CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

PROCEEDINGS OF THE CYBER SECURE NIGERIA

2019 Conference

THEME:
IMPLEMENTING CYBERSECURITY AND DATA PRIVACY PRACTICES IN NIGERIA

APRIL 9 - 11, 2019

VENUE:
CBN INTERNATIONAL TRAINING INSTITUTE, MAITAMA, ABUJA, NIGERIA

EDITORS:

DR MORUFU OLALERE
DR. OLAYEMI M. OLANIYI
DR. SHEFIU O. GANIYU
MR. OLUWAFEMI OSHO
DR. SHAFI'I M. ABDULHAMID
DR. HARUNA CHIROMA

Proceedings of the

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019
CBN International Training Institute,
Maitama, Abuja, Nigeria

Editors

Dr. Morufu Olalere
Dr. Olayemi M. Olaniyi
Dr. Shefiu O. Ganiyu
Mr. Oluwafemi Osho
Dr. Shafi'i M. Abdulhamid
Dr. Haruna Chiroma

FOREWORD

Cyber Security Experts Association of Nigeria (CSEAN) is a non-profit organisation made up of information security professionals, committed to a collective purpose and vision to be a vehicle championing the cause and awareness of information security best practices, acting as an agent of change to address cybercrime phenomena, through engaging intellectual minds, business and political leaders. As cyber security protagonists, we look to engage in healthy debates, to expand knowledge, awareness and understanding of the issues around cybercrime. Through workshops and seminars, we share knowledge and expertise, thereby contributing towards the growth of the information security industry in the country. We also create and support other forums to breed and equip future generations of information security professionals. In addition, the Association holds consultations and discussions with the government at different levels on tackling cybercrimes.

Conceived in late 2013, formed and registered with Corporate Affairs Commission in January 2014, CSEAN is focused on individual membership. We remain the only international cyber security association in Nigeria with membership comprising of Nigerian information security professionals both at home and in different parts of the world. Considering that cyber security is a global phenomenon, we intend to share knowledge and experience of our members in diaspora and home in tackling information security problems in Nigeria. We are convinced this would foster and promote the development of the information security industry and encourage the professional development of our members.

Cyber risks are growing and changing rapidly. Cybercriminals are stepping up their game and data breaches are becoming increasingly common and more severe. Hackers continue to devise new techniques to infiltrate the networks of government and organizations, to cause damage, access sensitive data, and steal intellectual property. As old sources of cyber threats fade, new sources will emerge to take their place.

The vast scope of cyber threats makes a compelling case for a multi-stakeholder collaboration in curbing domestic and international threats. The reality is that Nigeria is not immune from these threats. Hence, to effectively address the continually emerging threats, the CSEAN initiated the annual Cyber Secure Nigeria Conference, with the first edition held in 2015. The conference essentially provides a platform for discussion and sets precedence for constructive debates on various hot button topics around cyber security.

Remi Afon
President, CSEAN

WELCOME ADDRESS

In the constantly evolving digitalisation of the global economy, cybersecurity incidents are having bigger impact more than ever before. Cyber Security Experts Association of Nigeria (CSEAN) takes it as a responsibility to regularly increase and reinforce awareness of cybersecurity, including associated risks and threats, and provide solutions for increasing cybersecurity.

In this respect, I welcome you to the Cyber Secure Nigeria 2019 Conference, themed "Implementing Cybersecurity and Data Privacy Practices in Nigeria." Our focus is to further raise awareness on the need for individuals, institutions and government to pay adequate attention to cybersecurity and its associated risks. The annual conference is an important means that CSEAN uses to help Nigerians understand not only the risks that come with using the Internet, but also the importance of practicing safe online behaviours and digital privacy rights.

Our current network society (relating to social, political, economic and cultural changes) is a product of the digital revolution and the impact on society has been transformative at all levels, including communication, commerce, social interaction and access to knowledge. Despite the widespread cases of kidnapping, insurgency, terrorism, banditry, herder-farmer conflicts, militancy and political unrest threatening the social and economic development; we must also pay adequate attention to the ever growing effect of cybersecurity incidents.

The Conference intends to highlight ways in which academics, professionals and government can key into addressing the challenges of security incidents. We are pleased to have an impressive assembly of highly respected speakers to lead and provide direction, support and leadership for the dynamic conference programmes. As the industry speakers and academic presenters shed light on key challenges surrounding cybersecurity and data privacy practices, I encourage participants to raise issues and ask questions that may contribute to identifying necessary steps required to safeguard information asset, digital privacy rights and cybersecurity. Also, participants are urged to take advantage of the networking opportunities during the Conference to engage, interact and share ideas.

We want to thank the generosity of our partners, sponsors, media and friends. We thank them especially for their support of the Conference, their contributions are crucial to the development of Cyber Secure Nigeria. I extend my thanks to all the members of the Conference Organizing Committee and Program Committee. I also appreciate the CSEAN President, Mr Remi Afon, members of CSEAN National Executive Council and Secretariat staff for their time, contributions and commitment towards the success of this conference.

Thank you and have a fruitful conference!

John O. Odumesi
Chairman, Conference Organizing Committee

PROGRAMME OF EVENTS

Day 1 Tuesday, April 9, 2019	
Time	**Activity**
Opening Ceremony	
08:30 – 09.30	Arrivals & Registration
09:30 – 10.00	Arrival of the Special Guest of Honour
10:00 – 10.05	National Anthem & Opening Prayers
10:05 – 10.15	Welcome Address by Remi Afon (President, CSEAN)
10:15 – 10.25	Remarks by Special Guest: Mr. Ayor Ogon (The Clerk, Senate Committee on ICT/Cybercrime)
10:25 – 10.35	Remarks by Special Guest: Mr. Ibrahim Magu (Ag. Chairman, EFCC)
10:35 – 10.45	Remarks by Special Guest: Prof. Umar Garba Danbatta, (EVC/CEO, NCC)
10:45 – 10.55	Remarks by Special Guest: Dr. Adebayo Shittu (Hon. Minister of Communication)
10:55 – 11.05	Remarks by Special Guest: Mr. Boss Mustapha (Secretary to the Government of the Federation)
11:05 – 11.25	Keynote Address by Dr. Isa Pantami (DG, NITDA)
11:25 – 11.40	Opening Address by Special Guest of Honour: His Excellency Prof. Yemi Osibajo (Vice President Federal Republic of Nigeria)
11:40 – 11.45	Vote of Thanks by Dr. Ismaila Idris (Vice President, CSEAN)
11:45 – 12.00	Group Photograph with Special Guest of Honour
12:00 – 12.10	Departure of Dignitaries
12:10 – 12.30	Exhibition Break
Plenary Session 1: Fake News, Cyber Propaganda and Cyber Threats	
12:30 – 12.50	Presentation by Gold Sponsor: CyberDome
12:50 – 13.50	Topic 1: Sattire: Is Fake News the New Normal? *Ade Shoyinka (Senior Network Security Consultant, UK)* Topic 2: Impact of Fake News on Electioneering Process *Patrick Essien (Cyber Security Practitioner, Senate Committee on ICT & Cybercrime)* Topic 3: Security Operations and Management: Hustlers and Survivors in the Nigeria Cyber Threat Environment *Mobolaji Moyosore (Cyber Security Strategist & Thought Leader, Texas, USA)*
13:50 – 14.50	Lunch Break

Plenary Session 2: Blockchain, Privacy and Law Enforcement	
14:50 – 16.00	Presentations by Panelists/ Discussants with Moderated Discussion of Panelists/ Discussants with Audience Topic 1: Governing Blockchain: Privacy Coins and Law Enforcement *Adedeji Owonibi (Senior Partner, Fianacial Forensics A&D Forensics LLD)* Topic 2: Blockchain in Banking and Financial Services *Mohammed Wanka (Head IT, Premium Pensions)* Topic 3: Cybersecurity for Security and Defence Agencies *Tanwa Ashiru (Security & Defence Expert)*
16:00 – 16.05	Closing Remarks by Sakariyah Abdulkadir Bolanle (Chairman, CSEAN, Kwara Chapter)

Day 2 Wednesday, April 10, 2019	
Research Paper Presentation 1	
09:00 – 10.10	Topic 1: Securing Electronic-Health Record System Using Cryptographic Techniques *Shefiu Olusegun Ganiyu, Olayemi Mikail Olaniyi, and Orooniyi Tosin* Topic 2: IoT-Based Forensic System for Monitoring and Detecting Farmers-Herders Activities in Nigeria *Mohammed Ibrahim, Mohd Taufik Abdullah, and Muhammad Amin Ahmad* Topic 3: Security Risk Analysis and Management in a Non-Interest Banking Sector *Joseph A. Ojeniyi, Samuel A. Oyeniyi, and Shafi'i M. Abdulhamid* Topic 4: Cybersecurity in the Age of FinTech and Digital Business *Faya Moses and Nnubia Ogbuefi*
Plenary Session 3: Implementing Effective Privacy Controls and Policies	
10:10 – 10.50	Topic 1: Impact of PII breaches on Individuals and Organisations Glory O. Idehen (Assistant Director, Central Bank of Nigeria) Topic 2: How can the Nigerian legal framework address security challenges in a privacy-centric economy? George Tyendenzwa (Head, Cybercrime Prosecution Unit, Federal Ministry of Justice) Topic 3: National Vulnerability Landscape Dr. Uche Mbanaso (Centre for Cyberspace Studies, Nasarawa State University Keffi)
11:50 – 12.05	Exhibition Break
12:05 – 12.25	Presentation by Gold Sponsor: CyberDome
Breakout Session	
12:25 – 13.25	Threat Modeling Exercise
13:25 – 14.30	Lunch Break
14:30 – 15.00	Privacy and Cyberstalking Ismaelle Vixsama (Information Security Management Manager at Kudelski Security, Atlanta USA)
15:00 – 15.30	Dark Web Investigations Andrew Lewman (Dark Web Investigation Specialist, San Francisco, USA)
15:30 – 15.50	Practical Ways of Improving Patients' Confidence by Addressing Privacy Concerns in Nigeria Health Sector Tunji Igbalajobi (Principal Consultant, CyberCode Limited)
15:50 – 16.00	Vote of Thanks by Dr. Olayemi Olaniyi (Chairman, CSEAN, Minna Chapter)

Day 3 Thursday, April 11, 2019	
Research Paper Presentation 2	
09:00 – 10.20	Topic 1: Comparative Evaluation of Artificial Neural Network and Support Vector Machine for Money Laundering Detection *Nimatullah Yusuf and John. K Alhassan* Topic 2: Impact of Cyber Security Implementation in an Economy Driven by Cyber Physical System (CPS) *Olabode Gbenga Agboola* Topic 3: V-Authenticate: Voice Authentication System for Electorates Living with Disability *Olayemi M. Olaniyi, Jibril Bala, Juliana Ndunagu, Adamu Abubakar, and Ahmadtijani Is'Haq* Topic 4: Performance Analysis of Security Information and Event Management Solutions Detecting Web-Based Attacks *Morufu Olalere, Juliana Ndunagu, Shafi'i M. Abdulhamid, and Peter Odey*
10:20 – 11.30	CSEAN National EXCO General Election Manifestoes, Election to National EXCO Positions
11:30 – 11.45	Exhibition Break
11:45 – 12.05	Beware of that Email Attachment, It Could Be a Death Trap! *Oluwafemi Osho, Immaculatta Obar, and Ayanfeoluwa Oluyomi*
12:05 – 12.25	Regulations Not Enough, Compliance is Key *Simbiat Ozioma Sadiq*
12:25 – 14.00	Lunch
14:00 – 15.30	CSEAN Annual General Meeting (AGM)
15:30 – 15.35	Vote of Thanks by John Odumesi (Chairman, LOC)
Dinner/ Cocktail and Award Nite	
19:00 – 21.00	Dinner/ Cocktail and Award Nite

SPEAKERS

Andrew Lewmana
Senior Technology Executive, San Francisco, USA

Ismaelle Vixsama
Information Security Governance and Strategy Professional, Atlanta, USA

Abdul-Hakeem Ajijola
Chair, Working Group, Cyber Incidence Management & Critical Information Protection @ GFCE

Mobolaji Moyosore
Cyber Security Strategist & Thought Leader, Texas USA

Adedeji E. Owonibi
Senior Partner, Financial Forensics/Digital Currency Governance & Investigations

Simbiat Ozioma Sadiq
Cyber Security Analyst, Cyber Dome

Dr. Uche Mbanaso
Centre for Cyberspace Studies, Nasarawa State University, Keffi

Ade Shoyinka
Network Security Consultant and Chairman CSEAN UK Branch

Patrick Essien
Cyber Security Practitioner, Senate Committee on ICT & Cybercrime

Olatunji Igbalajobi
Principal Consultant,
CyberCode Limited and
Director of Programmes
CSEAN

Adeolu Fadele
Founder & Lead Researcher,
Cryptography Development
Initiative of Nigeria

Glory O. Idehen
Assistant Director, Central
Bank of Nigeria

Tanwa Ashiru
Chief Intelligence, Security
and Defense Analyst,
Bulwark Intelligence

COMMITTEES

Conference Organizing Committee

- John O. Odumesi (**Chair),** *Office of The National Security Adviser, Abuja, Nigeria*
- Elizabeth W. Kolade (**Secretary),** *Defence Space Administration, Abuja, Nigeria*
- Oluwafemi Osho, *Federal University of Technology, Minna, Nigeria*
- Agwu Kalu, *Iresa Limited, UK*
- Amos O. Bajeh, *University of Ilorin, Ilorin, Nigeria*
- Olayinka Scholastica, *Comcast Cable, USA*
- Helen George, *US Gov., US*
- Sakariyah A. Bolakale, *University of Ilorin, Ilorin, Nigeria*
- Olayemi M. Olaniyi, *Federal University of Technology, Minna, Nigeria*
- Kunle Alabi, *Emergence Networks, UK*
- Akinbo A. A., Cornerstone, *Member, Board of Trustees, African ICT Foundation*

Program Committee

- Dr. Morufu Olalere **(Chair)**, *Federal University of Technology, Nigeria*
- Dr. Shefiu O. Ganiyu **(Secretary)**, *Federal University of Technology, Nigeria*
- Prof. Sanjay Misra, *Covenant University, Nigeria*
- Prof. Ali Salamat, *UTM, Malaysia*
- Prof. Victor O. Waziri, *Federal University of Technology, Nigeria*
- Prof. Jemal Abawajy, *Deakin University, Australia*
- Dr. Olayemi M. Olaniyi, *Federal University of Technology, Nigeria*
- Dr. Mohd Taufik Abdullah, *UPM, Malaysia*
- Dr. Shafi'i M. Abdulhamid, *Federal University of Technology, Nigeria*
- Dr. Azizol Abdullah, *UPM, Malaysia*
- Dr. Haruna Chiroma, *FCE Technical, Nigeria*
- Dr. Ibrahim Abaker Jargio Hasheem, *APU, Malaysia*
- Mr. Oluwafemi Osho, *Federal University of Technology, Nigeria*

TABLE OF CONTENTS

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

Impact of Cyber Security Implementation in an Economy Driven by Cyber Physical System (CPS)

Olabode Gbenga Agboola

BS Fortis, London, United Kingdom

***Abstract*:** How ready is Nigeria government and the industries that run on cyber physical system to contain or remediate cyber-attack in their Industrial Control System (ICS other name for CPS)? Will Nigeria wait to experience what Saudi Aramco experienced in 2012? Arguably, manufacturing and oil and gas are the industries with the presence of cyber physical system not in Nigeria alone but globally. These two industries are the milk cow of Nigeria (Awolowo, 2018). Since early 1980's, the world has witnessed a drastic increase in the number of Cyber-attacks on Cyber-Physical Systems (CPS). One of the early attacks was the 'Logic Bomb' that set the Trans-Siberian Pipeline on explosion. The advent of Internet of Things (IoT) has increased the opportunities for attacks because the number of network and IP-based devices in some processing plants and facilities are increasing. Electronic intrusions into cyber physical system has reflected the impact of non-implementation of cyber security standard. How well has Information Security Management System (ISMS) been implemented in Nigeria as a whole? How well has it been implemented in manufacturing and oil and gas industries?

***Keywords*:** Cyber Physical Systems; Manufacturing, Oil and Gas Industry; ISMS, Implementation, Audit, ISO 27k

I. INTRODUCTION

Since early 1980's, the world has witnessed a drastic increase in the number of Cyber-attacks on Cyber-Physical Systems (CPS). One of the early attacks was the 'Logic Bomb' that set the Trans-Siberian Pipeline on explosion (Johann Rost, 2011). The advent of Internet of Things (IoT) has increased the opportunities for attacks because the number network and IP-based devices in processing facilities are increasing Electronic intrusions into critical industrial control systems has reflected the pervasive nature of these infrastructures and the unavailability or inefficiency of various available processes, systems and technologies to prevent the intrusions (Evans, 2011).

The outbreak of Stuxnet in 2010 was the beginning of attack on processing facilities through a computer worm that could adversely affect the plants physically. This Stuxnet worm adopted concealment methodology to cause destructions on the control systems of processing facilities (Wang, 2018). The sources of these attack and exposures range from social 'hactivists' who wanted to make political point, relevance and presence, teenager hackers, even to criminal organizations and ultimately to rival nation states (Fisher, 2014). A potential attack could be launched by any member of one of these groups and that could seriously damage a significant portion of critical industrial system (Krebs, 2012). Some asset owners of critical infrastructures and industrial partners reported more than 240 cybersecurity attacks and breaches in 2014 on their Industrial control systems (Cowan, 2015).

It is an established fact that there are lots of establishments in these two industries who have not adopted the strategic approach of using cyber security standards and policies to manage their Information systems and technology. As a matter of fact; The National Information Technology Development Agency (NITDA) estimated that Nigeria lost $450 million to cyber security related issues in 2015 alone. Nigeria and four other African countries have been listed among the world's highest risk countries in the Global Threat Impact Index released for May 2017, by Check Point Software Technologies Limited (Yadua, 2018).

Manufacturing industry in Nigeria places more importance on implementation of quality standards (ISO 9000) for the maintenance section of their plants because this directly affects their productivity and economy (Hammar, 2018). On the other hand, Oil and gas sector shows more passion for the implementation of quality management system (IS O/TS 29001) more than and standard because of the fact that their operational environment is hazardous so physical safety of men and assets are not negotiable. However, this oil and gas industry who have heavy CPS in their operations, pays little or no attention to an implementation of Information Security Management System (ISMS).

Some of the perspectives on the non-implementation of ISMS on the CPS environment as detailed in a survey conducted by SANS, it was an established fact that security breaches on CPS environment is on the rise. Interestingly, more than 27% of respondents are not aware of the information security management standards that their organization currently adopts. Some claimed that no Information Security Management System (ISMS) has been implemented for their CPS environment. Security of CPS environment especially in Oil and Gas industry and manufacturing will be difficult to be guaranteed when organisations have not implemented any ISMS standard or have not audited the already implemented ones – says SANS (Luallen, 2014).

II. IMPLEMENTING ISMS

ISMS implementation will surely play a role in securing CPS-based organization and this will ultimately become an assurance that Nigeria economy will not slope as a result of cyber vices. ISACA reckons with five stages of implementation of the information security management system. A typical ISMS implementation project should observe these five stages. The stages are: (1) Planning, (2) Definition, (3) Implementation proper, (4) Monitoring and (5) Review and Improvement (Gerhard Funk, 2017). The simple philosophy of incorporating ISMS into corporate control process will be followed as depicted the figure below.

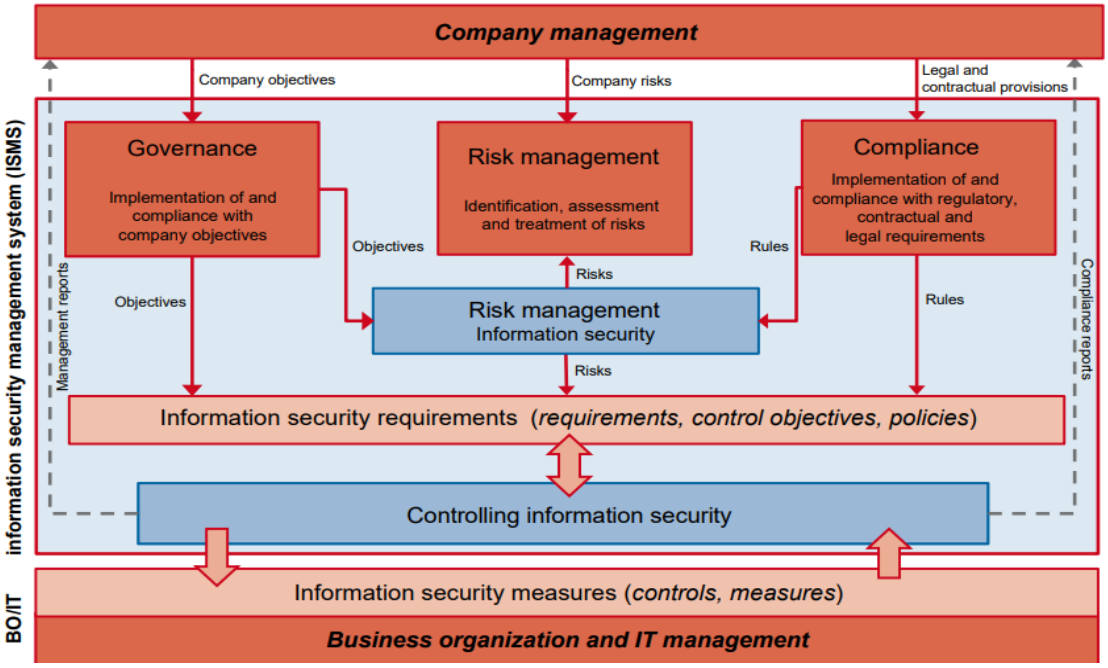


Figure 1. Incorporating the ISMS into corporate control processes

From the figure above (figure 1), the three basic corporate control processes are the governance, risk management and compliance. These three do not only form the basis for the information security risk management but also serves as input for it. All that goes into either implementation or auditing of ISMS are all embodied together as risk management.

It is an established fact that an adoption of cybersecurity framework or standard such as ISACA COBIT 5, CIS 20, NIST 800-53/CSF, ISO/IEC 27001/2, Oil and Natural Gas Subsector Cybersecurity Capability Maturity Model (ONG-C2M2) and other relevant ones can improve the efficiency of IT Service Management (ITSM) and SCADA systems service availability in organizations driven by CPS (ISACA, 2014). This literature will focus primarily on the Implementation of ISMS using ISO 27001/2 standard. ISMS can either be implemented as a project or as an audit project can be carried out for the organisations who have already implemented ISMS.

ISO/IEC 27001/2 technically refers to as an Information Security Management System (ISMS) which is an integration of all relevant activities that underpin information risks management. An ISMS is a detailed management framework through which organizations discover, analyze and address their information risks. Implemented ISMS guarantees that the security endeavours are refined to keep up with security threat changes, vulnerabilities and impacts to the business.

For simplicity of the implementation process, ISO 27001 standard are broken down into 14 domains which are; Context of the Organization, Leadership and Commitment, IS Objectives, IS Policy, Roles, Responsibilities and Competencies, Risk Management, Performance Monitoring & KPIs, Documentation, Communication, Competence and Awareness, Supplier Relationships, Internal Audit, Incident Management and Continuous Improvement (CORAL, 2012).

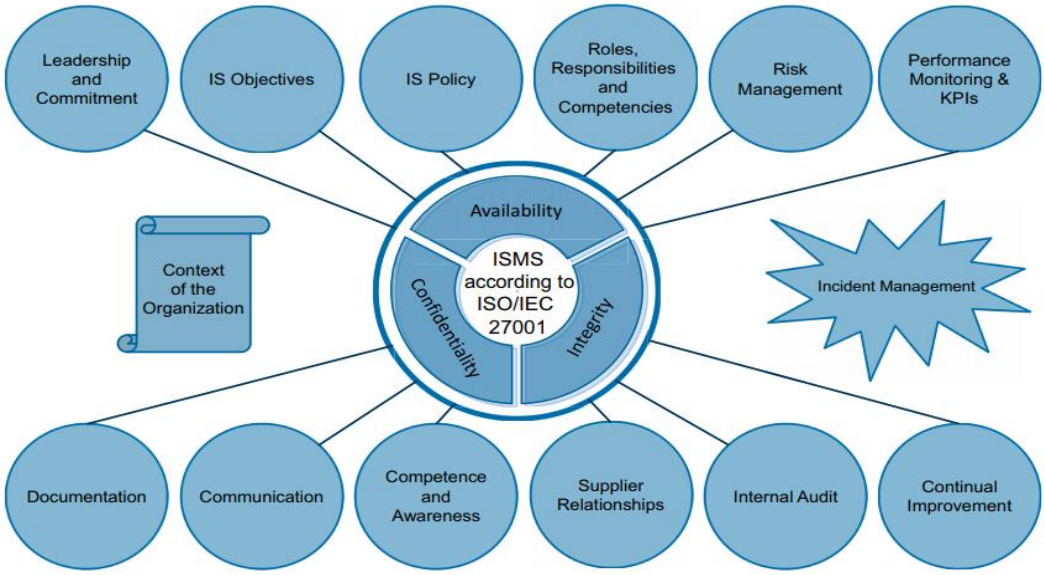


Figure 2. The 14 Domains of ISO 27001

As depicted in figure 2 above, these 14 components are interlinked to ensure information confidentiality, integrity and availability. According to Meng-meng Ren (Ren, 2015), a typical project such as ISMS implementation should adopt a standard model such as Plan-Do-Check-Act (PDCA) which is also called Deming wheel. The figure 3.below illustrates PDCA

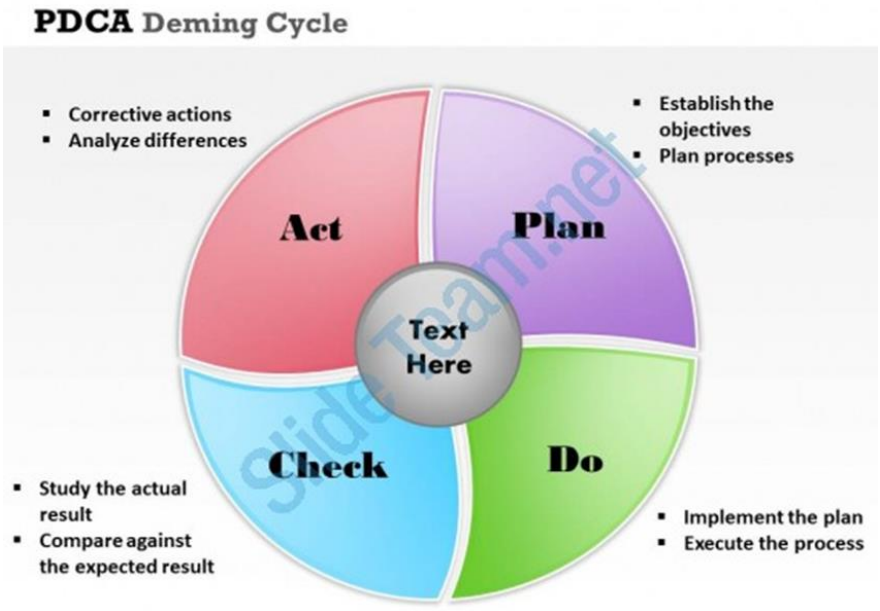


Figure 3.PDCA Deming Cycle

Plan: This is the process that set to establish the required policy, processes, objectives, and the necessary set of procedures needed to institute an information security management system. ISMS implementation has 5 planning stages. The 5 stages are: Scope definition, gap analysis, risk assessment, set up the required control and assets, and ensure the policy and procedure of ISMS are defined and agreed. The deliverables that will project the expected improvement in the implementation is covered in the planning stage (Candra, 2017).

Do: After the completion of the plan, the next stage is the implementation of the plan which takes place in the operation mode. The operation sets up the framework for routine and regular practicing of the policy, careful observation of the ISO 27001 controls, following the laid down processes and procedures of the management system and documenting deviations, gaps and potential non-conformances.

Check: Adopting ISMS for improving business continuity in organisations driven by CPS cannot be said to be perfect immediately after the implementation of the ISMS. A regular and consistent post implementation check is necessary to evaluate the effectiveness of the implementation and to discover areas that require modifications, Process performances measurement against the signed off and running policy, objectives and practical experience are what characterized this stage as well.

Act: For the purpose of continuous improvement, internal audit are necessary at some stages. Audit project is highly recommended at this stage. The result of the audit and the post audit management reviews are expected to trigger corrective and preventive actions. The level of maturity that the internal audit must have attained over a period of time will determine the readiness of the organization to subscribe to a full-fledged ISO 27001 certification audit provided the organization has financial capacity to embark on the project.

III. AUDITING ISMS

ISMS is an integral part of the overall corporate control process. In Figure 1, the monitoring and review element of the ISMS implementation process center mainly on the regular routine internal auditing. This section discusses the best in class approach to initiate, progress and complete a typical ISMS auditing endeavor same has been successfully adopted by deferent organisations globally. According to ISO standard number 19011, clause 3.1, an audit is defined as "sometimes called first party audits, are conducted by the organization itself, or on its behalf, for management review and other internal purposes (e.g. to confirm the effectiveness of the management system or to obtain information for the improvement of the management system). Internal audits can form the basis for an organization's self-declaration of conformity" (ISO, 2018). In the same ISO standard and clause referenced above, the parties involved in the overall activities of a typical audit are; the client, auditee, auditor, expert and audit team. A typical audit project spans over five stages. These are explained below;

Stage 1, using the risk management approach, ensure the context of the organization's security policy and objectives of ISMS is well stated in the audit project scope and understood (McCreanor, 2017).

Stage 2. There is a need to confirm that the organization is in alignment with his own policy and its strategic ambition of ensuring conformity to all standard's requirements and objectives. Thereafter, the audit team is expected to carry out an information security risk assessment and the resulting design of its ISMS (Henczel, 2015).

The following need be checked:

- Performance monitoring, measuring, reporting and reviewing against the objectives and targets
- The responsibility and the commitment of the management to the information security policy
- Evidence that monitoring of the informatics systems is regularly reviewed and a post-review follow up is carried out
- Confirm that the organization has a full knowledge of the importance of its IT environment

The following will help the audit team ascertain the preparedness of the organization

- The organization strategy reckons that IT is part of the corporate direction
- The formation of IT unit is structured in a way that appropriate segregation of duties exists and has got provision for continuity of key IT functions
- The IT unit has got a tested and proven policy and procedures for IT operations continuity in an event of unprecedented service or system failure.

Stage 3. Every single control of all the 114 controls in ISO 27001 standard need be reviewed against each and every relevant asset in the organization. The compliance status of each control will be adjudged to be any of the following;

- Fully conform
- Conform
- Minor non-conformity
- Major non-conformity

Stage 4. Articulate all the non-conformity and institute an action plan to treat all and ensure they ultimately turn to conformity (Calder, 2016)

State 5. Prepare and share post audit report and arrange for management review (Humphreys, 2016).

IV. CONCLUSION

Organisations driven by cyber physical systems rely on the deployment and administration of security systems such as firewalls, intrusion prevention and detection systems and other relevant security endeavours. Even with all these; there is a rapid expansion in the connectivity of cyber physical systems. The active threat actors are not decreasing, while the organizational priorities for cybersecurity have not been commensurate with system vulnerabilities and the waiting threats. These trends reveal the need for cyber security standard/framework implementations. Ultimately, the efforts to protect cyber physical systems assets reside in the hands of the business owner, asset owners and government, who can champion delivery of appropriate security measures; thus; ISMS implementations.

REFERENCES

Awolowo, O. (2018, August). *Nigeria's economy: services drive GDP, but oil still dominates exports.* Retrieved from Africa Check: https://africacheck.org/reports/nigerias-economy-services-drive-gdp-but-oil-still-dominates-exports/

Calder, A. (2016). An ISO 27001:2013 Implementation Overview. In *An ISO 27001:2013 Implementation Overview* (pp. 1-5). Ely, Cambridgeshire: IT Governance UK.

