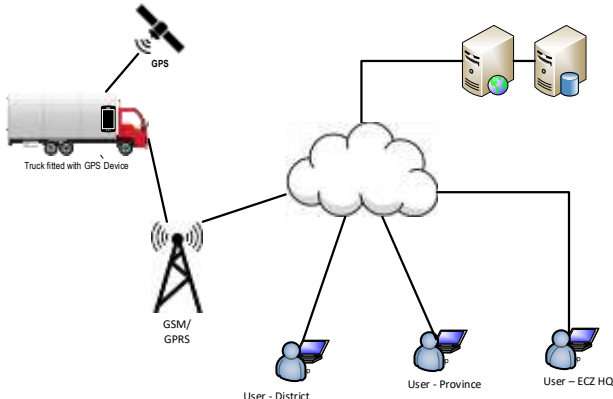


Fig. 4. System Architecture

The study also ascertained the manner in which examination papers were delivered and answer scripts collected from schools. It is clear that the DEBS office is highly involved in delivering question papers in examination centres while in most cases schools have to deliver the answer scripts on their own to the DEBS office or at some central/ zonal strongroom using their own transport. Participants however stated that they were satisfied with the security provided during transportation despite not having adequate transportation vehicles as 76.7 % indicated that they were satisfied while 19.0 % were not. The reasons given for not having adequate security were that they expected the armed state security personnel should be there to provide security both when transporting question papers and also answer scripts. It appears the emphasis on security is more on question papers and not answer scripts. Another important aspect mentioned in the use of a remote monitoring system on transportation vehicles (vehicles not secured). This entails that security need to be enhanced during or when transporting question papers and answerscripts.

The study established that there were security challenges during transportation of examinations. Cited among them were unreliable transport, insufficient transport from DEBS Office, lack of state security personnel to escort the examinations, poor road network and late delivery, also the absence of a remote monitoring system during transportation as show in Table 1. Costs were also a factor as schools are expected to contribute money for fuel. In addition other GRZ vehicles have to be used to beef up the transport situation.

TABLE I. CHALLENGES DURING TRANSPORTATION OF EXAMINATIONS

	Challenge	Frequency	Percent
1	Ministry of General Education should engage Zambia police so that they guard all the time	35	14.8
2	Lack of reliable transport	57	24.2

	Challenge	Frequency	Percent
3	Late delivery of examinations	5	2.1
4	Lack of Funds	11	4.7
5	Poor road network and long distances covered poses a risk	27	11.4
6	No Monitoring systems	6	2.5
7	No Challenges	95	40.3
	Total	236	100.0

A. Statistical Significance

The level of independence of variables like having designated transport and security during transportation between examination malpractices could not be established even after applying the Chi-square test where p-value, $p < 0.966$ for designated transport and for adequate security when transporting examination papers the p-value, $p < 0.081$. For both variables the p-value is above the significance level of 0.05 that we adopted in the study

TABLE II. SECURITY DURING TRANSPORTATION VS EXAMINATION MALPRACTICE

Have you had any examination malpractices in your area?		Yes	No	Not Sure	Total
Do you think that there is adequate security provided to secure examination materials during transportation to and from your school?	Count	98	80	33	211
	Expected Count	101.3	78.3	31.5	211.0
	% within adequate security in transit	46.4%	37.9%	15.6%	100.0%
	% within examination malpractices	74.2%	78.4%	80.5%	76.7%
	Standardized Residual	-.3	.2	.3	
	Count	28	21	4	53
	Expected Count	25.4	19.7	7.9	53.0
	% within adequate security in transit	52.8%	39.6%	7.5%	100.0%
	% within examination malpractices	21.2%	20.6%	9.8%	19.3%
	Standardized Residual	.5	.3	-1.4	
Total	Count	6	1	4	11
	Expected Count	5.3	4.1	1.6	11.0
	% within adequate security in transit	54.5%	9.1%	36.4%	100.0%
	% within examination malpractices	4.5%	1.0%	9.8%	4.0%
	Standardized Residual	.3	-1.5	1.8	
Total	Count	132	102	41	275
	Expected Count	132.0	102.0	41.0	275.0
	% within adequate security in transit	48.0%	37.1%	14.9%	100.0%

<i>Have you had any examination malpractices in your area?</i>		<i>Yes</i>	<i>No</i>	<i>Not Sure</i>	<i>Total</i>
	% within examination malpractices	100.0%	100.0%	100.0%	100.0%

***Pearson Chi-Square (X2) test statistics 8.316 (p-value = 0.081)

Table II. examined the relationship between adequate security when transporting examination question papers and answerscripts and examination malpractices. After applying the Chi-square test the p-value of $p < 0.081$ was obtained. This value is slightly above the significance level of 0.05 that we adopted in the study. This suggests that the connection between having adequate security when transporting examinations and examination leakages cannot be established. However, it is important to note that there are two types of errors that can be made when inferring statistical significance known as Type I (risk of rejecting the null hypothesis when it should be confirmed) and Type II (risk of confirming the null hypothesis when it should be rejected) errors. Our proposition is that Type I error was committed and hence our proposal to secure the transportation process.

Conclusion

The results of the baseline study were used to propose a more secure transportation model for examination materials in Zambia which integrates technologies that have been used in other fields such as GPS tracking, GSM/GPRS and cloud technologies. It is hoped that with enhanced monitoring and tracking of the movement of the examination materials, a more secure environment will be attained and information will be available in near-realtime for quick action should anything occur on the way like what happened in [25]. This in turn should reduce examination leakages that are related to insecure transportation of examinations.

ACKNOWLEDGMENT

We would like to thank the Director, Examinations Council of Zambia and Ministry of General Education for allowing us to carry out this research at both ECZ and the schools.

REFERENCES

[1] Government of Zambia, Constitution of Zambia (Amendment) No.2 of 2016, Lusaka: Government of Zambia, 2016.

[2] A. Shamsuzzoha and P. T. Helo, "Real-time Tracking and Tracing System: Potentials for the Logistics," in 2011 International Conference on Industrial Engineering and Operations Management, Kuala Lumpur, 2011.

[3] Examinations Council of Zambia, Guidelines for the Administration and Management of Examinations in Zambia, Lusaka: Examinations Council of Zambia, 2018.

[4] M. J. Kelly, Education in a Declining Economy, Washington DC: World Bank, 1991.

[5] Government of the Republic of Zambia, The Laws of Zambia - Examinations Council of Zambia Act, Lusaka: Government of the Republic of Zambia, 1983.

[6] Ministry of Education, Educating our Future, National Policy on Education, Lusaka: Zambia Educational Publishing House, 1996.

[7] Ministry of Communications and Transport, National Information & Communication Technology Policy, Lusaka: Ministry of Communications and Transport, 2006.

[8] Malawi Government, Malawi National Examinations Board Act (Cap30:04), Conduct of Examinations Regulations, Lilongwe: Malawi Government, 2008.

[9] J. R. Vacca, Computer and Information Security Handbook, New York: Morgan Kaufmann Publishers, 2009.

[10] Zambian Watchdog, "Grade seven exam papers stolen in Kawambwa," Zambian Watchdog, 27 October 2009. [Online]. Available: <https://www.zambianwatchdog.com/grade-seven-exam-papers-stolen-in-kawambwa/>. [Accessed 22 September 2018].

[11] Zambian Watchdog, "Police pick up 22 pupils, a teacher and guards for stealing exam papers," 5 November 2012. [Online]. Available: <https://www.zambianwatchdog.com/police-pick-up-22-pupils-a-teacher-and-guards-for-stealing-g7-exam-papers/>. [Accessed 08 August 2018].

[12] D. Mwendabai, "4 Pupils, guard arrested," Zambia Daily Mail, vol. 22, no. 286, p. 1, 29 November 2018.

[13] Examinations Council of Zambia, "2016 Examinations Performance Review Report – General Performance Analysis," Examinations Council of Zambia, Lusaka, 2017.

[14] N. Chadil, A. Russameesawang and P. Keeratiwintakorn, "Real-Time Tracking Management System Using GPS, GPRS and Google Earth," in IEEE Xplore, 2008.

[15] "GPS Tracking and Logistics Business," GPS Tracking Journal, February 2018. [Online]. Available: <https://www.gpsjournal.com/news/gps-tracking-and-logistics-business/>. [Accessed 17 October 2019]

[16] B. Ghribi and L. Logrippo, "Understanding GPRS: The GSM Packet Radio Service," Computer Networks, vol. 34, pp. 763-779, 2000.

[17] Google Earth, "Using Google Earth for infrastructure planning," [Online]. Available: https://www.eawag.ch/fileadmin/Domain1/Abteilungen/sandec/publikationen/SESP/Urban_Sanitation_general/Using_Google_Earth_for_Infra_Planning.pdf. [Accessed 22 October 2019].

[18] M. Mujinga, "Developing Economies and Cloud Security: A Study of Africa," Journal of Emerging Trends in Computing and Information Sciences, vol. 3, no. 8, pp. 1166-1172, August 2012.

[19] S. Ghule, R. Chikhale and K. Parmar, "Cloud Computing in Banking Services," International Journal of Scientific and Research Publications, vol. 4, no. 6, pp. 1-4, June 2014.

[20] D. N. Kuria, M. M. Ngigi, W. J. Wanjiru and R. K. Kasumuni, "Managing distribution of national examinations using geospatial technologies: A case study of Pumwani and Central divisions," International Journal of Computer Engineering Research, vol. 2, no. 5, pp. 82-92, 2011.

[21] "Smartlock Exam Management," June 2018. [Online]. Available: <https://www.specfile.co.za/wp-content/uploads/2018/06/smartlock-Exam-Management-2018.pdf>. [Accessed 28 August 2018].

[22] K. Leal and J. Leite, "Truck Full of CFA Exams Hijacked in Rio Crime Wave," 07 December 2017. [Online]. Available: <https://www.bloomberg.com/news/articles/2017-12-07/rio-crime-wave-hits-unexpected-target-a-truck-full-of-cfa-exams>. [Accessed 17 July 2018].

[23] Coventry News, "Singapore Chemistry A-Level test papers stolen in transit to British Examiner," 23 February 2018. [Online]. Available: <https://www.coventrytelegraph.net/news/coventry-news/exam-papers-stolen-warwick-university-3082226>. [Accessed 17 July 2018].

[24] R. Adams, "Exams rewritten after van containing papers stolen," 11 June 2015. [Online]. Available: <https://www.theguardian.com/education/2015/jun/11/exams-rewritten-after-van-containing-papers-stolen>. [Accessed 19 July 2018].

[25] Zambia Daily Nation, "Stolen Exam Papers Shock NGO," 25 September 2016. [Online]. Available: https://web.facebook.com/dailynationnews/posts/stolen-exam-papers-shock-ngo/894002080729778/?_rdc=1&_rdr. [Accessed 17 July 2018].

[26] A. Bryman, Social Research Methods, 5th ed., New York: Oxford University Press, 2016.

[27] C. L. Muyunda and J. Phiri, "A Wireless Sensor Network Based Grain Inventory Management System for Zambia's Food Reserve Agency," International Journal of Innovative Research in Science, Engineering and Technology, vol. 5, no. 3, pp. 3519-3526, 2016.

[28] Y. Agarwal, K. Jain and O. Karabasoglu, "Smart vehicle monitoring and assistance using cloud computing," International Journal of Transportation Science and Technology, vol. 7, pp. 60-73, 2018.

A Data Mining Model for Predicting and Forecasting Fraud in Banks

Knox Kamusweke
 Dept. of Computer Science
 University of Zambia
 P.O. Box 32379, Lusaka. Zambia
 Email: knox.kamusweke@cs.unza.zm

Mayumbo Nyirenda
 Dept. of Computer Science
 University of Zambia
 P.O. Box 32379, Lusaka. Zambia
 Email: mayumbo.nyirenda@cs.unza.zm

Monde Kabemba
 Dept. of Computer Science
 University of Zambia
 P.O. Box 32379, Lusaka. Zambia
 Email: monde.kalumbilo@cs.unza.zm

Abstract— Banks generate and handle millions of transactions across their platforms. These transactions contain significant patterns and trends which are hidden but needed for knowledge discovery and actionable insight. Uncovering these patterns and trends has always been a challenge for most financial institutions due to the large volumes of transactions and the ever changing patterns. This has made the patterns and trends more sophisticated to discover and has left many institutions without better knowledge and insight into the types of transactions taking place. This study aims to use data mining to construct a model that discovers hidden patterns and is able to predict and forecast fraud on the basis of those discovered patterns. The study was guided by two objectives; firstly, to find hidden patterns and trends in data, and secondly to construct a model that predicts and forecasts fraud on the basis of discovered patterns. Using the proposed model patterns of unusual transactions done with an upwards trend were discovered. Furthermore, the prediction score and forecasting indicated that these patterns of transactions were likely to continue presenting high chances of fraud happening hence the importance of more research into data mining for fraud detection.

Keywords: Data Mining; Fraud; Patterns, Trends, Forecasting

I. INTRODUCTION

Banks in Zambia generate and handle millions of transactions daily which are stored in large databases. These transactions have continued to grow tremendously across multiple platforms. One major challenge faced in these institutions is the lack of better knowledge and actionable insight about the patterns of transactions happening. Uncovering these patterns and trends has been a challenge due to large data sizes which demand scalable intelligent learning systems to analyse and handle.

There are many definitions of fraud that vary depending on the perspective of interested parties [1] [2]. In its simplest definition, fraud has been defined as any intentional act or omission designed to deceive others, resulting in the victim suffering a loss and/or the perpetrator achieving a gain [2]. Fraud commonly includes activities such as theft, corruption, conspiracy, embezzlement, money laundering and bribery [1]. Frauds is perpetrated by parties and organizations to obtain money, property, or services; to avoid payment or loss of services; or to secure personal or business advantage. The process of predicting and forecasting fraud can be very challenging due to the fraud patterns that keep changing. Calculating the actual costs of fraud for banks is a formidable task because of the stealthy nature of fraud. The detected fraud cases represent only the "tip of the ice berg" of all the frauds that go unnoticed [3]. In addition, fraud can be very complex and has

some temporal characteristics and patterns that have to be identified.

The Reserve Bank of India – RBI maintains data on frauds on the basis of area of operation under which the frauds have been perpetrated [4]. According to such data pertaining, top ten categories under which frauds have been reported by banks include the following credit cards, deposits – savings accounts, Internet Banking, housing loans, term loans, cheque/demand Drafts, Cash Transactions, Cash Credit A/C (Type of Overdraft A/C), Advances and ATM/Debit Cards.

Studies and research have proven that traditional fraud detection and prediction techniques like random checks, targeted audits, internal control systems, external audits, risk management systems and whistleblowing hotlines, might not be effectively applicable for the new trends in fraud [5]. To analyse data and determine various kinds of fraud-like patterns on large scale transaction data, data mining techniques have emerged to make it less vulnerable and provide reliable solutions to business [6] [5]. Although detecting fraud is considered a high priority for many financial institutions, the current literature lacks for an up-to-date, comprehensive and in-depth review that can help banks with decisions on selecting the appropriate data mining technique [7].

This research paper highlights the use of data mining for financial fraudulent pattern and trend discovery particularly in the banking sector and then constructs a model for predicting and forecasting fraud on the basis of uncovered patterns. In addition, it presents cases in which data mining techniques were successfully implemented to detect fraud. It further discusses the methodology used in our research before discussing our findings. The rest of the paper is organized as follows, section II provides a brief discussion of data mining, section III presents related works, section IV presents the approach taken to conduct the research, section V presents the experimental results and finally section VI presents a conclusion based on the findings of the research.

II. DATA MINING

A. Data Mining for Knowledge Discovery.

[8] Larose et al, defined data mining as the process of discovering useful patterns and trends in large data. The process must be automatic and the patterns discovered must be meaningful in that they lead to some advantages, usually an

economic advantage. Also known as Knowledge Discovery Data or Knowledge Mining, Data Mining (DM) involves the analysis of data from different perspectives and summarizing it into useful information [9] [10]. Data mining is the core of the knowledge discovery in databases (KDD) process, thus, data mining and KDD are often used interchangeably [11, 12]. Fig 1 shows the steps of extracting knowledge from data using data mining. The first three processes, that are data selection, data preprocessing and data transformation, are considered as data preparation processes. The last three processes including data mining, pattern evaluation and knowledge representation are integrated into one process called data mining [13].

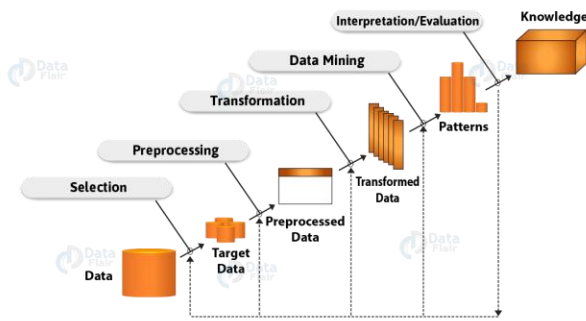


Fig 1. Process of Data Mining [13] [16].

In general, data mining techniques are classified into four categories according to the type of the machine learning;

1) **Supervised Learning** for Fraud Detection: This method uses supervised learning in which all the available records are classified as “fraudulent” and “non-fraudulent”. Then machines are trained to identify records according to this classification. However, these methods are only capable of identifying fraud that has already occurred and about which the system has been trained [5] [14].

2) **Unsupervised Learning** for Fraud Detection: This method only identifies the likelihood of some records to be more fraudulent than others without statistical analysis assurance [5] [14].

3) **Semi-supervised Learning** for Fraud Detection: This method addresses fraud in a similar way as supervised learning. However, in semi-supervised learning, the machine is provided with labeled data along with additional data that is not labeled with predetermined outcome [15].

4) **Reinforcement Learning** for Fraud Detection: This method differs from the ones above because models do not get trained with sample data but through trial and error to detect fraud [16].

B. Data Mining Models.

A data mining model is created by applying an algorithm to data. A data mining model gets data from a mining structure and then analyzes that data by using a data mining algorithm. The data mining models are of two types [13] [17]: Predictive and Descriptive. The predictive model makes predictions about

unknown data values by using the known values. Examples include Classification, Regression, Time Series Analysis, Prediction etc. The descriptive model identifies the patterns or relationships in data and explores the properties of the data examined. Examples are Clustering, Summarization, Association rule and Sequence discovery etc. Our study used both types of models to uncover the significant patterns using clustering and then construct a model to predict and forecast fraud using time series analysis on the basis of the discovered patterns (Fig 2). Some of the most significant DM techniques used in fraud discovery models include Neural Network, Decision Tree Analysis, K-Means Clustering, Predictive Analysis and Logistic Regression etc.

III. RELATED WORKS

Despite being a relatively new technology that has not fully matured, there are a number of industries such as banks, insurance companies and retail stores that are already using data mining on a regular basis [18] [19]. Data mining techniques like k-means clustering, logistic regression, decision tree, naïve Bayes, random forest etc. have been used in detecting financial fraud [7] [20]. One system that has been successful in detecting fraud is Falcon’s “fraud assessment system” being used by nine of the top ten credit card issuing banks, where it examines the transactions of 80% of cards held in US [4]. Mellon Bank is also using data mining for fraud detection and is able to protect itself and its customers’ funds from potential credit card fraud.

In their study, Vaishali [21], were able to detect fraud in credit card by using clustering approach. Their results were found by using K-Means clustering algorithm. Their algorithm formed four clusters being low cluster, high cluster, risky cluster and high risky cluster. They then tested transactions on five credit card numbers by applying K-Means clustering. Their results showed that some fraudulent transactions done in India, Ukraine and Ecuador with credit card numbers 456723, 234562, 176345 belonging to low cluster, high cluster and high risky cluster.

In the United Kingdom (UK), Provident Financial’s Home credit Division had no system to detect and prevent fraud. After applying data mining techniques, they have reduced frequency and magnitude of agent and customer fraud, saved money through early fraud detection, and saved investigator’s time and increased prosecution rate [18].

Peer Group Analysis, an unsupervised DM technique for monitoring customer behaviors over period of time is used in Australia [22] to monitor customer behaviors. For each individual that has a credit card account, a “Peer Group” of accounts are created that exhibit similar behavior. Over time, the behavior of an account is tracked by those accounts in its peer group. If an account has subsequent behavior which deviate strongly from its peer group, it is thus considered to have behaved anomalously and flagged as a potential fraudulent. Also in Australia, a “Break-Point” Analysis DM technique is being used to distinguish spending activities supported from transaction information in a single account. Current transactions are matched up with prior spending activities to spot features, such as rapid spending and an

increase in the level of spending, which would not essentially be captured without data mining [23].

In their study, S. Thiprungsri [24] used K-Means as a clustering procedure on the 40,080 claims that were paid in the first quarter of 2009. For the first set of clusters using 2 attributes, eight 8 clusters were formed. About 90% of claims were clustered into cluster 7 and 6% in cluster 0. Three clusters (1, 2 and 5) had membership of less than 1% and the numbers of claims in those clusters were 54, 84 and 31 respectively. Examining the characteristics of those less populated clusters, a couple of suspicious characteristic were mentioned. Claims in those clusters had high interest/beneficiary payment percentage and/or claims with a long period time from death dates to payment dates. The results showed that the total number of claims identified as possible anomalies from clusters was 169. Clusters with larger membership have higher numbers of possible anomalies.

[25] Kirkos et al. tried to identify firms that published fraudulent financial statements using Decision Trees, Artificial Neural Networks and Bayesian Networks. In the study, 76 financial tables with half of them fraudulent were used. As a result of training and tests with 10 selected features, the best classifier was Bayes Networks with 90.3% success.

[26] A large regional bank was looking to improve revenues, and improve the customer experience. As part of their growth efforts, they put increasing emphasis on reducing the occurrence of fraudulent debit card transactions. The bank was losing about \$100,000 per month and risked negatively impacting customers through rejected transactions and reissued debit cards. After employing data mining through predictive analytics, the bank quickly reclaimed an average of \$2 per account in fraudulent transactions. Given the bank's large customer base, this quickly added up to substantial savings.

IV. METHODOLOGY

The aim of this research is to construct a fraud prediction and forecasting model on the basis of discovered pattern and trends in the transactions. To achieve this, the first thing we did was to employ data mining technique on the transactions in order to discover hidden but intrinsic patterns and trends in the data. To better understand the usage scenarios, we will discuss our proposed approach based on the currently used approach at BANK X in Zambia. We withheld the name of the bank for privacy and sensitivity of data involved.

A. Case of BANK X

Like in most banks, fraud discovery through audit checks are targeted and sometimes following whistle blowing or engagement with external inspectors from government institutions such as Central Bank, Ministry of Finance and Auditor General's office. The current approach usually employs techniques such as sampling and random checks with very much predefined patterns to look for in the transactions. The auditors start by planning for the targeted audits checks. They set the audit objectives, which in many cases involve defining the patterns to look for. Manual integrity check and validations are done on the data using excel. The test data validity are done to check against any errors, missing

values, ensure only numeric data is in numeric fields, if data is in fields where it is expected, confirming control totals and ensure calculated fields deliver correct values using statistic commands. After completing their validity tests, the auditors then perform data analysis through tests necessary to achieve their objectives. During data analysis, the auditors define expressions, filters, and computed fields to apply in their analysis. They use expressions to normalize the data. After concluding their analysis, they report the significant findings developed in response to each audit objective. The current approach is biased, error prone due to heavy reliance on labour from humans and takes longer. The current approach is purely static-rule based because it looks for specific patterns in the transactions. It requires manual intervention at every stage of the process on manipulating variables and experts need to specify the set of patterns or characters to look for before starting.

B. Proposed Approach: A DATA MINING MODEL

We observed that one of the challenges with the current approach is that it remains a traditional packaged fraud detection approach because it is custom tuned. The current approach is static-rule based which looks for specific predefined patterns. This leaves a gap as fraud patterns are ever changing without known frequency in the growing data. It is for this reason that we base our model on the currently used approach. However, we incorporate data mining to replace the static rules set in the data analysis of the current approach. Therefore, preceding our steps in our approach with data mining was understanding of the application domain based on prior knowledge obtained from current approach.

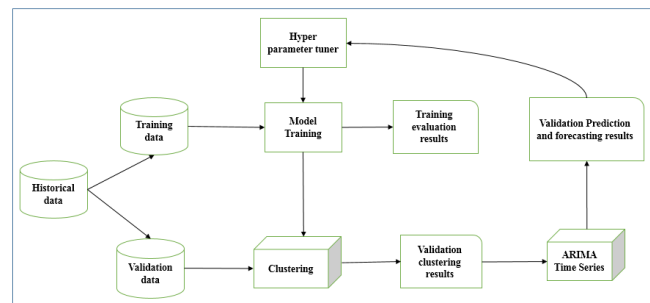


Fig 2: The proposed Data Mining model.

Our data mining approach was divided into two parts namely, data preparation and data mining.

1) Data preparation

The first part of our approach comprised data selection, data preprocessing and data transformation. We extracted transactional data from a database at BANK X from January 2017 to September 2019 and saved it in CSV files. This data was highly dimensional and unstructured without any labels.

We then preprocessed the data to make it clean, valid and consistent. One challenge we faced here was how to handle missing values. Missing data are a problem that continues to plague data analysis methods and this is common in data sources with large number of variables. The absence of information is

rarely beneficial but all things being equal, more information is always better for data mining. Therefore, we had to be careful about handling the thorny issue of missing values. To achieve this, we used a criteria of replacing the missing values with a value generated at random from our observed distribution of the variables.

Data was transformed to change all categorical feature into numeric. The variables in our preprocessed data had ranges varying from each other. Therefore, we had to normalize the variables in order to standardize the scale of effect each variable was to have on the results. Data transformation was also necessary in order to code our data and hide sensitive information such as original account numbers. Some coded fields were *ACCT_STATUS*, where Active=1, Inactive=2 and Dormant=3. *PAID_OFF_FLG* for loans was coded as YES=1 and NO=0, where *DR* for transactions was 0 and *CR* was 1 etc.

Our data was highly dimensional and to get rid of problems faced by multicollinearity and high number of dimensions, we implemented dimensionality reduction technique called *principal component analysis (PCA)*. PCA was used to help reduce the computational complexity for our model. Table 1 below shows individual principal component ratios in the dataset.

Table 1. Explained variance ratios of individual components

component	explained_variance	%
PC1	0.36884109	36.88
PC2	0.19318394	19.32
PC3	0.10752862	10.75
PC4	0.07421996	7.42
PC5	0.06245904	6.25
PC6	0.04909	4.91
PC7	0.04117287	4.12
PC8	0.02495984	2.50
PC9	0.02308855	2.31
PC10	0.01864124	1.86
PC11	0.01731766	1.73
PC12	0.01252785	1.25
PC13	0.00696933	0.70

2) Data Mining

This is the core step in our approach as it involved data mining itself, interpretation of discovered patterns and knowledge use. To address our first objective, we used clustering task to determine the intrinsic groupings in our unlabeled data and extract value from it. Kmeans clustering (Fig 3) was the algorithm used because it i.e. easy to implement. Time Series model was used AutoRegressive Integrated Moving Average (ARIMA) for our predictive analysis and give forecasting using TIME and transaction COUNT.

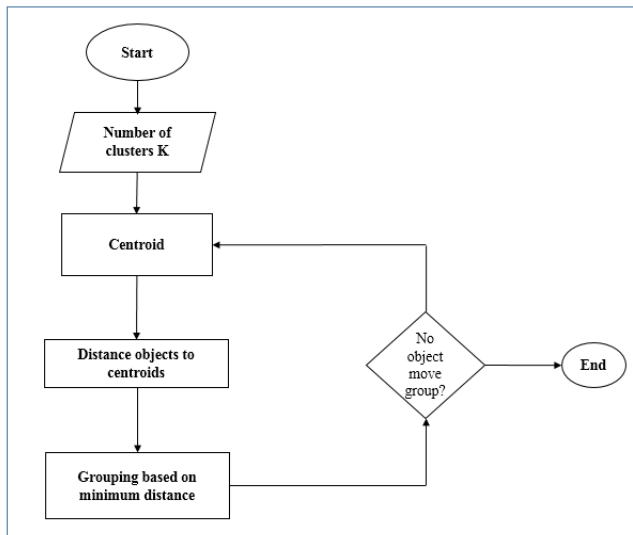


Fig 3. Workflow of our k-means clustering algorithm

To determine the optimal number of clusters for our Kmeans, we employed Silhouette analysis with scores in Table 2 below.

Table 2. Silhouette scores

#clusters	Silhouette_score
2	0.7498277640462251
3	0.7040624779460442
4	0.7123881956992123
5	0.7234924399047411
6	0.739829473008246
7	0.7230473525564337
8	0.7528808355974226
9	0.7572786439335218

V. EXPERIMENT AND RESULTS

A. Data set

To do our experimentation, we extracted the data from the core banking system using PL/SQL scripts and saved in CVS file format.

B. Setup

We then applied the procedure outlined earlier on in section V (B). PCA helped reduce number of dimensions in the data and K-means clustering algorithm to uncover the hidden patterns after which, constructed a time series ARIMA model for fraud prediction and forecasting. We did our implementations using Python v3.7.0 (64-bit), Jupiter notebook v5.7.8 running with Tensorflow2.0 Beta.

C. Discussion and Results

After applying our clustering with k-means, clusters were formed in the truncations which gave significant patterns of transactions (Fig 4 and Fig 5). Our results showed clusters with prevalence of transactions on paid-off loans, transactions on inactive/dormant accounts, balance transfers between loans, transactions without remarks, transactions with similar amounts passed on inactive/dormant accounts etc. When further checked

for pattern trend, the results indicated an uptrend for the discovered patterns (Fig 6).

(Table 4). The lower the values for MSE and RMSE the better the better model. Fig 7 shows plot for prediction score.