Candra, J. W. (2017). ISMS planning based on ISO/IEC 27001:2013 using analytical hierarchy process at gap analysis phase (Case study : XYZ institute). *2017 11th International Conference on Telecommunication Systems Services and Applications (TSSA).* Lombok, Indonesia: IEEE. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8272916

CORAL. (2012). *ISO_27001-2013__STANDARD.* Retrieved from Minimising Risks, Maximising ROI: https://www.coralesecure.com/pg/44/iso27001-2013standard.html

Cowan, J. (2015, March 12). *Energy sector tops list of US industries under cyber attack, says Homeland Security report.* Retrieved from IoTNOW: https://www.iot-now.com/2015/03/12/30962-energy-sector-stays-top-of-the-list-of-us-industries-under-cyber-attack-says-homeland-security-report/

Evans, D. (2011). *CISCO - The Internet of Things.* Retrieved from How the Next Evolution of the Internet is Changing Everything: https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

Fisher, D. (2014, May 21). ICS-CERT Confirms Public Utility Compromised Recently. United State. Retrieved from https://threatpost.com/ics-cert-confirms-public-utility-compromised-recently/106202/

Gerhard Funk. (2017). *A practical guideline for implementing an ISMS in accordance with the international standard ISO/IEC 27001:2013.* Retrieved from Implementation Guideline ISO/IEC 27001:2013. A publication of the ISACA Germany Chapter: https://www.isaca.de/sites/pf7360fd2c1.dev.team-wd.de/files/isaca_2017_implementation_guideline_isoiec27001_screen.pdf

Hammar, M. (2018). *ISO 9001 Academy*. Retrieved from Six Key Benefits of ISO 9001 Implementation: https://advisera.com/9001academy/knowledgebase/six-key-benefits-of-iso-9001-implementation/

Henczel, S. (2015). The Widening Horizons of Information Audit. *Qualitative and Quantitative Methods in Libraries.* At Senate House, University of London: Research Gate.

Humphreys, E. (2016). Implementing the ISO/IEC 27001:2013 ISMS Standard. In *Implementing the ISO/IEC 27001:2013 ISMS Standard* (pp. 1-224). Artech House.

ISACA. (2014). *Implementing the NIST.* Retrieved from Cybersecurity Framework: https://infosec.bryant.edu/docs/implementing_nist_framework.pdf

ISO. (2018). *Guidelines for auditing management systems.* Retrieved from ISO 19011:2018(en): https://www.iso.org/obp/ui/#iso:std:iso:19011:ed-3:v1:en

Johann Rost, R. L. (2011). The Dark Side of Software Engineering: Evil on Computing Projects. In W.-I. C. Press, *The Dark Side of Software Engineering: Evil on Computing Projects.* Wiley-IEEE Computer Society Press. Retrieved from https://www.wiley.com/WileyCDA/WileyTitle/productCd-0470597178,miniSiteCd-IEEE_CS2.html

Krebs, B. (2012, Oct 12). DHS Warns of 'Hacktivist' Threat Against Industrial Control Systems. United State. Retrieved from https://krebsonsecurity.com/2012/10/dhs-warns-of-hacktivist-threat-against-industrial-control-systems/

Luallen, M. (2014, April). *Breaches on the Rise in Control Systems.* Retrieved from A SANS Analyst Survey: https://ics.sans.org/media/sans-ics-security-survey-2014.pdf

McCreanor, N. (2017, November 27). *The five stages of a successful ISO 27001 audit.* Retrieved from IT Governance : https://www.itgovernance.eu/blog/en/the-five-stages-of-a-successful-iso-27001-audit

Ren, M.-M. (2015). The Application of PDCA Cycle Management in Project Management. *2015 International Conference on Computer Science and Applications (CSA).* Wuhan, China: IEEE. Retrieved from https://ieeexplore.ieee.org/document/7810878

Wang, W. (2018). Detection of data injection attack in industrial control system using long short term memory recurrent neural network. *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA).* Wuhan, China: IEEE. Retrieved from https://ieeexplore.ieee.org/abstract/document/8398169

Yadua, A. (2018). *Cybersecurity – Securing the digital organisation.* Retrieved from PWC Advisory Outlook: https://www.pwc.com/ng/en/assets/pdf/cyber-security-digital-org.pdf

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

Cybersecurity in the Age of FinTech and Digital Business

Faya Moses[1] and Nnubia Ogbuefi[2]

[1]Infusion Layers, Lagos, Nigeria
[2]O. J. Ogbuefi & Co., Enugu, Nigeria
[1]malanmoses@gmail.com, [2]nnogbuefi@gmail.com

Abstract: In 2016, the Cyber Security Experts Association of Nigeria (CSEAN) estimated that Nigeria loses N127,000,000,000(One Hundred and Twenty Seven Billion Naira) annually to cyberattacks. With the emergence of FinTech and Digital Businesses in Africa, the figure is bound to triple. Several FinTech and E-commerce startups overlook the importance of the position of a Chief Information Security Officer (CISO) in the early development stages. The concept of "waiting till there is a breach" can cripple businesses when a data breach finally occurs. Big companies strive to ensure that the private information of all its users are protected and these firms strive to be transparent on the information they access and how it is been used. The purpose of this Article is to throw into limelight the hurdles FinTech and Digital Business go through such as Regulations, Data protection, Cybersecurity and cyberattacks, and Intellectual Property thefts; and the best solutions to these hurdles.

***Keywords*:** Cybersecurity; FinTech; Digital Businesses; E-commerce; Data; Regulations; Cyberattacks.

I. INTRODUCTION

Financial Technology "FinTech"/Digital Businesses are a great ecosystem. It is already in the process of disrupting the old financial institution structure and also making people realize that money exchange, insurance, investment, and other financial service are going to be totally different in the next 10 years.

With FinTech, the change is wide-reaching; increasing operational efficiencies and productivity, providing alternative funding and, importantly, financial inclusion. Such developments include mobile payments, big data, natural voice recognition and response, use of crypto or virtual currencies, and artificial intelligence.

Investments in FinTech in Nigeria and other parts of Africa have moved from about $198 million in 2014, to $800 million currently. Though global investments in FinTech were put at $19 billion in 2015, indications have shown that investors are increasingly attracted to the industry's potential to tap Africa's huge un-served/underserved population (NIPC, 2016).

While this investment and transformation are taking place across the Finance industry, the momentum of growth is under attack wherein the security and safeguarding of assets and data are becoming increasingly important due to its proliferating threats which pose an ever-increasing danger to FinTech and the customers who use the services.

This essay, therefore, explains why FinTech companies must plan and build from the very beginning effective ways to address, data breaches, cybersecurity, and privacy issues.

II. THE EMERGENCE OF FINTECH AND DIGITAL BUSINESS INNOVATION

Today, FinTech is a cliché. It is the innovative approach by financial sectors in the world to deliver payment solutions and services to customers with the ease of a button. The mobile applications built by various startups and tech companies are instances.

What is FinTech? FinTech which stands for Financial Technology is a concept, a movement, a transformation, and a disruptive innovation that addresses the customers' issues and needs as it pertains to payment or financial services through automation of new and emerging technologies (Palermo, 2017) such as Artificial Intelligence, Machine Learning, Data Analytics, etc.

The concept of FinTech can be traced to the 1950s (Desai, 2015) when credit cards were introduced into society. The movement started shortly after the introduction of credit cards, to the installation of Automatic Teller Machines (ATMs) to replace cashiers and tellers in bank branches.

The transformation continued to the 1970s when electronic and thereafter, online stock trading was introduced in the Stock Exchange markets to improved record and data keeping of banking customers. The disruption came

towards the end of the 20th century and the beginning of the 21st century. The rise of the internet and E-commerce business drew digitized and faster financial services for users.

These services are categorized into robo-advisors for wealth, assets and retirement plans, crowdfund platforms, payment apps, mobile wallets, alternative lending platforms, and alternative investment opportunities.

These categories of financial services serve a singular purpose – addressing the needs of the consumers. This is rather different from traditional banking services, which is the institutional legacy.

Banks are programmed to offer services and enhancements that will improve the lending, savings, and investment opportunities of its customers through the traditional modality. But as Philippe Gelis, CEO of Kantox pointed out, "FinTech is changing the finance sector just like the internet changed the written press and the music industries.(Uzeb, 2018)"

Thus, from ALAT to Cowrywise, to Flutterwave, to Mpesa, to PayStack, to Remita, to VISA, to Zoona (Gulamhuseinwala & Bull, 2017), all these FinTech firms across the world share two basic characteristics, which are customer propositions' focus and novelty in the application of emerging technologies.

Just like the Baby boom, the FinTech boom is associated with the millennial (Villasenor, 2016). These are customers within the age of 22 – 34 years. These brackets grew up in the age of internet and tech-gadgets and as such will value simple, convenient, readily personalized and transparent services than the traditional system of long waits and queues.

The attraction to these services whether in wealth or financial management, retail banking or insurance is the ease in accessing them with mobile devices.

The birth of FinTech cannot be told without E-commerce. The e-commerce business models helped radicalize and revolutionize how consumers transact businesses.

Therefore, the era of rushing to the bank to make a cash deposit, while hoping that the transaction will be marked completed before the official closing time of a bank, has been relegated to the background with the advent of emerging technologies in the financial sector.

Now, mobile banking apps have helped customers simply make an order online and make payments without leaving a pinpoint location. Hence, digital businesses are online marketplaces where for product-focused or service-focused companies or firms strive to deliver personalized and seamless user experiences to consumers.

These experiences are either targeted through data and analytical research to ensure an increasingly recurring effect.

Therefore, the FinTech firms and digital businesses have created a ripple effect through the digitalization of traditional processes to draw consumers to its products and services through innovative, disruptive and emerging technologies.

The pertinent question to ask is if the emergence of FinTech would be porous to cyber attacks and if such will likely to affect the solid foundations of banking systems?

III. THE MISMATCH BETWEEN FINTECH AND REGULATION

The innovation with the FinTech world is happening at a light-speed, unlike their slow and laborious counterparts, thereby creating a wide gap between FinTech solutions and regulation. This gap is acute in FinTech and particularly so with respect to cybersecurity in the FinTech context (Oloyede, 2018).

Faced with this gap in Nigeria and with the rising nature of cybercrime around the world and Nigeria having its own fair share, leading to a huge financial loses, loss of trust with the potential risk of exploitation of design vulnerabilities, identity theft, malware attacks, and malicious use that post risk to security and safety of individual[8]; prompted the Central Bank of Nigeria (CBN) to release a set of guidelines sometimes in June 2018 to curb/fend off these attacks across Deposit Money Banks (DMBs) and Payment Service Providers (PSPs) as a requirement to increase cybersecurity (CBN, 2018).

The proposed guidelines by the CBN are set to come into force on January 1st, 2019. The increased sophistication of attacks has resulted in the huge financial losses, loss of trust with the potential risk of exploitation of design vulnerabilities, identity theft, malware attack, and malicious use that post risks to security and safety of individuals.

Good governance and regulation is profitable for FinTech as they increasingly rely on automated and electronic systems thereby posing as a risky venture for differently sophisticated cyber-attacks, regulators will have to calibrate its regulations and policies within the FinTech industry to ensure data protection and privacy, consumer protection and cybersecurity and ensuring FinTech firms be bound to maintain a level of security without stifling innovation

What Regulators should not do? Regulators should not rush into implementing hasty regulations that could end up stifling innovation; this does not suggest that Regulators should do nothing at all.

Regulators should be proactive; there should be a constant collaboration FinTech firms and Regulators, these will immensely assist regulators in understanding the technology surrounding FinTech, thereby giving the Regulators

clearer perspectives that can help ensure that any new regulations will not have collateral damage to the FinTech innovation ecosystem.

In addition, this engagement can benefit FinTech firms to have a better understanding of pertinent cybersecurity issues such as data protection and privacy.

IV. PROBLEMS OF FINTECH AND DIGITAL BUSINESS

There are several problems plaguing FinTech and Digital Businesses across the globe but in the course of this essay, only four (4) pertinent problems shall be addressed.

A. Regulations

A huge barrier to FinTech firms and Digital businesses globally are Regulations. Regulation Laws are constantly a hindrance to the growth and scaling up of these disruptions.

For instance, a firm may decide to expand its financial services beyond the boundaries of its country and the shores of its continent, only for the progress to be hindered by varying regulations laws.

In the USA alone, there are ten (10) Regulations laws (Chritiansen, *et al*, 2018) while in Nigeria, the Central bank of Nigeria (CBN) guidelines for regulation and licensing of financial firms are daunting. Also, a part of the Regulations barrier is legal fees for application and filing of the licenses. To ensure that startups gearing towards financial ease are not suffocated at the early stages of business, these Regulations ought to be relaxed.

B. Data

The fuss about data protection and privacy came into limelight with the data breach of millions of Facebook accounts. The data breach was linked between Facebook and Cambridge Analytica. It drew attention to what happens when private data are mined and sold on the dark web.

By May 2018, the European Union (EU) pushed for the General Data Protection Regulation (GDPR) which offers guidelines on how personal information of EU citizens is to be handled. It mandated all companies in the EU, including companies outside the EU but with clients/customers/servers located in the EU to comply with its laws or face penalties.

Thus, companies were mandated to be transparent on how personal information of its users are accessed and used either for analytical, marketing, advertising, or research purposes. Thereafter, the new European Payment Services Directive (PSD2) which mandated Payment Service Providers to be transparent to its customers on how their information is used or shared.

C. Cyber Security and Cyber Attacks

FBI Director, Robert Mueller in the RSA's Cyber Security Conference in 2012 stated that "I am convinced that there are only two types of companies: those that have been hacked and those that will be."

It is increasingly difficult for FinTech firms in the early stages of development to consider the risks of cyber-attacks and the issue of cybersecurity. Cyber Security Analysts are canvassing for a full-time position of a Chief Information Security Officer (CISO) in every FinTech startups or firms, and digital businesses, but mostly due to limited funds, companies rarely heed the call until they are hacked.

Firms ought to establish some level of protection especially when it pertains to personal information including credit cards information of its users. This is one of the crucial problems plaguing FinTech and Digital Businesses, especially in Africa.

D. Intellectual Property Theft

With the advent of innovative and emerging technologies comes increased theft of intellectual property rights. There are several cases pertaining to Patent trolls, theft of copyrights, and trade secrets. It is strongly canvassed that individuals and companies should ensure that their ideas are always protected.

It is no surprise that there is a strong clamor for Non-Disclosure Agreements and Transfer/Assignment of IPR Agreements where necessary. This is to restrict the pains and expenses associated with long legal battle or litigation.

These problems are not only restricted to FinTech firms but are also seen in E-Commerce businesses across the globe. Addressing these problems will be of tremendous impact on the growth of innovative disruptions in Africa and the world.

V. MEASURES IN SAFEGUARDING FINTECH AND DIGITAL BUSINESS

As the FinTech industry is evolving heavily, it is potentially vulnerable to attacks by malicious entities. FinTech companies must, therefore, endeavor to build and plan, from the very beginning effective ways to address cybersecurity, data security, and privacy protection right from the onset.

This is absolutely essential because these particular threats are proliferating and thus pose an increasing danger to FinTech companies and the customers who use their services.

So what is the best measure acceptable for the adaptation and protection of FinTech companies and Digital businesses for its customers?

Cybersecurity should be a top policy priority in the FinTech industry. Cyber-attacks have potential systemic financial stability risks and can discourage adoption of FinTech. Thus, there is an urgent need to adopt proactive measures that should be extended throughout the products and services lifecycles.

This will create room for the anticipation of wrong moves and porous measures; thereafter, put in place a robust, and effective measures to prevent, and mitigate itself from having serious problems in the areas of privacy protections, cybersecurity, denial of service attacks, insider threat, malware injection, insecure APIs, shared vulnerabilities and data security.

A. Proactive Measures

The proactive measure that should be adopted are:

- The development of a comprehensive cybersecurity framework that includes prevention, detection, monitoring, information sharing, financial and technology literacy, and recovery plans.
- The adoption of solemn responsibilities by FinTech companies and digital businesses to protect their overall architecture. Transactions usually take place across the interconnected global data communication enterprise which increases the overall vulnerability. New technologies should integrate security measures into their design.
- Regulatory oversight should ensure that the FinTech industry has cybersecurity implemented throughout its payments chain.
- Investments in technologies that prevent cybersecurity should be accompanied by training programs to increase awareness amongst staff, to prevent weak links which cybercriminals can exploit.
- Enrich financial security literacy through multiple platforms to reach consumers, investors, and small business owners who need it most. The FinTech industry is also expected to educate its customers on the "whys" and "whats."

For instance:

- Why they keep their bank account details and electronic devices?
- The Security and safety of FinTech applications.
- Why their data and personal information is at risk.
- What customers should look out for in terms of suspicious situations?
- What is okay to share and not okay to share?

Though educating customers on why it is necessary and how it helps mitigate risks will not only help protect the FinTech industry customers, but it will also strengthen their trust in the FinTech brand and its ability to protect customer's information

B. Clear laws and regulations to FinTech transactions and licensing requirements

The FinTech industry needs a dynamic regulatory architecture that can address risks as they emerge in the fast-changing landscape which can be relied upon. This will help identify gaps and restrictions in the law and regulations that hamper innovation and can enable development of a comprehensive and systematic roadmap for reforms.

Regulatory sandboxes can facilitate a better understanding of the risks posed by FinTech firms and enable the appropriate design of regulations.

VI. EMBRACING FINTECH

As previously stated, FinTech is a novel idea with a ripple effect on the banking and financial landscape across the globe together with immeasurable opportunities for emerging economies like Nigeria.

The CBN as a Regulator will have to find the right balance between protecting customers and creating non-stifling regulations. By doing so the CBN has undertaken a range of measures to promote financial inclusion, and one of them is supporting the FinTech ecosystem by the unveiling of a smart initiative to create a regulatory sandbox programme to support budding FinTech companies.

The objective of having a regulatory sandbox is to empower small companies, which we refer to as start-ups, innovators, technology companies and young Nigerians that have great ideas but lack the financial wherewithal to bring out their products or even integrate the ideas with the banking sector.

This is in recognition of the roles that FinTech play in the financial industry and the need for regulatory support, to assist them to play these roles within the market fit and security benchmark.

Regulatory sandboxes provide an environment of reduced regulatory constraints on innovative financial products and services.

They enable financial services innovators - both incumbents and startups - to test new products and services in a "safe area", providing greater flexibility or even exemptions from existing regulation.

Sandboxes can be highly valuable to financial services institutions in three important ways:

- They reduce the time and cost of getting innovation to market. They provide innovators with greater access to finance by reducing risks of client adoption and increasing returns on capital investment.
- They enable innovators to work with regulators to ensure new development of technologies and a business model aligns with regulations. The increasing reliance on automated and electronic systems FinTechs represents a risky venture because it requires them to be secure from cyber-attacks.
- Financial information is a high-value target for many cybercriminals, and it is imperative that both startups and established companies be bound to maintain a minimum level of security.

Fintech firms have increasingly attractive targets and typically have fewer resources dedicated to cybersecurity, as they prioritize growth and product-market fit. Regulators have to calibrate their policies and regulations to ensure an adequate level of cybersecurity and data privacy while encouraging innovation.

VII. CONCLUSION

There may be ongoing debates on the unexplored areas of FinTech, Digital Businesses, and Cybersecurity for a decade to come. But, the foreseeable option for emerging digitized businesses disrupting the financial sector should be security.

The importance of security cannot be overemphasized. When big companies globally get hacked, the smaller ones with porous firewalls becomes a testing ground. Thus, while there are innovative disruptions that address the customers' needs, there should be cybersecurity measures to protect customers' personal information.

- Why wait for the panic mode of a security breach when you can take preventive measures to erode and avoid them?
- Why make unnecessary expenses later on recovering hacked data when you can spend less protecting them?

Security, Protection, Privacy, and Transparency should be the bedrock of every FinTech firm and Digital business in Africa.

REFERENCES

Nigeria, Other Investments in FinTech Hit $800m (2016). Retrieved from https://www.nipc.gov.ng/nigeria-investments-fintech-hit-800m/

Palermo, F. (2017, December 18). The Emergence of Digital 2.0: What To Get Ready For In 2018. Retrieved from https://www.forbes.com/sites/forbestechcouncil/2017/12/18/the-emergence-of-digital-2-0-what-to-get-ready-for-in-2018/#25a82fcc1c94

Desai, F. (2015, December 13). The Evolution of FinTech. Retrieved from https://www.forbes.com/sites/falgunidesai/2015/12/13/the-evolution-of-fintech/#418735067175

Uzebu, C. (2018, January 30). The Rise and Rise of FinTech in Africa. Retrieved from https://www.cp-africa.com/2018/01/30/rise-rise-fintech-in-africa/

Gulamhuseinwala, I., Bull, T. (2017, July). The Rapid emergence of FinTech. Retrieved from https://www.ey.com/gl/en/industries/financial-services/fso-insights-the-rapid-emergence-of-fintech

Villasenor, J. (2016, August 25). Ensuring Cybersecurity in FinTech: Key Trends and Solutions. Retrieved from https://www.forbes.com/sites/johnvillasenor/2016/08/25/ensuring-cybersecurity-in-fintech-key-trends-and-solutions/#4c12337a35fd

Oloyede, R. (2018, December 3). Nigeria: Central Bank of Nigeria Issues Draft Cybersecurity Guideline For Deposit Money Money Banks (DMB) And Payment Service Provider (PSP) . Retrieved from http://www.mondaq.com/Nigeria/x/758332/Fiscal+Monetary+Policy/Central+Bank+Of+Nigeria+Issues+Draft+Cybersecurity+Guideline+For+Deposit+Money+Banks+DMB+And+Payment+Service+Provider+PSP

Exposure Draft of the Risk-Based Cyber-Security Framework and Guidelines for Deposit Money Banks and Payment service Providers (2018). Retrieved from https://www.cbn.gov.ng/Out/2018/BSD/RISK%20BASED%20CYBERSECURITY%20FRAMEWORK%20Exposure%20Draft%20June.pdf

Chritiansen, B. Maalouf, K., Brandt, P., Seve, M., Piquet, F., Sandman, J. & Seidner, G. (2018, February 16). A Look At US and EU Fintech Regulatory Frameworks. Retrieved from https://www.skadden.com/-/media/files/publications/2018/02/a_look_at_us_and_eu_fintech_regulatory_frameworks.pdf

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

Security Risk Analysis and Management in a Non-Interest Banking Sector: A Case Study of Jaiz Bank

Joseph A. Ojeniyi[1], Samuel A. Oyeniyi[2], and Shafi'i M. Abdulhamid[3]

Federal University of Technology, Minna, Nigeria

[1]ojeniyija@futminna.edu.ng, [2]sammykul01@gmail.com, and [3]shafii.abdulhamid@futminna.edu.ng

Abstract: The purpose of this article is to analyse security risk and management in all banking transaction, there by identifying the security risk in a non-interest bank sector. The method used are both primary and secondary techniques of data collection, which are sampling interviews questionnaires and content analysis. And the use of soda to do the analysis. In conclusion the research on risk analysis and facility management has proved that non interest bank manage their rusks better than other banks

***Keywords*:** Risk analysis; Risk management; Security; Non-interest banking.

I. INTRODUCTION

According to the central bank of Nigeria (CBN, n.d.), financial inclusion has assumed increasing recognition across the globe among policy makers. Nigeria's 46.7% population is still excluded from formal financial services. One of the reasons of exclusion stems from Islamic belief of half of the Muslim population who want to do away with Riba (Usury) in their daily activities. An alternative to conventional market is emerging in form of Islamic Banking and Non-Interest Banking. Risk management in banking contains a combination of processes and models, results of scientific research, that banks base on them to implement risk-based policies and practices. Banks are not any more practice traditional financial intermediation in low risk environment. A broad range of innovative and evolutional financial products, available globally at current time, have taken place and turned banking into a dynamic and active risk management process of assets and liabilities in a low regulated, high-risk environment. The basic reasons that made the risk-based practices to develop quickly are: banks have major incentives to move rapidly in that direction, regulations developed guidelines for risk measurement and for defining risk-based capital (equity) and the risk management toolbox of models enriched considerably, for all types of risks, providing tools making risk measures instrumental and their integration into bank processes feasible. This case study is about Jaiz Bank Plc; Nigeria's first full-fledged non-interest bank. As part of government financial inclusion policy, it has founded to tap the existing banking opportunity. Jaiz Bank is being positioned to be a national bank offering its services regardless of religious beliefs. The bank is an unquoted public company owned by over 26,000 shareholders spread over the six geographical zones of Nigeria the bank's balance sheet has grown from N12 billion in 2012 to about N62 billion, with asset financing of over N30 billion housing its head quarter in Federal Capital Abuja.

Jaiz Bank commenced operations with three branches in Abuja, Kaduna and Kano states in 2012, after it received license from the Central Bank of Nigeria (CBN) on the 11 November 2011 to operate as a non-interest bank. The bank plans to operate in all 36 states by upgrading and availing national operating license and increasing the share capital base to N15 billion (USD $78 million). Currently the bank operates in six states with eighteen branches. The Paid-up Capital as on 31st December 2014 was ₦ 11,829,700,000. The Central Bank of Nigeria has granted a national license and a waiver on the reduction of its liquidity ratio from 30% to 10%. The bank has massive expansion plan and intends to open 100 branches in first five years. (Jaiz bank website)

This research project aim is to analyses the risks management procedures of non-interest banking by giving analysis on risk management, and facility management. There are basically 3 types of Banks, commercial, development and central Bank, within the commercial Banking sector, there are some Banks operating on different principles, such as the JAIZ BANK which operates on the Principles of non-interest banking which is the case study of this research.(Jaiz bank website)

This research is aimed at ascertaining the security risk analysis and management strategies that can be adopted by non-interest banks in banking transaction to enable them stay afloat and remain profitable.

Different Banks adopt different strategies depending on their orientation, objectives and organizing principles, the main aim is to analyze and evaluate the security risk analysis and management of jaiz Bank plc. The main problems are analysis of risks and management of facilities by the non-interest Banks, how the Bank avoids the risks and manages facilities.

II. RELATED WORKS

The banking industry sector is characterized by intensive competition considering both the cost and the products. For this reason, the banks are forced to identify and adopt new and more efficient ways to fight their competitors and to gain more customers that will be retained and loyal. In this way, banks make efforts to reduce costs and make better offers by screening borrowers and differentiating the prices accordingly so as to maximize the profits and minimize the losses-risks. Generally, in the past the banks were product oriented which means that were interested in selling as more products as possible without paying much attention to their customers' ability to pay back. In this way, they were interested to take as much market share as possible so as to exploit the relative cost advantages and to continue to grow. Nevertheless, in the recent high speed and evolutional times, the expansion of business, deregulation and globalization of financial activities generated new financial products and increased level of competition even more. That made necessary an effective and structured risk management process in financial institutions.

Using bank accounting data for 22 countries in Asia over the period 1995–2009, (Lee, Yang, & Chang, 2014) applies the dynamic panel generalized method of moments technique to investigate the impacts of non- interest income on profitability and risk for 967 individual banks. We find that non-interest activities of Asian banks reduce risk, but do not increase profitability on a broad sample basis. Specifically, when considering bank specialization and a country's income level, the results become complicated. Non-interest activities decrease profitability as well as increases risk for savings banks. The impact is also different for commercial, cooperative, and investment banks either by increasing profitability or reducing risk. On the other hand, non-interest activities raise risk for banks in high income countries, while increasing profitability or reducing risk for banks in middle- and low-income countries.

(Lim, Woods, Humphrey, & Seow, 2016) used empirical evidence to examine the operational dynamics and paradoxical nature of risk management systems in the banking sector. It demonstrates how a core paradox of market versus regulatory demands and an accompanying variety of performance, learning and belonging paradoxes underlie evident tensions in the interaction between front and back office staff in banks. Organizational responses to such paradoxes are found to range from passive to proactive, reflecting differing organizational, departmental and individual risk culture(s), and performance management systems. Security is multidimensional in nature and diverse in practice. This diversity leads to difficulty in providing a single all-encompassing definition for the many applied domains of security. Security cannot be considered singular in concept definition, as definition is dependent on applied context.(Brooks, 2015)

(Kim & Yasuda, 2017) took advantage of the introduction phase of business risk disclosure in Japan as a natural experiment to examine the causal effects on firm risk. In contrast to risk factor disclosure that appeared partly in the Management Discussion and Analysis section (MD&A) in the United States, Japanese business risk disclosure is a new, independent disclosure regime, which began in the fiscal year ending March 2004. Also, they found out that the introduction of mandatory business risk disclosure has a negative impact on total risk. This suggests that an increase in business risk disclosure contributes to reduce a firm's cost of capital, which is contrary to the results found in previous research.

Risk shall be seen as the probable loss of income and assets' value. Only unexpected losses are included and expected losses are not included in the definition of risks (Khan 2004).

III. METHODOLOGY

A. Research Design

This research will adopt a semi structured survey design. The survey research method will have two identifying features. First, it is based on a sample of the population. Second, the data are collected by having each individual complete a questionnaire via GOOGLE FORMS. The researcher will obtain cross sectional data from the respondents by means of questionnaire.

B. Study Population

The researcher used questionnaire via GOOGLE FORMS to carry out the research work. The study is to find out and to analyse security risk and management in all banking transaction.

The result was gotten from GOOGLE FORMS which was administered to bank users form different location and part of Nigeria. The questionnaire was designed to analyse security risk and management in a non-interest banking transaction using jaiz bank as a case study.

C. Method of Data Collection

Data collected with the instrument (questionnaire) were edited to ensure accuracy. The quantitative data were analyzed using Statistical Package for the Social Sciences (SPSS) software. Statistical means such as percentages and frequency tables were used to reduce the raw data into manageable proportions.

D. Sampling Techniques

The type of sampling technique to be used for this research work is the Simple Random Sampling. Simple Random Sampling (SRS) is a subset of individuals chosen from a larger set of population. Each individual is chosen by chance.

IV. ANALYSIS

Jaiz Bank handles this risk properly; it covers every instance where the market risk is likely to occur, it also has a carefully outlined mitigation technique, where a down payment is made to avoid losses, risks are carefully analysed and mitigated in Islamic Banks to avoid losses, where as in the conventional Bank if the client decides to purchase the stated goods in this contract or decides to use the money for something else it is not the business of the Bank, the Bank's main concern is receiving its cash plus interest, it is absolutely clear that risks are assessed and mitigated better, or more priority in Islamic Banks because of the non-interest principle. This sharia noncompliance risk is a risk that is strange in the conventional Bank because they do not follow the Banking principles of sharia and are therefore not bothered about this kind of risks, while for Islamic Banks this is one of the major risks it tries to avoid, you can't have an Islamic Bank not complying to the rules and principles of the sharia supervisory board, the Bank avoids such risks by adhering to the principles of Banking as stated in the Quran and Hadiths. Banks face similar risk, be it Islamic or conventional and both Banks term their risks differently, a sharia compliance risk for Islamic.

Banks could be a separate risk for the conventional Bank, the Islamic Bank sees this risk as a sharia noncompliance risk, while the conventional Bank sees the risk as a market risk because, in conventional Banks mostly deal in terms of cash and therefore, buying a facility will be a market risk because the Bank is considering market prices. Conventional Banks are stricter in cases of default, the Islamic Bank tries to avoid default by proper scrutiny of the client. This balance sheet risk is the balances of the company's year end balances, the Islamic Banks will not have profit from interest in their year-end balance, as this goes against the number one principle of Islamic Banking i.e. prohibition of interest. Conventional Banks accept interests as it is one of the major sources of revenue for them, and profit from interest also reflect in their year-end balances.

V. RESULT

Like interest Banks, non-interest Banks are intermediaries and trustees of money of people. On contrary, it shares profit and loss with its depositors and introduces the element of mutuality in non-interest Banking. Interest banking follows Conventional interest-based principle, whereas, Non-interest Banking is based on interest free principle and principle of Profit-and-Loss (PLS) sharing in performing their businesses as intermediaries.

Interest banking (Conventional Banks) offer the facility of overdrawing from account of the customer on interest. One of its form isuse of credit card whereby limit of overdrawing for customer is set by the Bank. Credit card provides dual facility to customer including financing as well as facility of plastic money whereby customer can meet his requirement without carrying cash. As for facility of financing is concerned that is not offered by non-interest banking (Islamic Banks) except in the form of Murabaha (which means IFI shall deliver the desired commodity and not the cash) however facility to shop/meet requirement is provided through debit card whereby a customer can use his card if his account carries credit balance. Under conventional Banking a customer is charged with interest once the facility availed, however under Murabaha only profit is due when the commodity is delivered to the customer.

Furthermore in case of default customer is charged with further interest for the extra period under conventional system however extra charging is not allowed under Murabaha. Medium to long-term loans are provided for purchase or building of fixed assets by firms to expand or replace the existing assets.

Finally, clean borrowing is not allowed in Non-interest banking (Islamic Banks), Islamic Banks provide financing only to create assets, Islamic Banks do not offer credit cards, personal loans and overdrafts.

Below shows the gender distribution of the respondent.

Table 1. Sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MALE	24	52.2	52.2	52.2
	FEMALE	22	47.8	47.8	100.0
Total		46	100.0	100.0	

This table shows the number of individual that were given forms to fill in respect to their gender.
Below shows the age classification of the respondents, this show the level of literacy of the respondent.

Table 2. Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	17-24	9	19.6	19.6	19.6
	25-50	35	76.1	76.1	95.7
	51-above	2	4.3	4.3	100.0
Total		46	100.0	100.0	

This table shows the frequency of individual that were given forms to fill in respect to their age range

Table 3. How often do you make physical withdraw

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	L	8	17.4	17.4	17.4
	M	4	8.7	8.7	26.1
	H	15	32.6	32.6	58.7
	VH	19	41.3	41.3	100.0
Total		46	100.0	100.0	

Next question in the survey ask how often physical with-drawer is been made by an individual below shows the frequency table and also the percentage which was gotten.

Table 4. How often do you use internet banking?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	L	7	15.2	15.2	15.2
	M	2	4.3	4.3	19.6
	H	18	39.1	39.1	58.7
	VH	19	41.3	41.3	100.0
Total		46	100.0	100.0	

Here the survey ask how often internet banking is been used by an individual, above shows the frequency table and also the percentage which was gotten,

Table 5. How often do you change your internet banking password?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	L	6	13.0	13.0	13.0
	M	9	19.6	19.6	32.6
	H	21	45.7	45.7	78.3
	VH	10	21.7	21.7	100.0
Total		46	100.0	100.0	

Here the survey ask how often do individuals change their internet banking password, above shows the frequency table and also the percentage which was gotten,

Table 6. How often do you use other ATM for withdraws?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	L	8	17.4	17.4	17.4
	M	20	43.5	43.5	60.9
	H	18	39.1	39.1	100.0
Total		46	100.0	100.0	

Here the survey ask how often do individuals use other ATM for withdrawers, above shows the frequency table and also the percentage which was gotten,

Table 7. How often do you use mobile device for internet banking

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	L	7	15.2	15.2	15.2
	M	3	6.5	6.5	21.7
	H	17	37.0	37.0	58.7
	VH	19	41.3	41.3	100.0
Total		46	100.0	100.0	

Here the survey ask how often do individuals use mobile device for internet banking, above shows the frequency table and also the percentage which was gotten,

Table 8. How secure would you say is your bank?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	L	10	21.7	21.7	21.7
	M	8	17.4	17.4	39.1
	H	20	43.5	43.5	82.6
	VH	8	17.4	17.4	100.0
Total		46	100.0	100.0	

Here the survey ask how secure would they say their bank is, above shows the frequency table and also the percentage which was gotten,

Table 9. Would you recommend Jaiz Bank to a friend or colleague?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	L	24	52.2	52.2	52.2
	M	9	19.6	19.6	71.7
	H	7	15.2	15.2	87.0
	VH	6	13.0	13.0	100.0
Total		46	100.0	100.0	

Here the survey ask if individual could recommend Jaiz bank to a friend, above shows the frequency table and also the percentage which was gotten,

VI. CONCLUSION

This chapter is the summary of all work done from the introduction to the conclusion, this is a summary of the article as a whole, all work done from introduction to data analysis is been summarized.

This research on risk analysis and facility management has proved that non-interest bank manage their risks better than other Banks; it has also highlighted the basic risks faced by both non-interest banks and other banks, the difference in their modes of operation. The different types of risks where highlighted in chapter related works, the methods of data collection in methodology.

In conclusion to the study, non-interest Banks are better at risk mitigation than other Banks, non-interest Banks asses' risks properly and then a proper mitigation technique is assigned to that risk to avoid losses while most Banks also asses' risks their mitigation techniques are not as severe as that of non-interest Banks because they make profit from other sources of finance.

VII. RECOMMENDATION

This study recommends that more studies on risk analysis and management should be carried out both in the non-interest and conventional Banks and a comparative analysis should be done to assess the differences in facility management and how conventional Banks deal in asset financing.

Banks should alert customers on any suspicious and unusual transaction on their account

REFERENCES

Ali L., Ali. A. and Kwajah H. (2013). Comparison of Islamic and Conventional Banking on the Basis of Riba and Services, A case study of Peshawar Region.. . International review of management and business research journal, volume 2 issue 3 .

Brooks, D. J. (2015). What is security : Definition through knowledge categorization What is security : Defi nition through knowledge categorization, (July 2010). https://doi.org/10.1057/sj.2008.18

CBN (n.d.) www.cenbank.org/devfin/fininc.asp

Kim, H., & Yasuda, Y. (2017). Business risk disclosure and firm risk: Evidence from Japan. *Research in International Business and Finance*. https://doi.org/10.1016/j.ribaf.2017.07.172

Lee, C., Yang, S., & Chang, C. (2014). North American Journal of Non-interest income , profitability , and risk in banking industry : A cross-country analysis. *North American Journal of Economics and Finance*, *27*, 48–67. https://doi.org/10.1016/j.najef.2013.11.002

Lim, C. Y., Woods, M., Humphrey, C., & Seow, J. L. (2016). SC. *The British Accounting Review*. https://doi.org/10.1016/j.bar.2016.09.002

Wilson R. (Oct.'99 & Apr. 2000). challenges and opportunities for Islamic Banking and finance. Islamic Economic Studies,Vol. 7, Nos. 1 & 2, .

Khan T. (2004). Risk management in islamic conceptual framework.

Wohlfart L., B. L. (2010). Guideline: step by step development on facility management, a practitioner's guide.

Jaiz bank turning point, annual reports and accounts 2014

Zenith Bank plc, 2012 group annual reports and financial statement.

www.panafricancapitalplc.com Non-interest Islamic Banking in Nigeria..

www.google.com

www.wikipediaencyclopedia.com

www.investopedia.com

www.jaizBankplc.com

www.cenbank.org/devfin/fininc.asp

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference
April 9 – 11, 2019

IoT-Based Forensic System for Monitoring and Detecting Farmers-Herders Activities in Nigeria

Mohammed Ibrahim[1], Mohd Taufik Abdullah[2], and Muhammad Aminu Ahmad[3]

[1,3]Kaduna State University, Kaduna, Nigeria
[2]Universiti Putra Malaysia, Malaysia
[1]mibrahima47@gmail.com, [2]taufik@upm.edu.my, [3]muhdaminu@gmail.com

Abstract: Internet of things "IoT" application is expected to cut across various field in totality, its application in agriculture are either for monitoring plant or animals for the purpose of enhancing productivity and safety. However, in situation where both the safety of animals and the plant is required in a given agricultural zone, new IoT application is needed in monitoring that agriculture zone to avoid conflict between herders and farmers respectively. However, at the occurrence of incidents in a particular agricultural zone, law enforcement agencies requires intelligence, real and timely situation report for quick response to avoid loss of lives and properties. Hence, there is need for an IoT forensic system application that can detect the activities of the farmers and herders and send a real time report at the occurrence of incident. Therefore, this study proposed an IoT based forensic system that can monitor, detect and timely report illegal activities resulted from interwoven between plants and animals in a given agricultural zone. The proposed IoT based forensic system can be achieved in three segments: first segment is the installation of wireless sensor network (WSN) clusters that can aggregates field data, second segment is mounting of unmanned aerial vehicle(UAV) that can communicate the aggregated data to the cloud, the third segment is to disseminates the aggregated data to the law enforcement agents for prompt actions. Therefore, the IoT Based forensic System will assist law enforcement agent with real and complete picture of illegal activities taking place in a particular agricultural zone.

***Keywords*:** Animal; Crises; Farmers; Farmland; Internet of Things; Herders; Sensor.

I. INTRODUCTION

The revolutionary changes in internet and information technology enabled computers to capture and process large amounts of data, predict behaviors, activities and also to control the physical world (Alur et al., 2016). In this regard, the highlighted benefits of internet above can be achieved through the interconnection of embedded smart objects into our physical environment. The interconnection of embedded smart objects into the existing internet infrastructure resulted into a new era of Internet application known as "Internet of Things" or IoT. IoT promising concept enable access at anytime, anywhere, by anyone, anything, any service and any network. Similarly, communication in IoT enable Machine-to-Machine(M2M), Machine-to-People(M2P), People-to-People(P2P) and so on (Bagula, 2016). In short, IoT mean many different things, while IoT is cutting across various application domains, similarly fundamental research challenges will cut across those application domains(Alur et al., 2016).

(Miazi, Erasmus, Razzaque, Zennaro, & Bagula, 2016) categorized IoT application into industrial, environmental and societal applications domains. The environmental IoT applications domain focuses on activities related to safety, maintenance and development of all natural resources and the societal IoT applications directed towards the development and inclusion of societies, cities and people. Unlike in developed countries, developing countries need to monitor environment regularly to predict the climatic changes and to deal with the natural calamities(Miazi et al., 2016). However, lack of environmental monitoring in developing nation like Nigeria lead to climate-induced degradation of pasture, expansion of farms land and settlements have swallowed up grazing reserves and block traditional migration routes for herders(ICG, 2018). Consequently, movement of herders in searching of grazing area leads to the damage of farmer's crops by wrought herders' indiscriminate grazing. However, the confrontation as a result of the damages between the farmers and the herders resulted to the death of more than one thousand three hundreds Nigerians in the first half of 2018(ICG, 2018).

As part of its measure to stop the bloodshed, federal government deployed additional police and army units, and launched two military operations to curb the violence in six states. Nevertheless, even with these deployments killings still continues (ICG, 2018). In line with its recommendation 2017, Crises Group recommend for deployment of more police in affected areas, ensure that they are better equipped; improve local ties to gather better intelligence

and response to early warnings and distress calls. However, the key challenge is how to gather such intelligence information from farmers and herders activities and transmit in real time to the police for quick response to an incidents. Otherwise, damages can occur against a farm land without ascertaining the identity of the perpetrators that escape leaving innocents people into conflict that leads to loss of lives and properties.

II. The Need for IoT In Monitoring Farmers-herders Activities

IoT applications penetrate various areas including environmental and agricultural monitoring aspects. IoT is becoming a key feature of modern environmental management system(Miazi et al., 2016). It has many advantages for ecological research and monitoring of wild animal(Miazi et al., 2016). On the other hand, it is estimated that by 2050 up to two billions farms are likely to be connected to IoT (Nayyar and Puri, 2016). In line of the aforementioned, IoT can be effectively use in monitoring farmers and herders activities since Farmers activities relied solidly on the cultivation of plants and herders activities is all about grazing on plants and rearing animals.

III. Challenges In Monitoring Farmers herders Activities

Unlike in monitoring animals or plant specifically, in monitoring farmers-herders activities, both the activities of animals, plant, climatic condition and water availability most be put into consideration. In this regard, we categorized the challenge in to the following:

A. Farmland Monitoring Challenge

In tracking plant environmental activities via sensor we will be interested in the movement of the wind and sound of rainfall droplets which their pattern start from low to high and then to low frequencies that decay with time. Also we will consider the plants colour and the size, since the greener the plant the more attractive to the herders. Therefore, our sensor cluster need to capture various parameters from the farmland.

B. Animal Monitoring Challenges

In the face of the herders, cattle and sheep are the predominant animal species breeding by the herders, as such to sense their activities, we will be interested in their pattern of movement and sound. Flock of animals move in group and making random sound, therefore, sensing their activities will be based on the pattern of the amount of noise produce which is high at the beginning of events and deteriorate with time.

Hence, there is need for a mechanism that can distinguish between the plants and animals activities in order to detect the intrusions of animals into the farmlands. However, the current IoT monitoring applications are either animal or plant based and are mostly for enhancing productivity.

C. Climate Monitoring Challenges

Herders always move towards area of greener pasture and water availability. However, in Nigeria due to seasonal climatic conditional changes from raining to dry season, during dry season, herders move downwards to the southern part along the edges of river Benue and Niger in the north central part of Nigeria. In this regard, many farmlands situated at the edge of the two rivers are at the risk of being attacked as a result of intrusion from the wrought herders, therefore our IoT based forensic system must be capable of examining the seasonal change and make decision of where to monitor.

D. Incidents Monitoring Challenge

Farmers and herders are predominant in remote areas, they activities as such cannot be reach in real time since no technology is put in place to instantly monitor, detect and report incidents to the police. As such police and law enforcement agencies are tie with the locals to feed them with the situation report for any incident. However, due to human limitation, monitoring the entire farmlands cannot be feasible, as such damages can be done against farmlands by unknown herders. However, before the police intervention reprisal attacks can occur against the other party leading to dead of large number of people and loss of properties worth millions of Naira. Consequently, there is need for police and other relevant authorities to know what is happening in real time for quick response and intervention.

IV. Related Works

(Guo et al., 2015) reported that direct observation through field tracking and observing individuals and groups of animals still remain the main approach to study many species of animals. The paper furthered argued that the use of Radio Frequency Identification(RFID) required direct observation and monitoring which is time consuming and labour intensive. In this regard, (Guo et al., 2015) analyzed the advantages of the new IoT technology and

recommend that ecologist should make full use of the advantages of IoT technology in monitoring species of animals.

(Xu, Solmaz, Rahmatizadeh, Turgut, & Boloni, 2016) argued that in an unknown large area, it is difficult and sometimes not feasible to find wild animals and attach wearable tracking devices to them. Consequently, the paper divide the large observation area into small grids and treats each grids as cluster of sensor nodes. Data collected from the sensor nodes is then utilized by Unmanned Aerial Vehicle (UAV) and enable the UAV to explore and learn the small grids which can sense the animal position and activities.

The work of (Kirti & Sayali, 2018) combined the function of RFID with wireless camera sensors for tracking and identification of animals in digital zoo. This approach also addressed the problem of long distance identification of animals by embedding sensors into RFID tag which can be track using Global Position System (GPS).

The aforementioned techniques provide sound approaches in detecting and tracking animals, however, in detecting farmers-herders activities, we need a system that can detect and classify farmlands activities into legal or illegal and transmit those activities directly to the police for prompt action.

V. DEVELOPMENT OF IOT BASED FORENSIC SYSTEM

To develop an IoT based forensic system for detecting, monitoring and reporting farmers and herders activities, three main processes are required with the following objectives as follows:

A. Detection Process

The Objective of this phase is to gather and aggregate data from a farmland, in this regard, wireless sensor network (WSN) is required to place in strategic places within the farmland area. The sensor will detect the colour and the sound movement of the plant. Also, the sound of the wind movement as well as the rainfall droplets can be sense as part of the farmland legal activities. However, for the animals, the WSN will detect animals sound and their footsteps as illegal activities within farmland. The aggregated data collected from the WSN will then be forwarded to the unmanned aerial vehicle (UAV) for decision making.

B. Monitoring Process

Given the aggregated data collected from the WSN, the objective of this phase is to determine the priority area of the farmland that will be consider for monitoring. In this regard, UAV will play a significance role in deciding which part of the farmland will be given much consideration based on the aggregated data. The decision of the UAV can be achieve mathematically through modelling of plant and animal activities. Thereafter, K-clustering machine learning techniques will be apply to classify such activities as normal or abnormal. For abnormal activities, a direct report will be send to the cloud.

C. Reporting

Based on the aggregated data retrieve from the sensor and the decision made by the UAV, the information collected from the farmland will be transmitted directly to the police authority concern. In this regard, android based application will be develop that can interface the activities of the farmland with the police smart phones.

VI. CONCEPTUAL DESIGN ARCHITECTURE

Figure. 1 below shows the conceptual design of the proposed IoT based forensic system, the system is design on the assumption of the facts that the demarcated grazing reserves area and the cattle routes established by the federal government of Nigeria is put in place. The proposed system has three main components as shown below.

The first component which is agricultural zone comprises of grazing reserve area and farmland, where both herders and farming activities are legalize for breading of animals and crops. However, under this zone, cattle route demarcates the herders from farmers, in any circumstance by which farmer or herder activities penetrate outside their demarcated area, such activities is considered illegal. In this regard, sensor nodes will be deploy along the cattle route to detect such malicious activities. Data captured by the sensor will then be send to the next component for capturing and classification. The sensor data will be classify according to the pattern of the activities taking place in the farmland. Normal distribution is considered as the formula for modeling the pattern of the farmland activities, activities in the farmland are made up of rainfall, wind and plant movement which start gradually from lower to higher and to lower level with time. The farmland activities pattern is shown in Figure 2 below.

On the other hand, cattle activities are characterized by random movement with random sound which initialize with higher sound and decay with time. Figure. 3 below shows the pattern of cattle activities with time. This shows that the cattle penetrating into the farmlands will create a pattern of activities contrary to the normal activities detected by the wireless sensors network (WSN).

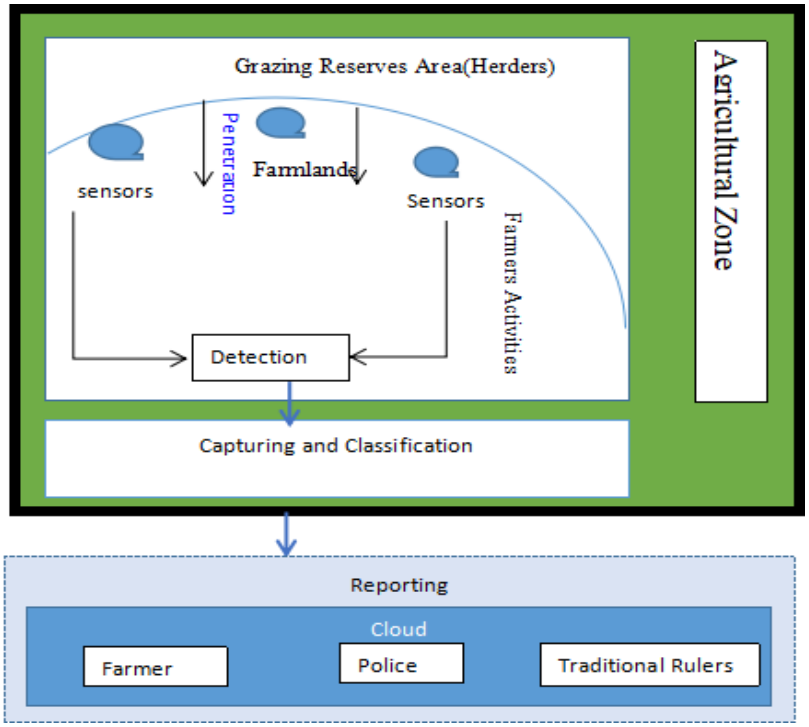


Figure 1. Conceptual design of the proposed IoT based forensic system for the investigation of farmers-headers activities

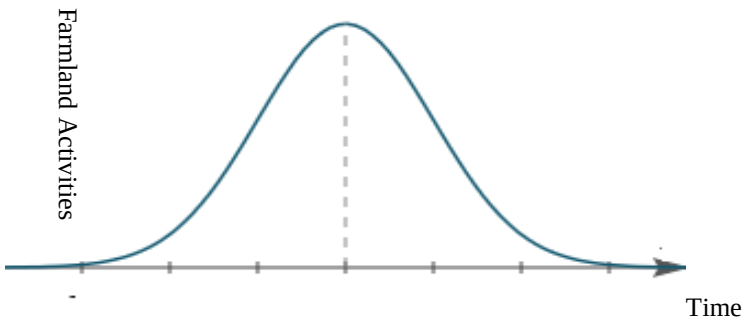


Figure 2: farmland activities pattern

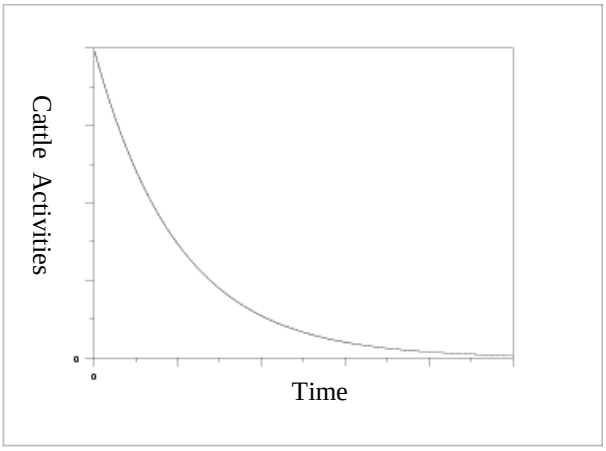


Figure 3: cattle activities pattern.

Therefore, based on the sense data from WSN the classification component categorizes the activities based on the pattern. If the sense data matches with the pattern in Figure. 3 above but found in farmland area, the classification components will forward a report to the cloud for immediate action. Finally, the reporting component will alert the farmer, police and traditional rulers via cloud service provider.

VII. IMPLEMENTATION TECHNIQUES

Our proposed system will be implemented using IoT technology, as illustrated in Figure. 4 below, the hardware requirement consist of WSN which can be deployed on the farmlands. WSNs consists of remote sensing nodes monitoring aspects of their environment (Jebril, Sali, Ismail, & Rasid, 2018). The data collected by the WSN will then be transmitted to UAV for further decision, UAV provides extremely flexible platform for WSNs. The emergence of UAV enable cost effective and appealing solutions for surveillance applications(Xu et al., 2016). In this regard, classification algorithm will be developed based on K-clustering machine learning algorithm to enable UAV make decisions of where to monitor and capture images for on-ward transmission to the long Range (LoRa) gateway. LoRa technology has the potential to be the next generation wireless communication standard with advantages of long communication range, low cost and reduced power consumption(Jebril et al., 2018). Once the LoRa gateway receive the images from the UAV, the images will be automatically transmitted to the cloud services providers which will instantly distributed to the police, farmers and traditional rulers for quick and incident response.

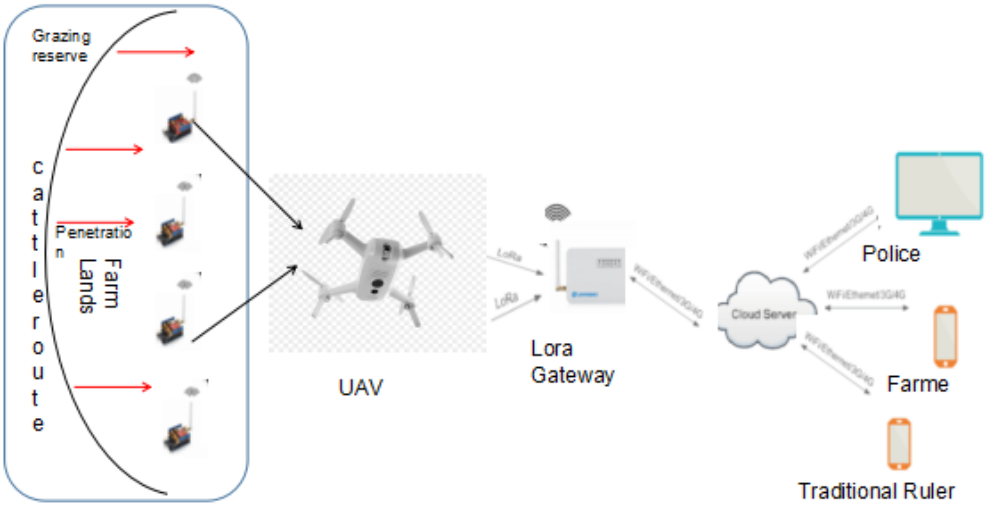


Figure 4: Hardware setup of the proposed IoT based forensic system for monitoring herders-farmer activities

VIII. BENEFIT AND USES

With the implementation of the proposed IoT based forensic system for detecting farmers-herders activities, farmers will be rest assured that their farms are under surveillance all the time. Also, the police will have firsthand real time information about the incidents taking place in the farmlands. The system will provide avenue for the police to quickly response to illegal farms intrusions and forestall any crises that may happen thereafter. Consequently, this will reduce the loss of lives and properties and restore peace and harmony in the affected areas, thereby reduces the government expenditure in providing security and restoration of loss of properties.

IX. CONCLUSION

Considering the impact of lack of technological approaches in monitoring environmental and agricultural zones in developing countries like Nigeria. This resulted to conflict between farmers and herders in striving for their survival in cultivation of plants and rearing of animals. In this regard, this study explored the advancement in internet technology known as Internet of Thing (IoT) to capture activities in real time from the farmlands and transmit to the authority concern for quick response to an incidents. Therefore, developing and implementation of the proposed IoT based forensic system for monitoring and detecting farmers and herders activities, will provide a platform for gathering intelligence and firsthand information with ease in preventing conflicts between the two giant farmers within a particular agricultural zone.

REFERENCES

Alur, R., Berger, E., Drobnis, A. W., Fix, L., Fu, K., Hager, G. D., . . . Patel, S. (2016). Systems computing challenges in the Internet of Things. *arXiv preprint arXiv:1604.02980*.

Bagula, B. A. (2016). Internet-of-Things and Big Data: Promises and Challenges for the Developing World. 19th UN CSTD session-Geneva.

Guo, S., Qiang, M., Luan, X., Xu, P., He, G., Yin, X., . . ., Chen, X. (2015). The application of the Internet of Things to animal ecology. *Integrative Zoology, 10*(6), 572-578.

International Crises Group (ICG): Stopping Nigeria's Spiralling Farmer-Herder Violence 2018, www.crisisgroup.org

Jebril, A., Sali, A., Ismail, A., & Rasid, M. (2018). Overcoming Limitations of LoRa Physical Layer in Image Transmission. Sensors, 18(10), 3257.

Kirti, W., & Sayali, P. (2018). Animal Tracking and Caring using RFID and IOT. *IOSR Journal of Computer Engineering (IOSR-JCE)*(2278-8727), PP 24-27.

Miazi, M. N. S., Erasmus, Z., Razzaque, M. A., Zennaro, M., & Bagula, A. (2016). *Enabling the Internet of Things in developing countries: Opportunities and challenges.* Paper presented at the Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on.

Nayyar, A., & Puri, V.(2016). Smart farming: IoT based smart sensors agriculture stick for live temperature and moisture monitoring using Arduino, cloud computing & solar technology. Paper presented at the international conference on communication and computing systems (ICCCS).

Xu, J., Solmaz, G., Rahmatizadeh, R., Turgut, D., & Boloni, L. (2016). Internet of things applications: animal monitoring with unmanned aerial vehicle. *arXiv preprint arXiv:1610.05287*.

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

Comparative Evaluation of Artificial Neural Network and Support Vector Machine for Money Laundering Detection

Nimatullah Yusuf[1] and John K. Alhassan[2]

[1]Department of Computer Science, Federal University of Technology Minna, Nigeria.
[2]Department of Computer Engineering, Federal University of Technology Minna, Nigeria.
[1]nimah93@gmail.com, [2]jkalhassan@futminna.edu.ng

Abstract: Money laundering ranks high amongst the factors that crumbles a nation's economy, In Nigeria, From recorded cases Politically exposed persons (PEP) and Yahoo boys are mostly the culprits of money laundering Preventive measures have been put in place to curtail the occurrence of money laundering in financial institutions by the Central Bank of Nigeria(CBN). Traditional approaches of money laundering detection have been proved to be time consuming and labour intensive. Hence Data mining approaches are used to automate money laundering detection (MLD). This paper aims at making a comparison of the effectiveness of Artificial neural network (ANN) and Support vector machine (SVM) in money laundering detection and to investigate suitable parameters to be used for the model. Based on two data mining frameworks; ANN and SVM., some sequence of processes are used in order to classify a transaction as either suspicious or not, the effect of number of neurons and layers on the model was determined, so was the effect of validation or no validation on the SVM. The model was evaluated based on various indicators such as; accuracy, sensitivity, specificity and precision to analyze the effectiveness in MLD of which both performed favorably well, But ANN surpassed SVM.

***Keywords*:** Data mining; Money laundering Detection; Artificial neural network; Support vector machine; Classification

I. INTRODUCTION

Across the world, financial transactions are carried out through different medium, over time, fraudulent activities by criminals have in one way or the other caused harm to financial institutions or their customers, or cause reputational damage, which is hard to repair and leads rapidly to loss of customers and market shares.

Money laundering is a procedure of concealing the origin of money or funds illegally obtained and accrued overtime. According to Collins dictionary, "Money laundering is the crime of processing stolen money through a legitimate business or sending it abroad to a foreign bank, to hide the fact that the money was illegally obtained." Money laundering has been observed to have great effect on the Nigeria economy, in defiance to the laws and polices enacted, financial and economic crime as money laundering still flourish in the country The Nigeria Deposit Insurance Corporation (NDIC) Decree No 22 of 1988 mandate of section 39 and 40 insured Nigerian banks to render returns on fraud, and other related malpractices occurring in their organization to the corporation, and to report personnel involved in fraudulent practices. It's been recorded that few banks comply. Amongst the objectives behind the Bank Verification Number (BVN) initiative is to reduce fraud (Vanguard News, 2015). Nigeria's parliament recently passed a bill meant to help authorities tackle money laundering and funding for terrorism by giving full control to financial intelligence unit (REUTERS, 2018).

Agu et al, 2016 stated some RED FLAGS pointers for money laundering which could be used by financial institutions to build (Money laundering detection) MLD framework. In order to determine accounts with or without potential characteristics of money laundering, transaction history of account holders has to be monitored (Ebere, 2016) Money laundering acting as a mechanism to aid terrorist financing and reducing government tax revenues is a worrisome issue for the Government (Suresh et al, 2016).Financial crimes in Nigeria have witnessed a number of publications, mostly affected, is the banking sector (Ebere, 2016). The lack of cooperation by banks to report money laundering culprits from their financial institutions to relevant bodies aggravate the situation (Anele, 2013b). Oluwadayisi and Mimiko (2016) analyzed the effects of money laundering; socio-economically, financially, politically, in manufacturing of domestic products and in the oil and gas sector, of which they concluded to have a great effect on the economy of Nigeria.

Hopwood et al, 2013 stated that Money laundering involves three phases

Placement phase; A money launderer introduces profits obtained illegally into the financial system ensuring its source is untraceable. Layering phase; the money launderer distributes the money around making tracing more complicated.

By passing it through a complex sequence of banking transfers or commercial transactions the launderer attempts to 'wash' the proceeds of his crime to evade detection or the suspicion of law enforcement agencies. Integration phase; the 'laundered' property is re-introduced into the legitimate economy by the launderer using legitimate means The effect of money laundering could crumble a nation's economy, hence proactive measures should be put in place to curb such offense and perpetrators be punished. MLD has attracted a great deal of concern and attention, different data mining techniques have emerged over time. A challenging factor in detection process is sorting of genuine transactions to detect the suspicious transactions in real-time. Financial institutions utilize statistical methods to detect and provide the information about the money laundering activities of the government sectors for disciplinary actions. Traditional approaches of money laundering detection have been proved to be time consuming and labour intensive.