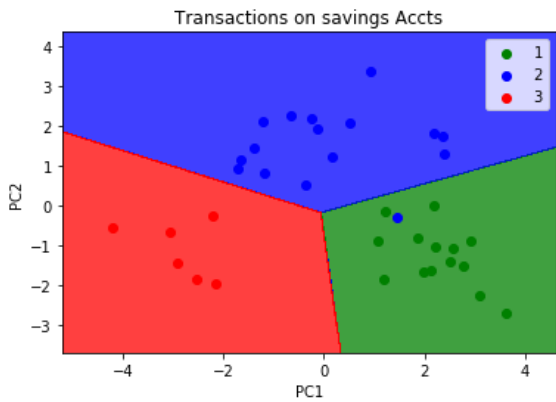


Fig 4. Discovered patterns of transactions observed on savings including inactive and dormant (1=active, 2=inactive, 3=dormant)

Table 4. Prediction results

Predicted Value	Expected Value	MSE_Predictions	RMSE_Score
0.861042	0.862600	0.000043	0.0472

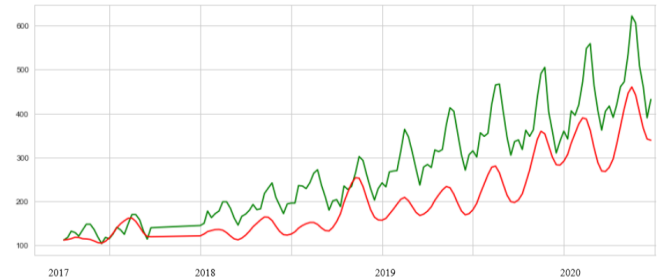


Fig 7. Our ARIMA prediction plot.

Fig 8 below shows results of ARIMA fitted values for forecasting.

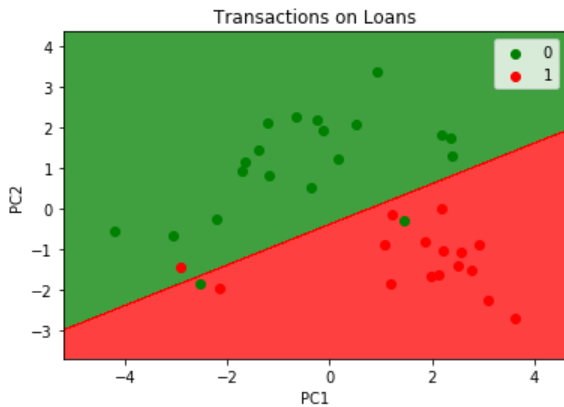


Fig 5. Discovered pattern of transactions on loans including those that are paid-off (0=running, 1=paid-off)

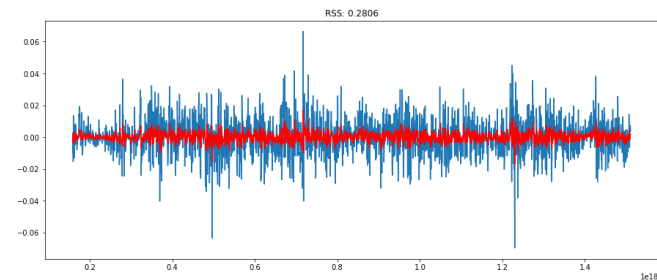


Fig 8. ARIMA forecasting.

When plotted for residual errors, our results gave a description around 0 (Fig 9).

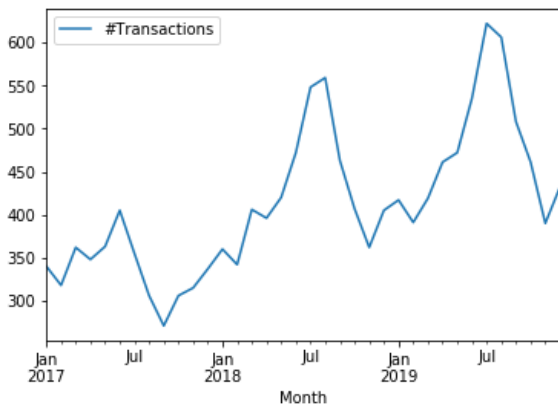


Fig 6. Discovered pattern trend (uptrend)

Based on the discovered patterns, we modeled a time series ARIMA model taking **time** from date as independent variable and **transaction count** as dependent variable for our prediction score

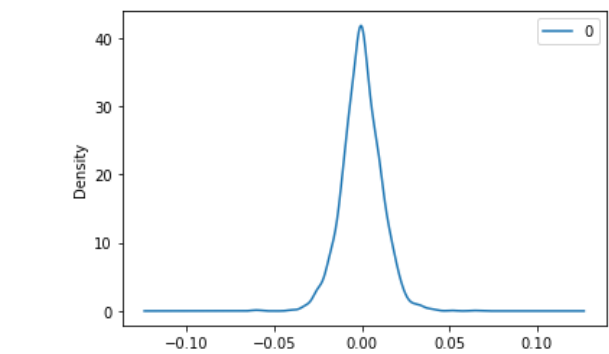


Fig 9. ARIMA residual plot

ARIMA model gave MSE = 0.000043, RMSE = 0.0472 and Residual plot around 0, from the prediction indicating a perfect score or no error therefore our model was accurate. From the pattern trend discovered, the model predicts these to continue.

VI. CONCLUSION AND FUTURE WORK

In this study, we developed a data mining model for fraud prediction and forecasting. We employed data mining tasks such as clustering and time series analysis from which we were able to get the patterns and trend. Some transactions may be legitimate and others not therefore, depending on the severity and business rules of the bank, sound decisions should be made and action taken accordingly.

We therefore, recommend that this model be implemented and adopted to predict and forecast fraud patterns for better knowledge discovery on the types of transactions taking place. The proposed model will provide actionable insight and profitable results through uncovering of hidden patterns of transactions and help alleviate chances of fraud.

For future work, we plan to incorporate Text Mining implementation in our model in order to also discover the common words that are used when carrying out transactions.

REFERENCES

[1] Fraud and Risk Management Working Group. *Fraud Risk Management: A Guide to Good Practice*. CIMA, 2008.

[2] JD. Ratley, BC. Melancon, DA. Richards. *Managing the Business Risk of Fraud: A Practical Guide*. IIA, AICPA, ACFE.

[3] C. Albrecht W. Albrecht, C. Albercht and M. Zimelman. *Fraud Examination*. South-Western Cengage Learning, Mason, USA, 2009.

[4] B. Rajdeepa and D. Nandhitha. "Fraud Detection in Banking Sector using Data Mining". *International Journal of Science and Research (IJSR)*, 4(7), July 2013.

[5] A. Shmais and Hani R. Data Mining for Fraud Detection. *Prince Sultan University, Saudi Arabia*.

[6] J. Agrawal S. Agrawal. Survey on Anomaly Detection using Data Mining Techniques. *Procedia Computer Science*, (60):708–713, 2015.

[7] M. Albashrawi. Detecting Financial Fraud Using Data Mining Techniques: A Decade Review from 2004 to 2015. *Journal of Data Science*, (14):553–570, December 2016.

[8] DT. Larose and CD. Larose. *DISCOVERING KNOWLEDGE IN DATA: An Introduction to Data Mining*. John Wiley & Sons, Inc, second edition, July 2014.

[9] J. Vohra and E. Jyoti. Data Mining Approach for Retail Knowledge Discovery. *International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, 6(3), March 2016.

[10] D. Bansal and L. Bhambhu. Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women. *International Journal of Computing and Technology (IJCAT)*, 3(9), September 2013.

[11] M. Mwanza. FRAUD DETECTION ON BIG DATA USING BUSINESS INTELIGENCE, DATA MINING TOOL. A CASE OF ZAMBIA REVENUE AUTHORITY. Master's thesis, School of Engineering, The University of Zambia, January 2017.

[12] Y. Fu. Data Mining: Tasks, Techniques, and Applications. *University of Missouri - Rolla*.

[13] WideSkills. Data Mining Tutorial: Data Mining Architecture and Process [online]. Available: <http://www.wideskills.com/data-mining-tutorial>, 2019.

[14] R. Bolton and D. Hand. Statistical Fraud Detection: A Review. *Statistical Science*, 3(17):235–255, 2002.

[15] K. Nevala. *The MACHINE LEARNING Primer*. SAS Institute Inc, 100 SAS Campus Drive, Cary, NC 27513-2414, USA, 2017.

[16] J. Hurwitz and D. Kirsch. *Machine Learning: For Dummies*. John Wiley & Sons, Inc, ibm limited edition edition, 2018.

[17] S. Deshpande and V. Thakare. DATA MINING SYSTEM AND APPLICATIONS: A REVIEW. *International Journal of Distributed and Parallel Systems (IJDPSS)*, 1(1), September 2010.

[18] B. Ramageri. DATA MINING TECHNIQUES AND APPLICATIONS. *Indian Journal of Computer Science and Engineering*, 1(4):301–305.

[19] C. Priyadharsini and Dr. A. Thanamani. An Overview of Knowledge Discovery Database and Data Mining Techniques. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(1), March 2014.

[20] E.W.T. Ngai, Y. Hu, Y.H. Wong, Y. Chen, X. Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, (50):559–569, August 2016.

[21] Vaishali. Fraud Detection in Credit Card by Clustering Approach. *International Journal of Computer Applications*, 98(3):0975–8887, July 2014.

[22] D. Hand C. Whitrow D. Weston, N. Adams and P. Juszczak. Plastic Card Fraud Detection using Peer Group Analysis. In *Marchel Publication*, Sydney, Australia, 2007.

[23] D. Montague. Fraud Prevention Techniques for Credit Card Fraud. In *Spring-field Press*, New York, 2006.

[24] S. Thiprungsri. Cluster Analysis for Anomaly Detection in Accounting Data. *Collected Papers of the 9th Annual Strategic and Emerging Technologies Research Workshop*, July 2010.

[25] E. Kirkos, C. Spathis, Y. Manolopoulos. Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Application*, (32):995–1003, 2007.

[26] MapR. Fraud Detection Solution at a Large Regional Bank. 2016.

Distributed Spatial Search Using Paillier Cryptosystem and a Distributed Ring Algorithm

Jimmy Katambo¹, Mayumbo Nyirenda², David Zulu³

^{1,2,3} Department of Computer Science, University of Zambia

¹jimmy.katambo@cs.unza.zm, ²mayumbo.nyirenda@cs.unza.zm, ³david.zulu@cs.unza.zm

Abstract—The problem of lack of anonymity and confidentiality can be experienced by those who collect statistical data online as well as those who provide the data. One end may be secure, for example, the one providing data and yet the other end, for example, the one collecting data, may not be secure. In another scenario both the data provider and collector may seek anonymity. Preventing decryption of data provided while providing aggregated results is the best solution for such scenarios. To achieve this, this paper proposes a protocol that puts into application Homomorphic Encryption and a Distributed Ring algorithm, to ensure anonymity of data of both parties involved in a spatial search that is a data provider and a searcher. Firstly, we identify a Homomorphic Encryption technique that can work best for a spatial search by reviewing literature on Homomorphic Encryption techniques. Among the Homomorphic Encryption techniques reviewed were Rivest, Shamir and Adleman (RSA), El Gamal cryptosystem, Goldwasser-Micali cryptosystem, Benaloh cryptosystem, Paillier cryptosystem and Fully Homomorphic Encryption (FHE). After a comprehensive study, Paillier Homomorphic Encryption technique was identified as the best approach to be employed in securing a spatial search. Secondly, we propose a protocol for distributed spatial searching using Paillier cryptosystem and distributed ring algorithm principles. Finally, a proof of concept prototype using the proposed approach was implemented. From initial experiments conducted using the proposed approach it is evident that the bigger cost comes from the communication over the network and less from the encryption algorithm and protocol itself. A 39% overhead when compared to the usefulness of the approach is outweighed making the solution highly practical and useful.

Keywords— Homomorphic Encryption, Paillier Cryptosystem, Confidentiality, Anonymity, Spatial search, Privacy, Distributed Systems, Ring Algorithm.

I. INTRODUCTION

Statistical information such as election results, number of HIV AIDS patients in certain geographical areas, data for use in cancer clustering and surveillance, etc. can be accessed by researchers online. The reason for accessing it are numerous including sharing and combining information, broadcasting it, identifying patterns for statistical analysis, logistical support and planning, market surveys and business initiatives, further scientific research to mention but a few. Approaches like that proposed by [25] [26] have made access to spatial searches a highly feasible approach. However, this access can be problematic to both a data provider and a searcher as a result of privacy breaches and sometimes careless release of sensitive information to the public. Research in Mathematical Cryptography [24] has opened many doors for researchers to study Homomorphic Encryption techniques to identify possible solutions so that anonymity is fully guaranteed on both ends of the equation,

i.e. anonymity of the person doing the searching and also anonymity of the data provider. Homomorphic Encryption enables servers to carry out sophisticated mathematical computations on encrypted records without acknowledging the original message [1]. Homomorphic Encryption allows users to make changes to the data without decrypting it [20]. The key principle in Homomorphic Encryption is we can achieve addition and subtraction of the data using Additive Homomorphism and we can also get multiplication and division of the data using Multiplicative Homomorphism. Paillier Cryptosystem was identified as the best for supporting a spatial search because it is computationally cheaper to be used in practice. Paillier's scheme is the most efficient among currently known additively homomorphic schemes, i.e. it requires simple operations in the encryption, decryption and addition procedures and hence achieves high performance. Another observation is that calculations on the encrypted data can be performed without necessarily reconstructing the original message and without having access to the private key. This article therefore focusses on proposing a protocol that puts into application Paillier Homomorphic encryption, to ensure anonymity of data of both parties involved in a spatial search, i.e. a data provider and a searcher. The work is presented in the rest of the paper as follows: Section II highlights various homomorphic encryption techniques; Section III discusses why the Paillier cryptosystem was chosen; Section IV presents the proposed solution and finally Section V presents a discussion and conclusion.

II. HOMOMORPHIC ENCRYPTION TECHNIQUES

A. Rivest, Shamir and Adleman (RSA)

RSA is a commonly adopted public key cryptosystem which was proposed by Rivest, Shamir and Adleman in 1977 at Massachusetts Institute of Technology (MIT) [2]. The security of RSA rests on the effort to factorize the big numbers of modulus. The size of modulus value is 1024 bits while the recommended length is 2048 bits as 640 bits key is not secure [3]. RSA uses two pairs of related keys (*public key*) $ku = \{e, n\}$ for encryption and (*private key*) $kr = \{d, p, q\}$ for decryption. A description of variables and operators in the stated equations is illustrated in full in the following steps below:

Step 1: p and q are two relatively prime and large random numbers.

Step 2: A positive integer n is defined as a product of p and q .

Step 3: Eulers value of

$$\phi(n) = (p-1)(q-1). \quad (1)$$

Step 4: Choose e such that $1 < e < M < n$ and

$$C = Me \text{ mod } n. \quad (2)$$

Step 5: In RSA e and n are public keys and d and (p, q) are private keys so the plaintext M is encrypted by: $1 < M < n$ and $C = Me \text{ mod } n$.

Step 6: The cipher text C is decrypted by

$$M = Cd \text{ mod } n \quad (3)$$

RSA is one of the earliest multiplicative homomorphic encryption schemes [5]. RSA has a multiplicative homomorphic property which can be used in many applications where there is a requirement for those applications to perform multiplicative operations on sensitive user data [5]. There are several techniques that improve the speed of the encryption procedure of RSA, however they also cause more computational complexity on decryption side [6]. Consequently, RSA is mathematically unstable like other cryptographic techniques. Different restrictions could be observed and several successful attacks are developed to break this algorithm [3], [4]. The major problem is related to its factorizing. The whole algorithm is broken once the process of factorization is done. When this happens, RSA has no more guarantees that the secrets it guards will remain secure [3].

B. El Gamal Cryptosystem:

According to [21], similar to RSA, the public key encryption scheme given by El-Gamal is a multiplicative homomorphic encryption cryptosystem. It was proposed by Taher El-Gamal in 1984, and its security relies on the hardness of the Diffi-Hellman problem [21]. El Gamal Cryptosystem [7] portrays a multiplicative homomorphic encryption propriety

C. Goldwasser-Micali Scheme

According to R. Shruthi, P. Sumana and A. K. Koundinya [8], the Goldwasser-Micali cryptosystem is an asymmetric key encryption algorithm developed by ShafiGoldwasser and Silvio Micali in 1982. Goldwasser-Micali has the distinction of being the first probabilistic public-key encryption scheme which is provably secure under standard cryptographic assumptions [8]. However, it is not an efficient cryptosystem, as ciphertexts may be several hundred times larger than the initial plaintext.

One of the drawbacks to the RSA encryption algorithm as originally defined is that it leaks a single plaintext bit in every ciphertext. The GM cryptosystem was the first cryptosystem to provably solve this problem. It was presented by Goldwasser and Micali along with a rigorous definition of security known as semantic security and a proof that the GM cryptosystem is semantically secure against plaintext attacks [8].

In their research [8], they show that Goldwasser-Micali takes slightly more time for encryption than RSA but it is also more secure against attacks. Secondly, they also show that the Decryption time of RSA remained almost constant between plaintext sizes 2 to 26 bytes while the Goldwasser-Micali had greater varying encryption times reaching a maximum of 27.5milli second plaintext of 6 bytes and minimum of 3.6milli seconds for plaintext of 20 bytes. Thirdly, they found out that RSA and Goldwasser Micali show a linear increase in the number of blocks of ciphertext generated for increasing plain text sizes [8]. However, the

increase is more pronounced in the case of Goldwasser Micali which can be attributed to the generation of a random value for every bit of plain text in Goldwasser micali [8].

In their conclusion of their paper [8], they demonstrate that Goldwasser and Micali develop a bit encryption function based on the hardness of quadratic residuosity problem. The method has many useful properties, but there is one major drawback: for a given security parameter N , the probabilistic encryption of each bit is N bits long, requires N random bits, and uses several operations on N bit integers. A major disadvantage of the Goldwasser-Micali scheme is the message expansion by a factor of $\lg n$ bits [8]. Some message expansion is unavoidable in a probabilistic encryption scheme because there are many ciphertexts corresponding to each plaintext[8]. The homomorphic property of Goldwasser-Micali public-key encryption scheme [9], has the ability to treat plaintext bit after bit, and the security is based on its semantic security, namely the quadratic residuosity assumption.

D. Benaloh Cryptosystem

Benaloh proposed an extension of the Goldwasser-Micali (GM) Cryptosystem by improving it to encrypt the message as a block instead of bit by bit [10]. Benaloh's proposal was based on the higher residuosity problem. Higher residuosity problem (x^n) [10] is the generalization of quadratic residuosity problems (x^2) that is used for the GM cryptosystem. Homomorphic property of Benaloh shows that any multiplication operation on encrypted data corresponds to the addition on plaintext [10]. As the encryption of the addition of the messages can directly be calculated from encrypted messages $E(m_1)$ and $E(m_2)$, the Benaloh cryptosystem is additively homomorphic [10].

E. Additive Homomorphic Encryption (Paillier Cryptosystem):

In 1999, Paillier [10] introduced another novel probabilistic encryption scheme based on composite residuosity problem. Composite residuosity problem is very similar to quadratic and higher residuosity problems that are used in GM and Benaloh cryptosystems [10]. It questions whether there exists an integer x such that $x^n \equiv a \pmod{n^2}$ for a given integer a . Other authors have referred to it as the Decisional Composite Residuosity Assumption (DCRA) [21]. This enables Paillier cryptosystem to have numerous applications such as threshold schemes and e-voting systems.

KeyGen Algorithm: For large primes p and q such that $\gcd(pq, (p-1)(q-1)) = 1$, compute $n = pq$ and $\lambda = \text{lcm}(p-1, q-1)$. Then, select a random integer $g \in \mathbb{Z}_{n^2}^*$ by checking whether $\gcd(L(g^\lambda \text{ mod } n^2), n) = 1$, where the function L is defined as $L(u) = (u-1)/n$ for every u from the subgroup $\mathbb{Z}_{n^2}^*$ which is a multiplicative subgroup of integers modulo n^2 instead of n like in the Benaloh cryptosystem [10].

Pick two large primes p and q and let $n=pq$. Let λ denote the Carmichael function, that is $\lambda(n) = lcm(p - 1, q - 1)$. Pick random $g \in Z_{n^2}^*$ such that $L(g^\lambda \bmod n^2)$ is invertible modulo n (where $L(u) = \frac{u-1}{n}$). n and g are public; p and q (or λ) are private. For plaintext x and resulting ciphertext y , select a random $r \in Z_n^*$. Then,

$$e_k(x, r) = g^m \cdot r^n \bmod n^2, \quad (4)$$

$$d_k(y) = \frac{L(c^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n \quad (5)$$

Finally, the public key is (n, g) and the secret key is (p, q) pair. The Homomorphism: Suppose x_1 and x_2 are plaintexts. Then,

$$\begin{aligned} e_k(x_1, r_1) e_k(x_2, r_2) &= g^{x_1} \cdot r_1^n \cdot g^{x_2} \cdot r_2^n \bmod n^2 \\ &= g^{x_1+x_2} \cdot (r_1 r_2)^n \bmod n^2 \\ &= e_k(x_1+x_2, r_1 r_2) \end{aligned} \quad (6)$$

To perform addition and multiplication on encrypted data stored in the cloud provider, the client must have two different key generators (one for RSA and one for Paillier) [7].

F. Fully Homomorphic Encryption Schemes:

In accordance with [10], an encryption scheme is called Fully Homomorphic Encryption (FHE) scheme if it allows an unlimited number of evaluation operations on the encrypted data and resulting output is within the ciphertext space. After almost 30 years from the introduction of privacy homomorphism concept [10], Gentry presented the first feasible proposal in his seminal PhD thesis to a long term open problem, which is obtaining an FHE scheme. Gentry’s proposed scheme gives not only an FHE scheme, but also a general framework to obtain an FHE scheme. Hence, a lot of researchers have attempted to design a secure and practical FHE scheme after Gentry’s work. Although Gentry’s proposed ideal lattice-based FHE scheme [10] is very promising, it also had a lot of bottlenecks such as its computational cost in terms of applicability in real life and some of its advanced mathematical concepts make it complex and hard to implement.

According to [19], another limitation is that FHE does not cater for multiple users. The practical applications which involve the running of enormously large and complex algorithmic computations homomorphically have a massive computational overhead, which makes the intermediate complex functional computations impractical [19]. Therefore, many new schemes and optimization have followed his work in order to address aforementioned bottlenecks. The security of new approaches to obtain a new FHE scheme is mostly based on the hard problems on lattices [10].

III. FACTORS THAT INFLUENCE A SPATIAL SEARCH

The most popular class of such services is k-nearest neighbor (kNN) queries where users search for geographical points of interests (e.g., restaurants, hospitals) and the corresponding directions and travel-times to these locations. Accordingly, numerous algorithms have been developed to efficiently compute the distance and route between objects in large road networks [22].

In the query processing of spatial-keyword search, indexing techniques for both text and geographic data are used [23].

In a GeoSN, a variety of spatial objects (e.g. restaurants, hotels, businesses) are marked on the map and annotated with user generated tags. GeoSN users can search for interesting spatial objects, and share information about their location and activities. More importantly, users with similar interests can plan for social activities collaboratively, such as going to somewhere for dining and shopping, or taking a cycling tour together. To make such plans, it is essential to identify a group of spatial objects, such as restaurants, shops and parks, which can maximally satisfy the users’ needs [23].

Scalability in Terms of Dataset Size. In order to simulate the real geo-social networking in which the number of objects and tags continuously increase, J. Zhong, X. Meng, X. Zhou and D. Liu [23] conducted a set of experiments to evaluate the scalability of three algorithms by varying the number of objects. A summarization of this discussion is presented in Table I.

TABLE I. FACTORS THAT INFLUENCE A SPATIAL SEARCH

S/N	FACTORS THAT INFLUENCE A SPATIAL SEARCH
1.	Point Data (Longitude/Latitude)
2.	Geocoding-When there are no coordinates, the search can still be done by gmap package
3.	Search Algorithms -such as the kNN (k-nearest neighbor)
4.	Time Dependent-how fast and how slow after running query
5.	Query Length-Handling queries of various length
6.	Scalability in Terms of Data Set Size

IV. MATERIALS AND METHODS

1) Cryptosystem Selection

The Paillier cryptosystem was selected for distributed spatial search for the following reasons:

- a) It has a smaller expansion rate and lower cost of encryption and decryption. To ensure that the system remains protected and secure, the lower bound of this expansion rate should be four. Improved schemes have developed with the expansion factor being lowered to increase efficiency [1]. For instance, Paillier cryptosystem allowed efficient decryption by enabling encryption of many bits during a single calculation with a better expansion rate of two [1].
- b) It is computationally cheaper and can be used in practice.
- c) Paillier’s scheme is the most efficient among currently known additively homomorphic schemes, i.e. it requires simple operations in the encryption, decryption and addition procedures and so achieves high performance.
- d) Calculations necessary to reconstruct the original message can be performed on the encrypted data even without having access to the private key.

- e) Analyses can be conducted on encrypted data with potentially little if any risk of revealing confidential information [12].
- f) It has a self-blinding property which property allows mapping a plaintext into possibly many different ciphertexts and the same plaintexts cannot be recognized from their ciphertexts [13].
- g) Since Paillier encryption is probabilistic, the encrypted files on the different peers are not linkable to each other for anyone not knowing the private decryption key [11].
- h) Paillier Encryption is probabilistic in that its encryption algorithm uses pseudorandom number generators. Hence, encrypting the same message several times will produce different ciphertexts, making it very cumbersome for even an informed adversary to compare encrypted messages in order to ascertain the original value that was encrypted [12].

The discussion presented is high level. Details can found in the literature cited. A summarization of this discussion is presented in Table II.

2) *Sample Application*

To illustrate how our platform works, we demonstrated by using a simple example of computing the mean. The example uses four data custodians namely P_1, P_2, P_3 and P_4 , but the approach applies in general when there are more than two P_i .

1. The client end has a searcher of data or a researcher.
2. The researcher or person doing the spatial search issues an online request for particular data by sending an encrypted random number which can be symbolic for a particular search request. He also sends a public key to all the locations identified, i.e. P_1, P_2, P_3 and P_4 when you consider the four distributed points as per the four locations. Hence the value of $n = 4$.
3. The request is then sent as an encrypted random value to the first point P_1 .
4. P_1 receives the encrypted value of the request, i.e. $E(R)$.
5. P_1 encrypts its attribute value and sends $E(R) \otimes E(P_1)$ to P_2 .
6. P_2 encrypts its attribute value and sends $E(R) \otimes E(P_1) \otimes E(P_2)$ to P_3 .
7. P_3 encrypts its attribute value and sends $E(R) \otimes E(P_1) \otimes E(P_2) \otimes E(P_3)$ to P_4 .

8. P_4 encrypts its attribute value and sends $E(R) \otimes E(P_1) \otimes E(P_2) \otimes E(P_3) \otimes E(P_4)$ to R.
9. The researcher subtracts the random number R from the value obtained after decrypting the received encrypted value to retrieve the sum.

Finally, by dividing by n the researcher retrieves the mean value of the selected attribute. This whole process is illustrated in Fig 1.

To discover the next machine to send the encrypted value to, each machine participating in the distributed computation checks its list of participating machines to find the next active machine. Extensive detail about this mechanism can be found in literature on ring algorithms in distributed systems.

Note that the encryption function is additive homomorphic, i.e., the multiplication of two ciphertexts will decrypt to the sum of their corresponding plaintexts [18]. For instance, in a voting contest, the server does the product of the ciphertexts to come up with the total number of votes and the polling assistant at a local polling station who is in charge of collecting final results decrypts the product (total votes cast) to arrive at the result. The results received from the server after decryption resembles the actual output that is expected, which proves the accuracy of encryption and decryption of Paillier cryptosystem [20].

In the method shown in Fig. 1, Paillier Encryption scheme is applied on both the client end and data provider end to ensure anonymity on both sides. The only key which is broadcast to everyone (data providers) by the researcher (client) is the Public Key.

Furthermore, a prototype implemented as distributed application was written in java using the proposed protocol. We demonstrated by using the Big Integer values in java how large mathematical computations can be carried out to transfer encrypted values, through a distributed system, that are sent as requests without decrypting them since we ensured that both ends didn't have the secret key. The fact that the Paillier Cryptosystem uses modulo arithmetic (mathematical calculations involving functions that return the remainder) and Big Integer values in Java makes it complicated and impossible for any third party or intruder to decrypt and reach the mean value of the selected attribute.

TABLE II. SECURITY ASSUMPTION OF THE HE SCHEMES AND COMPARISONS ON PROPERTIES.