Data mining approach to Money laundering detection automates the process of money laundering detection and helps in extracting hidden information from a large dataset which are then use in money laundering detection (Kannan and Somasundaram 2017). The efficiency of machine learning algorithm is highly influenced by the unique characteristics of financial money laundering data. Two most successful data mining techniques are the ANN and SVM.

II. LITERATURE REVIEW

Data mining techniques can be employed in finding transaction patterns that lead to money laundering by considering client risk assessment data, transaction risk measurement data and behavioral patterns in detecting money laundering patterns. Based on the similarities in the data, transactions are clustered (Manjunath 2015). Omar et al., 2018) stated that combining classifiers is a new outlook for improving the performance of classifiers. Camino et al., (2017) came up with a methodology for analyzing financial transactions using unsupervised learning methods, they applied anomaly detection algorithm to obtain suspicious ranking for user accounts. Kannan and Soumsundaram (2017) proposed auto-regression outlier algorithm for money laundering detection using Mean, Zero-mean, auto-regression a d Inter- quartile Range (IQ) operations on the extracted data for detection of money laundering activities. Soltani (2016) proposed a network-based algorithm to filter transactions, then uses clustering approach to detect suspicious money laundering patters within the network. The patterns are re-arranged, sorted and returned as output of the framework. Rohit and Patel, 2015 concluded that the best technique for Anti-money laundering detection is clustering techniques. Stefan Axelsson et al (2012), analyzed the implications of using machine learning techniques for money laundering detection in a data set consisting of synthetic financial transactions and aimed to detect anomalies inside a data set of mobile money financial transactions by using customer profiling to group transactions as suspicious or non-suspicious.

III. PROPOSED MONEY LAUNDERING DETECTION MODEL

The money laundering detection models proposed in this paper are based on two data mining frameworks; the Artificial Neural Network (ANN) and Support Vector Machine (SVM).

A. Artificial Neural Network (ANN)

Amongst the advantages of ANN is that its adaptive, secondly, ability to generate robust models, and third, modification of classification process with new training weights. ANN is predominantly applied to credit card, automobile insurance and corporate fraud (Sevda and Belafar. 2015).

Artificial Neural Network (ANN) has proven to be of strength for its speed of classification, tolerance to highly interdependent variables, accuracy and attempts for incremental learning. While its shortcomings include; learning rate due to number of attributes, tolerance to null values, noise and dealing with over fitting.

For less complex data, a layer can be used whereas for a more complex data, more than one layer is advised, as one layer might not give desired results. In this network, each element of the input vector **p** is connected to each neuron input through the weight matrix **W**. The *i*th neuron has a summer that gathers its weighted inputs and bias to form its own scalar output *n(i)*. The various *n(i)* taken together form an *S*-element net input vector **n**. Finally, the neuron layer outputs form a column vector **a**. the expression for **a** is shown in the figure below.

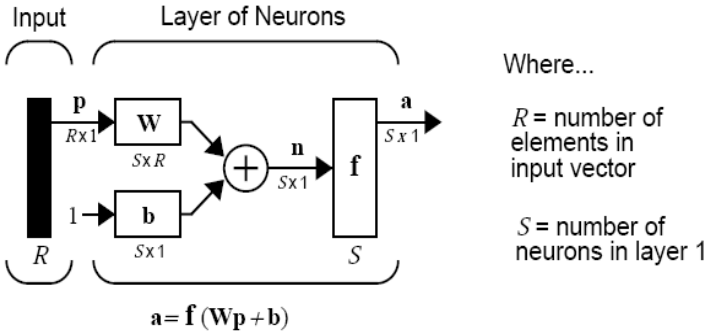


Figure 1. A Layer of neuron

Feed forward network was used, as it has been proven to be good for pattern recognition. We tried using multi-layer and single layer, Multi-layer network gave little or no effect on the performance, hence single-layer network was chosen, but with varying size of neurons.

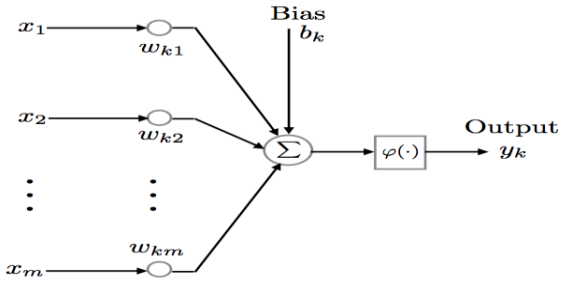


Figure II. Nonlinear model of a neuron k. (Simon, 2009)

$$\mu x = \sum_{j=1}^{m} wkjxj \quad (1)$$

$$yk = \varphi(uk + bk) \quad (2)$$

where wk1;wk2; ::::;wkm are the connecting weights of neuron k; uk is the summation of input-weights products; bk is the bias; $\varphi(.)$is the activation function; and yk is the output signal of the neuron k. The bias bk also can be regarded as a modible weight equal to wk0 with a fixed input of x0 = +1.

B. Support Vector Machine (SVM)

Support Vector Machine's strengths are for Accuracy in general includes; accuracy, speed of classification, tolerance to irrelevant and redundant attributes, whilst some of its weaknesses are learning rate due to amount of attributes and instances, model parameter handling, ability to explain, precision in classifications. (Omar et al 2018). To maintain the accuracy of detection rate and reduce numbers of false positive, AML system should be able to handle detection of unusual patterns from hidden transactions and forecast hidden cases with no effect on the performance accuracy and automate analysis of transactions. A typical AML typology method requires a set of training data (labeled) containing information of previously identified suspicious transaction and normal transactions (Zhiyuan Chen, 2018). According to Vapnik and Cortes, 1995 SVM is a learning approach of statistical evaluation which can either be used as solutions to classification and regression problems. Its main objective is to construct decision boundary with the largest distance to sample called maximum margin separator. The constructed hyper-plane that separates data is mathematically defined by the equation. SVM has been found to be successful when used for pattern classification problems. SVM performs well in classifying non-linearly separable groups; they do not require large training datasets. Out of a large dataset a subset of 1100 was extracted for training and testing, the results are discussed in the next session. Classifying the data correctly is a goal for calculating SVM. Kernels are used in non-linear SVM to move input data to a high dimensional space, which in turn makes the data linearly separable. Generally SVM problems have good generalization performance in pattern classification. Gaussian Radial Basis kernel of SVM algorithm was utilized, The width σ^2 in (1) is specified a priori.

$$K(x, x') = \exp\left(-\frac{||x-x'||^2}{2\sigma^2}\right) \quad (1)$$

From the classification learner application, Fine Gaussian SVM was used, since it had a better performance than the other types of SVM. Fine Gaussian SVM has fast prediction speed for binary classes, uses medium memory, hard to interpret yet highly flexible which decreases with kernel scale setting. Makes fine detailed distinction between classes.

These machine learning models used some sequence of processes in order to classify a transaction as either suspicious or not. The processes include data collection and processing, model design and parameters investigation, model training and testing and finally, performance evaluation of the models as shown in Figure III.

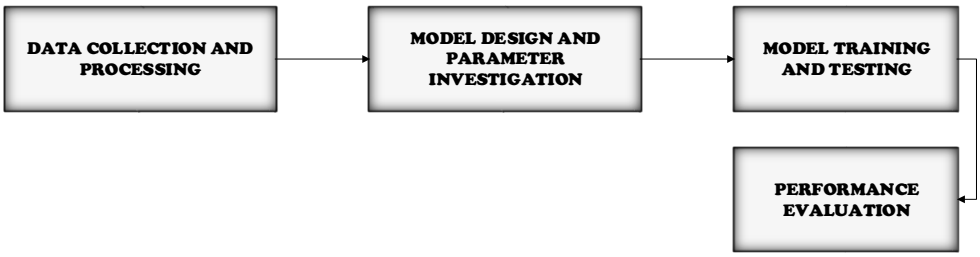


Figure III: Block diagram of the proposed money laundering detection framework

C. Data Collection and Processing

The data used for this study was gotten from an online repository (github). Existing literature used these parameters in detecting money laundering accounts; Amount received, amount withdrawn, debit/credit transaction frequency, risk value, individuals salary information, sender/receiver individual account history. The dataset was preprocessed first by extracting suspicious and non-suspicious transactions using the following rules (William-McKee, money_laundering challenge).

- First transaction receiver Id matches the second transaction sender Id
- Second transaction amount is between 90% and 100% of first transaction amount
- The two transactions occurred on the same day.

A total of 500 suspicious transactions were filtered from the dataset and to balance the ration of both class, 500 samples of normal transactions were added to form 1000 training and testing dataset. The target label for suspicious transactions is set to 1 while 0 for non suspicious (normal) transactions. The timestamp feature was converted to numeric data type after which normalization is done to prevent bias of the clasiffier around features with high values.

D. Model Design and Parameter Investigation

The model was trained in MATLAB iteratively with a set of suspicious transactions, after learning; it is used to classify a set of transactions into classes of suspicious and non suspicious transaction. In investigating parameters, from the data analyzed, most of the suspicious transactions occurred at the end of the month, some mid of the month. Timing might b a criteria for offenders, they could target month ending to disburse money (time for salary payment) or schematic ones could target early days or some other time of the month.

E. Model Training and Testing

The features were reduced to determine the effect on the performance of the model from the four features used. Time was exempted and trained which returned a result of 87.5% accuracy; hence three (3) features (Sender id, Receiver id and Amount) gave a percentage (%) difference from four features. Since the formerly listed attributes are important features in money laundering detection, they were not tampered with. In changing the size of the neurons or layer, over-fitting or under-fitting might occur, it is advised to interchange, and then repeat the process after a while, and so the system can learn well.

F. Performance Evaluation

The model was evaluated based on various indicators such as; accuracy, sensitivity, specificity and precision to analyze the effectiveness in money laundering detection (MLD). In evaluating a classification model's performance,

the number of correctly and incorrectly predicted records by the model is a determinant.. The counts are tabulated in a table known as a confusion matrix. (1) – (3) shows the mathematical representation of the performance evaluation metrics.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Sensitivity = \frac{Tp}{Tp+FN} \quad (2)$$

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

Where TP is correctly classified suspicious transactions, TN is correctly classified normal transaction, FP is wrongly classified normal transaction and FN is wrongly classified suspicious transactions. Cross validation is a technique for assessing the performance of an algorithm, it makes prediction using data not used at the training stage by partitioning the dataset and uses a subset to train while the rest are used for testing the algorithm. It is often a method of preventing overfitting during training.

IV. EXPERIMENTATION AND RESULTS

C. Parameter Investigation Result

Number of neurons were repeatedly interchanged to determine the effect on the model, low number of neurons returned average results, increasing to fifty (50) neurons have an impressive performance, but above 50 neurons, it was noticed that performance accuracy kept reducing.

Table I. ANN Performance

Algorithm	No of neurons	Accuracy
Artificial Neural Network	30	64.9%
	40	68.4%
	50	83.1%

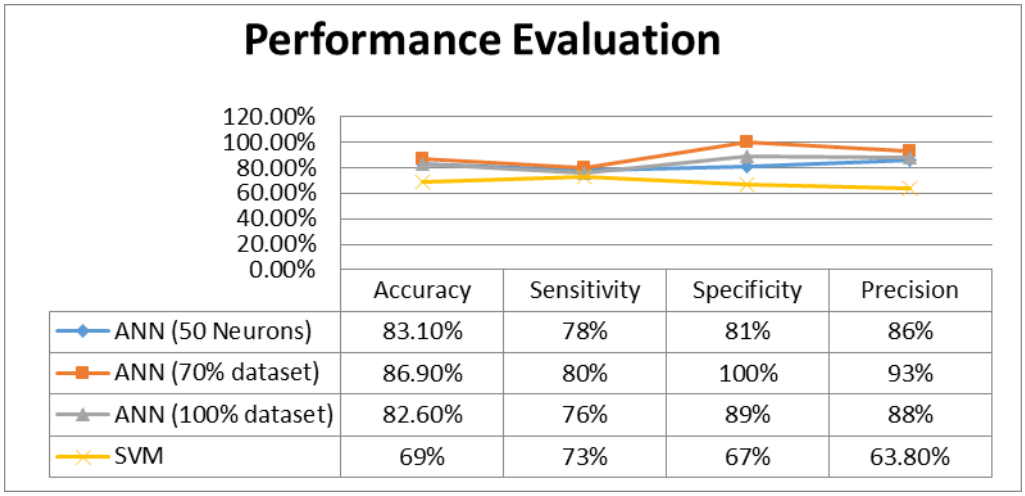
Table II. SVM Performance

Algorithm	Cross-validation	Accuracy
Fine Gaussian SVM	5-folds	69.0%
	10- folds	65.6%
	No validation	68.5%
	15 folds	67.2%
	20 folds	66.5%
	25 folds	66.6%
	30 folds	66.8%
	Hold-out validation with 30% held out	68.5%

Cross validation of 5-folds had the highest accuracy; hence, it performed better than others did, though the performance accuracy is not sequential. The cross-validation procedure can prevent the overfitting problem.

V. CONCLUSION

In conclusion, Money laundering detection could be challenging. An approach for analyzing financial transaction records and detecting laundered accounts from the records. The effect of number of neurons and layers on the model was determined, so was the effect of validation or no validation on the SVM. In this research, training the model with Bayesian regularization (BR) algorithms gave better result unlike Levenberg-Marquardt (LM) algorithm, and scaled conjugate. When the results of both SVM and ANN were compared, ANN outperformed SVM. In future, Dimensionality reduction is intended to be applied to the results to determine its effect. Since SVM performance is fairly accurate in MLD, we should combine both approaches; hopefully there will be improved performance.



	Accuracy	Sensitivity	Specificity	Precision
ANN (50 Neurons)	83.10%	78%	81%	86%
ANN (70% dataset)	86.90%	80%	100%	93%
ANN (100% dataset)	82.60%	76%	89%	88%
SVM	69%	73%	67%	63.80%

Figure IV: Performance Evaluation

REFERENCES

Agu, B. O., Enugu, U. N., & Onwuka, O. (2016). Combating money laundering and terrorist financing—The Nigerian experience. *International Journal of Business and Law Research*, *4*(1), 29–38.

Anele, K. K. (2013b). Money Laundering. In E. Azinge, & C. Nwabuzor (Eds.), Money Laundering Law and Policy (p. 34), Abuja: Nigerian Institute of Advanced Legal Studies

Ch. Suresh ,Dr. K. Thammi Reddy and N. Sweta (2016) A Hybrid Approach for Detecting Suspicious Accounts in Money Laundering Using Data Mining Techniques. I.J. Information Technology and Computer Science, 2016, 5, 37-43 Published Online May 2016 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijitcs.2016.05.04

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Ebere, C. C. (2016). Money Laundering and Forensic Accounting Skills in Nigerian. *Research Journal of Finance and Accounting, Vol.7*,(No.15,), 149-155.

Hopwood W., Young G. and Leiner Joy (2013) Forensic Accounting and Fraud Examination, 2nd Edition, USA, McGraw-Hill Companies Inc. pg. 12-13.

Kannan s, Somasundaram k, "Autoregressive based outlier algorithm to detect money laundering activities", Journal of money laundering control, vol. 20, pp. 1-7, 2017

Manjunath K.V. (2015). Data Mining Techniques for Anti Money Laundering. International Journal of Advanced Research in Science, Engineering, and Technology. Vol. 2, Issue 8.

Money Laundering dataset from https://github.com/William-McKee/money_laundering_challenge

Oluwadayisi, A. O., &Mimiko, M. O. (2016).Effects of money laundering on the economy of Nigeria.*Beijing Law Review*, *7*, 158-169.

Omar J. Sinayobye, Fred Kiwanuka, Swaib Kaawaase Kyanda. 2018. A State-of-the-Art Review of Machine Learning Techniques for Fraud Detection Research. In SEIA'18: SEIA '18: Symposium on Software Engineering inAfrica, May 27–28, 2018, Gothenburg, Sweden. ACM , 8 pages. *https://doi.org/10.1145/3195528.3195534*

Ramiro Camino, Radu.S, Leandro,M, and Petko Valtchev Finding Suspicious Activities in Financial Transactions and Distributed Ledgers "IEEE International Conference on Data Mining Workshops (ICDMW)" November 2017, DOI: 10.1109/ICDMW.2017.109

REUTERS, A. (2018, March 7th). *REUTERS*. Retrieved from https://www.reuters.com/article/us-nigeria-security-finance/nigerias-parliament-passes-anti-money-laundering-law-idUSKCN1GJ2DD

Reza Soltani et al (2016). A New Algorithm for Money Laundering Detection Based on Structural Similarity, IEEE, 2016

Rohit, Kamlesh D, and Dharmesh B Patel. 2015. 'Review on Detection of Suspicious Transaction in Anti-Money Laundering Using Data Mining Framework', International Journal for Innovative Research in Science and Technology, 1: 129-33.

Soltaniziba S. and Balafar M. A. (2015). The Study of Fraud Detection in Financial and Credit Institutions with Real Data. Computer Science and Engineering,10.5923/j.computer 2015 0502.02 5(2): 30-36

Stefan Axelsson, "Money Laundering Detection using Synthetic Data", The 27th annual workshop of the Swedish Artificial Intelligence Society (SAIS); 14–15 May 2012

S. S. Haykin, Neural networks and learning machines. Upper Saddle River, NJ: Pearson Education, third ed., 2009.

Vanguard news March, 2015. Importance of bank verification number https://www.vanguardngr.com/2015/03/importance-of-bank-verification-number/

Zhiyuan Chen, Le Dinh Van Khoa, Ee Na Teoh, Amril Nazir, Ettikan Kandasamy Karuppiah, Kim Sim Lam. Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection:a review February 2018 Knowledge and Information Systems https://doi.org/10.1007/s10115-017-1144-z

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

V-Authenticate: Voice Authentication System for Electorates Living with Disability

Olayemi M. Olaniyi[1], Jibril A. Bala[2], Juliana Ndunagu[3], Adamu Abubakar[4], and Ahmad Is'Haq[5]

[1,2,5]Federal University of Technology, Minna, Nigeria
[3]National Open University of Nigeria, Abuja, Nigeria
[4]Ibrahim Badamasi Babangida University, Lapai, Nigeria
[1]mikail.olaniyi@futminna.edu.ng

Abstract: Every ethnic, religious, and gender communities have electorates living with disability. They represent approximately 0.15 percent of the world's population roughly one out of every seven. When appropriate mechanism is not in place for such large volume of populace to fully participate in the electoral process, it deters democracy from giving this subset of populace a choice from how they wish to be governed. Electorate voice biometric modality could serve as credential to recognize legitimate voters. The voice recognition process relies on features influenced by the physical attribute of vocal tract and the behavioral features of the individual. Voice biometrics differs from other biometric techniques, in that speech samples are captured dynamically over a period of time. Incorporating technologies like the voice recognition system for authentication for the disabled is a huge step in increasing trust and inclusive participation in the democratic process. In this paper, we present V-authenticate, a voice authentication system to address the issue of valid voter's recognition and verification for the disabled electorate using their voice trait as biometric input parameter. The developed system was evaluated objectively using Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) qualitative metrics. The result of analysis system performance showed an average MSE value of 0.7 and PSNR was 42.9214 decibel showing the consistency of the applied algorithms to authenticate valid electorate living with disability. The system can be adapted by voting authority to enable electorate living with disability to participate in future electronic democratic decision making.

***Keywords*:** Voice; Security; Authentication; Voting; Democracy; Inclusion; Disability

I. INTRODUCTION

Democratic governance provides citizens the liberty to choose their leaders by means of election. One of the critical pillars to democratic process is voting (Okediran and Ganiyu, 2015). Voting provides individuals with a voice to influence decisions that affect their lives. However, electorates with disabilities have often been discriminated against in this context. In December 2006, historical response to the exclusion of people with disabilities from social and political processes by the United Nations (UN) General Assembly gave birth to the adoption of Convention on the Rights of Persons with Disabilities (CRPD) report (International Foundation for Electoral Systems, 2012). The UNCRPD report is an international human rights treaty, which promotes, protects, and ensures the fundamental human rights by persons with disabilities, particularly to public participation (International Foundation for Electoral Systems, 2014).

According to Article 1 of the UNCRPD treaty, the rights apply to everyone with a disability, including *"those who have long term physical, mental, intellectual or sensory impairments".* Article 29 of the treaty focuses on participation in political life. It ensures *"that persons with disabilities can effectively and fully participate in political and public life on an equal basis with others, directly or through freely chosen representatives, including the right and opportunity for persons with disabilities to vote and be elected"* (International Foundation for Electoral Systems, 2012; National Democratic Institute, 2012). As of August 2016, the UNCRPD has been signed and ratified by 160 countries. Of these, 23 African countries ratified the convention and protocol, and 16 African countries including Nigeria ratified the convention (Virendrakumar *et al.,* 2018a).The convention seeks to legislate laws and policies existing in these countries to ensure the participation of people with disabilities in general elections. Laws in countries such as Guinea, Mali, Nigeria, Ghana, Liberia, Burundi, Democratic Republic of Congo and Cote D'Ivoire allows voters with disabilities to request the assistance of a family member or a friend to cast their vote (Virendrakumar *et al.,* 2018a). Some African countries referred to the protection of human rights for the populace,

while fifteen countries specifically provided for the political participation of people with disabilities (Virendrakumar *et al.,* 2018b).

Overall, the UNCRPD document proclaimed equal opportunities for people with disabilities to exercise their franchise and referred to specific adjustments, mainly personal assistance, accessible communication and prioritization at the polling stations (Virendrakumar *et al.,* 2018a; Virendrakumar *et al.,* 2018b). In Nigeria for instance, the Independent National Electoral Commission's procedures allow people with disabilities in a targeted manner through priority access to polling units (especially visible pregnant women) to jump the queue in polling units when voting or registering to vote, use Braille Guide glass for visually impaired/blind registered voters and complete a special form(EC40H) (INEC, 2019).

According to National Population Commission, people living with disability in Nigeria is estimated to no fewer than 19 million (Premium Times, 2018). This margin has increased by an approximate number of 25 million (Haruna, 2017). The total adoption of People Living with Disability (PWD) framework by Independent National Electoral Commission (INEC) will ensure the voting process is more accessible to PWDs (International Republican Institute, 2018). Also, the legislative amendment of the Electoral Act by the Nigerian Senate empowering the INEC to introduce and implement any e-voting technology it deems suitable (Policy and Legal Advocacy Centre (PLAC), 2017; Verified Voting Foundation, 2017) will further enable improved participation by PWD.

Electronic voting characteristically is a multi-disciplinary subject studied by experts of different fields like engineering, cryptography, politics, law, economics and social sciences (Okediran *et al.,* 2011; Olaniyi *et al.,* 2015).Mostly e-voting is a challenging topic in cryptography and this arises primarily from the need to achieve voter anonymity from casted ballot, ensuring voter privacy without any violation and ensuring only eligible voters' votes have been counted (Cetinkaya & Cetinkaya, 2007). Another great challenge is the need for electronic voting system to enable disabled people in the sense that there is a critical need for the provision of a system to enable election participation by the disabled. The UNCRPD responds to this circumstance by providing a holistic solution to this need. Article 29 addresses the design and implementation of an electoral process that is non-discriminatory, while also requiring states to provide voters with disability-related accommodations and other facilitative measures to enable their equal right to vote (Virendrakumar *et al.,* 2018b). A number of secure electronic voting systems have been developed particularly for developing countries ecosystems using different techniques in the last decade (Enokela & Osuagwu, 2011, Aranuwa and Oriola, 2012; Abdulhamid *et al.,* 2013, Okediran and Ganiyu 2015; Olaniyi, *et al.,* 2016,Oke *et al*., 2017, Oke *et al.,* 2019). However, little attention has been given to design considerations to electorates with disability despite the fact that one out of every seven and no fewer than 25 million Nigerians are living with disabilities.

In this paper, we present V-authenticate, a voice recognition and verification system to address the issue of authentication for the disabled electorate by taking their voice as a measurable biometric trait. The rationale for the selection of voice trait for authentication is based on the fact that voice contains both physiological and behavioral factors unique to an individual. This measurable trait is difficult to duplicate since voice qualities are measured ranging from spectral magnitudes, dialect, style of speaking to pitch and format frequencies. Attempts to impersonate a voice or provide voice recordings to compromise authentication would fail due to the distinctive details of the voiceprint used for comparison and validation of valid voters (Authentify Inc, 2016). The remaining section of the paper is organized into three. Section II presents the system design methodology for the proposed system; Section III presents the results and discussions while Section IV concludes and opens scope for future research endeavors.

II. SYSTEM DESIGN METHODOLOGY

A. System Design Specification

In order to achieve a complete and functional system, the design and development of the system must be elucidated.

Figure 1 is a representation of the system architecture of the e-voting system showing the three phases of the electioneering process for the PWD: Pre-Election phase, Election phase and the Post-Election phase. It also includes the front and the back tier of the system.

This aspect involves the prerequisites essential for the accomplishment of the system based on the architectural design view for the purpose of implementation. The following are the design requirement specifications for each of the blocks in Figure 1:

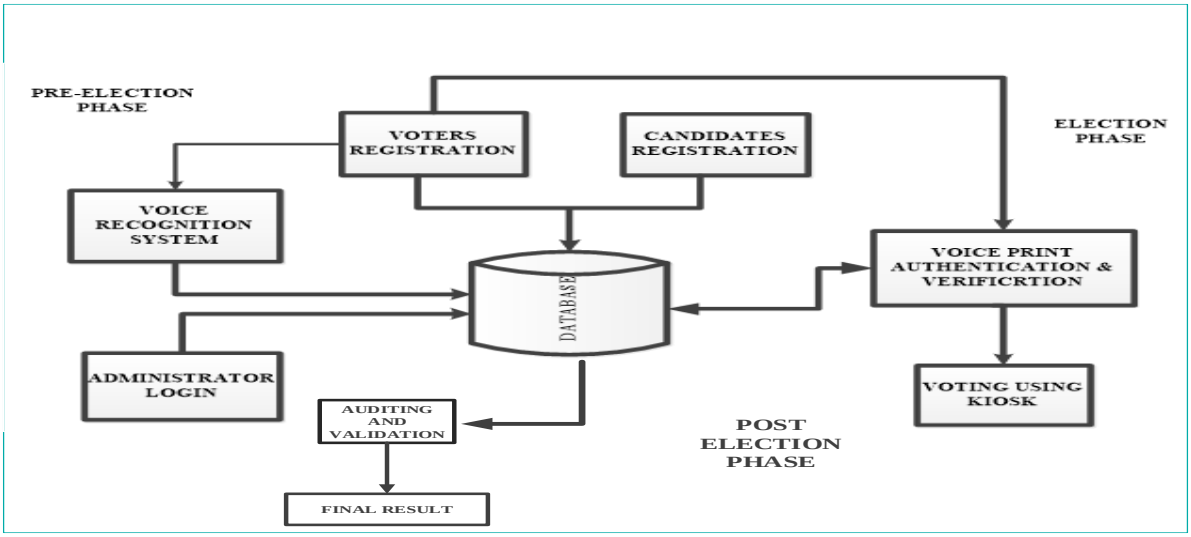


Figure 1: V-Authenticate System Architecture

- **Voice Recognition System:** The voice recognition system is a major part of the V-Authenticate-Franchising system for PWD. This system is responsible for the authentication and verification of voters as well as their votes. During the enrollment (registration) phase, voters register their personal information like age, date of birth, and address to the electoral official for record purposes, then they are required to give an input of their voice print (i.e. they speak into the microphone for voice print capture). Base level identification is done by assigning to every voter, a unique user identification number to be used in the identification of voters before the biometric capture.
- **Voting system:** The voting system is empowered with the voting software application that gives the interface of the voting process. It has a user friendly interface that enables the administrator/electoral official to easily conduct the registration for the individual voters. The voting system and the voice recognition system are heavily interwoven in the sense that they go hand in hand in their operations.
- **Database:** The database contains all the information provided by both the voter and the candidate. It guarantees the authorization of the voter by interfacing with the voting system application.

The proposed system by Figure 1 is divided into three different interwoven stages which are pre-election, election and post-election phases as mentioned earlier. The following are activities that are required to take place at each of the phases of the electioneering process:

- **Pre-election Phase**: This is the initial stage of the election process where the contestants, parties and the voters (PWD) are registered by the administrator during the enrollment phase. Details about the contestants are being inputted into the system and stored in the database; such information can be edited by the administrator if the need arises. During the voter's enrollment, a unique user ID that identifies to them alone is issued to the voters. Their voice prints are taken and stored in the database in anticipation for the Election Day. It is to importantly note that no two voters must have the same user ID and as such no two voters can have the same voice print because it is biometric.
- **Election Phase**: This is the stage of the electioneering process whereby the disabled voters come to the polling station to cast their votes, but first, they are authenticated by the system by providing their user ID and also their voice, meanwhile taking into cognizance that it has to be the same phrase uttered during the enrollment that is also required for the authentication process. During enrollment, speech sample is acquired in a controlled and supervised manner from the voter. The speaker recognition system has to process the speech signal in order to extract speaker discriminatory information from it to form the speaker model. During verification of the speech sample acquired from the user, the recognition system extract the features from the sample acquired and compare it against the models already stored beforehand for pattern matching/classification. Only eligible voters are allowed to vote as authentication and verification of identity is done before the voting takes place.
- **Post-Election Phase**: This is the final stage of the electioneering process, after the votes have been casted, they are being recorded and stored in the database for further processing. At this stage of the election

process, only the administrator has the rights to access the database. The votes are computed, validated and the results are generated for public declaration.

B. System Hardware and Software Design Specification

Hardware Requirement: These are the components part of the system that can be seen and touched. The major hardware component of the system is an external Universal serial bus (USB) Microphone, the kinobo Rikuto conference style USB microphone for voice input and recording purposes. It has captures frequency with a range of 20 Hz-20000Hz

Figure 2: Akiro Kinobo USB Microphone

Software Design Requirement: Based on the perspective of software engineering. The system is being stratified as stipulated as follows:

- **Voice Recognition Authentication and Verification:** In voice biometrics works disabled person's speech is digitized to produce a stored model voice print, or template. The voice digitization reduces each spoken word to segments composed of numerous dominant frequencies referred to as formants. Each segment has several tones that can be reconfigured in a digital format. The tones collectively identify the speaker's unique voice print which are stored in database. The voice recognition and verification involves three major steps which include:
  - i. Pre-processing of captured speech
  - ii. Feature extraction using Mel-Frequency Cepstral Coefficients (MFCC) Algorithm
  - iii. Feature Classification using Dynamic Time Warping (DTW)

Speaker recognition system consists of two main stages, the enrolment stage and the verification stage. These phases involve three main parts as shown in Figure 3. From Figure 3, at the time of voice enrollment, the voice sample is acquired in a controlled and supervised manner from the voter. The speaker recognition system has to process the speech signal in order to extract speaker discriminatory information from it.