HE SCHEME	HOMOMORPHIC NATURE	SECURITY ASSUMPTION	PROPERTIES OF THE SCHEMES			
			COMPUTATIONAL COST	PROBABILITY	EXPANSION FACTOR	ENCRYPTION/ DECRYPTION COSTS
RSA	Multiplicative	Integer factorization problem.	High	Deterministic	Higher	Low
El Gamal	Multiplicative	Diffi-Hellman problem	Low	√	High	Low
Goldwasser -Micali	Additive	Quadratic residuosity problem	Low	√	Moderate	Low

Benaloh	Additive	Higher residuosity problem	Low	√	Low	Low
FHE	Additive & Multiplicative	Sparse Subset Sum (SSSP) assumption	Higher	√	Complex	Complex/Higher
Paillier	Additive	Decisional Composite Residuosity Assumption (DCRA)	Low	√	Low	Lowest

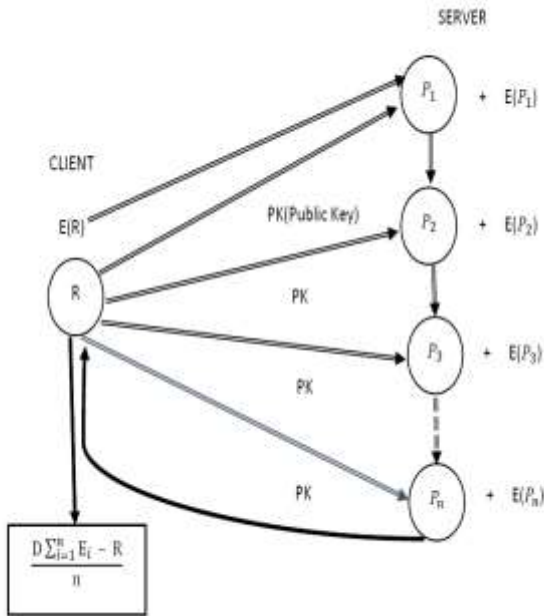


Figure 1: How the proposed protocol works that uses a distributed ring algorithm

Table III. shows the analysis of performance when we consider three scenarios as we did the research, i.e. the first one being doing the spatial search without the proposed protocol, the second one being doing the spatial search using the protocol with Paillier Encryption and the third option of running it with a protocol without Paillier encryption at all. From the results, it can be concluded that the encryption and protocol result in an increased running time. It can, however, be argued that they collectively introduce an overhead of 39% in processing time. The bigger expense comes from the network communication overhead.

After analyzing the encryption schemes using existing literature on Homomorphic Encryption, Paillier Encryption Scheme became the focus for the three experiments. In the first experiment, the spatial search was conducted without using the protocol and encryption. The average execution time after running the program ten times is shown in Table III. In the second experiment, the spatial search was conducted by using the protocol with Paillier Encryption Scheme. After running the program ten times, the average execution time is shown in Table III. In the third experiment, the spatial search using the protocol but without Paillier Encryption Scheme. The average result after running the program ten times is displayed in Table III.

Note that experiments were not conducted with other schemes here because the focus was on using Paillier Encryption scheme after selecting Paillier when the other

encryption schemes were analysed during the literature review of Homomorphic Encryption putting into consideration desirable qualities of a spatial search. It was during the literature review that the comparisons in Table II were made and Paillier was chosen as the best scheme to be used in our research.

TABLE III: ANALYSIS OF PERFORMANCE DURING A SPATIAL SEARCH.

DESCRIPTION	RUN (EXECUTION TIME)-JAVA PROGRAM	ANONYMITY/SECURITY
i. Spatial search without protocol and encryption	850 milliseconds	NIL
ii. Spatial search with protocol with Paillier Encryption	1182 milliseconds	Guaranteed
iii. Spatial search with protocol without Paillier Encryption	924 milliseconds	NIL

V. CONCLUSION

The risk of lack of anonymity and confidentiality is what a client or a data provider may experience. A key limitation is that both the user and the data provider focus on either getting information or providing data without being careful about their anonymity respectively. Therefore, protecting both sides of the equation, i.e. the client side and the data provider side may be of greater importance for spatial searches.

Homomorphic Encryption, particularly the Paillier Homomorphic encryption, supports a spatial search by providing anonymity of data of both parties involved in a spatial search while allowing analyses to be conducted on encrypted data in the encrypted space. This can be a reality when such a scheme as a Paillier cryptosystem is used to encrypt both running ends, i.e. the data provider end and the client end or a searcher. Paillier cryptosystem is cheap and computationally capable and viable compared to other cryptosystems. Analyses of large data sets such as election results can be achieved while ensuring anonymity of the one casting a vote and protecting the numbers of the votes cast. Even though the proposed solution introduces a 39% overhead this is outweighed by the benefits of the proposed approach.

REFERENCES

[1] M. Alkharji, H. Liu, M. A. Hammoshi, "A Comprehensive Study of Fully Homomorphic Encryption Schemes," International Journal of Advancements in Computing Technology (IJACT) Volume10, Number1, pp.1-24, Mar. 2018.

- [2] R. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," vol. 21 (2), pp.120-126, 1978.
- [3] Fahn, R., and M. J. B. Robshaw. Results from the RSA Factoring Challenge. Technical Report TR-501, Version 1.3, RSA Laboratories, 1995.
- [4] G.A.V.R. Rao, P.V. Lakshmi and N. Ravi Shankar, "RSA Public Key Cryptosystem using Modular Multiplication," International Journal of Computer Applications (0975 – 8887) Volume 80 – No5, p1-5, October 2013.
- [5] K. Mallaiah and S. Ramachandram, "Applicability of Homomorphic Encryption and CryptDB in Social and Business Applications: Securing Data Stored on the Third Party Servers while Processing through Applications," International Journal of Computer Applications (0975 – 8887) Volume 100– No.1, p1-15, August 2014.
- [6] Hwang, R. J., Su, F. F., Yeh, Y. S., & Chen, C. Y. (2005, March). An efficient decryption method for RSA cryptosystem. In Advanced Information Networking and Applications, 2005. AINA 2005. 19th International Conference on (Vol. 1, pp. 585-590). IEEE.
- [7] M. Tebaa and S. EL Hajji, "Secure Cloud Computing through Homomorphic Encryption," International Journal of Advancements in Computing Technology(IJACT) Volume5, Number16, December 2013.
- [8] R. Shruthi, P. Sumana and A. K. Koundinya, "Performance Analysis of Goldwasser-Micali Cryptosystem," International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 7, July 2013.
- [9] J. Bringer, H. Chabanne, M. Izabache'ne and D. Pointcheval, "An Application of the Goldwasser-Micali Cryptosystem to Biometric Authentication," Jun 2014. [Online]. Available. Accessed on: Oct 8, 2019.
- [10] A. Acar, H. Aksu, A. S. Uluagac and M. Conti, "A Survey on Homomorphic Encryption Schemes: Theory and Implementation," pp.1-35, Oct. 2017. [Online]. Available. Accessed on: Oct 9, 2019.
- [11] G. Danezis and C. Diaz, "Space-Efficient Private Search with Applications to Rateless Codes," pp. 1-15, 2007. [Online]. Available. Accessed on: Oct 9, 2019. [www.http://scholar.google.com](http://scholar.google.com)
- [12] G. M. Jacquez et al ., "Geospatial Cryptography: Enabling researches to access private, spatially referenced human subjects data for cancer control and prevention, pp. 1-26, Jul. 2017. [Online]. Available. Accessed on: Oct 8, 2019. [www.http://www.ncbi.nlm.nih.gov/pmc/articles/pmc5659297/](http://www.ncbi.nlm.nih.gov/pmc/articles/pmc5659297/)
- [13] I. San, N. At, I. Yakut, H. Polat, "Efficient Paillier Cryptoprocessor for privacy-preserving data mining," Feb. 2016, [www.http://www.doi.org/10.1002/sec.1442](http://www.doi.org/10.1002/sec.1442)
- [14] C. Amrutkar, R. Naik, I. Dacosta and P. Traynor, "Replacing Oblivious Computation with Private Search for Context Sensitive Communications," <https://pdfs.semanticscholar.org/4b0e/67efdb5792152d49655e2a4198613df7a8cf.pdf>
- [15] F. Baldimtsi and O. Ohrimenko, "Sorting and Searching Behind the Curtain," [Online]. Available. Accessed on: Oct 8, 2019. <https://pdfs.semanticscholar.org/238a/41e4283f2945870e3e7e4d79e0339086cb61.pdf>
- [16] Y. Hu, W.J. Martin and B. Sunar, "Enhanced Flexibility for Homomorphic Encryption Schemes via CRT," <https://pdfs.semanticscholar.org/cf17/7e93476c1cf70d9e6a8f8f25757641df2d42.pdf>
- [17] K. Peng, C. Boyd, E. Dawson, "A Multiplicative Homomorphic Sealed-Bid Auction Based on Goldwasser-Micali Encryption", Volume 3650, pp 374-388, Springer, 2005.
- [18] A. N. Khan, M. Y. Fan, M. I. Nazeer, R. A. Memon, A. Malik and M. A. Husain, "An Efficient Separable Reversible Data Hiding Using Paillier Cryptosystem for Preserving Privacy in Cloud Domain," pp. 1-17, Jun. 2019. [Online]. Available. Accessed on: Oct 8, 2019.
- [19] V. Biksham and V. Biksham, "Homomorphic Encryption Techniques for securing Data in Cloud Computing: A Survey," International Journal of Computer Applications (0975 - 8887) Volume 160 - No.6, February 2017.
- [20] J. Ashok, K. N. Dheeraj, C. Subhedar and R. Tiwari, "Homomorphic Encryption over Databases," International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8, June 2019.
- [21] M. Alkharji and H. Liu, "Homomorphic Encryption Algorithms and Schemes for Secure Computations in the Cloud," ICSCCT 2016 - International Conference on Secure Computation and Technology, Virginia International University, Fairfax, VA , 2016. Available. Accessed on: Oct 11, 2019.
- [22] U. Demiryurek, F. Banaei-Kashani and C. Shahabi (2010), "Efficient K-Nearest Neighbor Search in Time-Dependent Spatial Networks," International Conference on Database and Expert Systems Applications DEXA 2010: Database and Expert Systems Applications, pp 432-449. https://doi.org/10.1007/978-3-642-15364-8_36
- [23] J. Zhong, X. Meng, X. Zhou and D. Liu, "Co-spatial Searcher: Efficient Tag-Based Collaborative Spatial Search on Geo-social Network," DASFAA 2012, Part I, LNCS 7238, pp. 560–575, 2012. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.381.9314&rep=rep1&type=pdf>
- [24] N. Hamlin, " (2017) Number in Mathematical Cryptography," Open Journal of Discrete Mathematics, 7, 13-31, pp. 1-19, Jan. 2017. [Online]. Available. Accessed on: Nov 12, 2019. <http://dx.doi.org/10.4236/ojdm.2017.71003>
- [25] M. Nyirenda, H. Arimura and K. Ito, "Relaxing the data access bottleneck of geographic big-data analytics applications using distributed quad trees," 2016 5th International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, 2016, pp. 189-195.
- [26] M. Nyirenda and D. Zulu, "Speeding up construction of distributed quadtrees for big-data analytics applications using dilated integers and hashmaps," 2017 International Symposium on Networks, Computers and Communications (ISNCC), Marrakech, 2017, pp. 1-6.

The Use of Programming Environments Alice and Scratch to Teach Computer Programming to Children in Upper Primary

Makani Mweembe
 Computer Science Department
 The University of Zambia
 Lusaka, Zambia
 makanijkmweembe@gmail.com

Monica M. Kabemba
 Computer Science Department
 The University of Zambia
 Lusaka, Zambia
 monica.kalumbilo@cs.unza.zm

Jackson Phiri
 Computer Science Department
 The University of Zambia
 Lusaka, Zambia
 jackson.phiri@cs.unza.zm

Abstract—In the recent past there have been several debates as to whether computer programming should be taught to children from as early as five years old or not. Some scholars have argued that programming is too complex to be taught to children while other scholars have argued that just like reading and writing are important for children so is computer programming-computer programming is the new literacy. To this effect, programming environments such as Scratch, Alice, Minecraft, Logo and Snap! have been introduced as mediums that can be used to teach computer programming to children. In 2015 Zambia introduced ICT in primary schools, however the syllabus only exposes the learners to learning basic skills in office applications. This research focused on assessing the use of programming environments Scratch and Alice to teach children computer programming and proposed the addition of computer programming topics in the primary school ICT syllabus. Scratch was developed by MIT media lab to allow children to create their own stories, games and animations- and share their creations with the online community. Alice was created to make it easier to create animations, build interactive narratives, or programme simple games in 3D. 60 primary school pupils were taught computer programming in Scratch and Alice for 24 hours over a period of 3 months. The study used observation research during the 3 months period as data was recorded to measure the cognitive progress of the respondents. Additionally, interviews were carried out with a closed ended questionnaire was administered. Not only that, the ICT Curriculum Development Specialist was given a questionnaire. The review of the ICT curriculum was done to understand the practical activities in the ICT syllabus that would be supported by the programming environments. The use of the programming environments showed that learning to program was an enjoyable experience and that the pupils were able to understand programming concepts.

Keywords—Computer Programming, Computer programming Environment, Upper Primary, ICT syllabus, Alice, Scratch

I. INTRODUCTION

The perception that most people have of computer science specifically the field of computer programming is that it is complex. Programming is seen as an opportunity for students to develop the intellectual skills to tackle challenging problems. Recent studies have shown that children are able to learn while playing games (Prensky,2008) both knowingly and unknowingly. In the recent years, programming environments have been developed to simplify learning programming for children from as young as 5 years old. The philosophy on which

these environments have been developed is the philosophy of visualization and that children learn better from simple to complex. The basic idea is to incorporate learning into something that children perceive as fun (Meluso, Zheng, Spires & Lester,2012). Scratch is a programming language learning environment created with an idea to enable beginners to quickly create programs without having to worry to learn how to write them syntactically correct. It is intended to motivate further learning through playfully experimenting and creating projects such as games etc. (Maloney, Resnick, Silverman & Eastmond, 2010). In developed countries, learners are introduced to programming in school as early as 5 years old, whereas, in Zambia programming is not taught to children at primary school level. Starting in primary school, children from all backgrounds and every part of the UK should have the opportunity to learn some of the key ideas of computer science, understand computational thinking, learn to program and have the opportunity to progress to the next level of excellence in these activities.(The Observer, 2012).

In the recent years, computers have become a standard part of majority households (Seybert, 2012). This means that when children go to primary school they have exposure to computers. Therefore, when computers are used wisely to teach, they can improve the children learning experience by providing them with exciting problem-solving content suitable for their age. Children at this age, according to Piaget's theory of cognitive development (Wadsworth & Gray,2004) are entering the concrete operational state of cognitive development. In fact, many UK readers in their 30s will have been exposed to recursion, for example, because once upon a time many UK schools taught Logo programming, enabling children to learn how a mechanised turtle could be instructed to carry out complex manoeuvres. But in the end, most of those schools gave up teaching Logo and moved backwards to training kids to use Microsoft Word. Incidentally, the Logo story provides a good illustration of why teaching kids to write computer programs has to be an integral part of any new computer science curriculum. The reason is that there's no better way of helping someone to understand ideas such as recursion or algorithms than by getting them to write the code that will implement those concepts. (The Observer,2012). Programming is an important skill that enhances systematic thinking, problem solving and promotes creativity which are essential skills in life in the 21st Century.

The objective of the study was to assess the use of computer programming environments to support the primary

ICT syllabus and to add computer programming content to enrich the syllabus. The study brought together 60 respondents from 2 different schools of which one is from a low-density area and the other from a medium and high-density area. The respondents were in grades 5-7 and are between the ages of 10-13 years old. This study will contribute to the Zambian Ministry of Education’s policy of the Curriculum of curriculum development for ICT for primary schools. Additionally, this study will help the teachers to find easier ways of teaching computer programming and most importantly it will allow children to have the potential to become technology producers. The research sought to answer the following questions;

1. What ICT practical activities in the grade 5,6 and 7 can be supported by the computer programming environments?
2. What are the computer programming environments that would support the practical activities identified in question (1)?
3. What integrated environment would support the ICT curriculum and the computer programming environments?

II. METHODOLOGY

The study used two main methods to collect data which are observation and interviews. Two interviews were carried out, the first interview was with the ICT curriculum specialist from the Curriculum Development Center to understand the omission of computer programming in the syllabus and if it can be added. The second interview was carried out with 60 pupils in grades 5,6 and 7 to assess the use of programming environments (Scratch and Alice) to teach computer programming. Observation method was used during the 3 months period of teaching the 60 pupils to program in Scratch and Alice as data was recorded to measure the cognitive progress of the pupils.

Participants: The study had a total number of 61 respondents. 60 of the respondents were pupils in grades 5,6 and 7 who were randomly drawn from two schools which are in a low-density area and a medium-density area. 30 pupils were selected from each school, the school in the low-density area had 10 males and 20 females and the school in a high-density area had 19 males and 11 females. 1 respondent from Ministry of Education was purposefully selected.

Setting: Participants were grade 5,6 and 7 (between 10 and 13 years old) pupils enrolled at the school in the low-density area (A) and the school in the medium-density area (B). Lessons were conducted once in a week for 2 hours over a period of 3 months. Both schools had the same course description. Each class was taught by the same instructors, the author and the assistance of two teacher aids. The other participant was an expert who responded to the interview questions in a day.

Curriculum: Pupils in each class worked in a pair to complete tasks designed by the researcher. The first week focused on introducing children to computer science and its’ branches. Following that that they were introduced to the core programming topics which included; variables, loops, do together and condition statements that are shared by both Scratch and Alice.

III. RESULTS AND DISCUSSIONS

After reviewing the Zambian syllabus for ICT primary school, our findings were that programming is not part of the curriculum for primary school. The interview with the expert from Curriculum Development Centre reviewed that the ICT syllabus for grade 5,6 and 7 can be improved as computer programming environments would enrich the existing content as well as provide additional content to enrich the ICT syllabus. There are some topics that we identified from the syllabus that can be supported by programming environments, this will enrich the ICT primary curriculum.

Fig.1.1 shows some of practical activities from the syllabus that can be supported by computer programming frameworks.

5.11 Graphics	5.11.1 Graphic packages 5.11.2 Pattern creation	5.11.1.1 Open graphic package 5.11.1.2 Manipulate graphic image	<ul style="list-style-type: none"> • Opening graphic package and image • Resize, fill colour and change colour 	<ul style="list-style-type: none"> • Manipulation • Demonstration • Identification 	<ul style="list-style-type: none"> • Creativity • Application • Inquisitiveness
5.13 Saving	5.13.1 Saving files	5.13.1.1 Distinguish between Save and Save As 5.13.1.2 Name and save file in a specific location	<ul style="list-style-type: none"> • Save and save as • Naming and saving files 	<ul style="list-style-type: none"> • Manipulation • Demonstration • Identification 	<ul style="list-style-type: none"> • Application • Creative thinking • Exploration

Fig.1.1. Extracted from the Ministry of Education Technology Studies Primary School Syllabus.

We also assessed Scratch and Alice programming environments versus the curriculum and the results showed that the programming environments not only allowed learners to create stories, play games instead it teaches them creative thinking, logic, computational thinking and collaboration. These are the essential skills that the Zambian ICT primary school syllabus also seeks to achieve. After the learners using the Scratch and Alice they were able to create products and in return improved problem solving skills. It was also further discovered that Alice and Scratch programming environments equipped the learners with manipulation, identification and demonstration skills as outlined by the Zambian ICT syllabus. It was discovered that the programming environments could not only be limited to computer programming but could also be extended to other subjects like mathematics and science. This showed that programming environments would enrich the syllabus as they motivated learning through creative exploration.

An assessment on the usability of the programming environments Alice and Scratch was conducted. The assessment reviewed that 75% of the grade 5 respondents found it easy to complete tasks in Scratch and only 40% of the grade 5’s managed to complete tasks in Alice. 80% of the grade 6 and 7’s respondents were able to complete tasks in Scratch and 75% were able to complete tasks in Alice. 25% of the grade 5 respondents failed to complete tasks in Scratch while 60% failed to complete tasks in Alice. 20% of the grade 6 and 7 respondents failed to complete tasks in scratch and 25% could not complete tasks in Alice within the given time. 69% of the male participants when asked to create projects of their own choice created games, whereas 83% of the female participants created stories on

relationships and fashion. Fig.1.2. shows a story created by one of the respondents of a girl Abby who was thinking of what to wear to a dance show and a fairy appearing to say, 'Hello Abby'. Fig.1.3. shows one of the projects created by the male respondents in Alice.



Fig.1.2. Screenshot of a story by one of the female respondents showing the two sprites Abby and the Fairy.



Fig.1.3. Screenshot of the progress of a game created by one of the male participants of a car that should be moved and controlled by the arrow keys and the child as a driver.

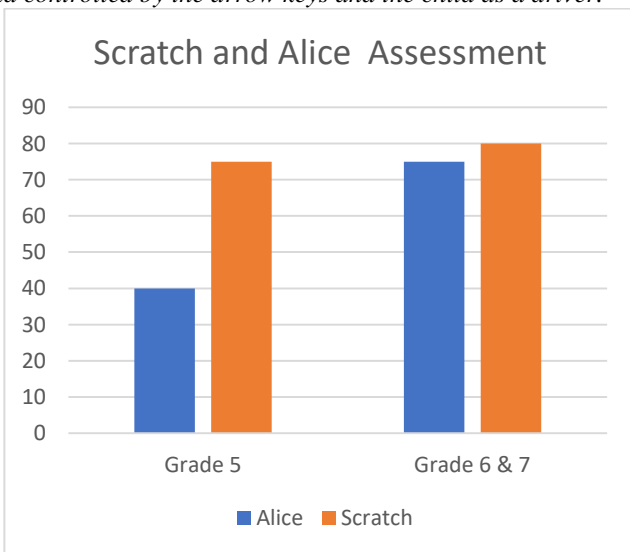


Fig.1.4. X-axis shows the grades of the respondents and the Y-axis shows the percentages of task completion.

The respondents from both schools gave feedback that Alice was more of a 'serious' environment unlike Scratch which they said had a playful interface, this explained why the grade 5's leaned more towards Scratch. Alice was embraced by the grade 6 and 7's as it was more challenging. With this established we made observations of the features that were in the two programming environments that would have contributed to the results in Fig.1.4. Below is a table that summarizes the features of Alice and Scratch.

Feature	Scratch	Alice
Functionality	Scratch is easier to use and understand. It has clear blocks to drag and drop. Eight (8) blocks include Motion, Looks, Sound, Pen, Variable, Control, Sensing, Operators.	Alice has a more complex interface and it has only 2 blocks classified as functions and procedures.
Stage Setup	By clicking on a sprite/object the respondents were able to write code and create events for that sprite.	In Alice all the code that is written was stored in 'MyFirstMethod' tab and for the respondents to create events for the objects they had to make Alice listen for certain actions like a click from the 'InitiazeEventListener' tab.
Availability	Scratch allowed the respondents to share their projects to the scratch online community immediately.	The option of sharing the project is there but proved difficult because it required the other peer to open it in Alice.
Printing	Scratch has the option of printing, but it prints the code as pictures which proved difficult got the respondents to review.	Alice allows for printing the code in html form. This allowed the respondents to be able to print and review the code.

Fig.1.5. shows a comparison of the features in Alice and Scratch

Fig.1.6. shows the results from the interview that was conducted with the 60 respondents. The results reviewed that 67% of the respondents from the school in the medium-density area found using Alice easy to learn about controls. 47% of the respondents from the school in the low-density area found using Alice to learn about the concept of controls easy.

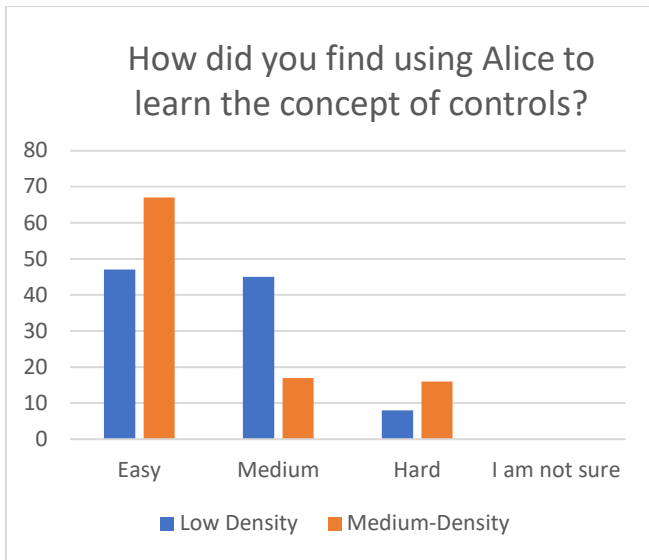


Fig.1.6. Shows a sample question from the interview on the concept of controls in Alice and how the respondents found using Alice for controls.

On the other hand, Fig.1.7 below shows how easy the respondents understood the concepts of repeats in Scratch. 67% of the respondents from the school in the low-density area found using Scratch to learn about repeats easy while only 27% of the respondents from the school in the medium density found it easy.

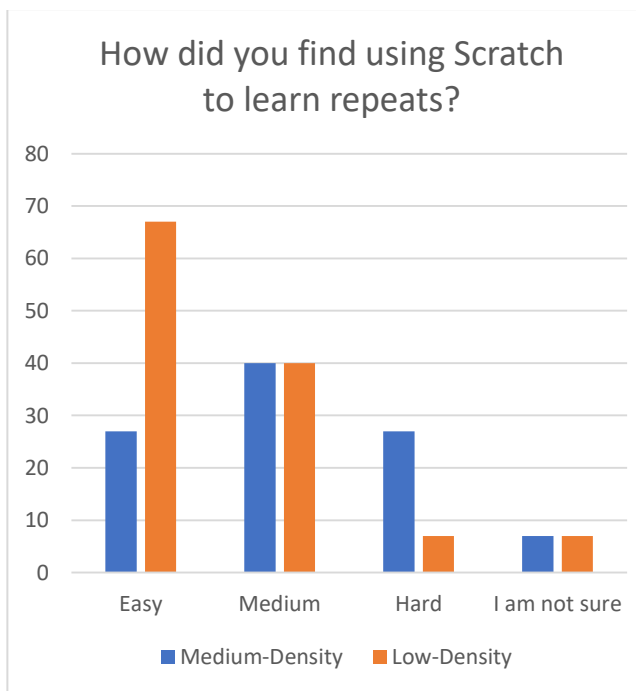


Fig.1.7. Shows a sample question from the interview on the concept of repeats in Scratch and how the respondents found using Scratch for repeats.

Fig.2.3 the survey reviewed that. 67% of respondents from the school in the low-density school were able to easily access the code anywhere whereas only 27% of

respondents from the school in the medium-density area could access the code anywhere.

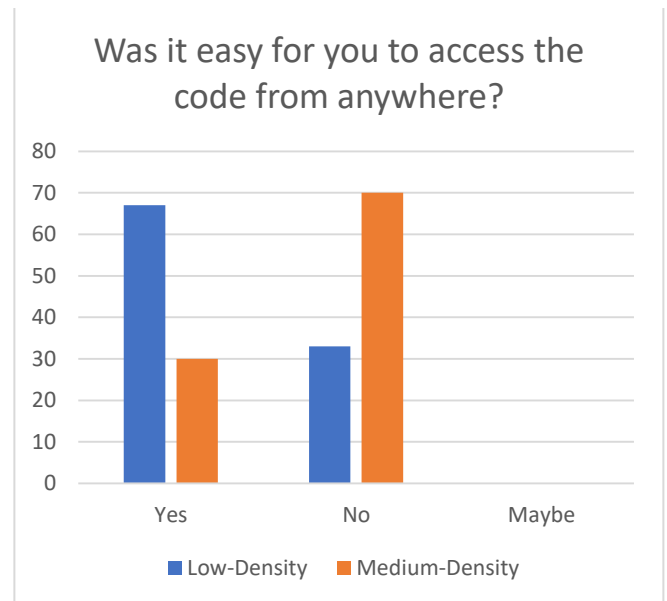


Fig.1.8. Shows a sample question from the interview on how easy it was to access the code anywhere in Scratch.

The survey further showed that 83% respondents from the school in the low-density area found it easy to navigate in Scratch whereas 67% respondents from medium-density school found it easy to navigate in Scratch

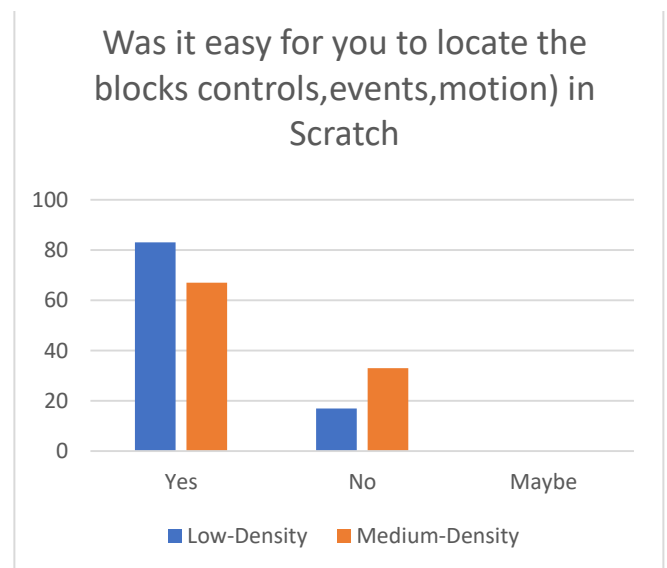


Fig.1.9. Shows a sample question from the interview on how easy it was to locate blocks in Scratch.

The results of the study suggest that the primary schools pupils in Grade 5,6 and 7 are able to use the programming environments Scratch and Alice to not only learn programming concepts but also learn logical thinking, problem solving and improve their creativity.

IV. CONCLUSION AND RECOMMENDATIONS

Programming should be added to the syllabus for learners in grade 5,6 and 7 and computer programming frameworks should be used to teach to enrich the syllabus. We are proposing both programming environments Alice and Scratch to be used. This is so because the findings suggest that Scratch is easier to understand and is suitable for grade 5, Alice is a bit more advanced and is suitable for grade 6 and 7. After assessing the use of programming environments and the syllabus, we are now going to use the recommended topics to teach the computer programming to children in upper primary.

Below is a sample of the addition that we are proposing to be added to the ICT syllabus for grade 5;

Theme 1:Stage setup

Sub-Topic:Dragging and dropping sprites

Sprites characterication

Specific Outcome:Sprites should have a complete character and outlook.

Theme 2:Animation

Sub-Topic:Computational thinking concepts of loops,events and parallelism.

Get introduced to the concepts of sequence

Specific Outcomes:Create an animated project.

Below is a sample of the addition that we are proposing to be added to the ICT syllabus for grade 6 and 7

Theme 1:Build a Scene

Sub-Topic:Dragging and dropping objects

Manipulating object joints

Using camera views

Specific Outcome:Build a complete scene such as a castle and a water scene and placing all the objects

Theme 2:Programming in Alice

Sub-Topic:Code editor overview

Using procedures overview

Using Do-Together

Specific Outcomes:Learners to program their first program of "Hello world"

It will be beneficial for learners to be exposed to Scratch first then later Alice which will expose them to understand object-oriented programming concepts. The use of the programming environments will be beneficial to the learners as they will not have to worry about syntax errors instead they will focus on grasping the concepts. The other benefit of using the programming environments is that the first positive experiences that the learners will have using the programming environments will motivate them to want to take computer programming as a career thus increasing their chances of being solution providers. Adding the use of programming environments to the syllabus will not only help the learners in ICT but they will be able to apply the knowledge gained from using the environments to other subjects like science, mathematics and humanities.

V. ACKNOWLEDGMENT

We would like to thank the Ministry of Education and Curriculum Development Centre for the help rendered. We would also like to thank the management of the schools that allowed us to work with their pupils. Special thanks go to the University of Zambia Department of Computer Science for the advice and expertise rendered to this project.

REFERENCES

- [1] Green,T. R. G.,& Petre, M.(1996). Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework. *Journal of Visual Languages & Computing*, 7(2), 131-174.
- [2] Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E. (2010). The Scratch programming language and environment. *ACM Transactions on Computing Education*, (TOCE), 10(4), 16.
- [3] Meluso, A., Zheng, m., Spires, H. A., & Lester, J. (2012). Enhancing 5th graders' science content knowledge and self-efficacy through game-based learning. *Computers and Education*, 59(2), 497-504.
- [4] Palfrey, J., & Gasser, U. (2008). Born digital: Understanding the first generation of digital natives. Basic books
- [5] Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas*. New York: Basic Books
- [6] Prensky, M. (2008). *Digital game-based learning*.
- [7] Seybert, H. (2012). Internet use in households and by individuals in 2012. *Eurostat Statistics in Focus*, 50/2012
- [8] Wadsworth, B., & Gray, W. M. (2004). Piaget's theory of cognitive and effective development. Pearson/A and B.

Understanding the Transactional Challenges Encountered in Village Banking While Exploring a Blockchain Enabled Solution

Zaima Banda
*Computer Science Department,
 University of Zambia,
 Lusaka, Zambia,
 zaima.banda@gmail.com.*

Mayumbo Nyirenda,
*Computer Science Department,
 University of Zambia,
 Lusaka, Zambia,
 mayumbo@gmail.com*

Abstract— Village Banking concept seems to be growing at a fast rate in the Zambian environment as most people are now joining simple groupings as a way of providing financial support to one another. They achieve this by making contributions on an agreed periodic basis towards the group's treasury. Individuals have a plethora of payment options to choose from, namely bank transfers, mobile wallet payments or cash. There are however a number challenges mainly attributed to the manual approach of handling finances in the village banking groups and a plethora of payment mechanisms. Our study takes cognizance of these payment modes to set up a payment ecosystem on a public blockchain ledger that will allow members to move funds into and out of the village banking platform with ease and in a manner where everyone in the group will be able to see and monitor the activities in the group clearly.

Keywords— Village banking, blockchain, payment ecosystem

I. INTRODUCTION

Village banks usually constitute of a group of low-income entrepreneurs who come together to share and guarantee one another's loans [1]. In this effort members of a group bring together their monies periodically. Based on these collections members then come to some decision on how the money can be used to empower the members of the group. Consequently, the main activities in a village banking group are making a contribution and receiving money from the group. There is a plethora of approaches for doing this. One of such approaches is the ever increasing in popularity use of Mobile money payments. Mobile payments refer to any payment made using a mobile device. In today's environment we see mobile devices as a fundamental part of our daily lives [2]. The ever-increasing smartphone dependence coupled with good internet service has resulted in development of various methods that allow consumers to pay conveniently through a phone. To keep a balanced ledger of finances based on different mobile payment platforms, blockchains [3] can be used as an incorruptible digital ledger of transactions. Thus, mobile devices coupled with blockchains can be used to automate collection of contributions as well managing them.

This study takes leverage of the existence and extensive footage and usage of mobile phones to provide a platform that will allow for maximization of a payment solution that is

bank and mobile money service provider independent. Transactions through this profile will be processed at a more affordable rate and will cut across different platforms. We solution an application that will connect different users and allow for seamless movement of funds, we hope to achieve this through a distributed blockchain public ledger that will provide an interface for different bank's client and mobile money users to connect and transact directly with each other/member in a village banking grouping. The ledger provides the advantage of transparency and redundancy in the group. Our solution basically provides an alternative payment method that makes it easy for individuals to exchange and move funds across different platforms irrespective of the bank or mobile money organization they are affiliated to. Because of these different platforms people usually face challenges when they have to transfer funds to others in a village banking setup, this therefore necessitated for us to conduct a survey of the different payment modes people use and then solution an application that will address their issues. Our solution provides an environment where all members are able to track and monitor how funds are moving in the group.

In this paper we begin with looking at some related works into village banking and blockchains related to our study, we then proceed into a discussion of our methodology in section III, we discuss the work done and results in section IV, and finally we give our concluding remarks in section V.

II. RELATED WORKS

Most recent studies show the need for financial inclusion. Village banking seems to address the issue of financial inclusion quite well as more of the unbanked are seemingly coming on board. Governments and donors looking to strengthen rural finance systems should consider the role that Village banking institutions can play, given that many Village banking institutions already have a strong rural presence (Glenn D. Westley, 2004). The main problem for people in underdeveloped countries is not only a shortage in capital resources, but also limited access to financial services, specifically bank and savings accounts (Kai Schmidt & Philipp Sandner, 2017).

A lot of success stories have been studied and its evident that village banks show steady but modest loan increases (Judith Painter and Barbara MkNelly, 1999). While (Glenn

D. Westley) shares some best practices on village banking as a way of showcasing the learnings and progress that have been made thus far by looking at some practical examples. Blockchains also put some emphasis on financial inclusion, they are considered as a development vehicle empowering people directly and mitigating power asymmetries (Kai Schmidt and Philipp Sandner. 2017). Blockchain can be regarded as a distributed ledger technology, which on the one hand provides a platform for participants to interact with each other and on the other hand serves as a transparent, comprehensible and trustworthy repository of data and information (Stefan Seebacher and et, 2018). Michael Pisa and Matt Juden in their paper entitled “Blockchain and Economic Development: Hype vs. Reality” of 2017 do provide a clear view of the blockchains technology potential in the context of development of facilitating faster and cheaper international payments, providing a secure digital infrastructure for verifying identities, making and disbursement of funds in a more secure and transparent manner. They contend that the blockchain technology can be regarded as a primary driver of long-term economic growth [10]

Conventional banks haven’t quite adopted blockchains as the solution is seen to be a threat to their existence, however, some recent studies seem to be investigating how banks can leverage from this solution, a lot of scholars are slowly trying to see how best the technology can be embraced.

Most literature reviewed do not tackle a solution that combines village banking and blockchains into one solution as a way of addressing the issue of financial inclusion. Village banking structures are less complicated than conventional banks and do provide a good platform for adoption into the banking sector albeit not the conventional banking as we know it. We, therefore, emphasize on the need for Implementation of a Zambian based Distributed ledger, that can be used as a platform for managing activities in village banks.

III. METHODOLOGY

The current payment modes used in village banking groupings are very manual and fragmented with different payment channels appearing in their own capacity and representing individual company’s wallets. The wallets are of two types, those originating from Banks and those originating from Mobile Money Operators. The fall back is usually manual cash, which is exchanged if individuals don’t have compatible bank accounts or compatible mobile money wallets.

To carry out this research we took a three-pronged approach that started with analysing the current modalities of making and receiving contributions in village banking groups. Secondly, we analysed the main business processes used in a village banking group, this made us understand the features that our solution should have. With that understanding, a blockchain powered solution was arrived at as the best method of responding to the challenges identified in the baseline study. Thirdly, we implemented a mobile application over a blockchain ledger that enhances management of finances in a village banking group.

We discuss the features of our solution in the results section.

A. Proposed Blockchain Model

Our Blockchain model addresses the shortfalls encountered in a manual village banking set up as the solution is provided over the internet through a distributed ledger. Banks, Mobile Money Operators and Card Schemes will be able to interface with our application to provide a platform where funds will be able to move across the different platforms with ease.

Figure 1 provides the modules or components that will connect to the blockchain ledger. Current payment modes available in the village banking groupings will be able to interface with our solution as shown in Figure1.

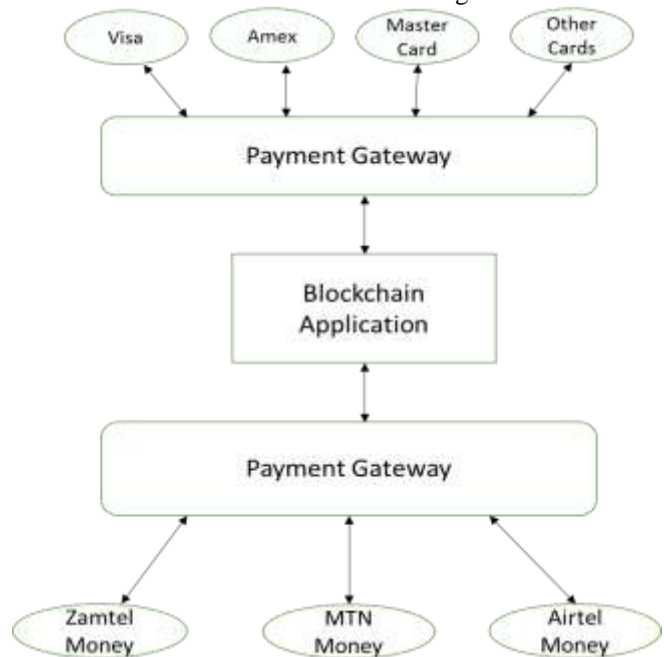


Fig. 1. Blockchain Model

The key features that this model addresses are discussed below:

1) Trust

Our blockchain solution represents a system that has a high degree of trust as transactions that are done between any two peers can be seen by everyone else. Movement and tracking of funds are done in a transparent manner. At any one time, it is easy to see the individual contributions that each member has made and at the same time check the total balance of the group.

2) Scalability and stability

The solution is provided over the internet and allows for as many individuals to be added to the system. This attribute makes our system stable and as such its able to grow as the numbers in the group grows and it can be accessed from anywhere.

3) Easy monitoring of activities

It is very easy for members in the group to know every detail of their operations as they are able to see the

transactions as they move from one individual to another, this will have the advantage of insuring that there is transparency and a higher degree of trust in the group.

4) *History and audit trail*

Any movement of funds will be seen and tracked in the system. This record will be maintained in the system and provide an audit trail from the first transaction done to the most recent. This audit trail or history will also help with query resolution in case of a dispute.

5) *Easy to move funds to peers in the group*

The system will allow a platform where members easily move funds from their bank accounts or mobile money wallets to the blockchain ledger. Each member in the group will have their own account reflective of what they have contributed. These funds will be stored in a digital currency and that will allow for easy movement between peers.

6) *Decision making is done consensually and not through one individual*

Good group dynamics should always be at play where finances relating to the whole group are concerned. Blockchains provide an avenue where consensus can be driven when lending out finances to ensure that the objectives and goals of the group are adhered to in all dealings.

IV. WORK DONE AND RESULTS

In this study we aimed to understand the different modes people use to transact in a village bank. We share some of the key results related to study in this section.

Our results indicated that there are three major transactions modes in use, these are Cash, mobile money transfers and bank transfers.

Figure 2 below captures these results with cash representing 25.5%, Bank transfers were at 22.5% while mobile money transfers stood at 27.5%. These results are a clear indication of the fact that mobile money is the preferred mode that individuals use in village banking.

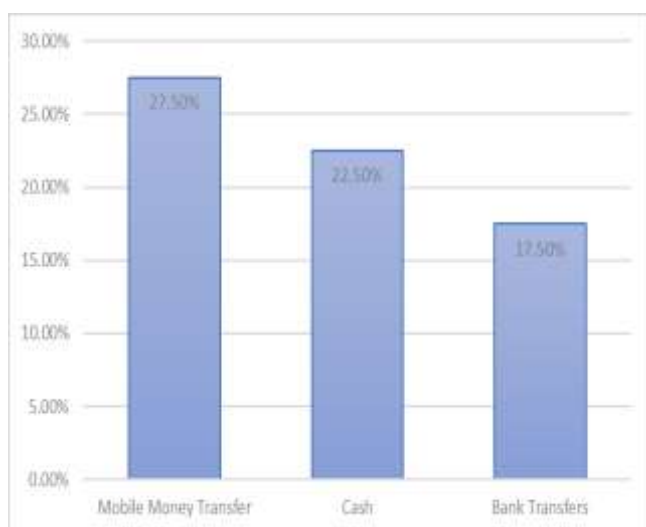


Fig. 2. Payment modes in Village Banks

A. *Key business processes in a village banking group*

The first step in our study was to understand the key business processes in village banking group. Using these key processes, we then proceeded to formulate a questionnaire to investigate the challenges faced in a village banking group. We first present these key operational and business activities that represent the core activities that take place in village banks.

Figure 3 below highlights the key business processes that are conducted at defined periodic times in a village banking group. The cycle is usually kick started by getting contributions from members. The group treasurer then collects the funds created and provides an update to the group. When all the contributions are in, a meeting is called and loan requests and other businesses are discussed, after which approvals or declines are made for the various loans requests received.



Fig. 3. Business processes/activities in village banks

Figure 4 highlights the key transactions in village banking, transactions flow between individuals and between the group and individuals.



Fig. 4. Key processes in village bank

B. Challenges encountered with moving funds in the current village banking groups

Irrespective, of whether a wallet or cash is used, transactions in the village banking group are done at a peer to peer individual level, other members or the entire group may not be privy with the transaction details while the two parties are transacting. Information would only be communicated to the group after facts. As with most manual systems, a set up based on word of mouth and trust would prove difficult as numbers in the group grow and if there are untrustworthy members in the group.

We tackle these challenges faced in village banking groups in this section.

These business and transactional activities present some associated challenges based on the feedback captured during the survey we conducted, we discuss them as follows:

1) Monitoring payments

This accounted for 11.8% of the total responses. This refers to how difficult it is to track how funds move between peers. The current setup is such that only two peers are able to know the details of any payment at any one time. While, individuals move their savings to the group’s coffers it is important to monitor how those funds are being gathered and to whom payments are made to. All payments and movement of savings need to be done in a more transparent manner rather than verbally.

2) Tracking collections

A better method of tracking contributed payments is needed to ensure that everybody in the group is contributing towards a common goal as specified in their original mandate. This issue represented 20.6% of the total responses.

3) Making Payments to member on a different platform (Bank and Mobile Money)

It is difficult to move payments that across disparate platforms, this accounted for 29.4 % distributed equally between making payments to other banks and making payments to individuals with a different mobile money provider. The challenges are usually as a result of extra associated costs in the form of costs, time and convenience. Village banking users will try to mitigate this issue by just dealing in cash. There is therefore a need for a system that will be convenient and allow for easy movement of payments.

4) How to deal with excess cash

This also came up as an issue and accounts for 20.6%. This issue is usually encountered when the group is not able to lend out any funds because there is no one ready to borrow the group’s funds. A situation that wouldnt be there if a platform were one village banking grouping would be able to lend to another group. Though not directly covered in this study as it will involve having agreements in place between two village banking grouping, we envision to address this by having two village banks establish agreements that will allow them to lend to each other.



Fig. 5. Respondents’ Preferences when funds are received

5) Approving payments to be lent out

Where finances are raised as a group even consensus must be obtained from a reasonable portion of representatives instead of one individual making the most decision. This accounted for 11.8% of the responses we gathered. There is therefore, need to have a consensus mechanism locked into the system that will provide approvals for all funds leaving the group coffers.

Figure 6 below summaries the key challenges captured in our study as discussed above.

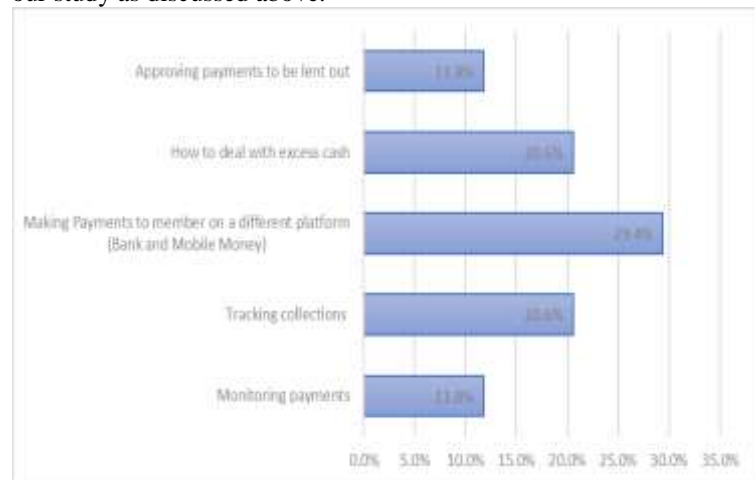


Fig. 6. Challenges encountered in Village BanksZ

Other type of challenges encountered in village banking are those related to the technology used, specifically, how a payment solution is accessed or how the provider of the payment solution provides the solution or how the payment solution is used. These have been summarized as ‘barriers’ and include issues like internet, charges, lack of proper mobile phone, bank system outages or mobile company outages. Figure 7 below highlights these barriers.

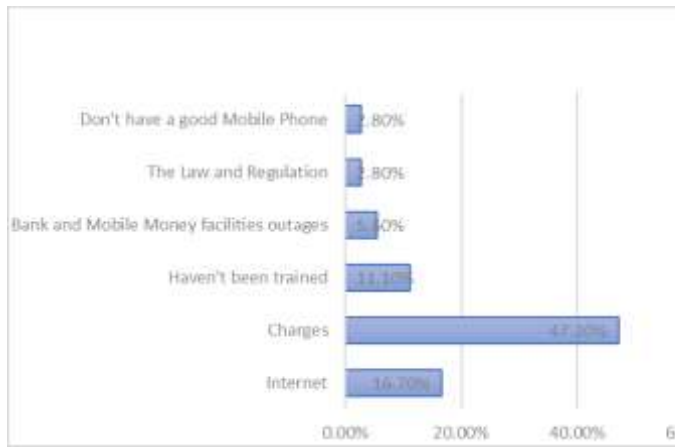


Fig. 7. Barriers with transferring funds in Village Banks

C. The Application

Most world economies have been moving away from hard currency in favor of electronic payment systems for many years before the arrival of bitcoin [11]. We note an energy system flow in its simplest form as follows:

$$\text{Hard Currency} \rightarrow \text{Electronic Payments} \rightarrow \text{Digital Currency} \quad (1)$$

Hard currency or cash and electronic payments are what are in use in village banks in today’s environment, our payment ecosystem will however accommodate the movement of funds both from banks and mobile network money operators into a digital form, which we are calling as the Zedcoin, and will be equivalent in value to the local currency, the Zambian kwacha.

$$1 \text{ Zedcoin} = 1 \text{ ZMK} \quad (2)$$

Mobile money transactions are the preferred mode that individuals in a village banking use as opposed to hard currency. Usually, distance and convenience of handing it is the major reason why people prefer to use mobile transfers as opposed to the other modes.

We consider equation (1) as a form of dynamic system in which the different modes of payments are competing for some state and obtainability. It is, therefore, imperative that we understand the environment in which they operate and how they compete so as to formulate a better solution. Logically, we see an environment in which cash is being destroyed by electronic payments which we postulate will be destroyed by digital currency in the foreseeable future. These modes of payments do converge at some point and allow for competitive coexistence to occur which will reach equilibrium at some point. Many competing species have been coexisting in the same habitat over time, resulting in a high species diversity [12]. The dynamics that will be at play will be modeled with adaptations using Lotka Volterra equations [13] expressed in the form:

$$\frac{dC}{dt} = aC - bCP \quad (3)$$

$$\frac{dP}{dt} = -dP + eCP \quad (4)$$

Where,

- C Represents volume of hard currency users,
- P Represents volume of Electronic payment users,
- a Is the rate of increase of hard currency users,
- b Represents volume of hard currency users,
- d Is the rate of increase for electronic payment users,
- e Electronic payment reduction rate,

Following, the Lotka Volterra model, hard currency would be the prey and electronic payment is the predator in the first instance (pre- solution implementation).

With a periodic systematic introduction of our solution we will see a situation whereby electronic payments become the prey and the digital currency we introduce becomes the predator.

With time we will experience an increase in the number of digital currency users until a time where equilibrium will be reached where:

$$\frac{dC}{dt} \cong 0; \frac{dP}{dt} \cong 0 \quad (4)$$

Using this model, we developed a blockchain based solution using open source Hyperledger for the server end and flutter and dart programming for the mobile application.

V. CONCLUSION

A Blockchain solutioned application addresses the shortcomings that are encountered in a manually managed village banking grouping. It is envisioned that if this solution is adopted at individual village banking groups, monitoring and compliance issues would be easier and manageable, this would therefore create a situation where two or more village banking groups can be linked together to constitute a bigger and more composite group of village banks whose savings can be amassed and put into bigger and more meaningful income generating investments. An electronically managed village banking group would therefore present more financial and social opportunities for a growing nation like Zambia. Individuals in the group would have better understanding of the monies they are saving and how those monies are being used and invested, this is quite contrary to the conventional banking norms at play at the moment. This, in our view, will encourage more people to jump on the village banking bandwagon. We postulate that our solution will revolutionize the financial sector as it combines village banking and blockchain technology to achieve a blockchain powered financial inclusion. Our application will be that vehicle used for this change we envision.

REFERENCES

- [1] FINCA, Village Banking, <https://finca.org/our-work/microfinance/financial-services/village-banking/>. [Accessed 29th May 2019]
- [2] Emily Sorensen, Different types of mobile payments explained, [Accessed 29th May 2019]
- [3] Anja Meironke1 et al, Business Process Compliance and Blockchain: How Does the Ethereum Blockchain Address Challenges of Business

- Process Compliance? 14th International Conference on Wirtschaftsinformatik, February 24-27, 2019, Siegen, Germany
- [4] Glenn D. Westley, "A Tale of Four Village Banking Programs, Best Practices in Latin America", Inter-American Development Bank, Washington, D. C. Sustainable Development Department
- [5] Kai Schmidt, Philipp Sandner, Solving Challenges in Developing Countries with Blockchain Technology, FSBC Working Paper, OCTOBER 2017
- [6] Painter, Judith and McKnelly, Barbara (1999) "Village Banking Dynamics Study: Evidence from Seven Programs," *Journal of Microfinance / ESR Review*: Vol. 1: Iss. 1, Article 6.
- [7] Stefan Seebacher and Maria Maleshkova, "A Model-driven Approach for the Description of Blockchain Business Networks", Proceedings of the 51st Hawaii International Conference on System Sciences, 2018
- [8] Max Raskin & David Yermack, "Digital currencies, decentralized ledgers, and the future of central Banking", Working Paper 22238, National Bureau Of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138, May 2016
- [9] Maica Krizna, A. Gavina et al, "Multi-species coexistence in Lotka-Volterra competitive systems with crowding effects", Article number: 1198 (2018)
- [10] Anja Meironke1 et al, Business Process Compliance and Blockchain: How Does the Ethereum Blockchain Address Challenges of Business Process Compliance?, 14th International Conference on Wirtschaftsinformatik, February 24-27, 2019, Siegen, Germany
- [11] Yining Hu et al, A Delay-Tolerant Payment Scheme Based on the Ethereum Blockchain <https://www.researchgate.net/publication/322851718>, January 2018
- [12] Michael Pisa and Matt Juden. 2017. "Blockchain and Economic Development: Hype vs. Reality." CGD, Policy Paper. Washington, DC: Center for Global Development.
- [13] N. Bacaer, "A Short History of Mathematical Population Dynamics", DOI 10.1007/978-0-85729-115-8 13, © Springer-Verlag London Limited 2011

Categorisation of Sexual Reproductive Health Short Messages Texts into Thematic Areas Using Text Mining

Tobias Makai
 Department of Computer Science University of Zambia
 Lusaka, Zambia
 tobias.makai@cs.unza.zm

Mayumbo Nyirenda
 Department of Computer Science University of Zambia
 Lusaka, Zambia
 mayumbo.nyirenda@cs.unza.zm

Abstract—In a quest to enhance behavioral change among adolescents in Zambia, the National HIV/AIDS/STI/TB Council in collaboration with UNICEF developed the Zambia U-Report platform to grant young people increased access to information on various Sexual Reproductive Health topics via Short Messaging Service (SMS) messages. Over the years, the platform has accumulated millions of text messages which need to be categorized into key thematic areas for better tracking of sexual reproductive health knowledge gaps among young people. Categorization of messages is currently done manually – a process that has proven to be both tedious and time-consuming. In this paper, we present a proof of concept for the automation of the categorization process of text messages that over 198,150 U-Report subscribers have been sending to the platform, into key thematic areas for better analysis using existing data and text mining techniques. The paper discusses the model currently being used to categorize information on the platform and the proposed prototype that automates the categorization of the unstructured text messages into key thematic areas. Through an exploratory research method using semi-structured interviews and observation, we established how the categorization of messages is currently done and obtained a list of categories that U-Report counsellors have over the years adopted. We collected a copy of the U-Report database and used the list of categories as a primary guide to build and train a text categorization model using supervised machine learning algorithms for data and text mining. Preliminary results show that an accuracy of 75% can be reached using a Support Vector Machine. Though not as accurate as the counsellors the faster categorization gives huge benefits.

Keywords—*Knowledge Discovery in Text (KDT), Machine Learning (ML), Multi-class Text Classification, Multi-label Text Classification, Natural Language Processing (NLP), Sexual Reproductive Health (SRH), Text analysis, Text Categorization, Text Classification, Text extraction, Text mining, Zambia U-Report*

I. INTRODUCTION

Zambia U-Report is a youth-led SMS-based HIV/AIDS response initiative platform accessible via the short code 878. As a platform for sharing Sexual Reproductive Health (SRH) information, U-Report enables subscribers to ask questions on issues that are not adequately explained in the system-generated SMS responses. The platform allows a team of counsellors to respond to specific messages sent by subscribers on any psychosocial and SRH issues. Launched in 2012 by the National HIV/AIDS/STI/TB Council (NAC) in collaboration with UNICEF to promote behavioral change among adolescents and youths through information sharing, Zambia U-Report has over the years accumulated a humongous pool of information that is not fully utilized for

decision making and impact assessment due to difficulties in categorizing it.

Since its launch, the platform has recorded more than 3 million user-to-platform (and vice versa) interactions with over 198,150 subscribers. Due to the humongous number of messages, it is difficult for the National HIV/AIDS/STI/TB Council (NAC) to effectively categorize U-Report data in order to identify the issues affecting young people the most as well as find out and track emerging issues outside SRH. Currently, categorization of data on the platform is done manually and requires counsellors to analyze each text message in order to allocate an appropriate category from a chosen list of thematic areas. This process is hectic and time-consuming especially that it is the counsellors hired for the purpose of providing counselling services to subscribers via the platform that are given the extra task of manually categorizing the messages.

This paper details the current categorization process and presents a proof of concept for the automated text classification.

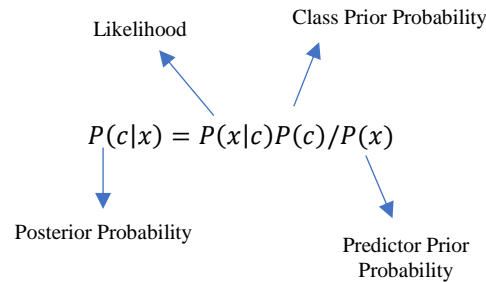
II. TEXT MINING TECHNIQUES

Text mining, also known as Intelligent Text Analysis, Text Data Mining or Knowledge-Discovery in Text (KDT), refers generally to the process of extracting interesting and non-trivial information and knowledge from unstructured text [1]. Textual data is unstructured, unclear and manipulation is difficult [2]. However, in today's world there are several text mining techniques for information extraction, categorization, summarization and topic tracing that can be employed to discover patterns in large pools of digitally available text data to inform decision making. Added to the available techniques are open source Machine Learning (ML) and Natural Language Processing (NLP) frameworks, libraries, toolkits and algorithms.

In related work, Rosa and Ellen [12] experimented with applying text classification methodologies to a new domain, that of 'micro-text' chat entries in military chat using different classifiers – among them Support Vector Machine (SVM) and Naïve Bayes (NB). Balabantaray et al. [13] designed an emotion classifier model trained using SVM to classify text extracted from Tweeter into emotional categories such as positive, negative and neutral to assess the possible emotions of the persons behind the individual tweets. Our work is similar because it also focused on text extraction. There have been many text classification algorithms that researchers have used to solve text classification problems.

1) Naïve Bayes

Naïve Bayes is one of the most popular and simple machine learning classification algorithms. It is a classification technique based on the Bayes' Theorem with an assumption of independence among predictors [6]. Bayesian classifiers assign the most likely class to a given example described by its feature vector [8]. The algorithm determines the probability of specified outcomes under the conditional independence assumption that each word in a text is independent of the rest. In other words, Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors [9].



$$P(c|X) = P(x_1|c) \times P(x_2|c) \times P(x_n|c) \times P(c)$$

Equation 1

Where

- $P(c|x)$ is the posterior probability of class (target) given predictor (attribute). $P(c)$ is the prior probability of class.
- $P(x|c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predictor.

2) Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges [6]. It is a computer algorithm that learns by example to assign labels to objects. For instance, an SVM can learn to recognize fraudulent credit card activity by examining hundreds or thousands of fraudulent and non-fraudulent credit card activity reports [10]. Classification is achieved by plotting the dataset features as individual coordinates on an n-dimensional space. Classification is performed by finding the hyper-plane providing a clear distinction between two classes.