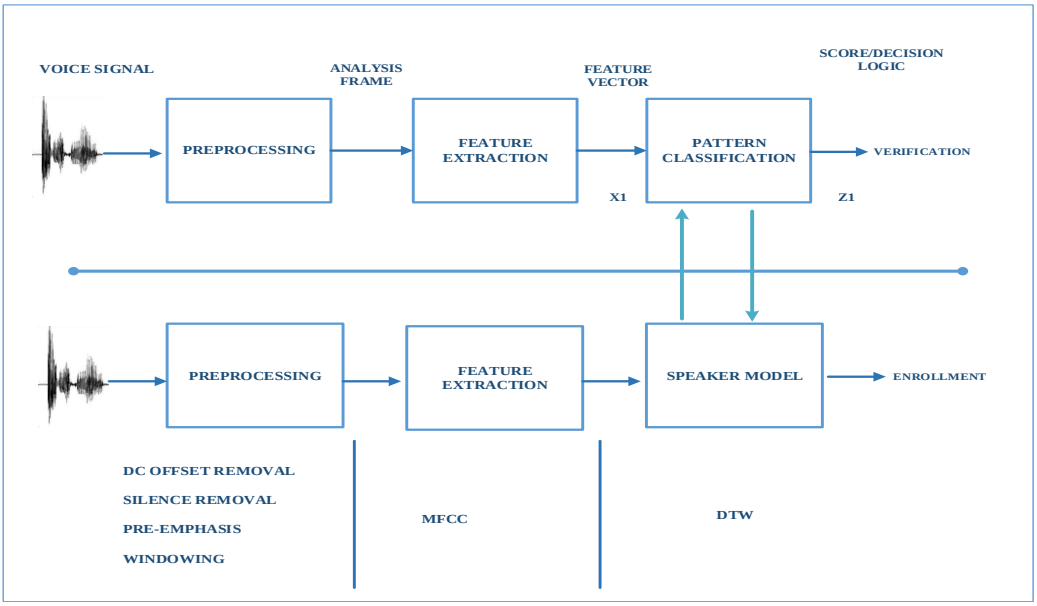


Figure 3: Representation of the voice recognition system

C. Pre-processing

Speech data performs in a discrete-time speech signal because such data are recorded by sampling the input. Therefore, some pre-processing techniques are required to make the discrete-time speech signal more flexible for further processes. There are four pre-processing techniques that come before feature extraction. These include DC offset removal, silence removal, pre-emphasis and windowing.

- DC Offset Removal: Speech data are discrete-time speech signal; it often carries some redundant constant offset called DC offset. This DC offset will affect quality of the information extracted from the speech signal. Consequently, we calculate the average value of the speech signal and subtracting this from DC offset from the original speech signal.
- Silence Removal: This process is performed in order to remove silence periods from the speech containing silence frames. So, the signal becomes more compact as shown in the Figure 3. Silence frames are audio frames of background noise with a low energy level with respect to voice segments.
- Pre-emphasizing: Pre-emphasis is a technique used in speech processing to enhance high frequencies of the signal. The main purpose of pre-emphasizing is to spectrally flatten the speech signal that is to increase the relative energy of its high-frequency spectrum.
- Windowing: This step requires processing the window of each individual frame in order to minimize the signal discontinuities at the beginning and end of each frame. A windowing function is used on each frame to smooth the signal and make it more amendable for spectral analysis. The concept here is to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. If the window is defined as Y_1 (n), where n is the number of samples in each frame, then the result of windowing is the signal

$$\mathbf{Y_1(n) = x\,(n)w(n),} \qquad \mathbf{0 \le n \le N - 1} \tag{1}$$

Typically, the Hamming Window is used, which is of the form

$$\boldsymbol{w(n) = 0.54 - 0.46\cos[2\pi n/N-1],} \quad \boldsymbol{0 \le n \le N - 1} \tag{2}$$

D. Feature Extraction

The voice feature was extracted using MFCC algorithm. Mel-frequency is the measure of the human perception of the frequency content of speech signals on the "Mel-scale". Mel-Frequency Cepstrum (MFC) stands for the power spectrum of the speech, based on a linear cosine transform of a log power spectrum, computed on the non-linear Mel-frequency. The MFCC features are obtained by taking the log of the outputs of a Mel-frequency filter bank, which is subsequently subjected to cepstrum analysis (Thakur, 2015)

The final MFCC feature vectors are obtained by retaining about 12-15 lowest Discrete Cosine Transform (DCT) coefficients. Each vector is independent of each other and ordering information is lost. The MFCCs are, therefore, the coefficients that collectively make up the MFC. The frequency bands in the MFC are equally spaced and from research findings in the psychophysical field, it has been established that the Mel scale approximates the auditory system of humans better than a linearly spaced frequency band. Melfrequency warping of the spectrum gives emphasis on low frequencies that are more important for speech perception by humans. The computational components of the MFCC algorithm are captured in Figure 4 (Thakur, 2015):

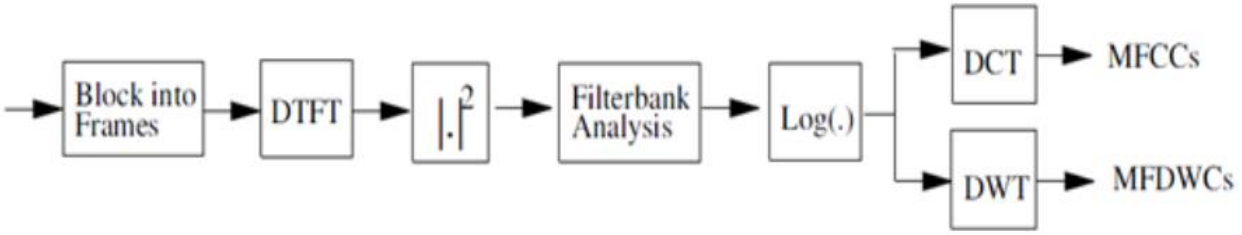


Figure 4: Extraction voice feature Using MFCC (Chang, 2012)

E. Feature Classification

The pattern classification task of speaker recognition involves computing a match score (a measure of the similarity of the input feature vectors) to some model. Speaker models are constructed from the features extracted from the speech signal. To enroll users into the system, a model of the voice, based on the extracted features, is generated and stored in a voting database. To authenticate a user/voter, the matching algorithm compares/scores the incoming speech signal with the model of the claimed user. The technique used for the pattern classification from Figure 3 is the Dynamic Time Warping (DTW).

Dynamic time warping is an algorithm for measuring similarity between two sequences which may vary in time or speed. DTW is an algorithm that focuses on matching two sequences of feature vectors by repetitively shrinking or expanding the time axis till an exact match is obtained between the two sequences. It is generally used to calculate the distance between the two-time series that vary in time. A real time application of DTW in the voice recognition is that, it should be able to recognize the user's voice even when spoken at different speeds. In order to check the similarity between two voice signals or the time series are warped non-linearly.

F. System Performance Evaluation Measures

Objective evaluation tests: The objective comparison of single channel speech of the voters captured was carried by evaluating the speech signals (original and processed speech) with Mean Square Error (MSE), and Peak Signal to Noise Ratio (PSNR) performance evaluation metrics. These metrics were computed based on mathematical comparison of the original and processed speech signals.

- **Peak Signal to Noise Ratio (PSNR):** This is used to estimate the difference between the recorded and the processed speech and is a function of the Mean Square Error (MSE). It is a ratio of the quality of the recorded voice sample against the processed voice sample calculated in decibels. The higher the PSNR (in equation 3) of the comparison, the better the analysis.

$$\text{PSNR} = 10 \log_{10} = R^2 \div \text{MSE} \tag{3}$$

Where R is length of the reconstructed signal

- **Mean Square Error (MSE):** This is an error metrics used to represent the cumulative square error between the original voice signal and the processed voice signal. The lower the value of the MSE, the lower the error rate between the samples which shows proper processing of the signal. It is calculated using the following relation in equation 4:

$$MSE = \frac{1}{4MN} \sum_{i=1}^{2M} \sum_{j=1}^{2N} \left(C_{ij} - S_{ij}\right)^2 \tag{4}$$

Generally, PSNR values below 30db signifies a fairly low quality. While, the value of MSE decreases when the two signals are similar to each other. A better quality signal would strive for 40db and above.

III. RESULTS AND DISCUSSION

The software prototype for the V-authenticate was developed Using MATLAB Graphical User Interface(GUI) called GUIDE, Figures 5, 6, 7,and 8 show the developed interface of the voting system which allows electorates with disability to franchise their choice in democratic governance.

Figure 5 shows the main voting platform while figure 6 depicts software platform for registration and eligible voter voice samples. The registration procedure is commenced for the proper documentation of the information about the voter in Figure 6.

Figure 7 shows the verification page which comes after the enrollment is done in figure 6. Here, the voter provides his/her user ID and the administrator loads it and the voter is then required to repeat the same phrase that was used during the enrollment stage for verification. If the user ID is correct, then the voter is allowed to record their voice for verification of identity after which they are redirected to the voting page to cast their votes.

In Figure 8, there is an automatic voice prompt that instructs the voter of the next step to be taken. The voter makes a choice via voice instruction and the vote is casted for voters preferred choice on the voting interface. Using the metrics outlined in Section III. The result of the tested developed system is shown in Table 1. Table 1 shows the voice sample of six voters' audio files evaluation with the various sizes, bit rates, vote size, and the audio sample size as well. It shows the voice sample analysis after evaluating their Mean Square Error (MSE) and Peak Signal to Noise Ratio in MATLAB software toolbox.

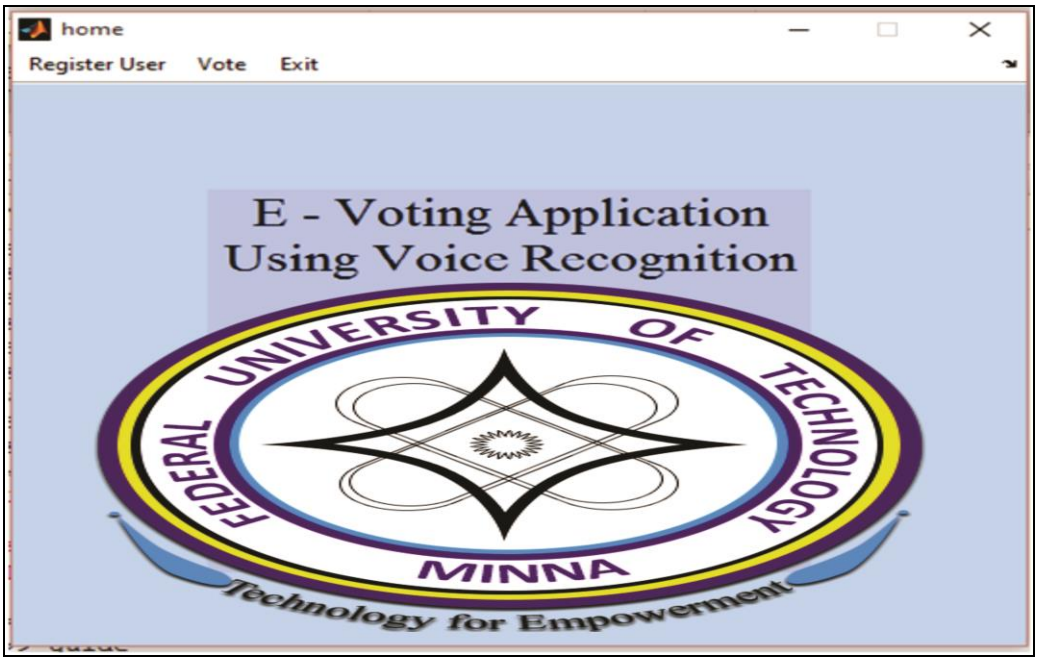


Figure 5: Electronic Voting System Home Page

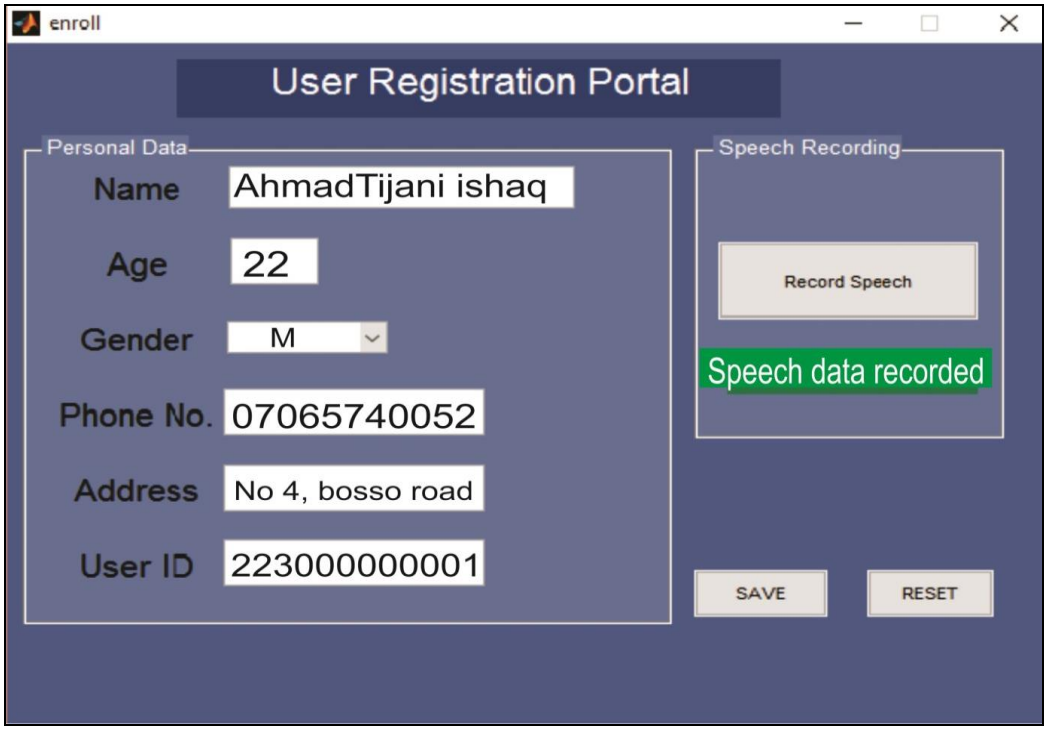


Figure 6: The voter enrollment page using Voice print input

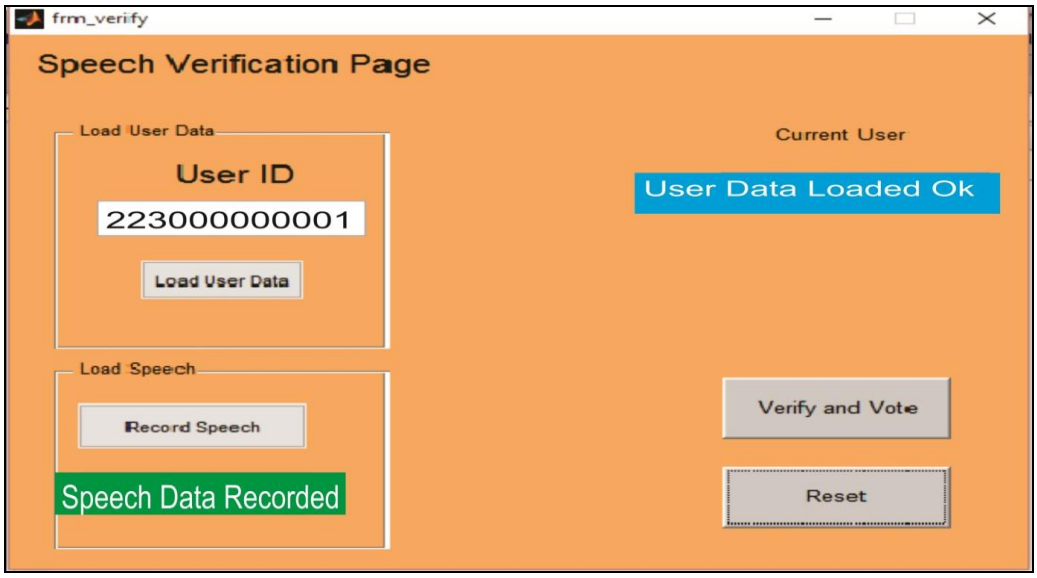


Figure 7: Voter Voice Verification page

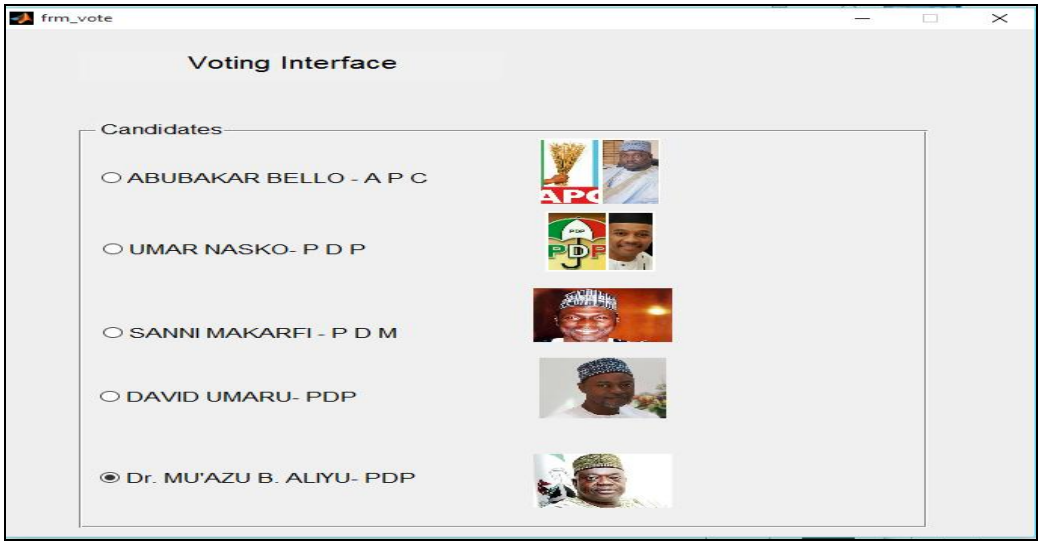


Figure 8: The voting interface

Table 1: MSE and PSNR of the Sampled Audio Voice Files

Voters Audio Voice File	Audio size (KB)	Bit Rate (Kbps)	Vote Size (kb)	MSE (db)	PSNR (db)
Tunde	20	120	1.62	0.312	41.5478
Saleem	29	135	1.81	1.002	38.2577
Taofeek	45	163	2.30	0.541	45.3698
Adam	38	139	2.21	0.126	56.3688
Ibrahim	46	143	2.36	1.320	40.2365
Niyi	27	134	1.75	1.112	35.7482

From Table 1, the average value of MSE value was 0.7 and PSNR was 42.9214 decibel. By the premises provided in Section II, PSNR values below 30db signifies a fairly low quality. While, the value of MSE decreases when the two signals are similar to each other. A better quality signal would strive for 40db above. It is clear from the results of performance measure in Table 1 (with computed average value of MSE and PSNR values) that V-authenticate- e-voting authentication system can be adapted to enable PWD to participate in future democratic decision making since the value is quantitatively higher than the standard benchmark proving the efficiency and effectiveness of the techniques adopted in Section II.

IV. CONCLUSION AND RECOMMENDATIONS

This work has successfully shown significant improvement on the existing secure voting authentication systems by focusing on design considerations for the disabled persons via voice biometric recognition. This work includes the three phases of the electioneering processes. The voice recognition system takes care of the authentication and verification of PWD voters. The signal analysis of the voice signal was accomplished by using MFCC technique on voice spectrum factors. This represents the exact vocal system for stored words and provide a high level of perception of the human voice and a better representation of the signal. This accounts for a higher resolution in the performance of recognition. Having tested and evaluated the developed system, it can be inferred or concluded that the system could be adopted by voting authority, like INEC, as part of measures to fully embrace people living with Disability (PWD) framework to ensure the voting process is more accessible to PWDs.

However, our contribution in this work principally authenticates voters based on ***"Isolated Word Recognition"*** metaphor. It is recommended the work can be extended to ***"Continuous Word Recognition"*** and ultimately create a Language Independent Recognition System capable of making the developed system more robust in terms of performance. Also, Statistical Models like Hidden Markov Models and Gaussian Mixture Models as well as learning models like Neural Networks can be incorporated in this direction to improve quality of voice signals required to verify valid voters with disability. This would make the system much tolerant to noise variations and associated residues and hence make it less error prone. Other open issues that can be looked into are:

- Addition of a cryptographic technique for digital vote signature to increase integrity of the votes.
- A multimodal biometrics for multifactor authentications for subjects in the same domain of application.

REFERENCES

Okediran O.O. & Ganiyu A.A. (2015). Framework of Electronic Voting in Nigeria.*International Journal of Computer Applications*.129(3).12-16.

International Foundation for Electoral Systems (IFES).2012. The Convention on the Rights of Persons with Disabilities and the Optional Protocol. Retrieved online at http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf 11th March 2019.

International Foundation for Electoral Systems. 2014. "Equal Access: How to Include Persons with Disabilties in Elections and Political Processes. Retrieved online at https://www.ifes.org/publications/equal-access-how-include-persons-disabilities-elections-and-political-processes. on 10th March 2019

National Democratic Institute.2012. Civic Update: Mainstreaming Persons with Disabilities. Retrieved online at https://www.ndi.org/Civic_Update_Jan_2012.pdf on 11th March 2019

Virendrakumar, B., Jolley, E., Badu, E., Murphy, R., Schmidt, E., 2018a. Disability-iclusive in Africa. Retreved Online at https://www.sightsavers.org/wp-content/uploads/2018/04/Disability-inclusive-elections-in-Africa-a-qualitative-systematic-review.pdf on 10 March 2019.

Virendrakumar, B., Jolley, E., Badu, E., Murphy, R., Schmidt, E., (2018b) Disability inclusive elections in

Africa: a systematic review of published and unpublished literature, Disability & Society, 33:4, 509-538, DOI: 10.1080/09687599.2018.1431108

Independent National Electoral Commission (INEC).2019. Regulations and guideline for the conduct of elections. Retrieved online at https://www.inecnigeria.org/wp-contents/uploads/2019/regulations and guidelines-2019.pdf on 7th March 2019

Nigerians with Disability Decree. 1993. Retrieved online at https://dredf.org/legal-advocacy/international-disability-rights/international-laws/nigeria-disability-decree/ on 8th March 2019.

Haruna M.A.2017.The Problems of Living with Disability in Nigeria. Journal of Law, Policy and Globalization.65(1).103-113

Premium Times .2018. Nigeria: 19 Million Nigerians living with disability .Retrieved online at https://allafrica.com/stories/201810080010.html on 12th March 2019.

International Republican Institute. 2018. Statement Of The Joint NDI/IRi Pre-Election Assessment Mission To Nigeria. Retrieved Online at: www.iri.org/sites/default/files/2018-7-20-ndi-irirnigeria-peam-statement.pdf 12th March 2019.

Policy and Legal Advocacy Centre (PLAC). (2017). *Report of the Ad-hoc Committee on the Legislative*

Agenda. Retrieved from http://placng.org/wp/wpc -co n t e n t / u p l o a d s / 2 0 1 6 / 0 6 / 8 TH - S E N A T E - D R A F TLEGISLATIVE-AGENDA-2.7 15.pdf, on 11th March 2019.

Verified Voting Foundation. (2017). *Senate Amends Electoral Act, Approves Electronic Voting.* Retrieved from https://thevotingnews.com/senate-amends-electoralact- approves-electronic-voting-ynaija/ on 12 April 2018.

Okediran O.O, Omediora E.O, Olabiyisi S.O, Ganiyu R.A. and Alo O.O. (2011). A framework For a multifaceted Electronic Voting system. *International Journal of Applied Science and Technology*, 135-142.

Cetinkaya, O. and Cetinkaya, D. (2007). Verification and Validation Issues in Electronic Voting. Electronic Journal of e-Government, 5(2), pp 117 – 126

Enokela, J. A. & Osuagwu, C C. (2011). An Algorithm for the Conduct of Multiple Simultaneous Multi-Party Elections Using a Microcontroller. *Pacific Journal of Science and Technology*. 12(2):253-259

Aranuwa, F.O. & Oriola, O. (2012). Improved Electoral Fraud Prevention Mechanism for Efficient.Electronic Voting. *African Journal of Computing & ICT, 5*(6), 70-77.2012

Abdulhamid, S. M., Adebayo, O. S. Ugiomoh, D. O. AbdulMalik. M. D. (2013). The Design and Development of Real-Time E-Voting System in Nigeria with Emphasis on Security and Result Veracity",. *I.J. Computer Network and Information Security*, 5(5), 9-18.DOI: 10.5815/ijcnis.2013.05.02

Olaniyi, O.M., T.A. Folorunso, Ahmed, A., Joseph, O., (2016). Design of Secure Electronic Voting System Using Fingerprint Biometrics and Crypto-Watermarking Approach. *I.J. Information Engineering and Electronic Business*, 5, 9-17.

Oke, B., Olaniyi, O.M.,Aboaba, A. A., & Arulogun, O.T. (2017). Developing Multifactor Authentication Technique for Secure Electronic Voting System. In S. Misra, V. O. Matthews, & A. Adewumi (Ed.), *IEEE International Conference on Computing, Networking and Informatics (ICCNI 2017)* (pp. 48-53). Ota: Covenant University, Canaanland, Ota, Ogun State, Nigeria.

Oke B. A., Olaniyi, O. M., Aboaba A. A., & Arulogun O. T. (2019) "Securing Electronic Voting System Using Crystographic Technique", *ATBU University Journal of Science, Technology & Education (JOSTE).* 7(1), pp. *88-105*

Authentify.inc. (2016). Voice Biometric Authentication. Chicago: Authentify.inc

Olaniyi, O.M, Folorunso, T.A, Abdullahi, A.M, Abdulsalam K. 2015. Developement of A secure E-voting system Using RFID and Enhanced LSB Audio Steganographic Technique. IOSR Journal of Computer Engineering 17(6), 86-97

Chang, W.W. 2012. Time frequency analysis and wavelet transform tutorial. Time frequency analysis for voiceprint (speaker) recognition. Retrieved online :https://www.scribd.com/doc/198845429/Voiceprint-Speaker

Thakur S. 2015. System Architecture for Secure Mobile Internet Voting. Doctor of Technology Thesis**,** Department of Information Technology in the Faculty of Accounting and Informatics, Durban University of Technology Durban, South Africa

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

Performance Analysis of Security Information and Event Management Solutions for Detection of Web-Based Attacks

Morufu Olalere[1], Juliana Ndunagu[2], Shafi'i M. Abdulhamid[3], and Peter Odey[4]

[1,3]Federal University of Technology, Minna, Nigeria

[2]National Open University, Abuja, Nigeria

[1]lerejide@futminna.edu.ng, [2]jndunagu@noun.edu.ng, [3]shafii.abdulhamid@futminna.edu.ng, [4]pewodey29@gmail.com

Abstract: With rising trends and forms of web application attacks such as SQL Injection, cross-site scripting and the likes, most organizations today deploy a security information and event management solution as a proactive measure for threat management to get a centralized view of the network security posture and for advanced reporting of incidents. The days of relying merely on perimeter controls are elapsed; it is no longer enough to just rely on firewalls, Intrusion Detection Systems, Intrusion Protection systems and antivirus alone. Security information and event management systems have become a crucial and essential component of complex enterprise networks. They typically aggregate and correlate incidents from different systems and platforms, and carry out a rule-based analysis to detect advanced threats. This paper detects, evaluates and analyzes the performance of various SIEM detecting web based attacks, noting the time of report of attack and behavioral patterns of each SIEM. An attack simulation experiment is performed on different SIEM tools to demonstrate the capabilities of SIEM in detecting any suspicious behavior of event logs and alerting the attacks in near real-time, then the best tool is recommended based on its ability to collect, filter, normalize, correlate, alert, and report attacks within minutes after attack incidents.

***Keywords*:** Security Information and Event Management; Web-Based Attacks; Malware; SQL Injection; Real-Time; Network; Enterprise; Small and Medium Businesses

I. INTRODUCTION

This paper, present an analysis based on the performance of various security information and event management (SIEM) solutions in detecting several web application based attacks presenting an extensive view on the best solution that can be deployed to mitigate these attacks. With the projection of over five billion internet of things devices by the end of 2020 and the trends in mobile technology development, most organization's networks comprise of large scaled information delivered by devices that offer huge amount of data around the world also cyber criminals are more experienced these days in their ability to break into an organization's networks and security teams need intelligent systems which provide meaningful data to detect advanced attacks and therefore require a successful apparatus for control, checking and mitigating cyber-attacks, this makes SIEM solutions a great tool for any security team since it provides an overall view of the entire network (Beechey, 2010).

Researchers and other experts knowing that web application attacks regularly bypasses signatures based systems of data and therefore, require additional data sources beyond simply detecting specific attack traffic which continue to impact thousands of websites and millions of users each year, have proposed different strategies to address the difficulties presented by these attacks but yet their methodologies either neglect to handle the full extent of the misfortune, since they only knows about just a subset of the extensive variety of systems accessible to attackers who are attempting to exploit web application vulnerabilities.

SIEMs gives real time analysis and additionally examination of security occasions which empowers fast remediation before an attack occurs, (Miller et al., 2011), it is capable of real-time monitoring of the network at all time to detect and alert in case it identifies an incident and a critical security issue. The main roles of the SIEM solution in an organization's network are to monitor the log data, collect and store it in a central console. The next step involves analyzing the log data, filtering alerts and build correlation rules (Seyed, 2016). The basic evaluation part of an SIEM system involves the evaluation of three elements. First is the central console, second is the monitoring entity, and finally the communication process between the monitoring entity and the central console (Pastrana et al., 2013). For the SIEM to function effectively, its design and development must ensure that the monitoring entity and the communication process supplies complete and integrated information to the central console. Carrying out a general performance analysis on various SIEM solutions can be very useful in this area by

collecting and correlating the data needed to identify patterns that signal attacks (Taylor, 2016) and present an overview of the organization's network.

The paper diagnoses the general capabilities and components of SIEMs and noticed that though most of the selected solution possess almost all the functionalities of SIEMs but still lack some in-depth functionalities that will be termed as technical and descriptive in analyzing events. The criteria for determining the best tool are based on three performance indicators that provides an understandable and a user friendly environment: (1) data examination and analysis, (2) real-time incidence reporting and (3) the percentage accuracy of the tool. Major components, such as support for SOC, forensic capabilities, use of threat intelligence, identity and access management capabilities and others are discussed under the various criteria.

II. Security Information and Event Management Solutions (SIEMS)

SIEM is a combination of security information management (SIM) and security event management (SEM) functionalities into a single security management system (Michael, 2012) designed to collect security logs from a wide variety of sources within an organization's network. While the SIM segment mainly emphasizes on the analysis of historical data intending to improve the long term storage performance and efficiency of information security infrastructures, SEM handles the aggregation of data into an understandable information investigating and providing mechanisms for which these incidents can be dealt with immediately (Henrik, 2009). SIEM solutions are required for handling of increased level of information security as well as the analysis and management of centralized log, giving an extensive view of the organization's network in monitoring both real-time events and a pile of long-term data to detect strange patterns of operation and alert the security team when needed (Zodik, 2016). Further, the systems respond quickly in case of an attack with accuracy up to 90% and speed within 60 seconds of event correlation and have the capability of generating compliance reports (Butler, 2009). SIEM systems allow users to build content, logic, conditions, and criteria. These are used with correlation rules deployed for faster identification and escalation of a security event or incidence. Data from different sources is collected and aggregated through agents. Noise or unwanted data is filtered and normalized to a proper format for analysis through correlation (Igor et al.,). Furthermore, SIEM works by deploying different sets of agents in a hierarchical manner with an aim of collecting security-related information and events from the end-user devices, system servers, and the network equipment Also, SIEM gathers security information for specialized security equipment, and tools such as intrusion detection systems, firewalls, and antivirus. The collected information is forwarded to a centralized control and management console (Seyed, 2016). The central console further performs inspections on the logs and flags any anomalies. Altogether, the roles of SIEM product is to collect, consolidate, correlate, communicate and control (Guillermo et al., 2015).

III. Web Application Attack

Major challenges in today's enterprise networks are web-based attacks (Halfond et al., 2005) such as distributed denial of service attack (DDoS), SQL injection, cross-site scripting (XSS) and the likes, which allows cyber attackers to take control of the internal resources of the network through vulnerable front-end gaining access to the database server and other resources of the network to ex-filtrate sensitive data. Web applications are programs running on a web browser and generally have a three-tier construction, which provides an interface between the web server and the end user to communicate: the presentation tier, common gateway interface (CGI) tier and the database tier (Inyong et al., 2012). Web browsers presents data generated at the server side in the form of html pages through the http protocol to the client side and any vulnerability on this architecture may lead to a web application attack. However, web server logs do not contain any data sent in the http header but it may contain valuable data, since most http forms and their parameters are submitted by post requests, this is a big deficiency for web server log files (Meyer, 2008) and also, SIEM solutions do not require session data but the ability to access this information can dramatically improve the capability of the system. The basic elements of session data in this case include the source IP, source port, destination IP, destination port, protocol (e.g., TCP, UDP, ICMP), timestamp, generally when the session began and measure of the amount of information exchanged during the session.

IV. Methodology

A. Configuring attack and defense Machine

With the aid of a virtual machine manager, multiple PCs instances are run, some collector agent of SIEM were configured on the target machine (defense machine) to collect data into SIEM managers for indexing and consolidation, accounts are created with the various vendors of SIEMs to allow for trial versions of their solutions to be used. Dashboards are created also in all the SIEMs with the aim of achieving expected results and security

monitoring items were added to present a graphical posture of the network. The simulated attack machine performs an attack while the target system detects the attack and subsequently prevents them. The solutions used allow the user to launch a web-based attack from a locally hosted web page. First, the log data is collected from different the web applications, which is then aggregated and normalized, the log data is then parsed and correlated, a process that involves putting pieces of an attack together to form a complete picture. In this step that contextual information about a network and common threats becomes more useful. The collected data is first stored locally in the organization's network before it is transferred to a central area for analysis and archiving

B. Designing the User Interface

A user interface for deploying the attack was developed on the attack machine consisting of html pages and a form which collects information from the user and posts them to the database, also a backend consisting of embedded PHP scripts and other associative scripts for form validation, processing of the information collected from the user interface and transferring them to the database accordingly.

C. SIEM used for the experiment

The selection of the seven (7) SIEM solutions used for the experiment are based on the major solutions that gives users the ability to quickly identify and react to inward and outside attacks, streamlines all parts of security tasks and gives understanding into machine information produced from security challenges, and some top general functionalities of SIEM as recommended by the Open Web Application Security Project (OWASP) top 10 projects which includes the capabilities of data aggregation, data correlation, alerting, provision of a dashboard, forensic analysis, data retention compliance, threat intelligence and so on. The selected SIEM for this examination are;

- Alien Vault Unified Security Management (USM) platform which arrangement uses the strong capacity of open source to enable clients contributes and get constant information and data about malicious hosts.
- Hewlett Packer ArcSight because of its rack-mount appliances which has a vast array of built-in capabilities that can collect, store and analyze all security data from a single interface. It is capable of analyzing millions of security events from firewalls, intrusion protection systems, end-point devices, and an array of other log- and data-producing devices. It boasts built-in security dashboards and audit reports that visualize threats and compliance and is able to protect against zero-day attacks, advanced persistent threats, breach attempts, insider attacks, malware and unauthorized user access
- Splunk Enterprise Security (ES) which allows for Real-Time Monitoring, clear visual picture of an organization’s security posture, easily customize views and drill down to the raw event, gain a security-specific view of your data to increase detection capabilities and optimize incident response, use ad-hoc search and static, dynamic and visual correlations to determine malicious activities, conduct breach and investigative analyses to trace the dynamic activities associated with advanced threats, Gain an understanding of threats and execute best practices for incident investigation and response and leverage the analytic stories of Splunk ES Content Update
- IBM QRadar Security which aids security teams to correctly detect and prioritize threats across the organization providing intelligent insights that enable them respond quickly to reduce the impact of incidents and correlates all this different information and aggregates related events into single alerts to accelerates incident analysis and remediation
- McAfee Enterprise Security that delivers actionable intelligence and integrations to prioritize, investigate, and respond to threats providing continuous visibility into threats and risk, actionable analysis to guide speedy investigations and orchestration of security remediation
- Logrhythm threat detection is one of the leading solution in security intelligence and analytics that enables organizations worldwide to detect, respond and quickly neutralize the damage caused by cyber attacks
- SolarWinds IT monitoring and management tools are used for system administration and network engineering

V. RESULT

Diagnosing the general capabilities and components of SIEMs, it is noticed that though the selected SIEMs possess almost all the functionalities of SIEMs but some still lack some in-depth functionalities that will be termed as technical and descriptive in analyzing logs. The criteria for determining the SIEMs to be recommended are based on these three (3) performance indicators that provide an understandable and a user friendly environment. They include data examination and analysis, real-time reporting and percentage accuracy of reporting attacks.

A. Data Examination and Analysis

Though SIEMs are aimed at automating all logs, analyzing and reporting alert, the best SIEMs should be understandable by humans to allow for easy examination and analysis of security events, such as supporting incident handling efforts. It should provide support for human readable examination and analysis of data which are mostly and basically the search capabilities and data visualization capabilities. Thus, these seven (7) rules were drawn out to test for the SIEMs with the best data examination and analysis functionality. They include

- Scalability of several solutions which determine how many concurrent queries can be run for a security operation center (SOC) operations and if overlapping intrusion protection systems is supported
- Provision for a security operation center (SOC) Support
- Capabilities to support identity and access management systems (IAM)
- The capabilities that supports scalability from small and medium business (SMB) to large implementations
- Selection of critical fields and scheduled summarization of events thereby providing a summary table
- Search Capabilities
- Data Visualization Capabilities and Database Activity Monitoring

The percentages are calculated using the equation (1)

$$(Sum\ of\ criteria\ support\ for\ a\ SIEM)/\ (Sum\ of\ criteria\ occurrences)\ \times 100. \quad (1)$$

The table below (Table 1) shows the tabular representation of the set rules (criteria) displaying the tools that support (with indicator 1) or does not support (with indicator 0) a rule

Table 1. Data Examination and Analysis

Criteria/Tool	AlienVault	ArcSight	Logrhythm	McAfee	Qradar	SolarWinds	Splunk
Support for SOC	1	1	1	1	1	1	1
Scalability	1	1	0	1	1	1	1
IAM Capability	1	1	1	1	1	1	1
Support for SMB	1	1	1	1	1	0	1
Summary Table	0	0	1	1	0	1	1
Search Capabilities	0	0	1	0	0	0	1
Data Visualization Capabilities	1	0	1	0	0	0	1
%Rate	**71.43%**	**57.14%**	**85.71%**	**71.43%**	**57.14%**	**57.14%**	**100.00%**

Indicators: 0 = Does not support; 1 = Support for

The figure (1) pictorially shows the distribution of the percentages of these rules in a pie chart.

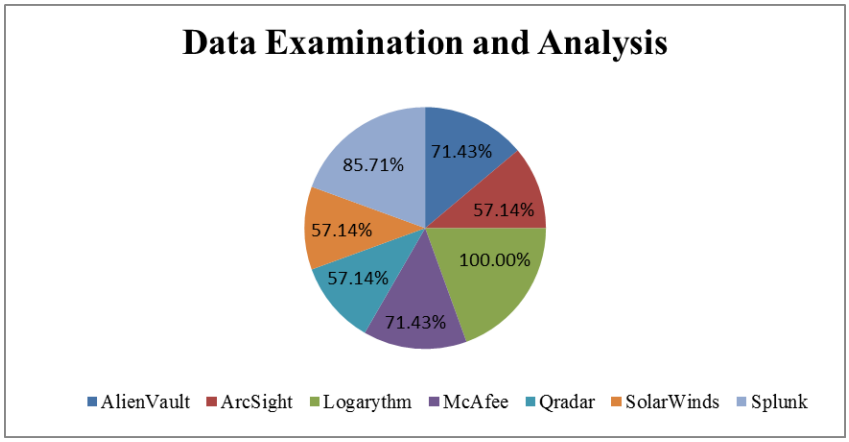


Figure 1. Pie chart showing Data examination and analysis

The product that has the most robust search capabilities is Splunk Enterprise and LogRhythm Security Intelligence Platform, which gives the user multiple types of searches and drill down capabilities. While other SIEMs, there is no information available on their search capabilities

B. Real-time Alert Reporting:

Comparing the timeliness of report of an attack, the security and effectiveness is based on implementation, that is, what the administrator wants or how the SIEMs is been configured with respect to environment specification. All the SIEM solutions reports attacks in near real time. ArcSight ESM does this through the ArcSight threat response manager add-on, Logrhythm uses its security intelligence platform, SolarWinds, through its log & event manager, McAfee ESM, Splunk Enterprise, QRadar, and AlienVault. Since all the tested tools report attacks in real time, the indicators for all is (1) meaning it is supported in all the tools

C. Accuracy

The third criterion is determining the percentage accuracy of these solutions reporting the attacks, this is done by determining the tool's ability to use of threat intelligence, forensic capabilities and the probability of generating false positive alerts.

1) Use of threat intelligence

Threat intelligence feeds holds important information about the characteristics of newly observed threats around the world, to enable the SIEM identify malicious activity more quickly and with greater confidence. All the SIEMs studied provide support for the use of threat intelligence feeds. Some feeds are provided by the SIEM vendors through a third party or they are being acquired directly by the customer from a third party as shown in table (2)

Table 2. Use of threat intelligence provided by SIEM vendors or third party

Tool	By SIEM vendor	By SIEM Users
ArcSight	×	●
AlienVault	●	●
Splunk ES	×	●
Loghrythm	●	●
McAfee	●	×
SolarWinds	×	●
QRadar	●	×

Indicator: ● = Support × = Does not support

Table (2), shows that McAfee ESM and QRadar offers threat intelligence provided by the SIEM vendor itself. SolarWinds, ArcSight and Splunk offer support for third-party threat intelligence feeds, and Logrhythm partners with five major threat intelligence vendors to allow customers to use one of their feeds or a combination of those feeds. Finally, the AlienVault OSSIM, being open source, has community-supported threat intelligence feeds available.

2) Forensic Capabilities

The collecting, identifying, and validating of logs to ensure its integrity and most times admissibility requires deploying a solution for protecting its sensitive data and detecting suspicious activity. Most of the SIEM products used in the experiment were able to correlates events in the internal systems, calculate risks, and generate reports showing patterns in chaotic log data. The information contains digital fingerprints of the attacks and traces of malicious activities. Considering the forensic capabilities of the SIEM studied, the following six (6) forensic features must be embedded in the SIEM especially those features required for all forensic products used by certified forensic investigators. They include:

a) No Intrusion

This feature ensures that data collected are not tampered with in any way by storing a copy of unmodified log entries as well as normal events in a backend database, having built-in functions for periodic backups and restores, having intrusion prevention mechanisms that can block the actions of an attacker who is attempting to corrupt logs, guarantee that the logs has not been modified by providing information for forensic purposes about changes made during collection and export and also must support role-based access to users access of certain data.

b) Integrity

The SIEM functionality should also allow collecting data and storing it for further forensic analysis in a tamper-proof form by using integrity mechanisms, such as running hash checks on blocks of stored log data.

c) High Performance

The SIEM used should be able to process an increasing number of separate events per second, so these systems require complex algorithms to process data as fast as possible since forensic analysis requires the collection of all possible information, as it all may be necessary for an investigation. Logrhythm though easy to deploy may not scale to support very high event volume environment.

d) Data Retention

The SIEM must have a long-term centralized storage of historical data to ensure the correlation of data over time and to retain data for forensic analysis this was determine by the amount of database storage, as some regulations may require data to be available for a particular length of time.

e) Data Relevance

The SIEM must contain features that allow users to reduce the amount of data provided for forensic analysis by providing a filtering system logs. More so, one should be able to narrow down data by keywords or times. Filters are important for forensic analysis because they provide users with only relevant data that relates to the incident to be investigated

f) Timestamps

Timestamps and timezone are the most sensible and valuable bit of information that are extracted from log data. Timestamps are essential for linking the events recorded by the SIEM to real-world facts. It is very important to have timestamps and timezone that are as precise as possible. Information are available with microsecond precision and they must be stored with precision and this original timestamp value and time zone that a timestamp is associated with must be available for forensic analysis and purposes.

g) Results and Documentation

The forensic analysis results provided by the SIEM solution must be shown in a familiar way and export query document in different format for it to be able to be admissible in court.

The percentage is calculated using formula (2)

(Sum of feature support for a SIEM)/ (Sum of feature occurrences) ×100

Table (3) and figure (3) below shows the percentage accuracy of all the SIEM solutions used for the study

Table 3. Forensic capabilities of SIEM with respect to Accuracy

Criteria/Tool	AlienVault	ArcSight	Logrhythm	McAfee	Qradar	SolarWinds	Splunk
No Intrusion	1	1	1	1	1	1	1
Integrity	1	1	1	1	1	1	1
Timestamp	1	1	1	1	1	1	1
Hiigh Performance	0	1	0	0	1	1	1
Data Retention	1	1	1	1	1	1	1
Data Relevance	0	0	1	0	1	0	1
Result and Documentation	1	1	1	1	1	1	1
%Rate	**71.43%**	**85.71%**	**85.71%**	**71.43%**	**100.00%**	**85.71%**	**100.00%**

Indicators: 0 = Does not support; 1 = Support for

3) False Positive Alert

Examining and relying upon the SIEMs to identify patterns which indicate threats, it was noticed that the patterns don't always tell the full story, it was mostly accurate but once in a while it will trigger incorrectly maybe based on the logic of the original rule definition, these are seen as false positive alert and it was noticed in all the SIEMs used. Notingly, the DDoS attack launched constantly produce authentication issues on all the SIEMs until some logic were added to the rules to remove these distracting noise. Though false positive alerts are not immediate security threats, it is very important to address the issues that causes them because it can distract the security team from real threatening events and ignoring them might leave your systems vulnerable to malicious attacks.

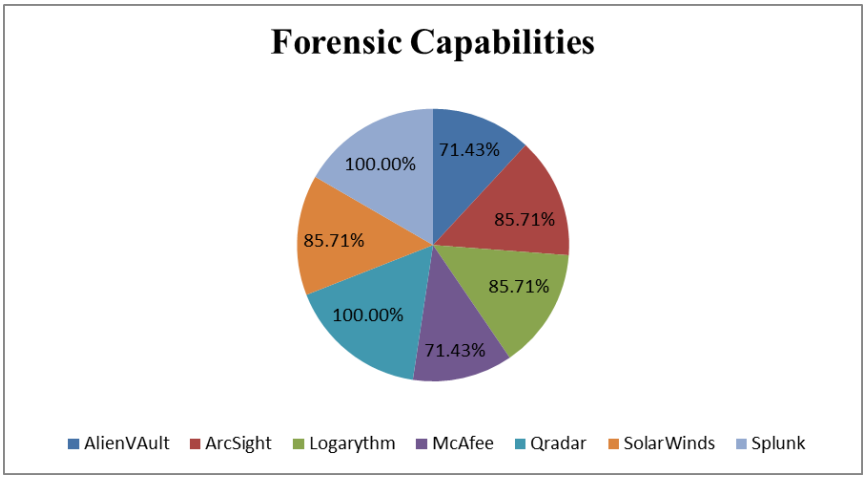


Figure 2. Pie chart showing distribution of forensic capabilities amongst tools

Table (4) and figure (4) displays a pictorial representation of these features amongst the tool used for this study.

Table 3. Percentage Accuracy

Criteria/Tool	AlienVAult	ArcSight	Logrhythm	McAfee	Qradar	SolarWinds	Splunk
Threat Intelligence	1	1	1	1	1	1	1
Forensic Capability	0.71	0.86	0.86	0.71	1.00	0.86	1.00
False positive alert	n/a	n/a	n/a	n/a	n/a	n/a	n/a
%Rate of Accuracy	**57.14%**	**61.90%**	**61.90%**	**57.14%**	**66.67%**	**61.90%**	**66.67%**

Indicator: 0 = Does not Support; 1 = Support for; n/a = Not Applicable

Since all the tools generate false positive alerts, it is not applicable in calculating the percentage rate but it is necessary in determining the accuracy of each tool.

Figure (4) below display the general representation and distribution of all the tools used in the study and how it responds to the performance indicators (data aggregation, real time alert reporting and accuracy of report of attack) detecting web based attacks. It summarizes that; the organization's security team will have to perform its own evaluation to determine the best tool to be used following their requirements and the need at hand.

VI. CONCLUSION

Detecting web-based attacks via web server logs is always challenging due to the likelihood of generating false positives alerts and the tendencies for attacker to encode attacks therefore, the analysis of SIEM solution is done using SIEMs that are strong indicators in detecting web based attacks. The examination is carried out using AlienVault, Logryhthm, Splunk, QRadar, ArcSight, McAfee ESM and Solarwinds, to detect web based attacks such as SQL injection, cross site scripting, distributed denial of service and malware attacks. The SIEM solution is deployed as a proactive measure for threat management, to get a centralized view of organization's security posture and for advanced reporting of security incidents collecting all relevant data in a central location and providing customizable altering and reporting that can provide significant value by helping to determine whether or not an incident occurred.

VII. RECOMMENDATION

Organization should perform its own evaluation, considering all the aspects of the SIEM that may be of importance to the organization since each SIEM implementation has to address a unique set of log sources and has to support different combinations of compliance reporting requirements, among other variations, the best SIEM system for one organization may not be suitable at all for another. Any organization interested in leveraging threat intelligence to improve the accuracy and performance of its SIEM should carefully investigate the quality of each available threat intelligence feed, particularly how often the threat intelligence is updated and how the vendor's confidence in each piece of intelligence is conveyed. For example, IBM Security QRadar SIEM provides relative

scores for each threat along with the threat category and this helps facilitate better decision making when responding to threats

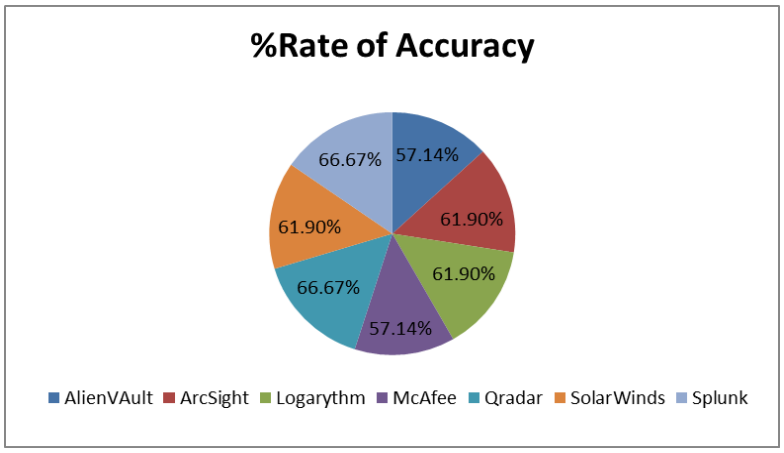
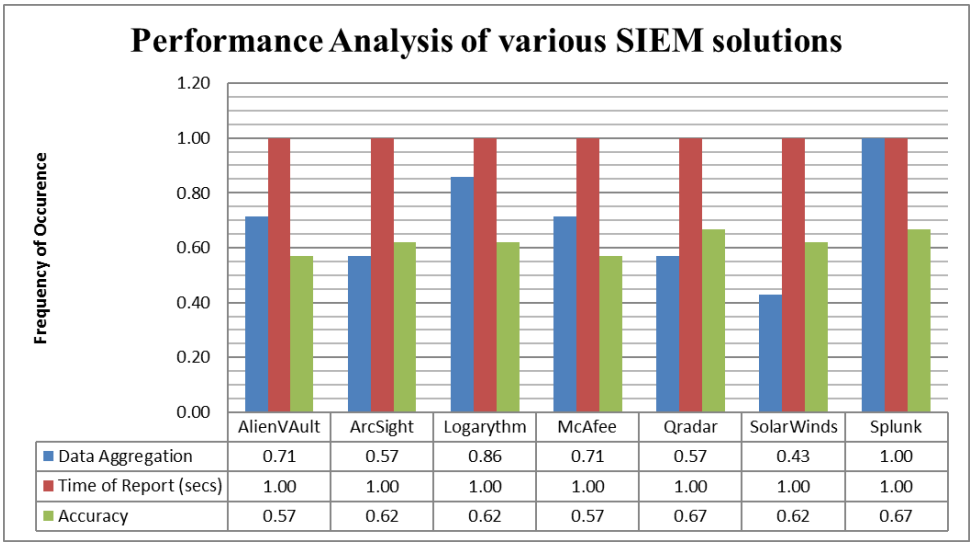


Figure 3. Percentage Accuracy of SIEM tools studied



	AlienVAult	ArcSight	Logarythm	McAfee	Qradar	SolarWinds	Splunk
Data Aggregation	0.71	0.57	0.86	0.71	0.57	0.43	1.00
Time of Report (secs)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Accuracy	0.57	0.62	0.62	0.57	0.67	0.62	0.67

REFERENCES

Beechey, J. (2010). *SIEM Based Intrusion Detection with Q1Labs Qradar*. *Information Security*.

Seyed Morteza Zeinali, Tallinn, 2006. Analysis of Security Information and Event Management (SIEM) Evasion and Detection Methods

D.Miller, S. Harris, A.Harper, S. VanDyke, Ch. Blask, *Security Information and Event Management (SIEM) Implementation*, McGraw-Hill, 2011

Halfond, W. G. J., Viegas, J., & Orso, A. (2008). A Classification of SQL Injection Attacks and Countermeasures. *Preventing Sql Code Injection By Combining Static and Runtime Analysis*. https://doi.org/doi=10.1.1.95.2968

Henrik Karlzénan (2009). An Analysis of Security Information and Event Management Systems

Igor Anastasov, Danco Davcev, „SIEM implementation for global and distributed environments, "%1 *Computer Applications and Information Systems (WCCAIS), 2014 World Congress*, 2014.

Inyong Lee, Soonki Jeong, Sangsoo Yeo, Jongsub Moon. A novel method for SQL injection attack detection based on removing SQL query attribute values

J. M. Butler, „Benchmarking Security Information Event Management," SANS, 2009.

J. Park, B. Noh, SQL injection attack detection: profiling of web application parameter using the sequence pairwise alignment, in: Information Security Applications, in: LNCS, vol. 4298, 2007, pp. 74–82.

Michael, K. (2012). The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice. *Computers & Security*, *31*(4), 634–635. https://doi.org/10.1016/j.cose.2012.03.005

S. Pastrana, J. Montero-Castillo, and A. Orfila, „Evading Idss And Firewalls As Fundamental Sources Of Information In Siems," p. Chapter 7, 2013.

Taylor, T. (2016). *Using context to improve network-based exploit kit detection*. *ProQuest Dissertations and Theses*. Retrieved from https://search.proquest.com/docview/1828257791?accountid=45153

The Open Web Application Security Project, OWASP TOP 10 Project. http://www.owasp.org/
Y. Huang, S. Huang, T. Lin, C. Tasi, Web application security assessment by fault injection and behavior monitoring, in: Proceedings of the 12th International Conference on World Wide Web, 2003, pp. 148–159.
Zodik, G. (2016). Cognitive and Contextual Enterprise Mobile Computing. In *Proceedings of the 9th India Software Engineering Conference on - ISEC '16* (pp. 11–12). New York, New York, USA: ACM Press. https://doi.org/10.1145/2856636.2876471

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

Securing Electronic-Health Record Systems Using Cryptographic Techniques

Shefiu Olusegun Ganiyu[1], Olayemi Mikail Olaniyi[2], and Orooniyi Tosin[3]

Federal University of Technology Minna, Niger State, Nigeria

[1]shefiu.ganiyu@futminna.edu.ng, [2]mikail.olaniyi@futminna.edu.ng, [3]orooniyi.tosin@st.futminna.edu.ng

Abstract: The successful storage, manipulation and retrieval of electronic medical records are some of the important functions of Electronic-Health Record (EHR) Systems. More so, the confidentiality and integrity of health information is of paramount importance to healthcare management organizations. Unfortunately, the occurrence of information security breaches with regards to EHR, which include loss of valuable data as a result of theft by unauthorized users, is increasingly becoming worrisome situations. Furthermore, these security challenges have led to cases of intentional or unintentional disclosure of vital patients' information among others. Interestingly, previous studies attempted to implement secure systems in domains like finance, communication and commerce using asymmetric or visual cryptography. However, based on reviewed literatures, this security techniques are yet to be combined and applied as worthwhile approach to address authorization challenges in the EHR domain. Thus, a prototype web-based e-health record system was developed and combination of asymmetric and visual cryptographies were incorporated into the system as authorization mechanism to ensure confidentiality and integrity of pertinent health information. Subsequently, the fortified E-health record system provided important functionalities of a typical EHR record system, as well as prevent unauthorized users from accessing vital information.

***Keywords*:** Heath record; Electronic-heath; Cryptography; e-health record system

I. INTRODUCTION

Electronic-health (e-health) is interchangeably used with Health Informatics, an interdisciplinary field that develops and applies theories, methods and processes for the generation, storage, retrieval, use and sharing of medical data, information, and knowledge (AMIA, 2019). One critical challenge of e-health is the issue of data storage through Electronic-heath Record (EHR), an electronic version of a patient's paper record. Characteristically, EHRs are designed to include the patient medical and treatment histories, diagnoses, medications, immunization dates, allergies, radiology images, and laboratory test results. They are also meant to provide secure patient medical information, only, to authorize users. However, sundry EHR's offer inadequate capacities to monitor or keep in check the exact health records to be shared, even with main family members and caregivers (Cushman, Froomkin, Cava, Abril, & Goodman, 2010).

Over the years, technology faces various modes of intrusion which include; the integration of incorrect data or destructive programs into information systems, theft resulting to loss of important data or programs from a system and overall takeover and manipulation of a system's set-up and performance. Usually, hackers carryout these security breaches as a means of satisfying their own agenda. More so, criminals use it to improve their own interests, as well as commercial organizations that utilize it as mechanisms for neutralizing rivals or terrorists whose attacks can spread over a wide geographical range, eventually producing related effects irrespective of the criminal (Onuiri, Idowu, & Komolafe, 2015).

Furthermore, the rate of information security issues with regards to Electronic-Health Records (EHR's) is increasingly becoming worrisome, this brings about the urgent need to mitigate these challenges so as to curb unauthorized manipulation, storage and retrieval of information, thereby, retaining required confidentiality, availability and integrity of the EHR's. Otherwise, any interception or interruption, may result to issues such as inaccurate diagnosis and medication, and in the worst case, death. Therefore, in this paper we present, a secure electronic health record using asymmetric and k of n visual cryptographies.

Visual Cryptography or Visual Secret Sharing (VC/VSS) is an information hiding technique that allows information (in form of images or text) to be encoded in a way that decoding is done by the human visual aid (Archana & Ambily, 2016). Developed by Naor and Shamir (1994), the technique involves sharing an encrypted secret image into n shares such that stacking a sufficient number of shares only reveals the secret image. In VSS, the

shares generated contains only black and white pixels which makes it difficult to easily gain any information about the secret image by viewing only one share. The secret image is revealed only by stacking sufficient number of shares. (Basavegowda & Seenappa 2013).

Asymmetric cryptography otherwise known as public key cryptography is a data security technique, which uses two distinct keys in cryptographic process. Unlike symmetric cryptography that uses same key for encryption and decryption, asymmetric process includes generations of the two distinct keys (private and public) from complex mathematical algorithm. The public key is made available to all message senders for encryption purpose, whereas, the private key is reserved as secret and only known to the receiver (owner of the public key) for message decryption. While key generation is faster with symmetric cryptography, the reverse is the case with asymmetric cryptography. Nevertheless, safeguarding a key from unauthorised user during transfer among legitimate users is the main setback for symmetric cryptography, while such challenge does not arise in asymmetric cryptography.

Our contribution significantly provides a medium through which the privacy of a patient's electronic healthcare record could be preserved by enhancing the security measures on the EHR's since they are sensitive to patients and their caregivers. The remaining section of paper is organized into five sections. Section II provides review of related works, Section III presents the methodology, Results and discussion are presented in Section IV, while Section V concludes the paper.

II. RELATED WORKS

Harman, Flite, and Bond (2012) conducted systematic analysis of health management system. The authors highlighted numerous advantages of electronic system over manual approach to health management. Amongst others, the authors identified reduced storage space, physical security of records and delay in processing health information as major drawbacks of manual approach. However, the authors emphatically raised concern over security of patient information by healthcare handlers. Thus, Harman et al. (2012) suggested the use of encryption as a proactive countermeasure to ameliorate security of health information in terms of confidentiality for devices employed in processing and disseminating such crucial information. In a similar systematic review, Fernández-Alemán, Señor, Lozoya, and Toval (2013) agreed with Harman et al. (2012) on the major benefits of E-Health management systems and the importance of encryption scheme, but emphasised the need for more concerted efforts to address emerging security and privacy threats, because access to E-health systems becomes ubiquitous among stakeholders (Li, Zou, Liu, & Chen, 2011).

Likewise, Onuiri et al. (2015), explored various cyberspace challenges with respect to safeguarding information, investigating cyberspace threats, identifying stakeholders and transmitting information in E-health record system. The authors opined the combination of cryptography and biometric authentication as mitigating strategy for threats associated with e-health systems. Empirically, Qureshi et al. (2014) studied the effect of E-health which comprised of Telemedicine and mobile health on distant or remote locations. Interestingly, the authors stated the important characteristics of a successful and useful e-health system to include user-friendliness, easy-to-use and well-integrated functionalities, as well as a backup, tracking and alert facilities. The authors observed a tremendous shift in e-health adoption, but reported low adoption in developing countries due to inadequate information on e-health websites. Thus, Qureshi et al. (2014) proposed that a qualitative, secured and well-structured e-health system will enhance the operational efficiency and sustainability of the healthcare institutions.

Also, Hawkes, Yasinsac and Cline (2000) analysed the possibility of securing financial documents with visual cryptography. The authors developed an application named VCRYPT which was described as a quick and uncomplicated visual cryptography technique. The VCRYPT is meant for privacy protection when data transits between offices, thereby preserving the integrity of document. According to Hawkes et al. (2000), previous visual cryptography systems suffered from the decoded images having a greying-effect or the deciphered image being blurry and much obscurer than the original image. Consequently, the VCRYPT application addressed this problem and also reduced the computational impact. Furthermore, in order to enhance the performance of visual cryptography irrespective of domain of application, D'Arco and De Prisco (2016) proposed the *deterministic* and *random grid models*. However, the authors reported issues relating to contrast, pixel expansion and randomness reduction in both models.

Recently, Okkali and Sandikkaya (2017) applied visual cryptography to facial recognition in surveillance system for privacy preservation. The system employed facial characteristics of targeted person is concealed and split into *n* shares which are put into distinct storage areas and under the control of distinct entities. More so, Kester (2013) developed a visual crytopgraphic encrytption system for medical image using pixel shuffling procedure. The system used algorithm that shuffled the red, blue and green (RGB) values of a pixel using an encryption key generted from the image. However, the system is limited to medical image which is only one of the the file formats involved in a comprehensive e-health record system. Again, Kester (2013) used same key for both encryption and decryption,

which could expose it to key theft, a known problem with symmetric key cryptography. Again, Sugiharto et al. (2018) developed a crypto system that is based on visual cryptography and Rivest, Shamir and Adleman (RSA) techniques for only image encryption and decryption. Thus, the use of asymmetry cryptography (RSA) by the authors eliminates issues relating to security of key, yet the study is limited to image cryptography.

Accordingly, Harman et al. (2012), Fernández-Alemán et al. (2013), Onuiri et al. (2015) and Qureshi et al. (2014) focus on systematic review of visual cryptography in E-health management system without actual implementation of the security scheme in the domain. The studies revealed the stakeholders, core functions and security expectations of e-health record system. Although, D'Arco and De Prisco (2016) developed variants of visual cryptography mechanism towards performance improvement, but Hawkes et al. (2000) and Okkali and Sandikkaya (2017) applied the security techniques to other domains of human endeavour outside E-health record system. Perhaps, Kester (2013) was among few studies, which applied visual cryrptography to medical image, an aspect of E-health record system, but used symmetric cryptography and covered only image. Similarly, Sugiharto et al. (2018) utilised asymmetric technique, but their study is limited to image, while text data and specific domain of application were not considered.