3) Logistic Regression

The use of logistic regression is routine in the social sciences when studying outcomes that are naturally or necessarily represented by binary variables [12]. Logic regression searches for Boolean (logical) combinations of the original predictors that best explain the variability in the outcome variable and, thus, reveals variables and interactions that are associated with the response and/or have predictive capabilities [11].

III. METHODOLOGY

To understand the current process, an exploratory research method was used and data was collected through interview sessions with the counsellors employed by the

National HIV/AIDS/STI/TB Council to categorize the messages. We also observed how they manually do the categorization. A semi-structured interview guide with a list of questions was developed and used during the interview sessions. Two (2) out of the three (3) counsellors were interviewed independently to ensure the data collected from one was devoid of any influence from responses provided by the other. Part of the data collection was aimed at understanding the actual tools used and the list of common categories adopted over time which served as a precursor to achieving our second objective which was to develop the prototype for automating the process. A copy of the Zambia U-Report database was then collected for use in building the model that automates the process. From the database, we extracted a dataset of 1,349 incoming text messages to which we added labels as part of the data pre-processing tasks preceding the creation of a model trainable using machine learning. The model training process involved separating the dataset into two (2) parts; 944 text messages for training and another 405 text messages for testing.

IV. CURRENT CATEGORISATION MODEL

The first objective for our study was to learn and establish a model currently used to categorize information on the Zambia U-Report platform into key thematic areas while the second objective was to develop a prototype that automates the categorization of U-Report information into key thematic areas. This section discusses the findings of the study on each objective in depth.

Through exploratory research methods using semi-structured interviews and observation, it was established that the National HIV/AIDS/TB/STI Council employed three (3) trained psychosocial counselors to provide counselling to adolescents and youths via the U-Report platform. The counsellors were given an additional task of categorizing each text message they respond to into an appropriate topic. Categorizing messages into their respective thematic areas is particularly useful in helping NAC identify the issues that are affecting young people in their daily lives in order to tailor interventions to their information needs.

Each counsellor categorizes the text messages by noting down a list of chosen categories and marking them accordingly using a pen and a notepad into a tally sheet as shown in Fig. 1. The counsellor assesses the message and counts it as part of one or more of the listed categories immediately after responding to it. Counting a particular text message as part of a given topic is purely determined by the counsellor based on the core theme of the message in accordance to their expert assessment or judgment. Some messages tend to fall under multiple categories due to how they are presented. For instance, the question 'Is HIV an STI?' is marked under the category 'HIV' as well as under 'STIs' whereas the question 'What is HIV?' is simply marked under the category 'HIV'.

Fig. 1. Primary Topic tally sheet for U-Report messages created on 21st April 2019.

As shown in Fig.1, there is a category labelled ‘Other’ for messages that do not fall under any of the listed categories. The ‘other’ category helps providers to identify new domains for some similar but frequently received messages with no category – it helps in discovering emerging issues. Despite being an already tedious process, it becomes more challenging for counsellors to perform the dual task of responding to the subscribers’ messages and marking them into their respective categories. At the end of each month, all counsellors sit down to collectively aggregate the tallies in each category to be able to generate a monthly report. The report is then submitted to a special committee of stakeholders known as the U-Report Core Group chaired by and hosted at NAC for use in decision making. In cases where the notepad is lost, damaged or misplaced, there is always a challenge reporting accurate numbers as the counsellors are required to trace the messages they responded to during the month and redo the categorization. Tracing the old messages on the platform is always difficult as there are no filters to aid the search process. Figure 2 shows a sample monthly aggregate for April 2019 representing the sum of tallies from all 3 counsellors.

	1st - 7th April	7th - 20th April	21st - 28th April
HIV	17	37	22
STIs	14	42	21
Pregnancy	4	15	5
Sex	11	34	16
Condoms	5	10	10
Circumcision	2	1	4
Menstruation	2	4	6
Birth Control	2	6	3
Relationships	3	14	5
ARVs	7	12	9
HIV Testing	4	12	2
PrEP/ PeP	1	8	9
2-for-5/ U-Report	12	15	6
Transmission	15	28	15
Prevention	6	17	8
Drugs/Treatment	1	10	14
Health	6	9	9
Diseases	9	24	14
PMTCT	3	7	2
Gen Questions	13	14	3
Others	4	14	10
Masturbation	1	16	-

Fig. 2. Sample monthly aggregate for April 2019 resulting from the manual categorization of the text messages.

It was observed that each counsellor takes a duration of approximately 15 minutes to complete their tasks on the first message and 5 minutes on subsequent messages. The duration is broken down as follows; approximately 10 minutes to jot down the list of categories onto a notepad, 3 minutes to respond to the message and 2 minutes on the actual categorization of the message as the counsellor has to first analyze the message before counting it under a category. Factoring in ambiguities that emanate from some shorthand text, the language used (local language), environmental distractions and platform downtimes, the duration sometimes increases. Table I shows our observation of the durations.

TABLE I
AVERAGE TIME SPENT BY EACH COUNSELLOR RESPONDING TO THE FIRST MESSAGE

Activity duration →	Counsellor 1	Counsellor 2	AVG
Listing topics	10.5mins	9.8mins	10.15mins
Responding to text message	2.7mins	3.4mins	3.05mins
Categorizing text message	1.7mins	2.4mins	2.05mins
Total			15.25mins

The current flow described above is summarized in Figure 3 below as a key result to achieving our first objective;

Automating the categorization process will therefore reduce the time counsellors spend on each text message received on the U-Report platform.

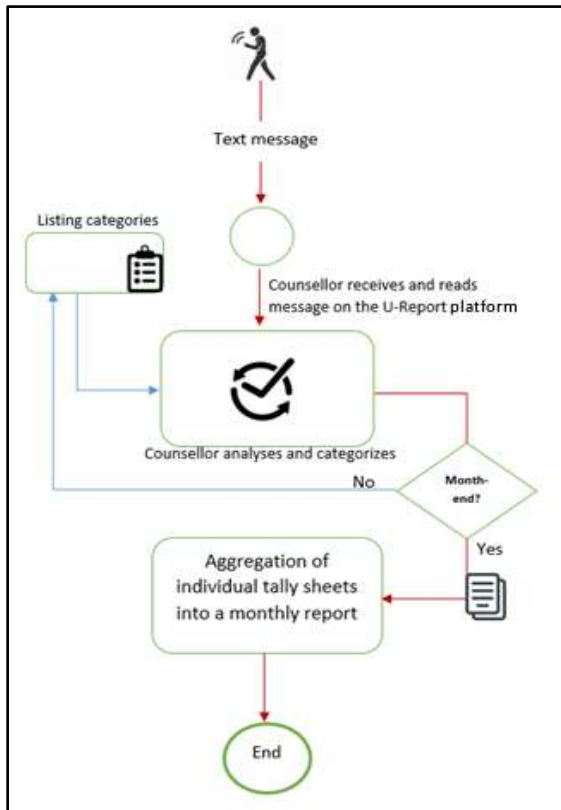


Fig. 3. Model for the current manual categorization of U-Report messages.

After a full understanding of the current categorization process and the challenges that come with it, we began building the prototype for the automation. We elected to use machine learning with natural language processing to achieve Text classification. Text classification, which involves assigning one or more predefined categories to a free text according to its content, plays a key role in organizing the massive sources of unstructured text information into an organized state [3]. Because our primary goal was to provide a solution that automates the current manual process and not alter it, we built our model using the labels and topics from the tally sheets provided by the counsellors.

Building a text classification model is a 3-step process. The steps include; Data preparation, Feature Extraction and Model training. The section details the steps we undertook to build and train the prototype.

1) *Data preparation*

The final three steps of the dataset preparation and subsequent stages were implemented using Scikit-learn – an open source machine learning library in the python programming language.

Data preparation is the process of loading a dataset and performing basic pre-processing [6]. The automation process began with the extraction of a copy of the U-Report database from which 1,349 text messages were extracted for training and testing the model. The text messages were then labelled in accordance with the guidance provided by counsellors. The dataset was split into 944 and 405 text messages for training and testing, respectively. Figure 4 illustrates the dataset preprocessing steps.

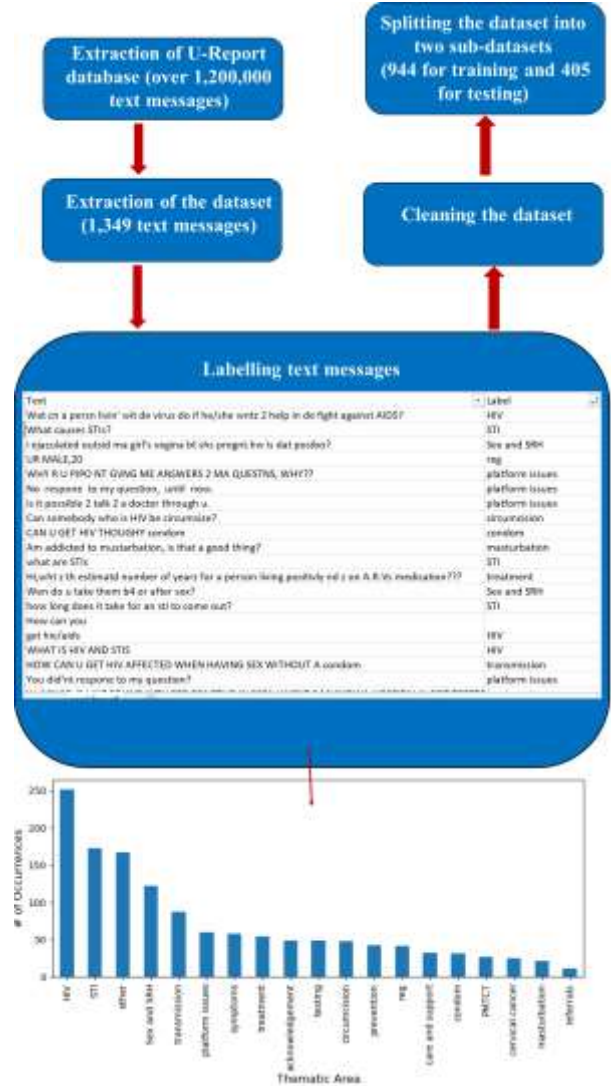


Fig. 4. Dataset preparation.

2) *The Feature Extraction and Classification Model*

In this step, raw text data is transformed into feature vectors and new features are created using the existing dataset [6]. To handle this process, our model was powered by three (3) text classification algorithms, namely; Naïve Bayes, Support Vector Machine and Logic Regression processed in python using the Scikit-learn machine learning library. We chose the 3 algorithms as our initial set of classifiers from which we can choose one that will achieve a

significant accuracy. Figure 5 illustrates the model training process.

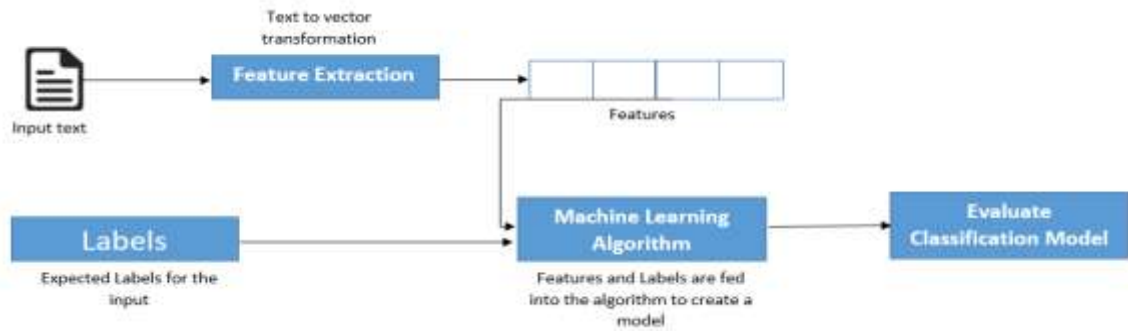


Fig. 5. Feature Extraction and Classification Model.

V. PRELIMINARY RESULTS

Preliminary results of the study are an evaluation of the prototype bordering on the topic prediction accuracies of the algorithms selected to train the model. These are listed in Table II.

TABLE III
EVALUATION OF ALGORITHMS' TOPIC PREDICTION ACCURACY

ALGORITHM	ACCURACY
NAIVE BAYES	0.53086
SUPPORT VECTOR MACHINE	0.75061
LOGISTIC REGRESSION	0.68642

The Support Vector Machine achieved the highest accuracy among the chosen algorithms at 75% while Naïve Bayes scored the lowest at 53%. These are preliminary percentages of correct predictions made by the classification model without insight into what types of errors each classifier is making.

VI. CONCLUSIONS AND RECOMMENDATIONS

This study presents a proof of concept that the automation of the categorization of messages on the Zambia U-Report platform is possible through supervised machine learning. Preliminary results have shown that the model can so far predict a thematic area for a particular text. Current focus is to continue training the model while increasing the size of the dataset. The most accurate classifier will be chosen to be used to meet the third objective of the study which is to compare the manual process against the automated model in terms of efficiency. For future work we plan to continue training the model. Only 0.1% of records have been used so far.

REFERENCES

- [1] V. Gupta and G. S. Lehal, "A Survey of Text Mining Techniques and Applications." *Journal of Emerging Technologies in Web Intelligence*, Vol. 1, No. 1, Aug 2009.
- [2] N. Yogapreethi and S.Maheswari, "A Review on Text Mining in Data Mining." *International Journal on Soft Computing (IJSC)* Vol.7, No. 2/3, Aug 2016
- [3] Y. Qin and Xiu-kun Wang, "Study on Multi-label Text Classification Based on SVM." *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, 2009
- [4] T. Wanga and H. Chiang, "Solving multi-label text categorization problem using support vector machine with membership function." *Neurocomputing* 74, 2011, pp. 3682–3689
- [5] A. Clare and R. D. King, "Knowledge Discovery in Multi-label Phenotype Data."
- [6] S. Bansal (2013, Apr. 23) *A Comprehensive Guide to Understand and Implement Text Classification in Python* [online]. Available: <https://www.analyticsvidhya.com/blog/2018/04/a-comprehensive-guide-to-understand-and-implement-text-classification-in-python/>
- [7] *An introduction to machine learning with scikit-learn* [online]. Available: <https://scikit-learn.org/stable/tutorial/basic/tutorial.html>
- [8] I. Rish, *An empirical study of the naive Bayes classifier* [online]. Available: <https://www.cc.gatech.edu/~isbell/reading/papers/Rish.pdf>
- [9] *Naïve Bayesian* [online]. Available: https://www.saedsayad.com/naive_bayesian.htm
- [10] W. S. Noble, "What is a support vector machine?" *Nature Biotechnology*, 24(12), 2006, pp.1565–1567
- [11] M. Banerjee, C. Filson, R. Xia and D. C. Miller, "Logic Regression for Provider Effects on Kidney Cancer Treatment Delivery" *Computational and Mathematical Methods in Medicine*, 2014, pp. 1–9.
- [12] K. D. Rosa and J. Ellen, "Text Classification Methodologies Applied to Micro-text in Military Chat" *International Conference on Machine Learning and Applications*, 2019
- [13] R. C. Balabantaray, M. Mohammad and N. Sharma "Multi-class twitter emotion classification: A new approach" *IJAIS* Vol 4, No.1, Sep 2012

Medical Equipment and Laboratory Services Support System

Joe Phiri ¹, Mayumbo Nyirenda ²

¹Department of Computer Science, University of Zambia, Lusaka, Zambia, ²Department of Computer Science, University of Zambia, Lusaka, Zambia

phirijoe26@hotmail.com, mnyirenda@unza.zm

Abstract—Medical equipment are key to the successful implementation of the antiretroviral therapy (ART) program and for uninterrupted services there is need for equipment to be monitored and managed through a structured program. The Ministry of Health and partner organizations has procured over the past 12 years approximately 700 chemistry, 900 hematology, and 1000 CD4 and over 2000 other medical equipment to support services at the different levels of health Care. This paper focuses on the development and implementation of a Medical Equipment Decision Support Information System for tracking the availability of consistently functional equipment throughout the 6 United States Agency for International Development (USAID) supported provinces of Zambia. USAID through the President Emergency Plan for AIDS Relief (PEPFAR) program has also provided equipment use and training support for 500 facilities in the six provinces to optimize outcomes of equipment placements and ensure quality management. Equipment monitoring, and tracking is critical to determining the overall consistent functionality and care of analyzers and has a significant bearing on the quality of services.

Keywords—Ministry of Health, web based application, decision support, business intelligence, cloud model, big data

I. INTRODUCTION

HIV/AIDS is considered the deadliest epidemic in the 21st Century. Zambia is not exempt, experiencing a national HIV prevalence rate of 20% among adults the age of 15 and 49. HIV/AIDS is ranked the number one killer, accounting for 30% deaths in Zambia. Within the Country over 1.2 million people are infected with HIV and 60,000 new infection cases are recorded every year [1]. This paper focuses on the development and implementation of a Medical Equipment and Laboratory Services Support System (MELS) for tracking the availability of functional ART labs within districts throughout the 6 provinces of Zambia (Copperbelt, Luapula, Central, Northern, Muchinga and North western), using placement of ART machinery by type (CD analyzer, Hematology analyzer, Chemistry analyzer, Immunology and Supporting equipment) and classification (major or minor), while also tracking the functional status, Planned Preventive Maintenance (PPM), breakdowns, repairs, gaps between breakdowns, and vendor response turnover time, the hope is to measure the effects on service delivery due to system unavailability or down time [2].

II. LITERATURE REVIEW

Several research studies have been conducted on medical healthcare decision support systems and researchers have

proposed and developed what could be, great methodologies in the design of a good health informatics decision support systems especially for patient care and logistics management. In this paper we have looked at some of the health information systems and research works around the subject of health informatics, and have learnt how other researchers and scientists have developed systems that has contributed to the improvement of disease surveillance and brought positive development to the health needs of mankind around the World [3].

We looked at Zambia's National Electronic Health Systems (smart care), Zambia's Electronic Logistics Management Information System (ELMIS), District Health Information System (DHIS2) and other health information systems that support HIV diagnosis, care, treatment, and disease surveillance in Zambia and across the globe. We also conducted research on other systems and research works through the internet search for published papers, technical reports and other key documents. The search was performed through PubMed database, google scholar and Google search engine. Key policy documents including guidelines, Strategic Plans, laboratory Standard Operating Procedures, legislations, unpublished monographs, and reports have also been reviewed [4].

A. Zambia's National Electronic Health Systems (Smartcare)

The SmartCare electronic health record system (EHR) has been developed and deployed by the Zambia Ministry of Health (MoH) in collaboration with the Centers for Disease Control and Prevention (CDC) and many other implementing partners [5]. SmartCare is:

- A fully integrated electronic health record system to provide continuity of care
- A clinical management information system at the facility and district (management/admin) level
- A key component in 'one National M&E system. SmartCare data is held at each facility in a distributed design; unlike centralized designs of most systems. Internet is not essential working with SmartCare. Uses client care cards, individual's health information is stored on a very compressed, secure care card to maintain continuity of care between visits, health services and health facilities. Smartcare Supports touch screen technology for data capture and GIS data visualization [7].

B. Zambia's Electronic Logistics Management Information System

Zambia, in collaboration with Tanzania, built an electronic logistics management information system (eLMIS). ELMIS is a web based application which is accessed by the user using a web browser. Data is stored at a central database [8]. ELMIS links health facilities with the central store to collect and distribute logistics data in real time. Knowing which medicines are used and which medicines are required helps supply chain managers provide continuity of supply for patients. The eLMIS enabled a transition from a paper-based system of data management to an electronic format which has fostered better, faster, and more accurate reporting of supply chain data, reduce stock outs of health commodities, and ultimately provide better access to medicines to improve health outcomes [9].

C. District Health Information Systems 2 (DHIS2)

The DHIS 2 is a tool for collection, validation, analysis, and presentation of aggregate statistical data, tailored to integrate health information management activities [10]. DHIS 2 is developed by the Health Information Systems Program (HISP). It is a modular web-based software package built with free and open source Java frameworks.

D. Implementing and measuring the level of laboratory service integration in a program setting in Nigeria

The article outlines a quantitative before-and-after study conducted in 122 Family Health International (FHI360) supported health facilities across Nigeria. A minimum service package was identified including management structure; trainings; equipment utilization and maintenance; information, commodity and quality management for laboratory integration.

E. African Regional report

The African Regional report published by the WHO-AFRO indicate public health laboratories responsibility for providing timely and reliable results primarily for the purpose of disease control and prevention. Clinical laboratories are responsible for providing accurate diagnosis of ongoing recent or past infections for appropriate case management [11]. Accurate and reliable Clinical laboratory testing is an important component to a public health approach for diseases management in resource limited settings. Laboratory data is essential for clinicians to accurately assess the status of patients' health, make accurate diagnoses, formulate treatment plans and subsequently monitor the effects of treatment [12].

F. Zambia National Laboratory policy and strategic plan

The document outlines the importance of health systems and delivery of focused health care. The medical laboratories are often a neglected component of health systems in poor resource countries. The laboratory services play a pivotal role in patient care and management and is a source of reliable information for policy development and health planning. There has however been insufficient long term investment in the strengthening of laboratory systems and services in many resource – poor settings with a high burden of major diseases such as HIV AIDS and TB [13]. Many poor resource countries with high prevalence of HIV AIDS

experience severe challenges with planning, organization, management, and delivery of accessible, quality laboratory services to support the national scale-up of testing and care [14].

The situation analysis of medical laboratory services in Zambia highlights a number of major constraints in providing quality laboratory services. These include a lack of basic inputs, such as supplies and equipment, poor infrastructure and limited human resource. In order to address these constraints, a national Medical Laboratory policy was developed in 1997 [15].

In 2006 – 2010 Zambia developed the Ministry of Health national planning strategic plan with a key objective of "Attainment of the Millennium Development Goals (MDGs) and national health priorities" the plan articulates issues on Laboratory strategy to support the ART scale up

G. The readiness of the national health laboratory system in supporting care and treatment of HIV/AIDS in Tanzania

The article outline Tanzanians well-established national health laboratory network sufficient to support HIV care and treatment services. However, laboratories at the primary health care level are constrained by inadequate resources and operate within a limited capacity. Improving the laboratory capacity in terms of number of qualified personnel, staff training on the national guidelines, laboratory diagnostic tools and coordination should be given a higher priority [16].

III. METHODOLOGY

The research assumed a retrospective description describing the current and previous systems and how their implementation have influenced the management of HIV AIDS in Zambia. Both qualitative and quantitative approaches have been undertaken to provide information and knowledge about this research.

A. Baseline study

The methods used for data collection in this study was through record inspection (requisition forms, Functionality forms, Status forms, delivery notes, EQA form and QA/QI forms), and interviews with the Zambia Prevention, Care and Treatment program national Lab technical advisor, provincial lab staff, MOH national chief equipment officer, principle biomedical scientist and provincial equipment officers.

B. Model Design Methodology

The Agile software development methodology was used, for its incremental and constant iteration between the MELS application design and the end users.

The MVC (Model, View, Controller) design pattern was followed, allowing for easy understand of the coding details of the MELS application. The prototype is designed to run as a three tier comprising the user interface Client Tier (View), the business logic Tier (Controller) and the database Tier (Model). The figure below outlines the 3-tier (MVC) design.

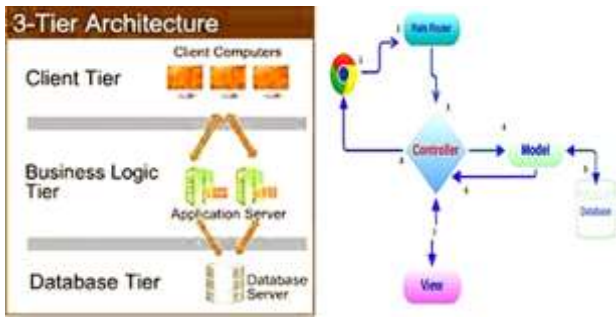


Fig. 1. MELS MVC – 3 Tier Architecture

The Client tier is made up of the Cascade style sheet, JavaScript and HTMLX files. The business logic tier is made up of the server side PHP programming language files and the database tier contains the system MYSQL database files written using the structured query language (SQL).

The NetBeans integrated development environment (IDE) was used to write and code the application and the APACHE WAMP server was used for remote testing and running of the application.

C. MELS flowchart daigram

Figure 2 below describe the systems flowchart. On a local server (APACHE), the application can be accessed using a web browser on the web address: localhost/mels. Users have to logon into the application using a correct combination of an email and password. The application supports different roles and system account types that can only be created and managed by the system administrator. Sensitive data is highly encrypted using a hash algorithm encryption method. The bootstrap framework have been used to make the MELS application mobile user friendly. The JavaScript has been used for the system security and system data validation.

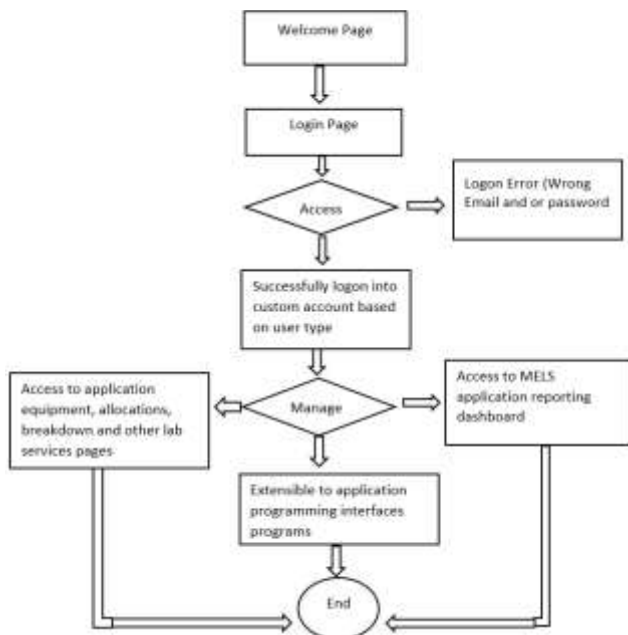


Fig. 2. MELS flowchart

Users have access to account management (edit user details) and only limited to the rights set and defined by the system administrator. Users can manage the creation of new records, update and delete the records. Users can access the systems reporting dashboards with well analyzed and visualized reports (table, graph and mapping reports) to aid decision making.

The MELS system also allow for system interoperability with other application. The prototype has been developed using open source technologies, PHP programming language, bootstrap framework, and MYSQL database and hundreds of libraries which are all available for use at no cost.

The MELS system require that users log out to end user session and to make sure that no unauthorized third person can have access to the system resources. Also the systems logs out and ends user session automatically every after 45 minutes if the account logged on stays idle for that long.

D. MELS use case

Figure 3 below shows the MELS application use case.

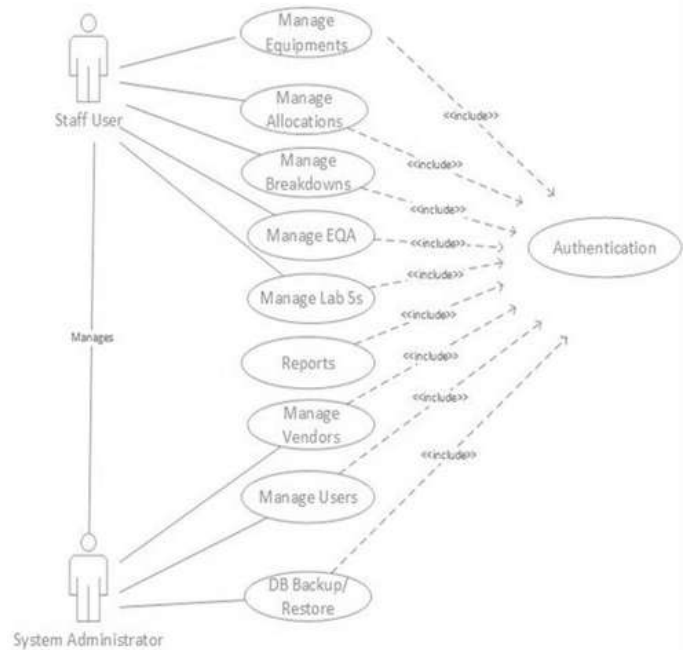


Fig. 3. MELS use case

The use case depicts the user interaction with the MELS application. The Systems Administrator manages the users (Creates, delete, updates, grants and revokes rights) and ensures application and databases are up and running at all time. The Users (MOH and Partner staff), each based on their roles and responsibilities have defined limited access to resources.

E. MELS Enhanced Entity Relationship Daigram (EERD)

Figure 4 below shows the MELS application enhanced entity relationship diagram.

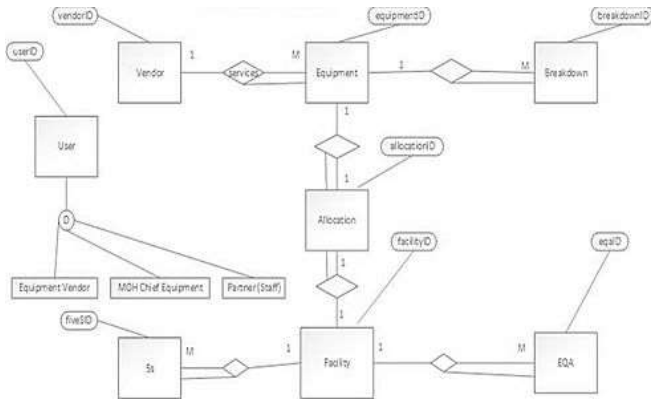


Fig. 4. MELS EERD

The EERD shows the key entities, attributes, degree of participation and relationships among the entities. The EERD is mapped into the logical schema and implemented as the database physical layer.