III. METHODOLOGY

No doubt, visual cryptography is among the underlying techniques for guaranteeing security of images that concerns human visual system in digital space. Thus, this study utilised the algorithm proposed by Kester (2013) for medical image encryption. However, instead of generating encryption key from the image as implemented by the researcher, we use key generated by RSA for visual cryptographic. Also, RSA was used to secure textual data in the E-health record. One major advantage of this approach is that, the proposed system can leverage on the soundness of RSA for textual cryptography. Similarly, since typical E-health record comprises of both image and textual contents, then using asymmetric key for both contents will eliminate the need to have disparate encryption keys for a single cryptosystem. In addition, this approach promotes seamless integration and ciphering of text and image from system implementation point of view. Figure 1 shows the algorithm for the encryption and decryption processes, while Figure 2 presents the high-level diagram for same processes.

1. *Begin*
2. *// encryption*
3. *Import E-health record (P_r) = {plain text (P_t), plain image (P_i)}*
4. *Convert {P_t, P_i} to binary equivalent {P'_t, P'_i}*
5. *Generate encryption key (E_k) and decryption key (D_k) with RSA*
6. *Encrypt P'_t with RSA as $a := (P'_t, E_k)$*
7. *Encrypt P'_i with visual cryptography as $v := (P'_i, E_k)$*
8. *Generate cyphered record (C_r) with $E_a = g(v,a)$*
9. *Store C_r in E-health database*
10. *// decryption*
11. *Retrieve C_r*
12. *Decompose C_r to {P'_t, P'_i} with $D_a = g^{-1}(C_r)$*
13. *Decrypt P'_t with $a^{-1} := (P'_t, D_k)$*
14. *Decrypt P'_i with $v^{-1} := (P'_i, D_k)$*
15. *Revert {P'_t, P'_i} to {P_t, P_i}*
16. *Export {P_t, P_i} as plain E-health record*
17. *End*

Figure 1: Algorithm for E-health Record Encryption and Decryption

A. System Architecture

The system architecture for the proposed E-health record system is presented in Figure 3. The architecture shows the stakeholders, information flow and data storage, as well as the positions of security mechanism to ensure confidentiality and integrity of medical records. Most of all, the stakeholders are patients whose health information needs to be secured and medical team who is the custodian of the records. As indicated in the diagram, all stakeholders require appropriate asymmetric key to encrypt or decrypt desired information. Equally important, all stakeholders must pass authentication challenge by supplying valid username and password before accessing the system.

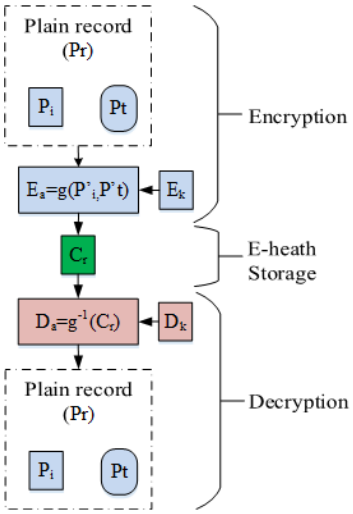


Figure 2: Encryption and decryption block diagram for e-health

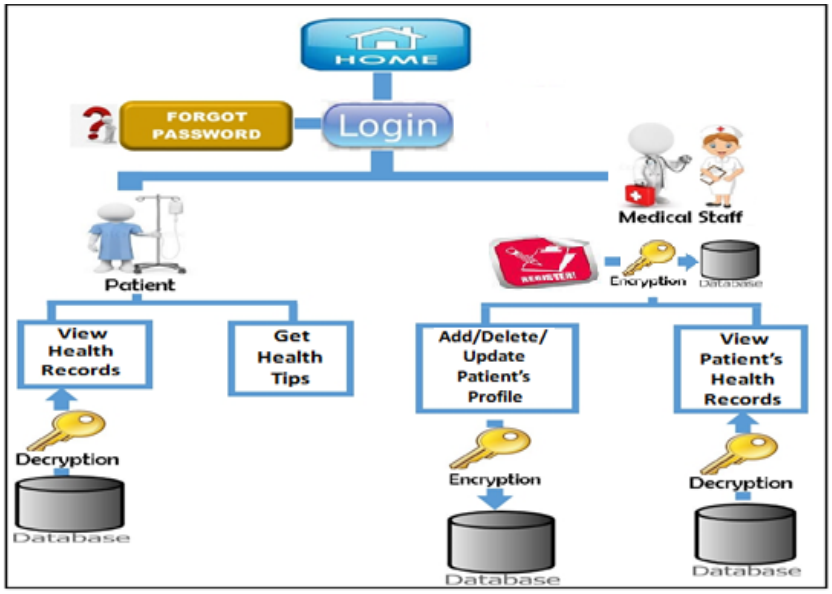


Figure 3: System architecture

B. System Development

The flow of information within the proposed system is presented in Figure 4. Thus, the flowchart guided the implementation of the system. Subsequently, the system was implemented as web application using Hypertext Preprocessor (PHP) platform and cryptography plugins. Then, MySQL database was used as database management system.

IV. RESULTS AND DISCUSSION

Figures 5 to 8 show some of the interfaces implemented for the E-health record system. For instance, all users are required to fill the form in Figure 5 as part of the sign-up process prior to login into the system. Thereafter, the interface presented in Figure 6 will be used for subsequent login by users depending on the role assigned to them at sign-up by system administrator. Again, Figure 7 depicts a typical health record for patient registered on the system. Similarly, Figure 8 reveals the access denial page displayed when user attempt to view health record with wrong access key.

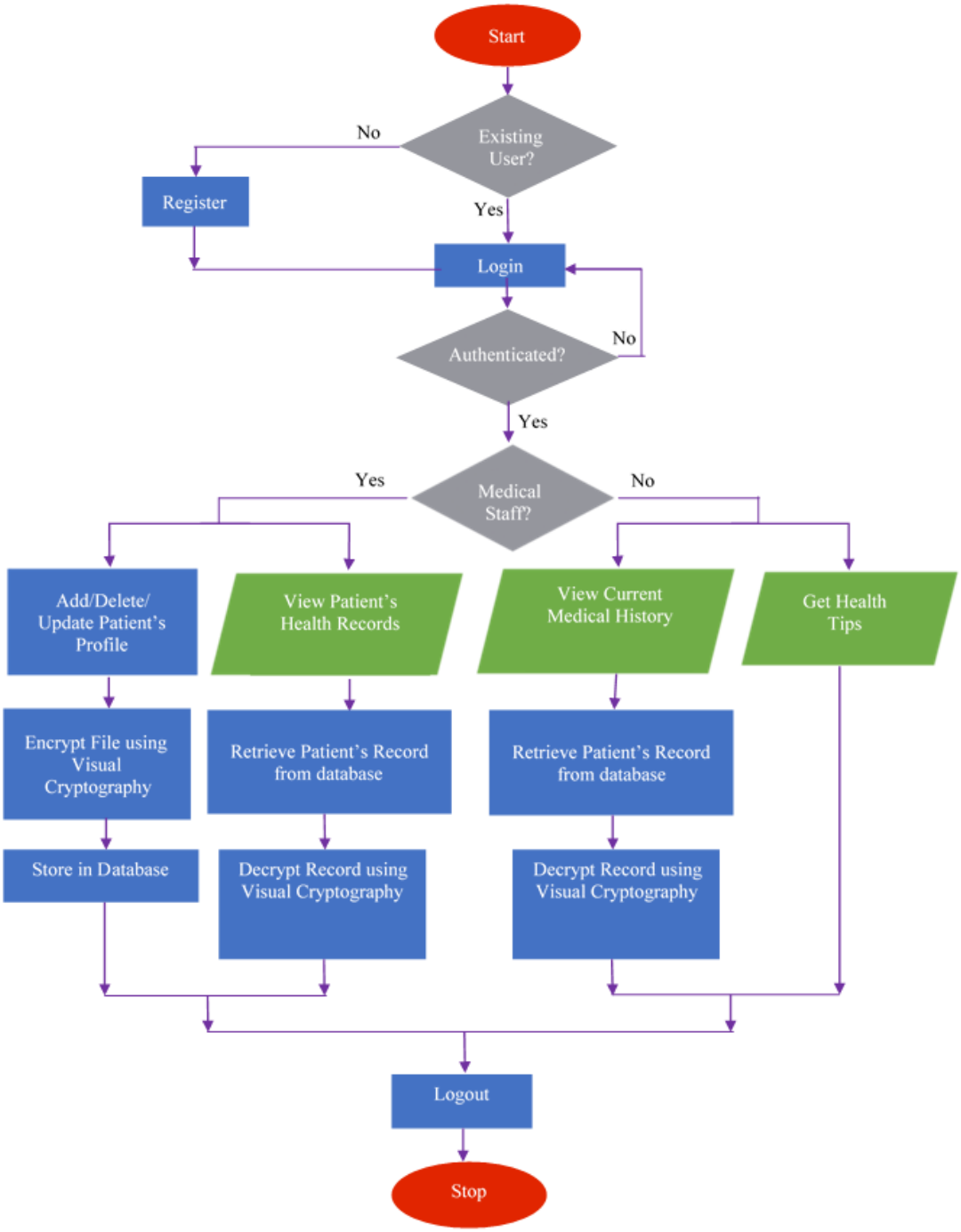


Figure 4: System flowchart

V. CONCLUSION

The synergistic combination of asymmetric and visual cryptography in the proposed EHR system has successfully provided a medium through which the privacy of a patient's electronic health record is maintained. This will no doubt ensure that patient's Electronic Health Records (EHR's) are securely stored and that only authorized persons are allowed to view them. It is still important that further studies be carried out to improve on the security of EHR's possibly by incorporating another security technique in addition to the Visual Cryptography technique. Also, quantitative evaluation should be conducted to measure the performance of the proposition in this paper.

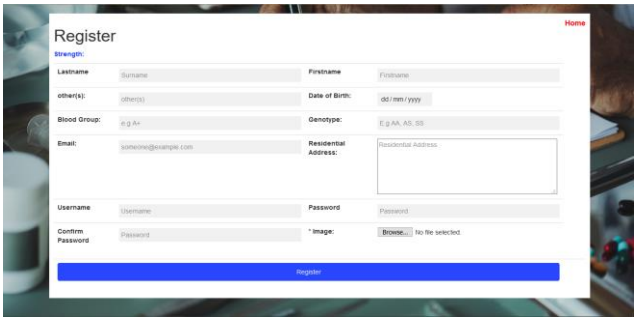


Figure 5: User registration

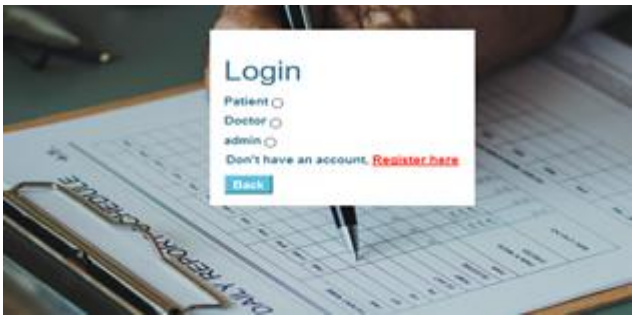


Figure 6: User authentication

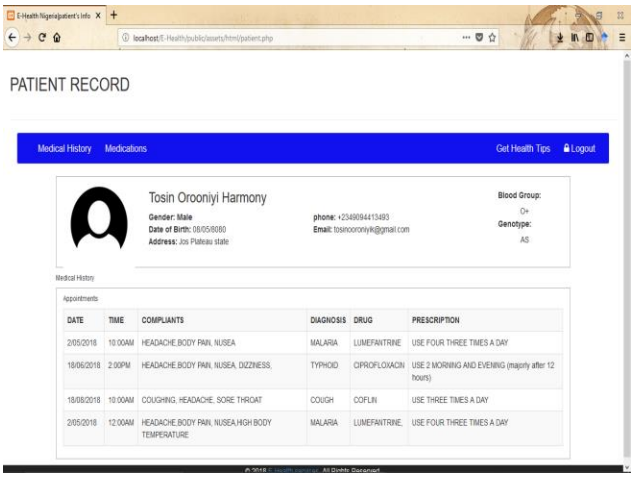


Figure 7: Patient record

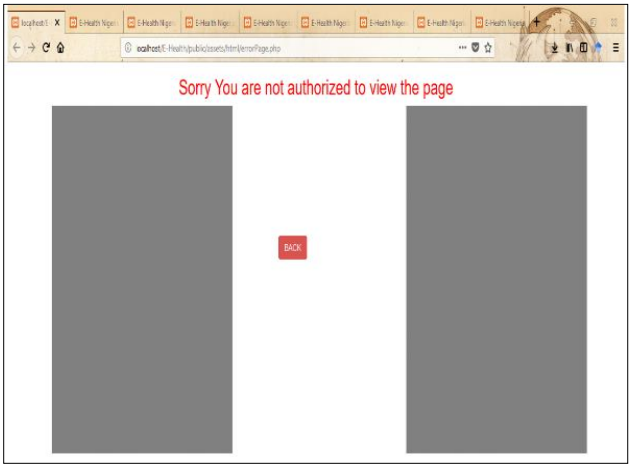


Figure 8: Access denied due to wrong secret key

REFERENCES

AMIA 2019. AMIA Health Informatics Core Competencies Retrieved online at https://www.amia.org/AMIA-Health-Informatics-Core-Competencies-for-CAHIIM.PDFom on 14th March 2019

Archana, P. S., Ambily, O. 2016. Visual cryptography in internet voting for extended security. International Journal of Engineering Research and General Science, 4(2), 365–368.

Basavegowda, R., Seenappa, S. 2013, Electronic Medical Report Security Using Visual Secret Sharing Scheme, 15th IEEE International Conference on Computer Modelling and Simulation. 78-83

Cushman, R., Froomkin, A. M., Cava, A., Abril, P., Goodman, K. W. 2010. Ethical, legal and social issues for personal health records and applications. Journal of Biomedical Informatics, 43(5), S51–S55. https://doi.org/10.1016/j.jbi.2010.05.003

D'Arco, P., De Prisco, R. 2016. Visual Cryptography Models, Issues, Applications and New Directions. In Innovative Security Solutions for Information Technology and Communications (pp. 20–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-47238-6

Fernández-alemán, J. L., Señor, I. C., Ángel, P., Lozoya, O., Toval, A. 2015. Security and privacy in electronic health records: A systematic literature review. Journal of Biomedical Informatics, 46(3), 541–562. https://doi.org/10.1016/j.jbi.2012.12.003

Harman, L. B., Flite, C. A., Bond, K. 2012. Electronic Health Records: Privacy, Confidentiality, and Security. American Medical Association Journal of Ethics, 14(9), 712–719. Retrieved from www.virtualmentor.org

Hawkes, L. W., Yasinsac, A., Cline, C. 2000. An Application of Visual Cryptography to Financial Documents. Tallahassee.

Kester, Q. 2013. A Visual Cryptographic Encryption Technique for Securing Medical Images. International Journal of Emerging Technology and Advanced Engineering, 3(6), 496–500.

Li, F., Zou, X., Liu, P., Chen, J. Y. 2011. New threats to health data privacy. In BMC Bioinformatics, 12, pp. 1–7.

Naor, M., Shamir, A. 1995. Visual Cryptography. Advances in Cryptology - EUROCRYPT'94, 1–12.

Okkali, A., Sandikkaya, M. T. 2017. Preserving Privacy Using Visual Cryptography in Surveillance Systems. In (UBMK'17) 2nd International Conference on Computer Science and Engineering, pp. 1141–1144.

Onuiri, E. E., Idowu, S. A., Komolafe, O. 2015. Electronic Health Record Systems and Cyber- Security Challenges. In International Conference on African Development Issues pp. 98–105.

Qureshi, Q. A., Khan, I., Shah, B., Nawaz, A., Waseem, M., Muhammad, F. 2014. E-Health System: A Study of Components and Practices in Developing Countries. Developing Country Studies, 4(16), 119–126.

Sugiharto, B.S., Kurniasih, N., Abdullah, D., Iswara, I.B., Napitupulu, D., Laritmas, S., Mouw, E., Ahmar, A.S., Kurniawati, N., Rahim, R. 2018. Visual Cryptography with RSA Algorithm for Color Image. International Journal of Engineering & Technology, 7 (2.5), 65-68.

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

Beware of that Email Attachment, It Could Be a Death Trap!

Oluwafemi Osho[1], Immaculatta Obar[2], and Ayanfeoluwa Oluyomi[3]

[1]Federal University of Technology, Minna, Nigeria
[1]femi.osho@futminna.edu.ng

Abstract: The emergence of email brought about many benefits in the area of communication. These include, among others, faster and less costly communication of messages. Malicious online users, however, take advantage of it to perpetrate different crimes, such as propagation of malware and gaining unauthorized access to confidential information. They use emails which often contain attachments and are classified, on most email service platforms, as spam. This study investigates spam emails with attachments to determine the actual intentions of the senders. It also seeks to identify patterns in their contents, composition and frequency. Samples of spam emails containing attachments were collected. After analysis, our findings revealed that in more than half of the email samples, malware were embedded in the attachments, most of which were Trojans that downloaded other malicious files to the computers of victims, to compromise their security. It was also discovered that attackers have devised means of customizing attachments and their email address domains to evoke positive disposition toward their malicious emails. Providers of email services and the research community need to improve on mechanisms for detecting and preventing malicious emails.

***Keywords*:** Spam; Malicious emails; Email attachment; Malware

I. INTRODUCTION

The usage of electronic mail (email) has been on a high rise because of its effectiveness in communication and cost-effectiveness (Eklund, 2003). An email is a digital letter sent over the internet (Tschabitscher, 2018). In 2018, about 3.8 billion users of email were exceeded, and an estimate of 281 billion emails were being sent per day (The Radicati Group, 2018). Because of the wide usage, like any other dynamic medium, email is liable to misuse, one of which is the sending of unwanted and unsolicited mails to large and random recipients. This is also known as spam emails. Spam can be used to spread malware and unwanted data (Li & Hsieh, 2006).

The type of information on the internet in recent years has been augmented to multimedia to appeal to more users. These can also be transmitted through email, which are known as email attachments. Some of these attachments can contain malicious contents, including malwares and links to malicious websites. These malicious emails can be sent to thousands of users, spreading across networks within hours. The malware contained in them will often run on any computer and elude many of the defences set against them (by having any title and file format). These could damage crucial parts of an organization's network (Buckingham, 2015; Das & Prasad, 2014).

The intentions of these emails often include, among other things, to illegally obtain financial information and users' account details. With the increase in the number of emails sent per day, there has also been an increase in the number of malicious ones, accounting for 3.9% of the total emails sent per day in 2013 (Kaspersky, 2013; Robinson, 2015).

Malicious emails often bypass most automated detection systems, for instance, evade reputation-based web defences that have been developed to identify suspicious URLs (Robinson, 2015). Nevertheless, with continuous improvement in the capacity of existing security mechanisms put in place by email service providers, most malicious emails are easily classified as spam.

Though there are many forms spam mails can take, they all have an underlying characteristics in terms of content, structure and distribution approaches. By analysing the content and structure of the emails, their patterns are better understood, which helps to build a sturdy email security intelligence. For email users, understanding the contents, structure and patterns of the malicious emails could reduce the likelihood of falling victim to attacks such as phishing, scamming, and malware infection.

This study aims at investigating spam emails containing attachments to ascertain the malicious goals of the senders and identify any patterns in their contents, composition, and frequency.

The study is significant in many ways. First, it demonstrates why online users should be wary of certain emails that come with attachments. Identifying malicious intents of spam emails with malicious attachments, which often appear benign, could help increase awareness of unsuspecting or novice email users of the malicious activities of the senders. An understanding of the different ways the emails are composed and sent could enhance the capacity of users to detect these emails.

The rest of the study is organized as follows: section II reviews related concepts. The methodology used is described in section III. In section IV and V respectively, we present and discuss the findings. The study is concluded in section VI.

II. LITERATURE REVIEW

A. Email

Electronic mail (abbreviated as e-mail, email, E-Mail, etc.) is a computerized system for sending electronic messages from one computer to one or more recipients. It is the direct rendition of the conventional mail system. The size ranges from 1kb to many megabytes, bigger than other communication systems. Emails have become a popular and powerful mode of communication. It provides its users with low cost messages to a large number of people. It has other advantages: it is instantaneous, easy to use and cost effective (Hershkop, 2006).

Emails follow the normal process of internet data transfer. The TCP breaks down the message into packets, the IP transfers the packets to their destination and the TCP reconstruct the message in the destination. Email attachments cannot be handled directly by the internet, so they are encoded in an encoding scheme (popular ones are MIME and Uuenode) by the sender's application and decoded by the receiver's application. Gateways translate email formats from a network to another for proper delivery (Ali, Elazim, & Abdelaziz, 2017; Gookin, 2009).

The actual structure of an email consists of several elements. First, there is the header information. This is the electronic equivalent of all the information on the outside of an envelope, plus a bit more. Then there is the body of the message. This is the functional equivalent of a letter inside the envelope. And then there are possible attachments, which can come in many different forms (Angell & Heslop, 1994).

B. Malicious Emails

Email can be used as a vector for malware. Typically, the malware is either embedded in email attachments or a link to malicious websites is contained in the email. Once clicked, the user is redirected to a malicious website which automatically unleash the malware attack on the victim's computer system (Ryan & Kamachi, 2014; Tran, Alazab, & Broadhurst, 2013). The most common payloads include viruses, Trojan horse, worms, spywares, adware (Rouse, 2012). Our study focuses only on emails with malicious attachments.

Another category of malicious electronic messages are emails that contain links to websites where victims are required to perform actions purported to be beneficial to them. An example of this type of email is phishing email.

Phishing is a crime that directly targets and deceive the user into divulging confidential and personal information. In most of this scam, a relationship is built by the scammer with the user and once trust is established, the scammer starts to exploit the user (Humphrey, 2011; Oluyomi, Osho, & Shuaib, 2018).

In the first quarter of 2013, spam e-mail constituted, on average, 66.5% of all e-mails sent. Of these, 3.3% contained malicious attachments. About 6 billion e-mails with malicious attachment were sent daily in this period (Blanzieri & Bryl, 2008; The Radicati Group, 2015). This is made easier by the use of botnets by attackers. Botnets are networks of exploited computers controlled by a 'botmaster' who can lease it to attackers. An estimates of 85% of world's spam e-mails are sent from botnets (Broadhurst et al., 2013).

Attackers also make use of other assistive technologies. Generally, the richer the messaging media, the more opportunity there is to camouflage malicious content within the rich content. A link to a malicious software download site may be camouflaged as a link to a news article, to a blog, or to what appears to be a legitimate information site. It is actually easy to disguise the true nature of an electronic link by displaying one that appears to be quite innocuous. The problem is compounded with other elements; for example, there are URL abbreviation technologies, such as bitly or tinyurl, that compress long or ungainly links to something a bit more manageable (or that fit within the character limitations of some messaging services) (Ryan & Kamachi, 2014).

III. METHODOLOGY

To ascertain the real intents of senders of spam emails with malicious attachments and identify any patterns in the content, composition and frequency of the emails, we collected and analysed some spam emails with attachments.

A. Collection of Emails

To gather the emails, we created two email accounts, one each on Gmail and Yahoo. These accounts were posted on different forums to make it as visible as possible to spammers, because most of these spammers harvest emails from such forums.

We also contacted friends and colleagues to forward such emails to any of the two accounts. Fifteen people eventually participated. The collection was carried out between March and August, 2016.

It was noticed that once Gmail and Yahoo discovered any emails to be malicious, downloading and forwarding options were disabled. One way round this is to download the base64-encoded version of the email. This involves the following steps:

- Open the email (Gmail or Yahoo!) containing the malicious attachment.
- Save the email as pdf (i.e. print it to pdf), using the name of the attachment.
- For Gmail, click the right drop-down arrow, and select 'Show original.' For Yahoo mail, click 'More', and select 'View Full Header' from the drop-down menu. (In current Gmail app, the drop-down arrow has been replaced with an icon with three vertical dots. In the case of Yahoo mail, 'More' has been replaced with a three-horizontal-dot icon. And when clicked, 'View raw message' is selected.)
- A popup will appear with what appears to be garbled text. This is actually your email and attachment as a text file.
- Copy the entire text to clipboard. Paste it in a text file (preferably a Notepad), and save it with any generic name (preferably the name of the attachment file) ending in *.txt.

For those who participated in forwarding malicious emails, they were requested to follow the same steps and place the files (i.e. all saved emails using the above steps) in a folder, zip it, and forward them to us.

B. Analysis of Emails

A total of 68 emails were received. To determine if the attachments contained malware, we used VirusTotal, a free online tool, containing more than 50 different virus scanners, for analysing suspicious files and URLs (Chronicle, n.d.). An email was considered malicious once at least one of the scanners reported it as such. After analysis, 39 (57.4%) of the emails were found to be malicious. 35 (89.7%) of these contained attachments that were .zip files, while the remaining 4 (10.3%) were DOCM files.

To identify any patterns in the emails, we observed the sending date, number of similar emails received daily, and how the attachments were titled.

IV. FINDINGS

A. Purported Intent of the Emails

Based on the purported intent of the emails, as deduced from the message content, we found out that slightly more than half of the emails (51.3%) were purported to be transaction-based. An example of this type is displayed in Figure 1.

Figure 1. Sample transaction-based email

It was also observed that 3 (7.7%) of the emails had only attachments. Table 1 presents the different categories of the emails.

Table 1 Categories of malicious emails

Category	Frequency	Percent (%)
Transaction	20	51.3
Proposal	1	2.6
Report	5	12.8
Invitation	1	2.6
No texts	3	7.7
Miscellaneous	9	23.1
Total	39	100.1

B. Actual Intents of the Emails

Findings revealed that all the malicious attachments were Trojans. Table 2 contains the different categories of Trojans.

Most of the malware (79.5%) were Trojans that download malicious files to victims' computers to compromise their security. The downloaded files could, among other things, steal information from and deliver ransomware to the systems of the email users. The main families of Trojans identified in this category include JS:Trojan.JS.Downloader, Generic.JS.Downloader, Trojan.GenericKD, and Trojan.JS.Agent.

Another family of Trojans identified was the W97M.Downloader. This accounted for 10.3% of the malware discovered. The malware were malicious Word macro Trojans also capable of downloading other malicious malware.

Other categories were ransomware, redirecting, and root-access Trojans. In these categories, we identified the JS.TeslaCrypt, JS:Trojan.Script, and HEUR:Trojan-Downloader.Script Trojans respectively.

Table 2. Purpose of malware

Trojan Type	Function	Frequency	Percent (%)
Trojan Downloader	Download malicious files	31	79.5
Word Macro Trojan	Malicious Word macro that downloads additional malware	4	10.3
Ransomware Trojan	Encrypts files on the machine and demands ransom payment	1	2.6
Redirect Trojan	Redirects user to a malicious site	2	5.1
Root Access Trojan	Attempts to gain 'root' access to user's computer	1	2.6

C. Email Patterns

A careful observation of the received emails revealed some patterns:

- **Similar emails, same day**: Attackers often sent similar emails, within the same day. In many cases, the title and/or contents are similar. However, the sender address would usually be different. The malicious attachments were also given different names. From the sample malicious emails used in the study, there were 6 instances were at least two emails sent on the same day had such pattern. For example, the contents of two emails received, on May 4, by the same recipient are displayed in Figures 2 and 3.
- **No texts, only attachment**: Some of the emails were sent without any texts but only attachments. In the study, 3 of such mails were received. Figure 4 presents one of such mail.
- **Attachments with receiver's name**: Attackers would often customize the malicious attachments using titles that include the intended recipient's name. In this study, 24 (61.5%) of the malicious emails had their attachments saved with the name of the receiver. A sample is presented in Figure 5.
- **Sender's email domain similar to Recipient's:** Surprisingly, some attackers have devised means of using email addresses that have similar domain as those of the recipients. In our study, we noticed one such emails (Figure 6). In that mail, the sender's email address was qwer3@futminna.edu.ng while that of the receiver was femi.osho@futminna.edu.ng.

V. DISCUSSION

The aim of the study was two-fold. The first was to investigate attachment-containing spam emails, often disguised as benign, which contains malicious contents. We also sought to identify any patterns in the contents, composition, and frequency of such emails.

Our findings revealed that more than half (57.4%) of the email samples that were collected and analysed contained malware. This percentage of malicious emails is close to that of the Botnet dataset reported in (Alazab & Broadhurst, 2016). All the malicious attachments, after analysis, contained Trojan malware, most especially those that, once they infect a victim's system, download other malicious files, and were mostly stored in zipped files.

From the message of those malicious samples, most contained texts that suggested some form of transaction had taken or was to take place between the sender and the recipient. The intention of the senders of such emails obviously was to disguise their real intents, hoping that unsuspecting or novice email users would fall prey. In very few cases, the emails had no textual contents, but only attachments.

In addition to masking their actual intents by making the email message as benign as possible, we identified some other tactics used by the attackers. A sender may customize the title of the attachments partly using the name of the intended recipient. In one case, the domain name in the sender's email address was the same as that of the recipient. We suggest that these measures were targeted at subduing any form of suspicion from the receiver. In other words, by customizing the attachment or email address, the sender hoped to elicit a less suspicion disposition to the benign-looking email, thereby earning the trust of the intended receiver. A receiver who receives an email containing attachment saved with the receiver's name could be expected to feel less suspicious of the content of such attachment than would have been felt if such attachment had a different name.

VI. CONCLUSION

Attackers exploit numerous vectors to gain unauthorized access to users' devices to steal information and perpetrate other types of cyberattacks. One of such vectors is email. Our study focused on analysing spam emails with attachments to identify the malicious intents of the senders. Although our sample size was not large, our analysis of the email samples revealed interesting results. Our findings revealed that attackers would often conceal malware in email attachments, with the emails crafted to look benign. They also employed other measures, such as customization of attachments and email address domains, to deceive intended recipients.

The findings of this study underscore the need for more improvement in existing technical security measures used to detect malicious emails. While significant progress has been made in classifying them as spam, further research efforts could lead to the development of more effective solutions to detect them as containing malware, and, as such, prevent them from being delivered.

How attackers customize their email address domains to be the same as those of recipients need further investigation. Future studies are needed to identify the tools and techniques used, in order to develop appropriate countermeasures. Also, the rationale behind attackers sending similar emails from different sources on the same day needs to be explored further.