IV. RESULTS AND DISCUSSION

Based on MOH secondary data, it has been discovered that 90% of MOH staff have received computer training, 90% of health facilities have computer systems set up and installed either by MOH or implementing partners. It can be anticipated that if the prototype is implemented, 90% of the medical equipment used in the ART can be captured and assigned to the correct facilities of each district in the system. The Ministry of Health with its implementing partners can track the type and classification of equipment availability in each facility, district and province. Staff can track the Planned Preventive Maintenance for each piece of equipment, track equipment history usage (breakdowns and repairs) for each equipment, track current status in real-time and determine the sites where the systems are fully functional for the purpose of sample referral. The prototype can also be used to measure the turnaround time of the vendors during breakdown and repair process.

F. MELS access on local server

Figure 5 below shows an access login screen. Users need to provide their correct access credentials email and password. Access to resources is limited based on user roles.

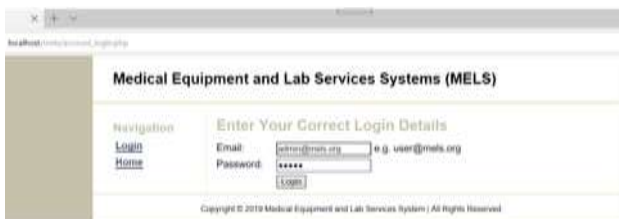


Fig. 5. MELS access on local server

G. Equipment dashboard reports

Figure 6 below shows a dashboard report (aggregated and grouped by equipment type) indicating the availability and placement of the equipment on a national level. The sample data is based on MOH/partner secondary data loaded into MELS for the period 2017. A total of 500 equipment was captured across the 30~ health facility from Copperbelt, Central, Northern,

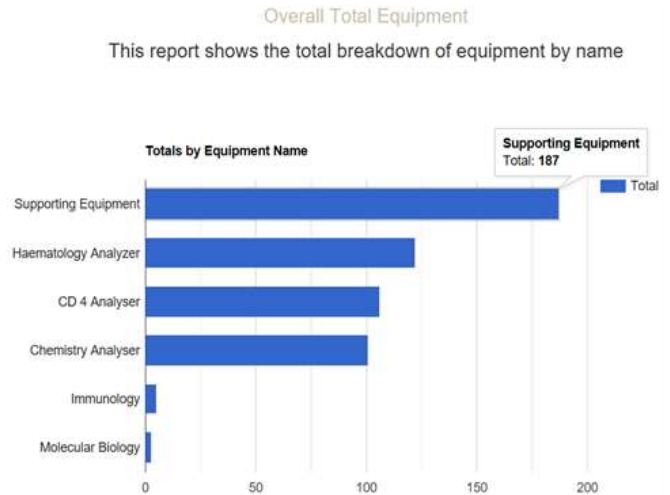


Fig. 6. Equipment dashboard report

Luapula and North Western. Supporting equipment account for 37%, Hematology analyzer 24%, CD4 analyzer 21%, chemistry analyzer 20%, while immunology and molecular biology equipment's accounting for 1% and ~1% respectively.

H. Repairs vs breakdown report

Figure 5 below shows a line trend for the equipment breakdown against the repairs for the 500 sample equipment for the whole year 2017. On average 12 equipment's breakdown against an average 7 repair rate on a monthly basis.

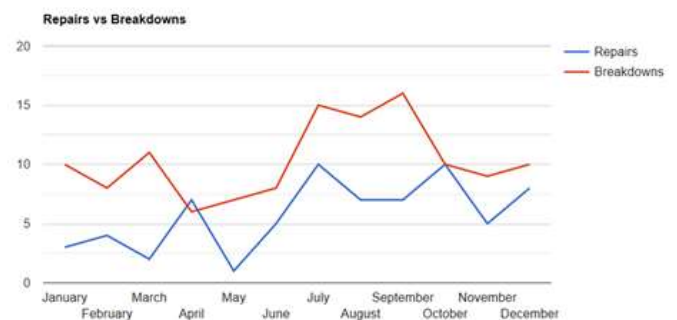


Fig. 7. Repairs and breakdown report

CONCLUSIONS AND RECOMMENDATIONS

The prototype described provides a tool that can be used for quantifying the likelihood of improved outcome while planning to reach the UNAIDS three 90s. It can be anticipated that with an enhanced equipment management approach, tracking and monitoring can be improved. It can also be expected that quality patient management can be an outcome with an enhanced equipment management approach. Future research can be done on how best to

implement software and hardware into a method that will fully support DSS cloud-based computing technology services for managing equipment in a resource limited setting such as Zambia.

ACKNOWLEDGMENT

We would like to thank the Ministry of Health and USAID ZPCT program for the support rendered in producing this work. Special gratitude goes to Dr. Mayumbo Nyirenda and the UNZA Department of Computer Science team for all the guidance and leadership provided during the course of this research.

REFERENCES

- [1] H Mbah, Negedu-Momoh “Implementing and measuring the level of laboratory service integration in a program setting in Nigeria” international Journal, September 2014.
- [2] A Desale, B Taye, G Belay, A Nigatu “Assessment of laboratory logistics management information system practice for HIV/AIDS and tuberculosis laboratory commodities in selected public health facilities in Addis Ababa, Ethiopia” international journal, June 2013.
- [3] “Zambia Ministry of Health national planning strategic plan 2006 - 2010” national policy strategies
- [4] Cassim N, Coetzee L, Motlonye B, Mpele N, Glencross DK “Piloting a national laboratory electronic programme status reporting system” international journal, April 2016.
- [5] Nader CM, Tsevat J, Justice AC, Mrus JM, Levin F, Kozal MJ, Mattocks K, Farber S “Development of an electronic medical record-based clinical decision support tool to improve HIV symptom management” international journal, July 2009.
- [6] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.
- [7] Martin R, Hearn T, Ridderhof J, Demby A. Implementation of a quality systems approach for laboratory practice in resource-constrained countries. *AIDS*. 2005;19:S59–65.PubMedView ArticleGoogle Scholar

An Insight of Smart Health Care as Growing Application Need in Zambia

Darius Bwalya
 Directorate of Information
 Communication and Technology
 Copperbelt University
 Kitwe, Zambia.
 Email: darius.bwalya@cbu.ac.zm

Christopher Chembe
 School of Science, Engineering and
 Technology
 Mulungushi University
 Kabwe, Zambia
 Email: cchembe@mu.ac.zm

Victor Kasonde
 Directorate of Information
 Communication and Technology
 Copperbelt University
 Kitwe, Zambia.
 Email: vkasonde@cbu.ac.zm

Abstract— With the current trend of world population exploding, there is a rising constraint on the provision of quality healthcare. This exactly describes the nation Zambia in which according to the United Nations Department of Economic and Social Affairs: Population Division, the current population is now well over 18 Million against a limited number of medical staff, creating a situation where access to a doctor is characterized by long waits which escalate in hours in most cases. Add to that management of chronic illness. It is undeniably clear that there is a great need to explore other viable healthcare alternatives. The aim of this paper is therefore to highlight what the components of Smart Health Care (Cloud Computing and IoTs) are and how the two technologies can collaborate as well as the challenges arising from this and how they can be overcome. The approach has been to systematically investigate literature on the two technologies smart health care and Internet of Things(IOT) and the CloudIoT paradigm. Analysis of the literature results revealed that there is an emerging form of healthcare provision called Smart Health Care in which Cloud Computing and Internet of Things (IoT) is expected to complement each other, creating a great possibility to automate health care. Finally the paper highlights how this technology can be adopted in Zambia in its endeavor to improve provision of healthcare by discussing the research conducted on overcoming challenges identified.

Keywords— *Cloud Computing, IoT, CloudIoT, Smart Health Care, Wireless Body Area Network(WBAN)*

I. INTRODUCTION

The population explosion currently being experienced world over is placing a great constraint on the provision of high quality health care for all. Zambia is a typical example whose current population as reported by United Nations Department of Economic and Social Affairs: Population Division, has exploded to over 18million currently against limited number of medical doctors. According to the World Health Organization[1], the ideal doctor to patient ratio is 1 to 5,000 respectively whereas in Zambia the ratio is shockingly at 1 doctor per 12,000 patients.

This situation results in an unacceptably long waits before having access to the doctor which in many cases is over an hour. Appointments can last even a year or more. Chronic illnesses that require constant management are increasingly common. Add to that the difficulties in life that comes with getting aged. It is undeniably clear that there is a

great need to explore other viable healthcare alternatives. Thanks to ICT technology, there is an emerging form of healthcare provision called Smart Health Care in which Cloud Computing and Internet of Things (IoT)[2] [3]is expected to complement each other, creating a great possibility to automate healthcare. Before Zambia adopts Smart health Care, it must be understood and appreciated by the stakeholders. This paper is therefore a systematic literature investigation of an insight of Smart health Care provision. The paper highlights what the components of Smart Health Care (Cloud Computing and IoTs) are and how the two technologies can collaborate as well as the challenges arising from this. Finally the paper shows how this technology can be adopted in Zambia in its endeavor to improve provision of healthcare. It is now imperative that important aspects of the research are explored to provide an insightful perspective these being: Cloud Computing, IoT, CloudIoT paradigm, Smart Health Care.

II. RELATED WORK

A. Insight on Cloud Computing

Cloud Computing revolves around servers on the internet providing services to clients in terms of infrastructure, Platform and software. The services highlighted are classified into service models as being: Infrastructure-as-a-Service(IaaS), Platform-as-a-Service(PaaS) and Software-as-a-Service(SaaS) respectively. Perhaps one of the all-encompassing definition of Cloud Computing is one provided by Mell & Grance in [4] when they define it as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. The main service models are now briefly explained:

- (a) Infrastructure-as-a-Service (SaaS):- Under this model clients or customers are provided with infrastructure services by the service providers including networks, storage, computing allowing a client or cloud user to deploy and run application as well as system software. The benefit arising from this is that the cloud user does not need to control or manage the infrastructure but has full control over applications and system software. Additionally the user will have nothing to do with the knowledge of where his or her information or

application is stored and actually running(Transparency concept).

- (b) Software-as-a-Service(SaaS):-This service model implies that the cloud user or customer uses applications of the service provider running on the cloud. Mode of access of these applications could be via interfaces such as web browsers and other Application portals. Control and knowledge of servers, network, processing scalability involved, storage is not necessary which turns out to be the benefit.
- (c) Platform-as-a-Service(PaaS):-This service model according to [4]provides capability to the consumer by means of deploying consumer-created and or acquired applications created using programming languages, libraries and tools supported by the provider onto the cloud infrastructure. Similar benefits that apply to other models apply here also being hiding or separating from the user management and control of network, servers, OS or even storage apart from user deployed and perhaps configurations of the host environment.

The service models are now summarized in figure 1.0 below

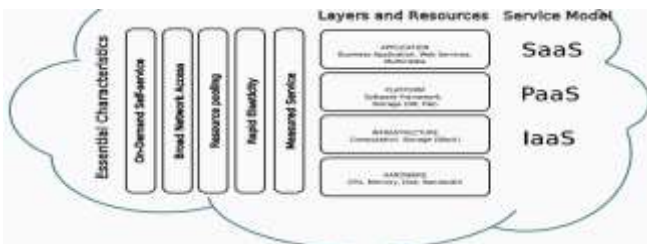


Fig.1 [3] an overview of cloud computing paradigm

Benefits and Barriers

[5] highlighted some of the benefits as follows:

- Great potential in lowering the cost of entry for smaller firms trying to benefit from compute-intensive business analytics that were hither to available only to the largest of corporations.
- It can provide an almost immediate access to hardware resources, with no upfront capital investments for users, leading to a faster time to market in many businesses.
- Makes it possible for emergent of new applications and services that were not possible before.

[5]however also noted some of the major concerns of going cloud, warning that there may be :

- Security and privacy concerns for maintaining privacy of an individual’s information.
- Reliability of connectivity: adoption depend on having optimum internet connectivity as well as power. This also implies that services must available 24/7 basis.
- Political issues to do with global boundaries: For cloud computing to continually evolve into a borderless and global tool, it needs to be separated

from politics. Currently, some major global technological and political powers are making laws that can have a negative impact on the development of the global cloud.

Despite the challenges [5] noted that Cloud computing represents a huge opportunity to many third-world countries that have been so far left behind in the ICT revolution.

B. An Insight on Internet of Things(IoTs)

This is an emerging technology that is receiving attention and which is rapidly progressing and a good number of literature has been published on the same [6] [7] . [8] defines IoT as a network of physical devices and other items, embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data. IoT is a technology composed of three main components identified by Stergiou and others in [9] and is characterized by having to deal with huge volumes of data and a countless different sources of the data even as Mircea, Stoica and Ghilic-Micu well noted in [7]. It one of technology that is now referred to as disruptive technology [10]. It is becoming part of our lives and represents one of the most disruptive technologies giving rise to ubiquitous and pervasive computing scenarios [3].

C. An Insight on the CloudIoT paradigm(CloudIoT)

Cloud Computing and IoT are two separate technologies much has been published about them separately. The CloudIoT paradigm seeks to combine these two technologies in ways that will bring about technological revolution. The major drivers to combining cloud computing and IoT have been well noted by many researchers[11]. The main reasons for integration stems from the fact that the two technologies are complementary in nature that is, having capabilities that lack from each other. For instance Mircea et al. in [7] was apt to note that IoT and Cloud Computing are emerging technologies with features of their own, which may create added value when used together Table 1 below is a summery by [3]of the main reasons for the integration:

TABLE I. COMPLEMENTARY ASPECTS OF IOT AND CLOUD COMPUTING

	IoT	Cloud Computing
Displacement	Pervasive	centralized
Components	Real world	Virtual resources
Computational Capacities	Limited Processing capability	Virtually unlimited Processing capability
Storage	Limited or none	Virtually unlimited storage
Role of the Internet	Point of Convergence	Service delivery
Big data	Big data source	Means to manage big data
Reachability	Limited reachability	Ubiquitous

From the table above it can be deduced that the major drivers for integration of IoT and Cloud Computing being referred to as CloudIoT are as follows:

- **Storage:** The Things have limited or even none storage capabilities in which case they can benefit from the virtually unlimited storage provided by the on the cloud infrastructure.
- **Processing:** With limited computation capabilities, the Things can utilize the virtually unlimited processing capabilities provided on the cloud.
- **Power:** since the things have limited power they can still benefit from the power provided on the cloud.
- **Security:** This too can one advantage that the things can benefit from the cloud.
- **Reliability:** Because cloud computing is form of distributed systems there is high level failure transparency because of pooled interconnected resources which seamlessly ensure reliability which the things can benefit from.

On the other hand the cloud can benefit from the CloudIoT by extending its scope to deal with real life objects in a more dynamic and distributed way and providing services for countless devices in assorted real life scenarios [12]. The cloudIoT paradigm give rise to the possibility of smarthealth care application scenario which is just one among many other application scenarios that CloudIoT makes possible. In this paper the focus is on Smart health care which will now be discussed.

D. Smart health Care as a CloudIoT Application Scenario

This involves management of a patient’s condition by using ICTs application resulting in improved better quality of services. [13], stated the benefits of this application scenario for instance, that smart health care support patients to adhere to prescribed medication increasing chances of recovery. They add that smart healthcare makes it possible for better treatments by for instance providing of smart pill boxes with appealing features like opening at precise time and recording of pill intake time.



Fig. 2 A typical illustration of a possible Smart Health application [8]

There is too a promise of good hospital management. There is even an emerging network application in IoT called Wireless Body Area Network (WBAN). Khan et al.[14]defines this form of technology as being a special purpose sensor network designed to operate autonomously to connect assorted sensors and appliances, located inside and outside of human body. Fig 2 shows how the things referred to as wearables can be used to collect various internal conditions such as blood pressure, glucose level, body temperature e.t.c and transmit that data to Biobanks for storage or to other gateway to some other services. The physician can then, using a mobile device or desktop PC, access or monitor patient’s condition and advise accordingly or perhaps schedule and appointment as need arises. At times a patient could be alerted by means of the mobile devices using the information gathered from the wearables and be warned of possible worsening condition or of missing prescription at an appropriate time. IoTs can be used in what is now referred to as Ambient Assisted Living(AAL)[15] for people with chronic diseases and the elderly. However there are a number of challenges that comes with the Smart healthcare and the next section discusses these.

III. CHALLENGES OF ADOPTION OF SMART HEALTH CARE AS A CLOUDIoT APPLICATION SCENARIO

Because CloudIoT paradigm as already alluded to is a disruptive technology that brings about entirely new applications scenarios, it comes with it new fresh challenges to deal with which have a potential to impede successful implementation. Researchers in [12] identified the many ones as being:

1. **Security and privacy:-** As data are transmitted from the things, implementation of only authorized users to have access to the confidential data which obviously is so important to patients must be guaranteed and additionally when data moves to the cloud trust issues arise [12]. Possibility of attacks must be put into consideration too.
2. **Heterogeneity:-** In the CloudIoT, there can be so much versatility when it comes to the types of devices, platforms, Oss and services that may exist.
3. **Big data:-** it is expected that there will be some 50billion IoT devices by 2020 [12] and probably you would know what that implies as to the volume and the velocity(frequency of data production) that will be expected to be transmitted. This brings about storage constraints, bandwidth as well as analysis complexities. This highlights why high speed preferably real time processing of data implemented especially where health care is involved a suggestion advocated for in [14].
4. **Large scale:-** since there is interaction of billion devices, this poses some new challenges difficult to deal with such as computation capabilities and the IoTs challenge of connectivity because of their wide distribution.
5. **Standardization:-** “Although a number of proposed standardizations have been put forth by the scientific society for the deployment of IoT and Cloud approaches, it is obvious that architectures, standard protocols, and APIs are required to allow for interconnection between

heterogeneous smart things and the generation of new services, which make up the Cloudbased IoT paradigm”[12].

- 6. *Energy efficiency*:- Most IoTs have very limited amounts of power as table 1 highlighted especially that data is expected to be transmitted between the cloud and the things.

Concerning these challenges a number of research has been done to handle these challenges and now these will be the focus of discussion in the summary Table II below.

TABLE II. STUDIES CONDUCTED TO HANDLE THE CHALLENGES OF CLOUD AND IOT INTEGRATION

Challenge	Study	Problem	Solution
Security and Privacy			
	[9]	Security issues in both cloud and IoT	Part of AES presented for improvement of security issue, resulting from integration.
	[16]	Blockchain-based approaches provide decentralized security and privacy, yet they involve significant energy, delay, and computational overhead that is not suitable for most resource-constrained IoT devices	Proposed BC-based smart home framework was proved secure by thoroughly analysing its security with respect to the fundamental security goals of confidentiality, integrity, and availability.
	[17]	New security and privacy threats arise that were not present in centrally-managed cloud computing	Presents a summary of up-to-date research contributions and to outline future research direction to solve different challenges in privacy and security in the fog computing.
	[18]	Data sent to the Cloud may be used incorrectly or copied without permission	Proposes encryption of data by access. Solution also allows data owner to restrict access to certain period by destroying data and keys after this time in order to change them; In addition, this approach improves availability, integrity and confidentiality of information

			stored in the cloud by building a Cloud of Clouds
Heterogeneity			
	[19]	The need to have a cross platforms for a unified access to and sharing of sensing or actuating resources	Description of IoT landscape revealing challenges which need to be address also authors presents a SymbloTe project which has focus on interoperability platforms.
	[6]	Expected interconnection of trillion number of a variety of devices and sensors	An intensive exploration of IoT architecture is reported focusing on advances made then categorizing and classifying of IoT architectures based on applications, business objectives, architectural requirements, network topologies among others and then identify key future requirements.
	[20]	Success of IoT expected to depend so much on connecting and integration of existing resources.	The solution is a generic platform which demonstrates that IoT is all about interoperability, from connected cloud computing using RFID, NFC, M2M as well as digital content and context-aware services.
Large Scale Computation			
	[21]	Requirement of localized computation in order to ensure performance, security, and lower costs in IoTs	Aura, an advanced localized IoT based cloud computing model is presented in which computations are done with high flexibility, giving rise to full control to start,stop,migrate, and restart computations in nearby devices as clients move between different physical locations.
	[22]	Gaining an understanding of challenges	By means of experiments the study revealed

		arise with running computational jobs on mobile devices using different types of workloads	that mobile devices offer free time slots to servers in the cloud based on data analytics system.
Standardization			
	[23]	Most IoT service providers building manufacturer specific hardware devices to the relevant cloud services thereby lacking consistency. The need for a standard model to perform popular IoT backend tasks, such as processing, storage, and firmware updates, is becoming more significant.	IoT standardization model is explained with components Platform, connectivity Killer Applications, Business Model and Challenges of IoT are identified.
	[24]	Issues to do with a variety of languages, protocols and standards. Lack of single platform of standardization.	Review challenges of standardization and existing state of the art standards to highlight and overcome it by presenting some technical solutions.
	[25]	Challenges on technologies, applications, and standardization a Chinese perspective	Proposes an open and general IoT architecture made up of three platforms to meet the architecture challenge as well opportunity and prospects of IoT.
Energy efficiency			
	[6]	Frequent interchange of datasets among things impacting negatively on performance	Paper develops a novel multi-cloud IoT service composition algorithm (E2C2) that seeks to creating an energy-aware composition plan by searching for and integrating least possible number of IoT services.

From summary table it is indeed clear that a number of strides have been made in seeking to address the main hindrances to cloudIoT, a basic foundation on which Smart Health Care application scenario is based.

IV. LITERATURE SEARCH METHODOLOGY

A search strategy focused on the keywords on which the research topic is centered, these being Cloud Computing, IoT, CloudIoT, Smart Health Care, Wireless Body Area Network. Queries on Scholar.google.com were done. For instance to make sure much of the relevant links are harvested query expressions such as “Smart Health Care” OR “smart healthcare”, “CloudIoT” OR “Cloud IoT” OR “Cloud Computing”. For IoTs terms like “Smart Objects” OR “Internet of things” OR “IoT” were used.

To aggregate the research works on the challenges of integrating cloud computing and IoT or CloudIoT the same last key phrases “.....challenges of Cloud IoT” were used for all six of them. For example “security and privacy challenges of CloudIoT” which would harvest the related relevant search results.

A. Inclusion Criteria

Focus was placed on research work done between 5 to 10 years representing most recent research work. Any paper that never matched any key terms was rejected.

Databases explored include but no limited to ieeexplore.ieee.org, Elsevier.com, emeraldinsight.com, dl.acm.org, researchgate.com, link.springer.com, sciencedirect.com, emerald.com/insight/Semanticscholar.org, Researchgate, ieeexplore.ieee.org, and grey sources from general www.google.com.

B. Exclusion Criteria

Every search conducted would extract superfluous links and so the strategy was to exclude obviously all the results that never had key terms and or concepts highlighted in the methodology. The year of publications was determining factor too with studies done over 10 years ago excluded from consideration.

V. RESULTS

A dozen of research papers were accessed although the limiting factor was not having enough subscription to some other databases. Some main repositories examined and their results are summarized in table III.

TABLE III. RESEARCH RESULTS SUMMERIZED

Key Search Term	Repository Name	Article Result Count
CloudIoT	Researchgate	25
	Sciencedirect	19
	ieeexplore.ieee.org	9
	Semanticscholar.org	231
Smart healthcare	ieeexplore.ieee.org	2,666
	Researchgate	100
	Sciencedirect	17,569
	Semanticscholar.org	461,000
challenges of Cloud IoT	Semanticscholar.org	43,100
	sciencedirect	4,528
	emerald.com/insight/	441
	Researchgate	
Cloud Computing	Sciencedirect	100,943
	link.springer.com	137,445

The literature has highlighted a big gap of knowledge that exists due to lack of enough publications of information on attempts to explain possibility of implementing smart health care in Zambia. It is however gratifying to note that this is the area that has attracted a lot of research as results already show.

VI. CONCLUSION

From this in depth understanding of Smart health care as an application scenario of a disruptive technology (CloudIoT paradigm) with a great potential to revolutionize management of not only diseases but also hospital management, it is expected that there will be a new dawn of high quality unprecedented health care. Much work has been published has already noted which Zambia can take advantage of as it does not have to be a trial methodology. Right now researchers in India have made headways in actually implementing smart health care on smart solution to child vaccination reported in [26]. For instance the concern of security and privacy which is a very significant impedance to integrating cloud and IoT which is a key in implementing smart health has attracted useful research highlighted in table 2 which should be a reason for having implementation confidence. Of most importance, smart health care will very much decongest hospitals and make illness management more personal, improve medication prescription adherence via alerts from smart objects or things and assist the aged in Ambient Assisted Living. It is definitely a technology worthy trying.

VII. REFERENCES

- [1] World Health Organization, "The world health report 2000: health systems: improving performance," 2000.
- [2] A. Darwish and A. E. Hassanien, "Cyber physical systems design, methodology, and integration: the current status and future outlook," *J. Ambient Intell. Humaniz. Comput.*, vol. 9, no. 5, pp. 1541–1556, Oct. 2018.
- [3] A. Botta, W. de Donato, V. Persico, and A. Pescapé, "Integration of Cloud computing and Internet of Things: A survey," *Future Gener. Comput. Syst.*, vol. 56, pp. 684–700, Mar. 2016.
- [4] P. Mell, "The NIST definition of cloud computing," 2011.
- [5] M. G. Avram, "Advantages and Challenges of Adopting Cloud Computing from an Enterprise Perspective," *Procedia Technol.*, vol. 12, pp. 529–534, 2014.
- [6] I. Yaqoob *et al.*, "Internet of Things Architecture: Recent Advances, Taxonomy, Requirements, and Open Challenges," *IEEE Wirel. Commun.*, vol. 24, no. 3, pp. 10–16, Jun. 2017.
- [7] M. Mircea, M. Stoica, and B. Ghilic-Micu, "Using Cloud Computing to Address Challenges Raised by the Internet of Things," in *Connected Environments for the Internet of Things*, Z. Mahmood, Ed. Cham: Springer International Publishing, 2017, pp. 63–82.
- [8] D. V. Dimitrov, "Medical Internet of Things and Big Data in Healthcare," *Healthc. Inform. Res.*, vol. 22, no. 3, p. 156, 2016.
- [9] C. Stergiou, K. E. Psannis, B.-G. Kim, and B. Gupta, "Secure integration of IoT and Cloud Computing," *Future Gener. Comput. Syst.*, vol. 78, pp. 964–975, Jan. 2018.
- [10] M. Weyrich and C. Ebert, "Reference architectures for the internet of things," *IEEE Softw.*, vol. 1, pp. 112–116.
- [11] J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, "Security and Privacy for Cloud-Based IoT: Challenges," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 26–33, Jan. 2017.
- [12] H. F. Atlam, A. Alenezi, A. Alharthi, R. J. Walters, and G. B. Wills, "Integration of Cloud Computing with Internet of Things: Challenges and Open Issues," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Exeter, 2017, pp. 670–675.
- [13] V. Stantchev, A. Barnawi, S. Ghulam, J. Schubert, and G. Tamm, "Smart items, fog and cloud computing as enablers of servitization in healthcare," *Sens. Transducers*, vol. 185, no. 2, p. 121, 2015.
- [14] A. A. Mutlag, M. K. Abd Ghani, N. Arunkumar, M. A. Mohammed, and O. Mohd, "Enabling technologies for fog computing in healthcare IoT systems," *Future Gener. Comput. Syst.*, vol. 90, pp. 62–78, Jan. 2019.
- [15] A. El murabet, A. Abtoy, A. Touhafi, and A. Tahiri, "Ambient Assisted living system's models and architectures: A survey of the state of the art," *J. King Saud Univ. - Comput. Inf. Sci.*, Apr. 2018.
- [16] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for IoT security and privacy: The case study of a smart home," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Kona, Big Island, HI, USA, 2017, pp. 618–623.
- [17] M. Mukherjee *et al.*, "Security and Privacy in Fog Computing: Challenges," *IEEE Access*, vol. 5, pp. 19293–19304, 2017.
- [18] F. A. Guechi and R. Maamri, "Secure Self-Destruction of Shared Data in Multi-CloudIoT," in *2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)*, Prague, 2017, pp. 161–168.
- [19] S. Soursos, I. P. Zarko, P. Zwickl, I. Gojmerac, G. Bianchi, and G. Carrozzo, "Towards the cross-domain interoperability of IoT platforms," in *2016 European Conference on Networks and Communications (EuCNC)*, Athens, Greece, 2016, pp. 398–402.
- [20] G. Suci, S. Halunga, A. Vulpe, and V. Suci, "Generic platform for IoT and cloud computing interoperability study," in *International Symposium on Signals, Circuits and Systems ISSCS2013*, Iasi, Romania, 2013, pp. 1–4.
- [21] R. Hasan, Md. M. Hossain, and R. Khan, "Aura: An IoT Based Cloud Infrastructure for Localized Mobile Computation Outsourcing," in *2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, San Francisco, CA, USA, 2015, pp. 183–188.
- [22] S. Dey, A. Mukherjee, H. S. Paul, and A. Pal, "Challenges of Using Edge Devices in IoT Computation Grids," in *2013 International Conference on Parallel and Distributed Systems*, Seoul, Korea (South), 2013, pp. 564–569.
- [23] A. Banafa, "IoT standardization and implementation challenges," *IEEE Org Newsl.*, 2016.
- [24] S. A. Al-Qaseemi, H. A. Almulhim, M. F. Almulhim, and S. R. Chaudhry, "IoT architecture challenges and issues: Lack of standardization," in *2016 Future Technologies Conference (FTC)*, San Francisco, CA, USA, 2016, pp. 731–738.
- [25] Shanzhi Chen, Hui Xu, Dake Liu, Bo Hu, and Hucheng Wang, "A Vision of IoT: Applications, Challenges, and Opportunities With China Perspective," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 349–359, Aug. 2014.
- [26] S. Kumari, A. Haripriya, A. Aruna, D. S. Vidya, and M. N. Nithy, "Immunize — Baby steps for smart healthcare: Smart solutions to child vaccination," in *2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)*, Coimbatore, 2017, pp. 1–4.

Augmented Reality Aided Navigation System Model for Support of Student Registration Process

Andrew Biemba
Department of Computer Science
University of Zambia
Lusaka, Zambia
andrew.biemba@gmail.com

Mayumbo Nyirenda
Department of Computer Science
University of Zambia
Lusaka, Zambia
mayumbo@gmail.com

Abstract— Universities worldwide are expanding structurally year by year as the number of students gets bigger and bigger with every enrolment period. Due to the enormous size, there is a high chance that some students would not be able to find their way around campus with parents and newly enrolled students being the ones greatly affected. One of the key processes that universities undertake with every enrolment period is the student registration process. Student Registration is a systematic process that involves but is not limited to fees settlement with the bank, course selection and submission for approval by responsible department. In this study, the student registration process was modelled and integrated with a proof of concept Navigation system that makes use of the new technology, Augmented Reality. This project takes advantage of the exceptional data processing abilities that are possessed by smartphones in conjunction with their various services such as Global Positioning System (GPS), wireless networks, and compass to model a Navigation System that makes extensive use of Augmented Reality to support the student registration process. Augmented Reality (AR) is a type of interactive, reality-based display environment that takes the capabilities of computer-generated display, sound, text and effects to enhance the user's real-world experience. The Global Positioning System (GPS) is a space-based satellite navigation system that provides location and time information in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites. A navigation system is a system that aids in navigation. A compass is an instrument used for navigation and orientation that shows direction relative to the geographic cardinal directions (or points).

Keywords— Augmented Reality, Global Positioning System, Orientation, Virtual Trace, Navigation, Compass, Graphical User Interface

I. INTRODUCTION

Augmented Reality is a technology based on virtual reality where virtual data is superimposed onto real life images, resulting in a mixture of the two environments. AR fills the area between the real world and the virtual world. It takes the information features of the virtual world and makes them easily accessible and relevant to the user. AR is not only a visually based system, another medium that is commonly augmented onto reality is the dimension of sound. For example, a visitor at a museum may use headphones that inform him of the displays as he stands in front of each one. An intelligent system would also ask the user how long they intend to spend at the museum to calculate how much information to give about each exhibit. Often AR augments data stored in many media and relays the information to the user dependent on their position in relation to the information source. The research underway at the University of

Nottingham concentrates on using map information stored digitally for graphical overlay on views of the real world. A database is used that contains known positional information of the subsurface features. The AR system, knowing the position and orientation of the user can calculate which part of ground he is looking at and relate the database information accordingly. Added information can also be displayed as text or graphically, removing the requirement of continuous referral to a legend, allowing feature attributes to be much more accessible [7]. AR is one of the most promising technological innovations and has already been adopted within several industries, most prominently the manufacturing industry. AR enables the overlaying of digitally created information into the real-world environment. To access this kind of information, users require camera or GPS enabled devices [8].

A case study was undertaken at the University of Zambia (UNZA) Institute of Distance Education (IDE). The idea is about providing a guide and offering directions, through a smartphone's Graphical User Interface (GUI), that assists students to easily locate places during the student registration process. The steps involved in the student registration process were identified and a model for this process was created. Key places and features that are involved in the student registration process were also identified and a model/and or proof of concept of the Augmented Reality Aided Navigation System to support the student registration process was created. The model was simulated to use the smart phone's inbuilt camera and GPS to display a user's surround information in real time on the smart phone. The proposed system model combines the GPS location-based technology and virtual trace technology to provide the user with basic information about a building they are looking for or one in their immediate surrounding. Virtual Trace is a powerful ground-breaking technology that allows any drawing or view to be overlaid on top of or beside any other view.

II. LITERATURE REVIEW

AR allows flexibility in use that is attractive to education. AR technology can be utilized through a variety of mediums including desktops, mobile devices, and smartphones. The technology is portable and adaptable to a variety of scenarios. AR can be used to enhance content and instruction within the traditional classroom, supplement instruction in the special education classroom, extend content into the world outside the classroom, and be combined with other technologies to enrich their individual applications [9].

Augmented reality allows 3D Virtual Images to appear in a 3D real environment in real time and allow more

interactivity for the users. In some ways, Augmented Reality navigation may be taken in with comparison to GPS navigation, but it concentrates more on enhancing reality than simply generating a map [2]. Augmented Reality (AR) is an area of research that aims to enhance the real world by overlaying computer-generated data on top of it. Azuma identifies three key characteristics of AR systems: (i) mixing virtual images with the real world, (ii) three-dimensional registration of digital data and (iii) interactivity in real time. The first AR experience with these characteristics was developed over 40 years ago, but mainstream adoption has been limited by the available technologies [6].

Mobile computing, ubiquitous computing and augmented reality have provided technology explorers a platform to develop new applications. Augmented reality is upcoming solution for envision and navigation with the wide ranging high-end processors combined with mobile devices having GPS and camera functionalities. It is widely being used on hybrid processors, dual core processors and 1 GHz processors. Over the years, pageant systems which enlightened researchers' imagination were developed [4].

Augmented reality (AR) is a growing phenomenon on mobile devices, reflected by the increase in mobile computing in recent years and the common ubiquity of Internet access across the world. The NMC Horizon Report for 2011 named augmented reality as the highest-rated topic by its Advisory Board, with widespread time-to-adoption being only two to three years (Johnson et al., 2011). What was once seen by many as being a mere gimmick with few applications outside of training, marketing/PR or sport and entertainment, is now becoming more mainstream with real opportunities for it to be used for educational purposes. One of the most compelling affordances of AR is its resonance with immediate surroundings and the way in which information can be overlaid on top of these surroundings, enabling us not only to learn about our environment but also giving us the tools to annotate it [5].

Geo-based Augmented Reality is the combination of Augmented Reality, Location Based System and GPS giving birth to such an advanced and innovative technology. Its fundamental concept is to retrieve the location of device and superimpose the information on screen about the point of interests. Location Based System and GPS are the two primary concept which relies on geolocation information and the fact is Location Based System uses one or more geolocations such as GPS which gives location whereas LBS (Location based Service) use this location to create value just like in Facebook which stores the user's location once and increment the count after each location tag so that the other users got to know which of his/her friends have been here before. Geolocation is closely related to GPS, but a slight distinguishable concept may break them apart. Global Positioning System also known as GPS gives the geographic coordinates whereas geolocation gives the meaningful locations rather than just set of coordinates, for example, street address. Geo-based AR focuses more on determining the position of an object or person and overlaying the geographical coordinates of the same on screen of dedicated devices in more interactive ways rather than just working with positioning services [3].

Navicam has since become a popular topic of AR applications. A notable example of a wearable tourist guide application is the GUIDE project described by Davies et al.

Supported & sponsored in part by ZICTA under grant GRZ891005.

(1999) that presents roaming tourists with location-based information on Lancaster. Augmented Reality has also been identified as an important means of Computer Supported Collaborative Work (CSCW) as demonstrated by Billingham et al. (1996) in the Shared Space project [10].

III. MATERIALS AND METHODS

The data collection methods/tools used included expert judgement/advice, a survey at UNZA-IDE and recorded documentation of the student registration process. The published steps of the student registration process on the university website were critically scrutinized from the first to the last step.

The modeled system uses the smart phone's inbuilt camera and GPS to display a user's surround information in real time on the smart phone. The system combines the G.P.S location-based technology and virtual trace technology to provide the user with basic information about a building they are looking for or one in their immediate surrounding. The resource libraries used include ARtag and ARToolkit which provide a fast and accurate way of calculating co-ordinates.

IV. METHODOLOGY

The RAD methodology was employed due to the fact that it is a dynamic methodology which will allow reviews and changes to the system requirements whenever deemed necessary during system development.

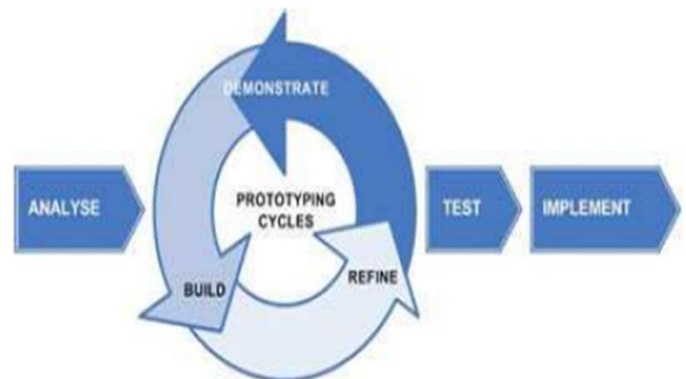


Figure 1 (Behrang [2]): Rapid Application Development Model

The figure below is a simplified Use Case Diagram showing how the User interacts with the camera and display screen information without having access to the internal GPS.

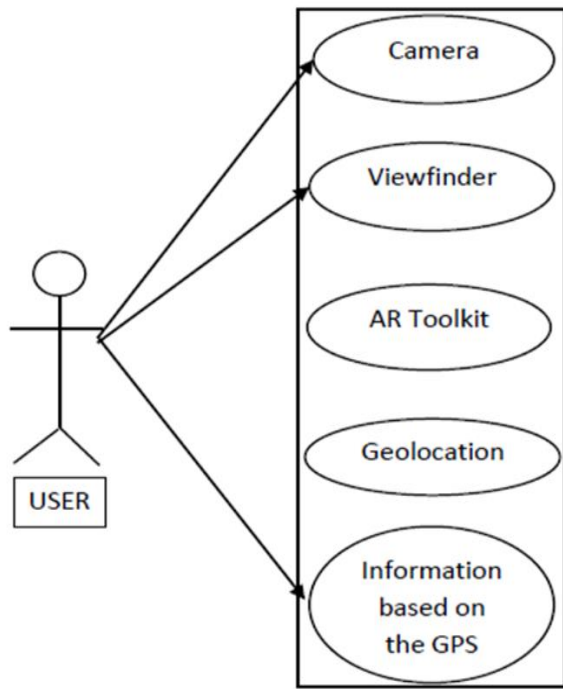


Figure 2 (Prakhar [3]): User Interaction with Components in functioning of app.

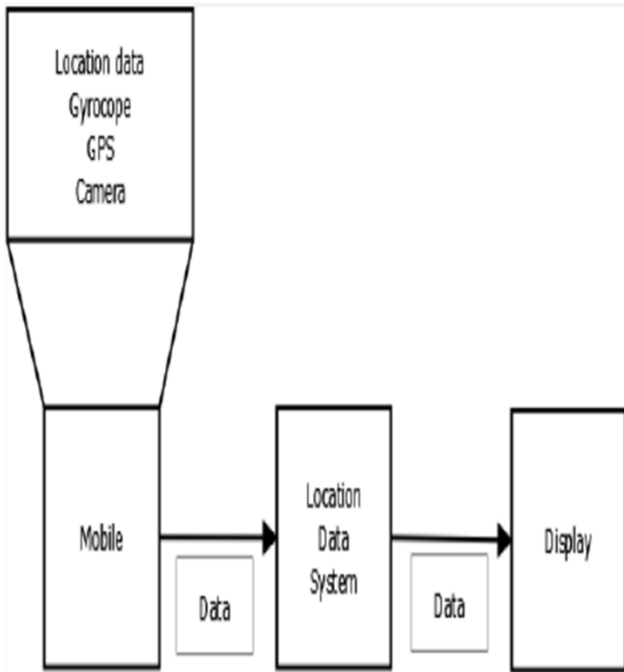


Figure 3 (Chee [1]). Mobile Device augmented reality flowchart.

V. RESULTS AND DISCUSSION

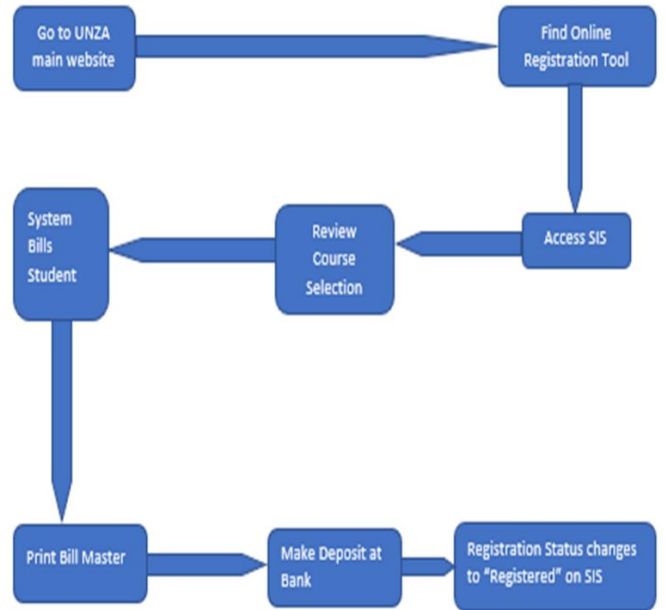


Fig. 4: The Student Registration Process Model

The table below shows the key places that were identified as playing a part in the student registration process at IDE.

DEPARTMENT/PLACE TO BE VISITED	ISSUE OR STEP IN REGISTRATION PROCESS
Accounts Department	Incorrect Fees Correction
Registry	submission of acceptance letter, passport-sized photos, etc.
Learner Support Department	Registration process clarification, guidance, fee schedule, academic calendar, admission letter, orientation module, etc.
Production Development Programme (PDP)	Collection and uploading of modules on student portal/platform
Zanaco Bank	Cash deposit

Table 1: key places/departments involved in the student registration process.

Figure 5 below shows a web-based HTML implementation of AR using ARToolKit. The ARToolKit reads from webcam or smartphone’s camera and launches the ar.js in the background which tracks the marker position and movement in space.

```

1 </script> <!-- Points to source of various assets,
2 i.e, marker descriptions, camera calibrations -->
3 <body style='margin : 0px; overflow: hidden;'> <!-- body
4 styling -->
5 <a-scene embedded artoolkit='sourceType: webcam;'>
6 <!-- Using ARToolKit to read from webcam and launch the
7 ar.js in the background and then track the position of
8 the marker in space -->
9 <a-tetrahedron position = '0 0.5 0' color = "green">
10 </a-tetrahedron> <!-- Adding element and defining
11 it's position relative to marker position -->
12 <a-marker-camera preset='kanji'></a-marker-camera>
13 <!-- Camera setup & Marker description -->
14 </a-scene>
15 </body>

```

Figure 5: Web based AR HTML code for the web.

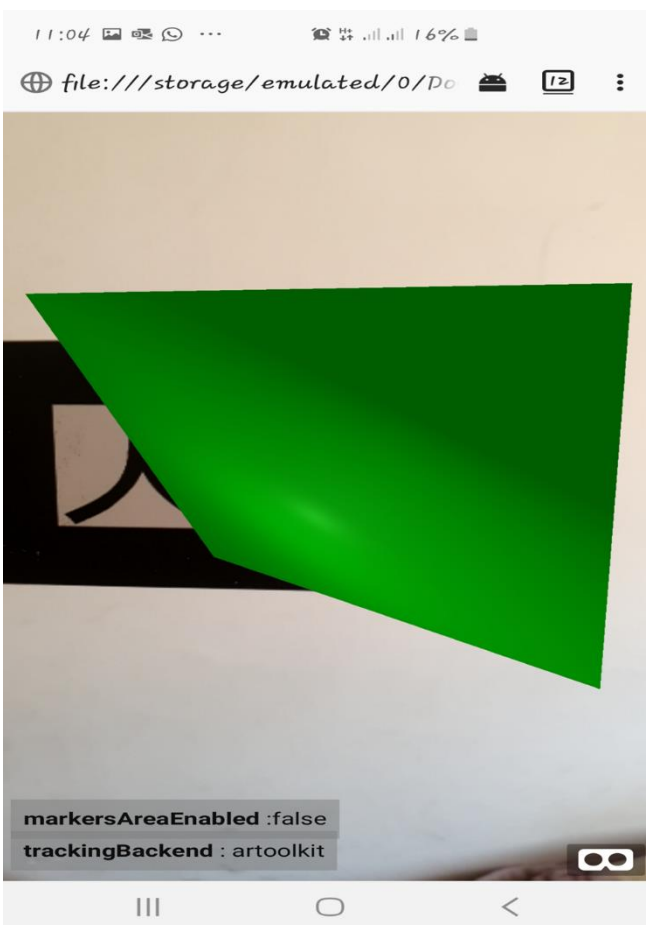


Figure 6: A tetrahedron laid on top of camera view in Marker-based AR

Figure 6 above shows a live proof of concept image taken from my smartphone's camera. The virtual image (a green

tetrahedron in this case) is superimposed on the marker and is able to follow the marker. The next step is to combine this AR code with GPS location so that we have a hybrid augmented reality-based navigation model that will calculate coordinates and give correct graphic to aid the user in finding desired location.

VI. CONCLUSION & RECOMMENDATIONS

This study proposed a Proof of Concept Augmented Reality Aided Navigation System Model to support the Student Registration Process at the University of Zambia (UNZA) Institute of Distance Education (IDE). We recommend that relevant authorities and academic bodies consider this proof of concept and organize resources to implement a full-fledged AR Based Navigation system that will not only help in the student registration process but also be of used as a navigation tool for a lot of other student activities.

VII. ACKNOWLEDGEMENTS

We would like to thank the University of Zambia Computer Science Department and the Institute of Distance Education for supporting the research project and giving expert advice and direction to the scope of the project.

VIII. REFERENCES

- [1] Chee Oh Chung et al, "Augmented Reality Navigation System on Android," International Journal of Electrical and Computer Engineering (IJECE), Vol. 6, No. 1, February 2016, pp. 406-412, Seogu, DaeJeon, Korea.
- [2] Behrang Parhizkar et al, "Mobile AR GPS-Based for Navigation Purpose," Conference Paper, January 2012, Cyberjaya, Malaysia.
- [3] Prakhar Saxena, "Geo-Location Based Augmented Reality Application," IJRET: International Journal of Research in Engineering and Technology, Volume 04, Issue 07, July 2015, U.P, India.
- [4] Ashwini S. Kasbekar et al, "Location Based Augmented Reality GEO Tag and GEO Locator: An Android Application," International Journal of Electrical, Electronics and Data Communication, Volume 3, Issue 5, pp. 47-51, May 2015, Maharashtra, India.
- [5] Elizabeth FitzGerald et al, "Augmented Reality and Mobile Learning: The State of the Art," Open Research Online: CEUR Workshop Proceedings, Vol-955, pp. 62-69, 2012, UK.
- [6] D. Schmalstieg et al, "Virtual Realities," Springer-Verlag/Wien, 2011.
- [7] Dr. Gethin W. ROBERTS et al, "The Use of Augmented Reality, GPS and INS for Subsurface Data Visualisation," Nottingham, UK, April 19-26, 2002.
- [8] Timothy Jung et al, "Acceptance of GPS-based Augmented Reality Tourism Applications," Manchester, UK, February 2016.
- [9] Misty Antonioli et al, "Augmented Reality Applications in Education," The Journal of Technology Studies, vol. 70, pp. 96-107, August 2015.

[10] Gerhard Reitmayr and Dieter Schmalstieg, "Collaborative Augmented Reality for Outdoor Navigation and Information Browsing," Interactive Media Systems Group, Vienna, pp. 31-41, 2013.



Andrew Biemba. (M'87) received the B.S. degree in Computer Science from The University of Zambia, Lusaka, Zambia, in 2014. He is currently pursuing an M.S. degree in Computer Science at the same university.

From 2015, he has worked as a consultant for International Business Machines (IBM) specialized in Management Information Systems (MIS) and Reporting and offered systems support services to Zambia Telecommunications (ZAMTEL) Company Limited and Airtel, Bharti, India. He is now providing Data Warehousing & Business Intelligence (BI) solutions to African Banking Corporation Zambia Limited trading as Atlas Mara. His research interest includes Augmented Reality, Navigation Systems, Design and development of Machine Learning Algorithms for Panorama Applications, Computer Vision, Data Analytics, Big Data, Data Mining and Warehousing.

Assessing Automated Flood Disaster Alert Systems in Zambia: Case of Mbeta Island in Sioma, Western Province

¹ Clifford Siatala, ² Jameson Mbale

¹Copperbelt University, Kitwe, Zambia, ²Director Research, Copperbelt University Kitwe, Zambia

¹csiatala@gmail.com, ²jameson.mbale@cbu.ac.zm

Abstract - Globally, as advancement and mainstreaming of Technology become more pervasive, information Communication Technologies (ICTs) have become agents of change transforming the way we conduct and perceive human activities to sustain and promote the quality of life. It is also not surprising that today's technology has become a sole candidate for change allowing for developing efficient mechanisms for communication to creating systems for software compliance and disaster recovery. Using a positivist approach with methodological triangulation at data collection and analysis stage, this study assesses automated flood disaster alert systems in Zambia with specific reference to Mbeta Island in Sioma. Findings indicate that there is need to adopt a new and robust disaster mitigation system that would allow for integration of community needs and everyday challenges in relation to natural and man-made disasters.

Background to the Study

Entering the fourth industrial revolution, Mankind has witnessed the rapid rate at which technology has transformed the way we conduct and perceive human activities in an endeavor to sustain and promote the quality of life. While it is also possible to use it in enhancing business processes and functions, today's ICTs has been integrated in modern life styles and is becoming pervasive in everything from telephone networking, automated security systems, software compliance, backups to disaster recovery platforms and Internet of things. Numerous scholars argue that the emerging avenues such as Wi-Fi, devices with ready Internet connectivity, the cloud and big data has created a conducive environment to nurture the development of effective and efficient processes that may improve many aspects of life. Additionally, one may argue that it is no longer feasible or rapidly becoming obsolete to solely rely on manual systems to predict potential threats such as natural and man-made disasters. Thus, automated systems in monitoring and predicting disasters especially in the advent of climate change cannot be overemphasized.

Similarly, emerging interest and research in business and artificial intelligence systems integrating ICTs have become mainstream tools in achieving or creating optimized, efficient and effective services. Given the increased attention to natural disasters with possible

relationship to climate change, various economies, organizations or institutions have embarked on the creation of deliberate policy frameworks that are promoting funding for research on uncovering new knowledge to act as key components in adoption and integration of ICTs. Studies by Mahonty, et al (2012) and Bilamage, B. (2013) and Singh, J. P. et al (2017) argues that ICTs mediated systems have become useful in early detection and prediction of disasters, communication and dissemination of information on disasters. This is being argued to act as an effective platform for information exchange and sharing among people and with government and non-government agencies in time of disasters especially with known possibilities in modern geographical information systems (GIS), remote sensing and satellite-based communication systems.

Further, studies by Kelvin, O.K et al. (2012) and Maier, (2011) and Trimi (2012) argue that ICTs have become key in risk reduction in entities such as companies, organization or communities through the use of mobile cellular networks coverage. Also, big data analytics have emerged as a powerful set of inference platform necessitated by technologies that allow for swift execution of commands or interventions from a sea of data to address and mitigate disasters.

The importance of timely disaster warning systems in mitigating negative impacts on communities or societies can never be underestimated. For example, developed countries have created systems with robust predictability to enable reduction in loss of life due to disasters compared to their counterparts in developing countries. A key reason for this has been the implementation of effective disaster warning systems and evacuation procedures in developed countries and the absence of such measures in developing countries especially African countries. In implementing early warnings systems, communities play an important role in the use of ICTs in realizing potential warning hazards that can be reported through automated and integrated systems that escalates emergency notifications to appropriate local government authorities, rescue units and mitigation teams Samarajiva et al., (2010). While it is possible for people to predict disasters in communities, it is within their control to report early warnings of disasters to mitigation teams before such potential threats turn into real disasters.



Literature Review (Related works)

Currently, there is an increasing literature on a number of studies that have been done on emerging systems that are intended to process disaster information. These systems aggregate disaster information in order to streamline notification, response and recovery efforts. Among them include

- ✓ Integrated Public Alert and Warning System (IPAWS)
- ✓ Global Disaster Alert and Coordination System (GDACS)
- ✓ Emergence Alert System (EAS)
- ✓ Real-Time Early Response System (RTERS)

For instance, the Integrated Public Alert and Warning System (IPAWS) leverages its relationships with state and local Government agencies, universities, researches, broadcast stations and two-way radio dispatch. IPAWS system is capable of linking alerts of disasters related information on Tornado, Hurricane, Tsunami and earthquake information from universities, research partners and local government authorities. These alerts are processed at alert system group (ASG), a research partner that sends alert information directly to local FM radiobroadcast stations for public awareness. Alerts are delivered to the American people through the media such as TV, web widgets, FM, AM, satellite radio, web browsers, websites, and digital analogy. Jung, H (2015)

The United Nations in collaboration with the European Commission maintained a Global Disaster Alert and Coordination System (GDACS). GDACS Corporation established partnerships with scientific monitoring organizations and aggregates worldwide disaster information following a natural disaster. The primary users of GDACS system are governments and disaster response organizations. However, there is no direct process in place for individuals to receive GDACS alert notifications. Users access disaster information on the GDACS website worldwide.

The EAS system is an emergence system that uses broadcasters, cable television systems, wireless cable systems and satellite digital service providers to send alerts. EAS system is designed in such a way that it utilizes the direct broadcast satellite service providers to make their communications facilities available to the President during a national emergency. Maity et al (2016) explained that state and local authorities deliver important emergency information such as amber alerts and severe weather warnings targeted to specific geographical regions or areas

in the US. Maity 2016 further stated that the EAS system generalized notifications at a national level whereby information filtered up to higher level and also transmitted through a variety of means and mediums for public safety. According to Maity (2016) the EAS system was designed to issue notification alerts information through available communication facilities to the local authorities and the President of the United States.

Real time early response systems (RTERS) incorporated real time information as feedback data for closing control loop and for generating real time situation assessment. RTERS arose as a useful tool for supporting people in their decision making process during natural disasters. RTERS's design integrated a component of decision support systems (DSS). The DSS systems collect real time information from sensors around the world, process that information and make suggestions that allows the accomplishment of decisions by decision makers in disasters. The system is customized to help in emergence situations and provided a real world environment representing human behaviour. The systems representations of the real world are fundamental tools in RTERSs because they help to estimate and to predict scenarios for making decisions.

Methodology

The study was informed by positivist approach complemented by methodological triangulation during data collection and data analysis. Additionally, the nature of the study necessitated use of case study method to allow for probing the study area from different perspective. Also, the study utilised none-probability sampling specifically snow ball to select respondents for questionnaires and purposive sampling on selecting respondents for interviews.

Furthermore, in order to ensure a systematic assessment of the study area, it was necessary to outline the logical structure of the inquiry including the scope that indicated a broader perspective of the research project. (Yin, 1994, Babbie & Mouton, 2001; Creswell 2009). Additionally, the study considered a number of relevant steps critical to an empirical inquiry such as identifying a sampling criteria, data collection instruments, data analysis, validity and reliability, among other things. The summarised circle of the study design is shown in Figure1 below.

Variable	Output	Reason
Effort Expectancy	43 (65.1%) of respondents disagreed perceiving functionality of AFDAS clearly and easy to understand	As Mbeta Island had not experienced flooding since 2017, the surrounding communities were not able to see the system in full operation and consequently their difficulties in appreciating it
Performance Expectancy	46.5 (69.4%) of respondents disagreed that the automated flood disaster alert system would become useful	
Facilitating Conditions	46.5 (69.4%) of respondents disagreed that AFDAS had necessary infrastructure to support its use	The study showed gaps in the system lack of robustness in incorporating a dedicated interface that would allow for interacting with the system (e.g. toll free facility for communities interaction with system)
Social Influences	73 (71%) of respondents disagreed that the automated flood disaster alert system had changed their farming processes	As Mbeta Island had not experienced flooding since 2017, the surrounding communities were not able to see the system in full operation and consequently their difficulties in appreciating it

The study was focused on establishing the relationship of the five variables (Independent variables) where four were analysed quantitatively using frequencies and percentages to establish their impact on the dependent variable while the fifth one was qualitative that was measured through one to one interviews. From the study, Table 1 below shows the summarised findings based the evidence collected and inferences drawn.

The study findings were further analysed using regression. This was critical to establish the appropriate impact or relationship of each independent variable on the dependent variable. However prior to regressing the data sets were tested for normality in determining the impact of all variables in the study shown in Table 2 below:

Variable	R Square	Sig Value
Effort Expectancy	.062	.401
Performance Expectancy	.118	.094
Facilitating Conditions	.044	.588
Social Influences	.026	.082

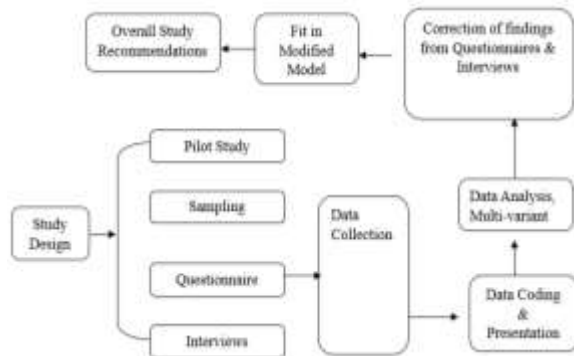


Figure 1 Research Circle
Source (Yin, 1994; Babbie & Mouton, 2001; Creswell 2009)

Study Findings

It is important to underscore the fact that establishing the relationship and appropriate impact levels of each variable on the dependent variable showed the relevance of each variable in ensuring that AFDAS is fully utilised by both communities and government agencies. The relevance of a variable was indicated by how high or low the impact registered as shown by their respective R Squared values. According performance expectancy showed the highest impact while social influences the lowest.

Conclusion and Recommendations

Conclusions

Natural and man-made disasters are wide spread and reflect vulnerabilities and susceptibility of people affected

when the early warnings of impending disasters are ignored.

The study findings availed a number of gaps and potential areas for improvement in ensuring that the adopted system in alerting communities on impending disasters such as floods, fires, road accidents, terror attacks, agricultural droughts etc. Also, it was established that leveraging the potential of ICTs for information sharing, escalation and handling warnings and disasters was critical.

Further, mechanisms to respond to disasters via resource coordinators such as DMMU ought to be alerted by effective and user-friendly information channels. Swift responsive systems have now become a hallmark of efficient relief to mitigate the impact of disasters in the 21st century. Thus, it must be pointed out that communities encountering disasters must display a greater level of knowledge and skills of systems implemented to present their challenges in order for them to appreciate such facilities.

Recommendations

According to study findings and the emerging picture from the empirical evidence, there is need to adopt a new and robust disaster mitigation system that would allow for integration of community needs and everyday challenges in relation to natural and man-made disasters. The following recommendations were made:

In consistent with the findings, there is need for embarking on investment or improvement of the existing automated flood disaster alert system to ensure that it becomes more robust. Such a system must integrate new features to improve interaction with the community.

Additionally, the researcher would recommend the following:

- There is need for a robust early warning disaster system that should incorporate Geo-sensing that entail detecting soil moisture levels that will consequently help in predicting floods and agricultural droughts using geo-sensors in distributed flood prone areas and fitted with a GSM network gadget to automatically send pre-coded message to the system for critical flood and droughts alerting DMMU and other stakeholders.
- Secondly, the system should be able to profile the location of the reported disaster/hazard using GPS google maps and reporting mobile device as a beacon to the location of the disaster or early warning.
- Thirdly, the system should be able to process any disaster type; fire, livestock and community disease outbreaks, traffic accidents, terror attacks, droughts via a downloadable mobile client-side app that integrates with the main aggregating system. The client side app used as a platform to

relay early warning messages to the main aggregating system deployed at DMMU. With prevailing social norms of texting apps such as Facebook, WhatsApp, twitter indicates that using the system app will be without challenge by mobile users of all ages and gender in relaying appropriate early warning messages to the main aggregating system to track disaster locations for and mitigation purposes.

- Fourthly, the system should leverage on a toll free GSM network platform for the local communities that should entail that anyone can access the system even when they have no airtime to relay disaster messages to the main aggregating system. In far flung rural areas where there may not be network installations, utilizing deployable specialized cell phone handset technology that can establish stand-alone communications capabilities can be accessed via indunas or headmen for use in times of disasters
- Fifth, the system should be able to alert the public of impending disasters by disseminating and sharing alert information by LDMC, DDMC and DMMU for public awareness, and disaster responses, by NGOs, aid organizations, LDMC, DDMC, DMMU, Cooperating partners, donors, Funders and other stakeholders
- Sixth, the system should be able to provide a Database pool of disaster related information for other purposes. It should be able to share online information and integrate with other systems such as websites, Data mining, Online Analytical processing (OLAP), that may wish to pull online information related to disasters and for other purposes such as researches, disaster intelligence and other information related to early warnings, incidents, patterns, trends, shared understanding, decision making and stakeholders needing to act to achieve common goals in disaster related issues.
- There is need to provide contextualized and easy to use resources such as dedicated ICT early warning systems integrated with mobile phones for community accessibility. In doing so, these resources would be utilized to reduce risks and impacts of disasters.

References

Maity, S., Ayanam, V.S., Christopher, S. and Parthiban, B., American Megatrends Inc, 2016. *Emergency alert system using management controller in network*. U.S. Patent 9,378,638.

Blum, J.R., Eichhorn, A., Smith, S., Sterle-Contala, M. and Cooperstock, J.R., 2014.

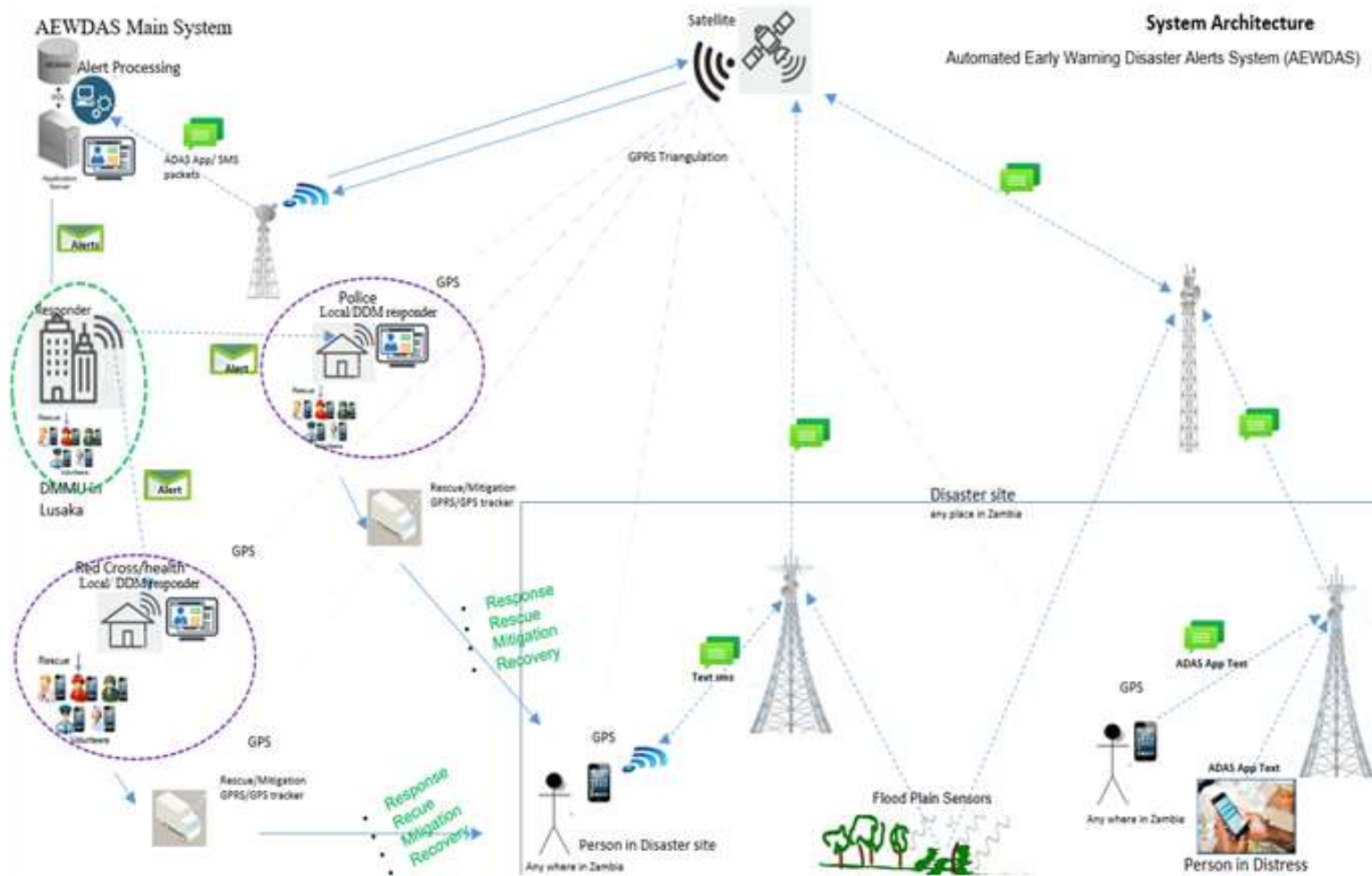
Real-time emergency response: improved management of real-time information during crisis situations. *Journal on Multimodal User Interfaces*, 8(2), pp.161-173.

Phiri, A., 2015. *Creating a model in community based disaster risk management for informal settlements: a case of Kanyama Settlement, Lusaka-Zambia* (Doctoral dissertation).

Hristidis, V., Chen, S.C., Li, T., Luis, S. and Deng, Y., 2010. Survey of data management and analysis in disaster situations. *Journal of Systems and Software*, 83(10), pp.1701-1714.

Rhea, et al. 2008 “*Status of disaster risk reduction in the sub-Saharan Africa region*. The World Bank).

Joshi, P.K., Munsu, M. and Joshi, A., 2010. Chapter 8 Teleprocessing for climate change adaptation and disaster risk reduction. In *Climate Change Adaptation and Disaster Risk Reduction: Issues and Challenges* (pp. 171-192). Emerald Group Publishing Limited.



System Architecture (AEWDAS)

A Survey on the State, Adoption and use of ICT in self-study by Students:

A Case of the Copperbelt University, Zambia

Nchimunya Chaamwe

Information and Technology Department
The Copperbelt University
Kitwe, Zambia
Chimz@cbu.ac.zm

Abstract— The past decade has seen a proliferation of ICT tools in Universities. Universities all over the world including Zambia have taken advantage of this increase by integrating ICTs in the teaching and learning of students. The internet is one such tool that has revolutionized the way learning and teaching is being conducted in Universities. Little research has been done though to verify the process of how university students adopt and use ICTs for their self-study in Zambian Universities. The research looked at Student experience in using ICTs in their learning, especially during their self-study at The Copperbelt University (CBU) in Zambia. The results showed that most students have adopted the use of ICTs in their learning. Of the surveyed students, 94.7% of them indicated that they consulted the internet sometimes or most of the times when studying. This is despite the obstacles they faced when accessing the internet. The paper recommended the improvement in internet connectivity and the increase in wireless access points to cover a wide internet access area.

Keywords—University; students; ICTs; self-study

I. INTRODUCTION

The penetration of ICTs in Zambia saw the Zambian government through the Zambia ICT policy commit itself to deploy ICTs at all levels of the Zambian educational system in order to improve and expand access to education, training and research facilities [1]. The Copperbelt University, one of the biggest public Universities in Zambia, with a population of over 10,000 students, is a beneficiary to this policy. The University boasts of a computer connected to broadband internet for almost each and every member of staff. The University has a total of 8 schools and a total of 8 computer labs. Most of the computers in these labs are also connected to the internet. Students who do not have personal computers or laptops have access to the computers in these labs. Apart from the wired internet, the University rolled out wireless access points in a number of places within the university. Students and members of staff can use their personal mobile devices to access the internet from these points.

Studies have shown that ICTs can be used to support the teaching and learning in schools as well as in universities. The Copperbelt University has appreciated this research evidence by purchasing and installing teaching aids such as projectors and interactive screens in some classrooms. In addition the

University Library has subscribed to a number of e-resources and the University populace can access these materials remotely without physically visiting the library. With such an increased emphasis on ICT, and a large investment in its infrastructure, students are expected to be competent and effective in adopting it; students are not only required to learn, but also to be able to use these resources in their self study. However, despite all this effort by the University management, the question that this research sought to answer is: are students making full use of the ICT in their self study?

ICTs offer a number of potential benefits for Students in Universities and Colleges, for example the authors of [2] in their study of ICT Adoption Impact on Students' Academic Performance discovered a close link between ICT adoption and academic performance. Their study revealed that students who had adopted ICTs in their learning and study performed better than those who did not. However, adopting ICTs in the study and learning process by students is not an easy task because there are a number of factors that may inhibit the adoption process. Empirical experience and also evidence from literature on ICT adoption has widely demonstrated that the uptake of ICTs is one of the most challenging tasks faced by decision-makers since it requires people to understand, absorb, internalise and adapt to new requirements [3].

There is currently a significant lack of related work that examines the factors that influence the adoption and diffusion of ICTs by students in their self-study in Zambian Universities. Key differences that exist between Universities in developed countries and developing countries like Zambia may result in different adoption behaviours of ICTs by students. Differences between Universities in developed countries and Universities in developing countries such as ICT infrastructure, University culture, economic status and social norms, may result in ICT adoption being different.

In summary, there exists a gap in knowledge about the factors that determine the adoption of ICTs by students in Universities in Zambia. Hence this study seeks to bridge this gap by developing a research model that provides insights into the uptake of ICTs by students in their studying and learning experiences.

II. LITERATURE REVIEW

A. Theoretical Frameworks

Researchers have argued that the rate of adoption of ICT by academics and students in their teaching and learning experiences is influenced by a number of factors ranging from attitude to competence. Frameworks have been proposed explaining the factors responsible for the adoption and use of ICTs by consumers.

The technology acceptance model (TAM) and the extended technology acceptance model (TAM2) have been widely used in studies examining why consumers used or adopted new communication technologies. TAM examines the mediating role of perceived ease of use and perceived usefulness in their relation between systems characteristics (external variables) and the probability of system use (an indicator of system success). A key purpose of TAM is to provide a basis for tracing the impact of external variables on internal beliefs, attitudes, and intentions. It suggests that perceived ease of use (PEOU), and perceived usefulness (PU) are the two most important factors in explaining system use [4]. TAM is as shown in figure 1 below.

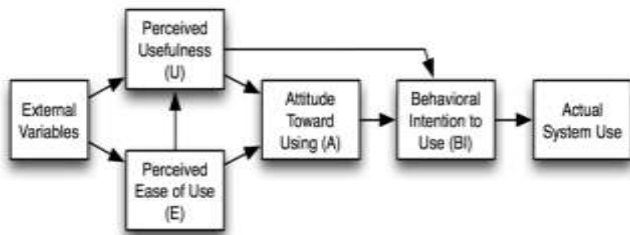


Fig. 1. Technology Acceptance Model (TAM)

B. Similar Studies

This section introduces and discusses similar studies that considered the adoption of ICTs by students for their learning and studying at higher learning institutions. Studies have indicated that the majority of the studies on ICT adoption have concentrated so much on the educator and not the learner. The assumption is that if the educator adopts ICTs in their teaching, then a learner will automatically adopt. This is so because teachers are responsible for the transformation in the ways in which teaching is carried out in the classroom and ICTs if adopted will provide opportunities for greater flexibility, interactivity and accessibility for engaging teaching and learning at the individual and group levels [5]. This may explain why there is scanty research attention on adoption and usage of ICTs by students in their learning and self-studying.

Against the background that ICTs adoption by students for learning and studying purposes is scanty, especially in Zambia, we conducted a survey to assess the levels of adoption at the Copperbelt University in Zambia. The literature reviewed showed a number of obstacles hindering students from adopting ICTs in their learning and self-study for both developed and developing countries. The obstacles

ranges from lack of ICT adoption in the syllabus, poor or unavailable ICTs to attitudes by students.

A study conducted to assess Internet usage, challenges, and attitudes among University Students at the University of Jordan in Jordan showed that most students actually access Internet before they attended university and that there is a positive attitude towards Internet but, they used it mainly for social websites, chatting and information gathering, the slow speed of the Internet connection and the lack of adopting ICT in courses syllabus were some of the constraints that students faced in the adoption of ICT in their learning [6].

The authors of [7] in their study that looked at the influence of students' ICT Skills and their adoption of Mobile Learning which they conducted in three Universities in New Zealand showed that the perceived ease of use and usefulness of mobile technology would mediate the relationship between ICT skills and the intention of students to adopt mobile learning. The survey also found that students' intention to adopt mobile learning was influenced by specific types of ICT skills. In particular, it was found that advanced skills in mobile technology and basic ICT skills both played significant roles in the intention to adopt mobile learning.

A survey was carried out to determine state, adoption and use of ICTs by students and academic staff at Mzuzu University in Malawi. The results showed a number of obstacles to the adoption and use of ICTs and they included poor network infrastructure, the limited number of computers, and high cost of internet access, persistent power outages, and the lack of relevant ICT skills, among others [8].

Another study was done in Nigerian Universities that looked at Information and Communication Technology adoption and use among students. The following factors were identified as constraints responsible for the poor adoption and use of ICT by students: erratic power supply, frequent breakdown of ICT facilities, and controlled access to ICT facilities were identified as some of the factors militating against the use of ICT in learning [9].

The authors of [10] in their study conducted in Kenya to consider the status of E-Learning in Public Universities, showed poor ICTs adoption by students and the following were identified as some of the factors contributing to students not adopting ICTs in their learning and studying: Lecturers were not using ICTs, Lack of computers and laptops to access soft copies especially during weekends and evenings, poor Internet accessibility and connectivity especially outside the university premises, expenses involved in Internet connectivity; and convenience of reading hard copies anywhere, anytime..

Further to the above, the authors of [11] reviewed some articles to identify factors that inhibit adoption of ICT in higher educational institutions. The review identified lack of institutional support, financial support, time to learn new technologies, access to computing, technical support and training. Other obstacles identified were technology reliability, uncertainty about its worth, resistance to change, negative

attitude, awareness, relevance, ease of use, attitude of organization, and computer literacy.

III. RESEARCH METHODOLOGY

The research was a survey with a total sample of 80 students. This number though small represented the students that frequents the wireless access points dotted in a few places at the Copperbelt University. It has been argued within the university circles that students who frequent these wireless access points do so to chat on social media. Data for the study was collected using a standardized questionnaire which comprised both closed and open-ended questions. The data collected was analysed using SPSS software. The population of the study was the Copperbelt University students. The sampling techniques adopted in this study included a simple random technique. The study targeted a total of 80 respondents, however due to study limitations; the study gathered a total of 57 responses which represents 71.25% response rate. These provided insights on the pace of ICT adoption by students in self-study in a University.

IV. RESULTS AND DISCUSSION

A. Access to computers

The study wanted to find out how students accessed computers. The findings from the survey showed that computer ownership and access among the students that were surveyed was high. The results as shown in table I shows that most students (63.2%) had in possession a personal Desktop computer or personal laptop. Computers made available by the university in Computer Labs (29.8%) were the second highest used, while PC or laptop owned by a colleague were the least used (7%). It is encouraging to note that a good number of students owned their own computers and this reduced the pressure on the university to provide computer access to all the students at the University. Studies in ICT adoption have shown that computer ownership and high access can positively and significantly affect the level of ICT adoption in the teaching and learning process. [12].

TABLE I: ACCESS TO COMPUTERS BY STUDENTS

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
PC	36	63.2	63.2
PC owned by friend	4	7.0	70.2
School lab	17	29.8	100.0
Total	57	100.0	

B. Internet Access

According to [13], Internet access provides great opportunities for the improvement in the quality of education for individual learners and contribute to national economic and social

welfare. It does this by providing new ways of teaching and learning, better access to a much wider range of information and resources. The Copperbelt University realises the importance of the internet in enhancing the teaching and learning experiences by offering both wired and wireless internet access to both students and members of staff. The desktop computers in all the labs are connected to the Internet through wired connection. In an effort to increase internet connectivity the university has placed wireless access points in strategic areas so that those with mobile devices like laptops could access the internet from these areas. The high levels of Computer accessibility by students entailed more students having access to the internet using the same computers. The results from the study showed that all the surveyed students had access to the internet. The study sought to find out means by which students accessed the internet. As shown in table II below most students (42%) accessed the internet through their personal bundles only, although a good number access through the University network only, either wired or wirelessly (33.4%). Most of those who accessed their internet using their personal bundles also accessed through the university network either wirelessly or wired (19.3%). Another means of accessing internet for students was through commercial internet cybercafés (5.3%).

Findings on this aspect of this survey are similar to those in a survey conducted by [8] involving students and academic members of staff at Mzuzu University in Malawi. This study noted extensive private internet access among respondents (66.7%) which was also the case at Mzuzu University although statistics were slightly higher (77.3 %). The main difference between these two studies is that whereas a good number of students at Mzuzu accessed the internet from Café's, only 5.3% of the surveyed students at CBU accessed the internet from a café'.

TABLE II: MODE OF INTERNET ACCESS BY STUDENTS

	Frequency	Percent	Cumulative Percent
Bundles	24	42.1	42.1
Internet Cafe	3	5.3	47.4
Wire CBU and Bundles	1	1.8	49.1
Wired CBU	5	8.8	57.9
Wireless CBU	14	24.6	82.5
Wireless CBU and Bundles	10	17.5	100.0
Total	57	100.0	

C. ICT skills

Having access to the internet by students is one thing and having the skills to using it in their learning is another thing. The study therefore sought to find out the ICT skill levels of the students for this is one the factors contributing to the adoption of ICTs in learning by students. Studies have shown

that those with high ICT skill are more likely to adopt ICTs in their learning. [14]. Students were asked to rate their ICT skills using either poor, average, Good or Very good. ICT skills can be measured by being able to use digital tools, media, and resources efficiently and responsibly, to solve practical tasks, find and process information, design digital products, and communicate content. ICT skills also include developing digital judgment by acquiring knowledge and good strategies for the use of the Internet [15]

The results in table III below showed that 63.2 % had either good or very good ICT skills, followed by 29.8% who rated themselves average. Only 7% of the students had a poor mastery of ICT skills. The results showed that most of the students had some skill in using computers (97%) and therefore according to [14], which found that advanced skill in mobile technology and basic ICT skills both played significant roles in the intention to adopt mobile and ICT supported learning, the surveyed students were expected to adopt ICTs in their self-study.

TABLE III: STUDENTS' ICT SKILLS

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
Average	17	29.8	29.8
Good	18	31.6	61.4
Poor	4	7.0	68.4
Very Good	18	31.6	100.0
Total	57	100.0	

D. Internet Usage

The study sought to determine the frequency of Internet access and usage by students at the university. Students were asked to state either daily, twice a week, weekly or monthly. The results in table IV shows that 87.7 % accessed the internet daily, followed by 7% accessed at least twice a week, 3.5% accessed the internet weekly and 1.8% accessed the internet at least monthly. The high percentage in the daily usage could be attributed to the high penetration of mobile phones among students that could be used to access the internet.

TABLE IVIII: INTERNET USAGE BY STUDENTS

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
Daily	50	87.7	87.7
Monthly	1	1.8	89.5
Twice a week	4	7.0	96.5
Weekly	2	3.5	100.0
Total	57	100.0	

E. Obstacles in accessing the internet

An attempt was made to explore the obstacles faced by students in and when accessing the internet. The reasons given by students ranged from lack of access to slow connectivity. Table V Showed that 64.9% of the students had access to the internet but experienced slow connectivity. The problem of slow connectivity was not only experienced by users of the university network, it was also experienced by students who were using their own data bundles to access the internet. 33.3% of the students indicated that the internet was not readily available. This could be attributed to the fact despite having their own computers, the wireless access points provided by the university covered a small radius. The other reason could be the cost of bundles for those using their own bundles to access the internet.

TABLE IV: OBSTACLES IN ACCESSING THE INTERNET

<i>Obstacle</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
Availability	19	33.3	17.5
Inadequate PCs	1	1.8	35.1
Slow Conn	37	64.9	100.0
Total	57	100.0	

F. Internet Consultation

The students were asked to indicate as to how often they consulted the internet when studying. They were required to indicate whether they consulted most of the times, sometimes, rarely or never. The results in table VI showed that most of the students consulted the internet most of the times when studying. 57.9 % of the students indicated that they consulted most of times, 36.8% indicated that they consulted sometimes and the rest indicated they rarely or never consulted the internet when studying. Despite the obstacles faced by students when accessing the internet, most students consulted internet resources in their studies. The results showed a positive attitude by students towards the usage of the internet for study purposes. These results agree with the research that was conducted by [16], that aimed at investigating the attitudes of students at the Islamia University of Bahawalpur, Pakistan towards learning through the Internet.

TABLE VI: LEVELS OF INTERNET CONSULTATION BY STUDENTS

<i>Internet Consultation</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
Most times	33	57.9	57.9
rarely	3	5.3	63.2
Sometimes	21	36.8	100.0
Total	57	100.0	

G. E-Library resources awareness

The Copperbelt University has subscribed to a number of e-Resources. The study wanted to establish the levels of awareness by students on the availability of these e- Libraries. The study revealed as shown in table VII below that 68.4% of the students were not aware of these resources. Only 31.6% had some knowledge of these libraries. This showed that most students consulted general internet resources in their self-study rather than the subscribed for e-resources. This lack of awareness could be as a result of poor or no information dissemination by the university management. It could also be attributed to lack of recommendations to these resources by academic members of staff who themselves may not be using these e-resources in their teaching and research.

TABLE VII: UNIVERSITY SUBSCRIBED E-LIBRARY AWARENESS

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
No	39	68.4	68.4
Yes	17	31.6	100.0
Total	57	100.0	

H. Other uses of the Internet

The study sought to find out other uses of the internet by students apart from using it for studies. The results showed that the majority of the students used the internet also for chatting purposes on social media (61.4%). The results showed that 80.7% of the students used the internet for other things other than studying. This high percentage could explain the daily internet access by the majority of the respondents (87.7%). The remaining 19.3% indicated that they used the internet for the purpose of studying alone. Table VIII below outlines the other uses of the internet by students.

TABLE VIII: OTHER USES OF THE INTERNET BY STUDENTS

	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
	1	1.8	1.8
Chat	8	14.0	15.8
Chat, download	6	10.5	26.3
Chat, email	5	8.8	35.1
Chat, email, download	9	15.8	50.9
Chat, News	1	1.8	52.6
Chat, shop, games, download	4	7.0	59.6
Chat, email, games,	1	1.8	61.4
Download	6	10.5	71.9

Downloads and news	1	1.8	73.7
Email	1	1.8	75.4
Email and news	1	1.8	77.2
Email, download	1	1.8	78.9
News	1	1.8	80.7
Study	11	19.3	100.0
Total	57	100.0	

V. CONCLUSION AND RECOMMENDATIONS

The study looked at the state, adoption and use of ICTs by students at the Copperbelt University in their self-study. An analysis of responses from students revealed that all the students surveyed at the Copperbelt University have access to computing facilities. 63.2% of the surveyed students had in possession their own computer. A good number of students accessed computing facilities provided by the university and a few (7%) used computers owned by their friends. This high rate of access translated into all the students that were surveyed having access to the internet. A good number of students accessing their own computers translated into a good number using their own bundles when accessing the internet (42.2%). The high levels of internet access (100%) translated into a good number of the respondents indicating that they consulted the internet most of the times when studying (57.9%). The study shows that despite the willingness of students in the university to integrate ICTs in their study and learning process, some obstacles were identified that make it difficult for the students to achieve these goals. The study revealed also that students used the internet for a variety of other things apart from using it for studying. The study showed that despite students consulting the internet whilst studying, 68.4% of the surveyed students were not aware of the E-resources the university was subscribing to.

The researcher recommends that the University management should take steps to improve the state of ICTs at the institution by among others increasing network coverage, purchase more computers for use by students. It is again recommended that management through the directorate of ICT should improve the connectivity of the existing network to make it more reliable. It is also recommended that management through the Library should disseminate information to students and members of staff on the availability of e-Libraries the university subscribes to. The library should further facilitate installation of a proxy server to ensure that students and academic staff who use private networks to access internet are able to access the e-libraries. It has also been observed in this study that a majority of the students who use the internet for academic purposes also use the internet for chatting purposes and other non-academic uses which may consume a lot of their time. Care should therefore be taken by these students to create a balance between constructive use of ICT and ordinary internet usage.

References

- [1] National Information & Communication Technology Policy, April, 2006. Available online <http://unpan1.un.org/intradoc/groups/public/documents/unpan/unpan032690.pdf>.
- [2] S B Wael, et al, "ICT Adoption Impact on Students' Academic Performance: Evidence from Saudi Universities", Education Research International, Vol 2018, Article ID 1240197, 2018.
- [3] M. Magni, and F. Pennarola, "Intra-organizational relationships and technology acceptance", *International Journal of Information Management*, Vol 28, no 6, pp517-523, 2008.
- [4] Legris, P, Ingham, J & Collette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model *Information & Management*. 40 (3), p 191–204.
- [5] E. L Japhet and A Usman, "Factors that influence teachers' adoption and integration of ICT in teaching/learning process", , Educational Media International Journal, Vol 55, no 1, pp79-105, 2018
- [6] T Almarabeh et al, "Internet Usage, Challenges, and Attitudes among University Students: Case Study of the University of Jordan" *Journal of Software Engineering and Applications*, Vol.9 No.12, 2016.
- [7] K. M. Callum, and L. Jeffrey , "The Influence of Students' ICT Skills And Their Adoption Of Mobile Learning", *The Australasian Journal of Educational Technology*, Vol 29, no: 3, pp303 -315, 2013
- [8] A. H. Chaputula, "State, adoption and use of ICTs by students and academic staff at Mzuzu University, Malawi", *Program*, Vol. 46 No: 4, pp.364-382, 2012
- [9] O, A Oluseun Et Al, "Information and Communication Technology Adoption and Use Among Students Of A Nigerian University For Distance Learning". *Library Philosophy and Practice* (e-journal). 2015
- [10] G. L. Makokha and D. N. Mutisya, "Status of E-Learning in Public Universities in Kenya", *International Review of Research in Open and Distributed Learning* Vol 17, No 3, April – 2016
- [11] S. N. Attuquayefio and H. Addo, "Review of Obstacles Which Inhibit ICT Adoption in Higher Education", *European Scientific Journal*, vol.10, No.13 , May, 2014
- [12] M. Cox Et Al, "What Factors Support or Prevent Teachers from Using ICT in their Classrooms?", *Education Line*, Available online: <https://www.leeds.ac.uk/educol/documents/00001304.htm>
- [13] Internet Society, "Internet Access and Education: Key considerations for policy makers", 2007 https://www.internetsociety.org/wp-content/uploads/2017/11/Internet-Access-Education_2017120.pdf
- [14] K. M. Callum, L. Jeffrey, "The Influence of Students' ICT Skills And Their Adoption Of Mobile Learning", *The Australasian Journal of Educational Technology*, Vol 29, no: 3, pp303 -315, 2013
- [15] Norwegian Directorate for Education and Training. (2012). Framework for basic skills. Retrieved from http://www.udir.no/PageFiles/66463/Framework_FOR_BASIC_SKILLS.pdf?epslanguage=no
- [16] K. S. Ahmad Et Al "Internet access, use and gratification among university students: A case Study of the Islamia University of Bahawalpur, Pakistan. *Chinese Librarianship*", *International Electronic Journal*, Vol. 31., 2011 URL: <http://www.iclc.us/cliej/cl31KKB.pdf>