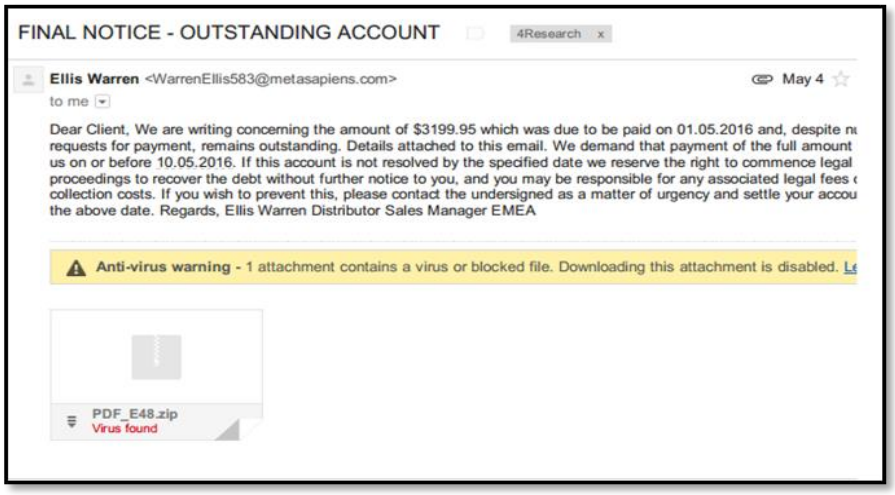


Figure 2. Similar mail, same day (a)

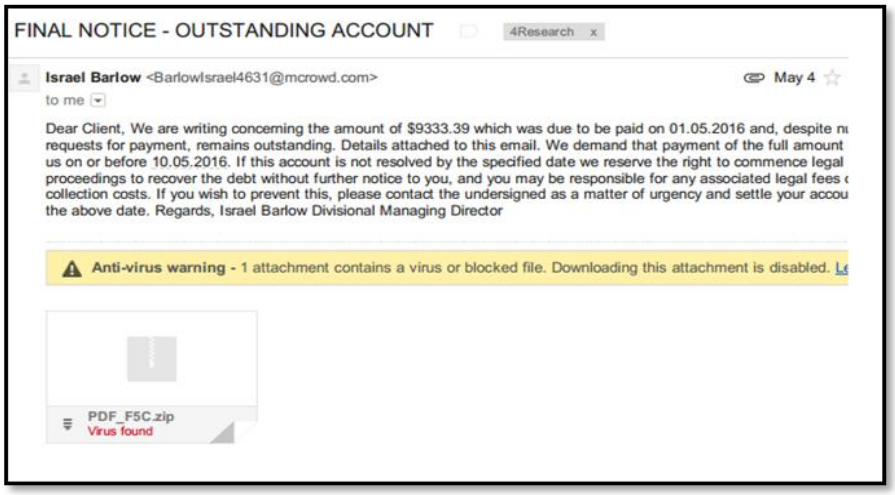


Figure 3. Similar mail, same day (b)

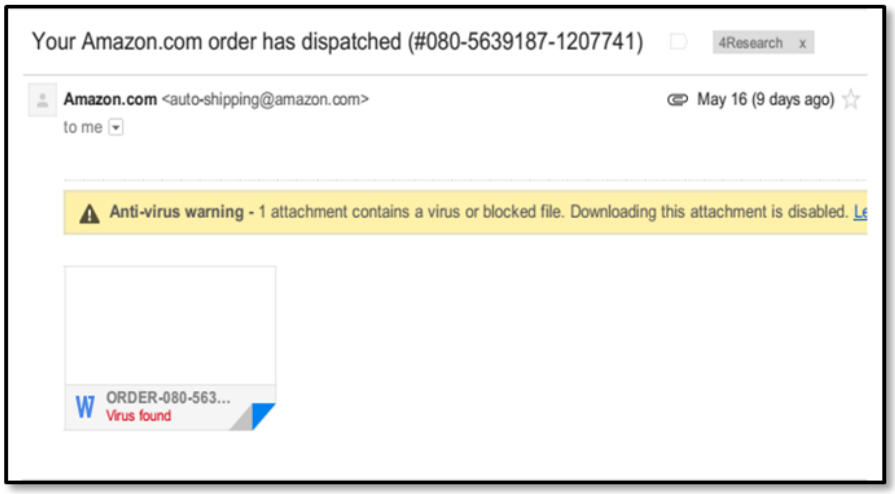


Figure 4. E-mail without texts

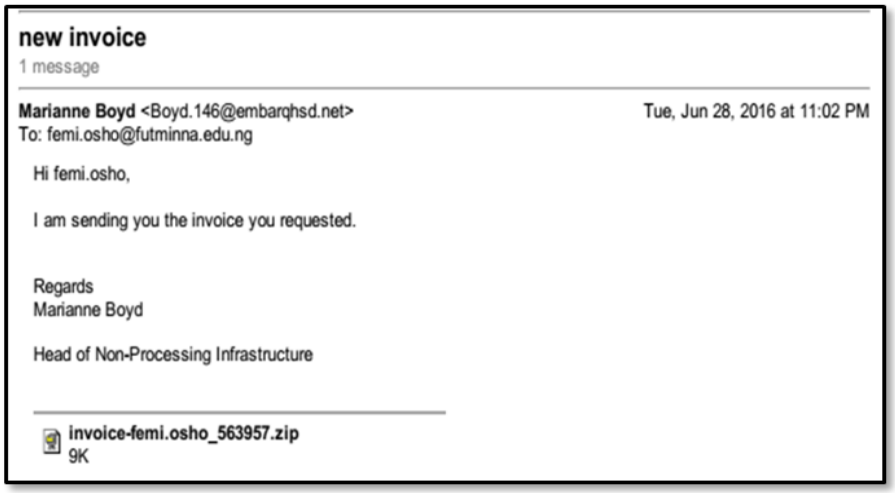


Figure 5. E-mail with attachment customized with recipient's name

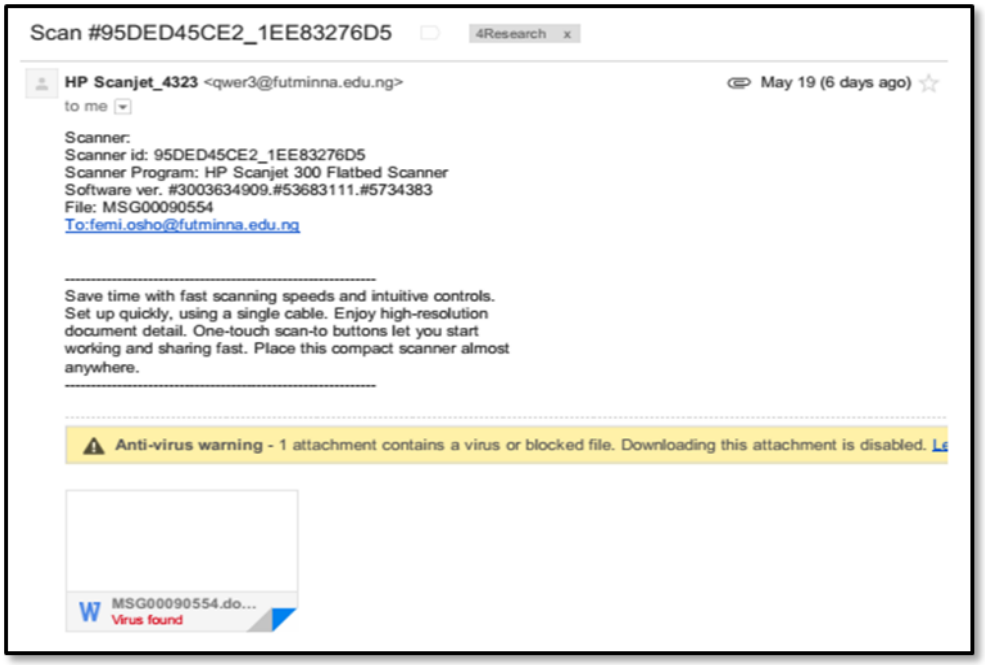


Figure 6. Similar sender and recipient email address domains

REFERENCES

Alazab, M., & Broadhurst, R. (2016). Spam and Criminal Activity. *Trends and Issues in Crime and Criminal Justice (Australian Institute of Criminology)*, 1–14.

Ali, E. S., Elazim, S. M. A., & Abdelaziz, A. Y. (2017). Ant Lion Optimization Algorithm for optimal location and sizing of renewable distributed generations. In *Renewable Energy* (Vol. 101, pp. 1311–1324). Elsevier Ltd. https://doi.org/10.1016/j.renene.2016.09.023

Angell, D. F., & Heslop, B. (1994). *Elements of e-mail style: Communicate effectively via electronic mail*. Addison-Wesley Longman Publishing Co., Inc.

Blanzieri, E., & Bryl, A. (2008). A survey of learning-based techniques of email spam filtering. *Artificial Intelligence Review*, *29*(1), 63–92.

Broadhurst, R., Grabosky, P., Alazab, M., Bouhours, B., Chon, S., & Da, C. (2013). *Crime in cyberspace: offenders and the role of organized crime groups*.

Buckingham, A. (2015). Beware: Malicious email attachments are making a return | ITProPortal. Retrieved January 13, 2019, from https://www.itproportal.com/2015/09/07/beware-malicious-email-attachments-making-return/

Chronicle. (n.d.). VirusTotal. Retrieved from https://www.virustotal.com/#/home/upload

Das, M., & Prasad, V. (2014). Analysis of an Image Spam in Email Based on Content Analysis. *International Journal on Natural Language Computing (IJNLC)*, *3*(3), 129–140.

Eklund, C. (2003). Spam – from nuisance to Internet infestation. In R. Kantola (Ed.), *Peer to Peer and SPAM in the Internet* (pp. 126–134). Helsinki University of Technology.

Gookin, D. (2009). *Laptops for dummies*. John Wiley & Sons.

Hershkop, S. (2006). *Behavior-based email analysis with application to spam detection*. Columbia University.

Humphrey, J. (2011). Spokane Woman Falls Victim To Nigerian Boyfriend Scam - KXLY. Retrieved January 13, 2019, from https://www.kxly.com/news/spokane-woman-falls-victim-to-nigerian-boyfriend-scam/177210816

Kaspersky. (2013). Spam Statistics Report Q3-2013. Retrieved January 28, 2019, from https://usa.kaspersky.com/resource-center/threats/spam-statistics-report-q3-2013#.Vz3O2PkrLIU

Li, F., & Hsieh, M.-H. (2006). An Empirical Study of Clustering Behavior of Spammers and Group-based Anti-Spam Strategies. In *CEAS* (Vol. 2006, pp. 21–28). https://doi.org/http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.210.4820

Oluyomi, A., Osho, O., & Shuaib, M. (2018). Evaluation of Classification Algorithms for Phishing URL Detection. In *2nd International Conference on Information and Communication Technology and Its Applications* (pp. 243–249).

Robinson, R. M. (2015). Malicious Attachments Make a Comeback as Top Attack vector. Retrieved from https://securityintelligence.com/malicious-attachments-make-a-comeback-as-top-attack-vector/

Rouse, M. (2012). Attack vector. Retrieved January 13, 2019, from https://searchsecurity.techtarget.com/definition/attack-vector

Ryan, J., & Kamachi, C. (2014). *Detecting and Combating Malicious Email*. Syngress.

The Radicati Group. (2015). *Email Statistics Report, 2015-2019*. Palo Alto, California.

The Radicati Group. (2018). *Email Statistics Report, 2018-2022*.

Tran, K., Alazab, M., & Broadhurst, R. (2013). Towards a Feature Rich Model for Predicting Spam Emails containing Malicious Attachments and URLs. In *Eleventh Australasian Data Mining Conference (AusDM 2013)*. Canberra.

Tschabitscher, H. (2018). What Is Email? Retrieved January 28, 2019, from https://www.lifewire.com/electronic-mail-overview-1164107

CSEAN
CYBER SECURITY EXPERTS ASSOCIATION OF NIGERIA

Cyber Secure Nigeria 2019 Conference

April 9 – 11, 2019

Double Compression Heuristics in Digital Image Forensics

Lawrence O. Oyaniyi[1], Aderonke F. Thompson[2], Oluyomi K. Akinyokun[3], and Boniface K. Alese[4]

[1]Federal Polytechnic, Auchi, Nigeria
[2,3]Federal University of Technology, Akure, Nigeria
[4]University of Mines and technology, Tekoa, Ghana
[1]lanreoyaniyi@gmail.com, [2]afthompson@futa.edu.ng, [3]okakinyokun@futa.edu.ng, [4]bkalese@umat.edu.gh

Abstract: Image trustworthiness has become a challenge these days, resulting into lack of affordable veritable tool that ensures viable admissibility as evidence in the court of law. Existing tools in this category are characterized with high cost. Thus, in this study, we present a cost-effective approach that could assist forensic experts in establishing the reliability of image by checking discrepancy in EXIF metadata and detecting the presence of double compression artifact. Experimental set up using Discrete Cosine Transform and Quantization techniques shows that our approach has an improved outcome over some existing techniques for image authenticity check required in digital forensics investigation.

***Keywords*:** JPEG; EXIF; Metadata; Quantization; Compression; Double Quantization

I. INTRODUCTION

In today's world, photography has become almost everyone occupation. This is as a result of advance in technology which brought about availability of handy and pocket-sized digital camera at an avoidable price especially, the availability of camera with mobile phone.

Many people use these pictures for reminiscence; others use it for website decoration while some use it as evidence to support claim. The high potential of visual media and the ease in which they are captured, distributed and store is such that they are used to convey information.

Digital images have become one of the efficient and major information carriers in our modern daily lives. While the efficiency of information exchange has become of the major source of joy to people, the security and trustworthy of digital images have become a crucial issue due to the ease of malicious processing, e.g., implanting secret messages for covert communications, altering origin and content of images with popular image editing software. These malicious usages could give way to serious problems if they are taken advantage of by terrorist organizations, treated as evidence in court, or published by mass media for information dissemination.

An adage says "A picture worth a thousand words", in recent years, this trust in picture has been eroded due to availability of advance software that assist in editing image with no prior knowledge about its usage which has made image manipulation easy.

The development of a tool in verifying the authenticity and integrity of any digital image is highly important particularly those that are used as evidence in a court of law, as part of medical records, as news items or as part of financial record (Hitesh & Mohit, 2015).

Here, we proposed an innovative approach to identifying tampered image by examining the EXIF parameter and identify the effect of Double Compression on any image.

II. RELATED WORKS

Here, we review previous research work in relation to digital image tampering detection.

Qing et al (2016) presented an approach on a convolutional neural network (CNN) in detecting double JPEG Compression in digital image. Their approach classified histograms of DCT coefficient which distinguishes between single compressed and double compressed area. Doctored image localization was obtained in relation to the results of classification. The results from their experiment showed that their algorithm performed credibly in detecting double JPEG compression and in localization of doctored part especially if the quality factor of the previous compression is higher than the present quality factor.

Taimori et al (2016) developed a forensic tool that can detect double compression clue. The authors suggested a dimensionality reduction algorithm to visualize the behavior of a big database which comprises of both singly and doubly compressed images. Based on this approach, they proposed bottom-up, top-down and combined top-down/bottom-up learning strategies. Their technique was able to distinguish between single compressed images from double compressed images. It also estimates the first quantization in double compression as well as localizing tampered regions in a forgery examination. The result of their experiments shows that their approach performed wonderfully.

Simone et al (2015) presented a forensic approach where they analyze distribution of the first significant digits of the DCT coefficients. Their methodology was optimized in identifying the number of compression stages that has been applied to an image. Their method was designed to detect multiple compressions to a tune of up to four consecutive stages. Their result explained that the proposed approach extends and outperforms the previously established algorithms which were proposed for detecting double JPEG compression.

Tiziano et al (2012) developed an algorithm which can detect nonaligned double JPEG compression. They established that their scheme was able to approximate the quantization step of the DC coefficient and the grid shift of the primary JPEG compression, which gives room for detailed analysis of possibly tampered images.

In the work of Yi-Lei et al (2011), an approach which can discover traces of recompression was presented. They assumed the image under examination follows JPEG format. Authors formulated the periodic characteristics of JPEG images in spatial and transform domains. They designed an efficient detection approach that effectively detects either block-aligned or misaligned recompression using theoretical analysis. The results of their experiment demonstrated the effectiveness and validity of the proposed approach.

Also, Yujin et al (2011) presented a methodology which effectively detects Shifted Double JPEG (SD-JPEG) compression. The authors used Markovian transition probability matrix. They discovered that statistical artifacts were left by the SD-JPEG compression among the features of the JPEG 2-D arrays. They generated difference JPEG 2-D arrays in four directions. These features were used in order to enhance them. They were thresholded by a predefined threshold for reducing computational cost. They model JPEG 2-D arrays using Markovian transition probability matrix in order to utilize the second order statistics. They used elements of the transition probability matrices as features for SD-JPEG compression detection. In this research work, Support Vector Machine was used to classify them. Experimental results demonstrates that their approach is efficient.

Dong et al (2011) model the distribution of the mode based first digits of DCT (Discrete Cosine Transform) coefficients. The researcher uses Markov transition probability matrix and utilize its stationary distribution as features for double compression detection. Their experimental results show the effectiveness of the proposed method. They also presented comparison for validation purpose which shows that their approach has an improvement over second order statistical model.

III. JPEG Compression

A. JPEG Compression Algorithm.

In the modern days, JPEG is the most widely and commonly used compression algorithm. It is a standard that comprises of three basic steps:

- Discrete Cosine Transform (DCT): Sample images are grouped into 8 x 8 blocks in raster scan order from left to right as well as top to bottom and shifted from unsigned to signed integers, then follow by computation of DCT of each blocks. Let $f(x, y)$ represent an 8x8 image block. The DCT of each block takes the form

$$F(w_x, w_{xy}) = \frac{1}{4} c(w_x) c(w_y) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x, y) \cos \frac{(2x+1)\omega_x \pi}{16} \cos \frac{(2y+1)\omega_y \pi}{16}$$

$$\omega_x, \omega_y = 0, \ldots, 7, \qquad (1)$$

where $c(\omega) = {}^{1}/_{\sqrt{2}}$, for $\omega = 0$, and $c(\omega) = 1$ otherwise.

- Quantization: The obtained DCT coefficients are uniformly quantized. The purpose of quantization is to obtain a desired visual quality. Quantization is a point-wise operation where each DCT coefficient is divided by quantization step and rounded to the nearest integer:

$$F_q(w_x, w_{xy}) = \left\lfloor \frac{F(w_x, w_{xy})}{s(w_x, w_y)} + \frac{1}{2} \right\rfloor, \quad \omega_x, \omega_y = 0, \ldots, 7, \tag{2}$$

Where $s(w_x, w_{xy})$ is a frequency depended quantization step. The relationship between the JPEG quality, Q and the quantization steps $s(w_x, w_{xy})$ takes the form:

$$s(w_x, w_y) = \begin{cases} \max\left(\left\lfloor \frac{200 - 2Q}{100} C(w_x, w_{xy}) + \frac{1}{2} \right\rfloor, 1\right) & , 50 \leq Q \leq 100 \\ \left\lfloor \frac{50}{Q} C(w_x, w_y) + \frac{1}{2} \right\rfloor & , 0 < Q < 50 \end{cases} \tag{3}$$

where $C(w_x, w_y)$ are experimentally determined to be:

$$C(w_x, w_y) = \begin{matrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{matrix} \tag{4}$$

- Entropy Encoding: Entropy encoding involves lossless entropy compression which transforms the quantized DCT coefficients into a stream of compressed data. Huffman coding is the most frequently used procedure. Also, arithmetic coding can also be used.

B. Double Quantization

Double quantization has been discussed by Lin et al (2009), Junfeng et al (2006) and Alin (2004). An image is said to undergo double compression or double quantize if it is compressed first with a quality factor and recompressed with another quality factor where the first quality factor is different from the subsequent quality factor.

C. Effect of Double Compression

Here, we provide the analysis of Double quantization effect on an image. Below, we will investigate how histogram of an image changes after double quantization. Let us consider a distinct 1-D signal $x[t]$. Quantization is described as a point-wise operation denoted by one-parameter functions:

$$q_a(u) = \left\lfloor \frac{u}{a} \right\rfloor, \tag{5}$$

where a denotes quantization step (a positive integer), and u represents a value within x[t]. De-quantization is the inverse operation where the quantized values are turned back to their original form: $q_a^{-1}(u) = au$. While $q_a(u)$ is not reversible, de-quantization is not a reverse function of quantization. Also, double quantization is described as a point-wise operation denoted by two-parameter functions:

$$q_{ab}(u) = \left\lfloor \left\lfloor \frac{u}{b} \right\rfloor \frac{b}{a} \right\rfloor, \tag{6}$$

where a and b denote the quantization steps. Double quantization is performed with the following sequence of three steps: quantization with quantization step b, followed by de-quantization with the same quantization step b, then quantized with step a.

In order to explain the effect of double quantization on histogram, the original signal x[t] is quantized in two different ways, and the histograms of single and the two quantized signals are shown below. Figure 1 show the histogram of the normally distributed original signal while figure 2 depicts the histogram of single quantized signal with step 2 and figure 3 shows the histogram of double quantized signal with steps 3 and quantized with 2. It was observed that some bins are empty in the histogram of the double quantized signal.

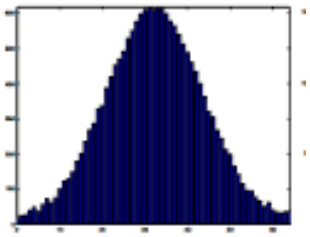


Figure 1: Histogram of the normally distributed signal

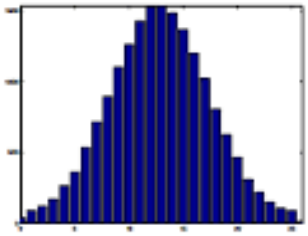


Figure 2: Histogram of single quantized signal

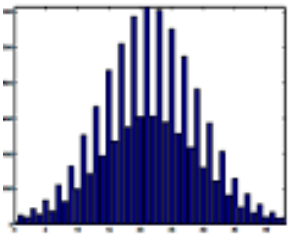


Figure 3: Histogram of double quantized signal

To understand why periodic artifacts are present in the histogram of double quantized signal, we will look at the dependency between original and double quantize signal histogram. Let $x_a\lfloor t \rfloor$ denotes the signal of singly quantized signal, also, let $h(u)$ and $h_a(v)$ denotes histogram of original and quantized signals and let h(u) represents the number of samples of $x[t]$ that take the value u. Therefore, $x_a\lfloor t \rfloor = q_a(x\lfloor t \rfloor)$ is the relationship between quantized and original signal. $q_a(.)$ is a many-to-one function, therefore, many values from $x[t]$ will correlate onto the same value in $x_a\lfloor t \rfloor$. Let v denote a value in $x_a\lfloor t \rfloor$, then the values in $x[t]$ that relate to it are in the interval $[av, av + (a-1)]$. The relationship between $h(u)$h(u) and $h_a(v)$is given as:

$$h_a(v) = \sum_{k=0}^{a-1} h(av + k) \tag{7}$$

Also, let examine another scenario in which a double quantized signal, represented by $x_{ab}[t]$, and its histogram represented by $h_{ab}(v)$. The correlation between the double quantized and original signals is given by: $x_{ab}[t] = q_{ab}(x[t])$. In disparity with case of single quantization, the number of bins of h that contribute to a bin of h_{ab} rely on the double quantized bin value. Let v denote a value in the range of $x_{ab}[t]$. Let U_{min} and u_{max} denotes the smallest and largest values in the range of x[t] that map to v, then it satisfy the equation:

$$\left\lfloor \left\lfloor \frac{u}{b} \right\rfloor \frac{b}{a} \right\rfloor = v. \tag{8}$$

Hence,

$$v \le \left\lfloor \frac{u}{b} \right\rfloor \frac{b}{a} < v + 1 \Leftrightarrow \frac{a}{b} v \le \left\lfloor \frac{u}{b} \right\rfloor < \frac{a}{b}(v+1). \tag{9}$$

We can rewrite equation (9) to include integers only using ceiling function since $\left\lfloor u/b \right\rfloor$ is an integer:

$$\left\lceil \frac{a}{b} v \right\rceil \le \left\lfloor \frac{u}{b} \right\rfloor \le \left\lceil \frac{a}{b}(v+1) \right\rceil - 1. \tag{10}$$

From Equation (10) it can be deduced that u_{min} must satisfy:

$$\left\lfloor \frac{u_{min}}{b} \right\rfloor = \left\lceil \frac{a}{b} v \right\rceil \Leftrightarrow u_{min} = \left\lceil \frac{a}{b} v \right\rceil b, \tag{11}$$

While u_{max} must satisfy:

$$\left\lfloor \frac{u_{max}}{b} \right\rfloor = \left\lceil \frac{a}{b}(v+1) \right\rceil - 1 \Leftrightarrow u_{max} = \left(\left\lceil \frac{a}{b}(v+1) \right\rceil - 1 \right) b + (b-1)$$
$$= \left\lceil \frac{a}{b}((v+1) \right\rceil b - 1 \tag{12}$$

Since double quantization is a monotonically increasing function, it follows that all the values between u_{min} and u_{max} will map to v through double quantization. The relationship between the original and double quantized histograms takes the form:

$$h_{ab}(v) = \sum_{u=min}^{umax} h(u). \tag{13}$$

The number of original bins contributing to bin v in the double quantized histogram depends on v, and let $n(v)$ denote this number. Using Equation (2.11) and Equation (2.12) the value of $n(v)$ is given by

$$n(v) = u_{maz} - u_{min} + 1 = b\left(\left\lceil \frac{a}{b}(v+1) \right\rceil - \left\lceil \frac{a}{b} v \right\rceil \right). \tag{14}$$

$n(v)$ is a periodic function with period b, i.e., $n(v) = n(v + kb)$, where k is any integer. This periodicity is the reason that periodic artifacts appear in histograms of double quantized signals.

From Equation 14, double quantization artifacts show above (figure 3) can be explained. Consider first the case of double quantization using steps b = 3 followed by a = 2. The number of original histogram bins contributing to the double quantized histogram bins of the form $(3k + 2)$, with k integer, is given by:

$$n(3k + 2) = 3\left(\left\lceil \frac{2}{3} 3k + 3) \right\rceil - \left\lceil \frac{2}{3}(3k+2) \right\rceil \right)$$
$$= 3\left(\lceil 2k + 2 \rceil - \left\lceil 2k - \tfrac{4}{3} \right\rceil \right) \quad = \left(3(2k + 2 - 2k - \left\lceil \tfrac{4}{3} \right\rceil = \right) 0 \tag{15}$$

This is consistent with the observation that every $(3k + 2)^{nd}$ (k integer) bin of the double quantized histogram is empty. In the other example of double quantization, when steps b= 2 followed by a = 3 are employed in Equation (2.14), it can be shown that $n(2k) = 4$ and $n(2k + 1) = 2$, with k integer. Again this is consistent.

1V. THE PROPOSED METHODS

A. EXIF Metadata Analysis

We extracted EXIF metadata from the image which can be found in the JPEG header. JPEG header file stores different useful information about the image and camera. According to EXIF standard, the following are the standard image file directories where metadata are organized: Primary, Thumbnail, Exif, GPS and Interoperability.

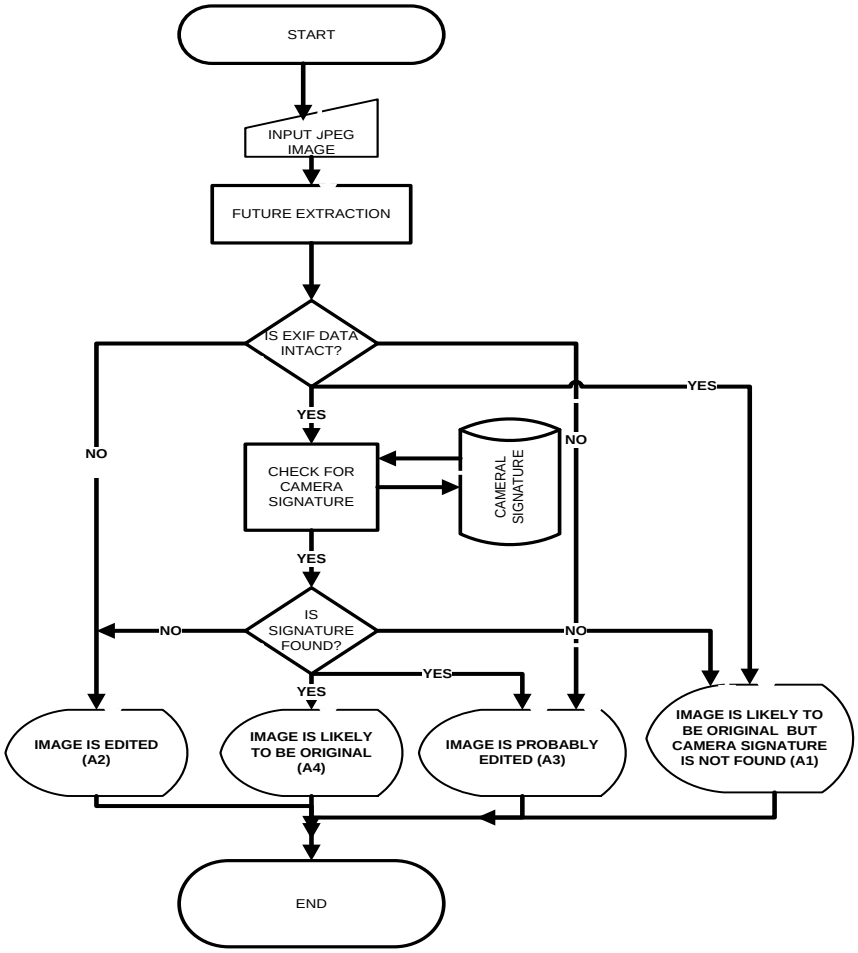


Figure 4: EXIF Analysis Flowchart

For any given image, we extract the EXIF parameters and check for discrepancies in the date of creation, date of modification as well as device software. If there are traces of photo editing software and discrepancies in the EXIF parameters or either, then the software report that the image is edited, otherwise, it is not edited. Also, we extract camera signature and compare with over 3000 known camera signature as contained in JPEGSNOOP database with a view of reporting the device make and model that was used to capture the image. The report from this phase is categorized into four:

- Image is likely to be original but no camera signature is found A1
- Image is edited A2
- Image is probably edited A3 and
- Image is likely to be original A4

where A1, A2, A3, A4 are variables that stored the result of the analysis that are later used for further analysis.

B. Detecting the Effect of Double Quantization

At this phase, our effort is to detect the presence and the effect of double JPEG compression in any given image. The image under investigation was decompressed to extract Discrete Cosine Transform (DCT) coefficients and the

quantization table from the JPEG file. An image is said to undergo double compression or double quantized if it is compressed in the first instance with a quality factor Q_1 and then compressed again with another quality factor Q_2.

Let us consider a compressed image with quality factor Q and denote the coefficients of DCT frequency k with $C_k\ (x, y)$; and $a(k)$ denote the step obtained from the quantization table which can be obtained from the JPEG file. Thereafter, DCT coefficients histogram was computed as in equation 16 – 18.

$$h_{ab}(v) = \sum_{u=min}^{umax} h(u). \tag{16}$$

$$u_{max} = \left(\left\lceil \frac{a}{b}(v+1) \right\rceil - 1\right) b + (b-1) = \left\lceil \frac{a}{b}(v+1) \right\rceil b - 1 \tag{17}$$

$$u_{min} = \left\lceil \frac{a}{b} v \right\rceil b \tag{18}$$

In order to obtain the peaks values of the DCT coefficients, the histogram h_{ab} and u_{ab} were subjected to Fourier transform to yield equations 4 and 5:

$$H_{ab}(\omega) = h_{ab}(v) \tag{19}$$

$$U_{ab}(\omega) = u_{ab}(v) \tag{20}$$

Furthermore, double quantization peaks which are located in $U_{ab}(\omega)$ were selected by selecting values above a predetermined threshold using equation 6.

$$P = \{\omega | U_{ab} \geq E_t\} \tag{21}$$

where: P denotes set of peak located in the frequency domain;

E_t denotes an empirically chosen threshold.

In addition, decaying trend that may be present in the Fourier transform of the DCT coefficient were removed using a two-parameter generalized Laplace model as presented in equation 22.

$$L(\omega; \alpha, \beta) = e^{-|\omega|^{\alpha}/\beta} \tag{22}$$

This was computed by removing the generalized Laplace model as stated above, $L(\omega; \beta^*, \beta^*)$, from $H_{ab}(\omega)$. Then, we avoid negative value by taking its square. This is formally given as expressed in equation 23.

$$D_{ab}(\omega) = (H_{ab}(\omega) - L(\omega; \beta^*, \beta^*))^2 \tag{23}$$

Finally, periodicity measure was computed from $D_{ab}(\omega)$ (in equation 8) as a weighted average at the peak location.

$$M(h_{ab}; a, b) = \frac{\sum_{\omega} W(\omega) D_{ab}(\omega)}{\sum_{w} W(\omega)}, \tag{24}$$

where: the weights $W(\omega)$ are given by:

$$w(\omega) = \begin{cases} U_{ab} & \omega \epsilon P \\ 0 & \omega \notin P \end{cases} \tag{25}$$

Thus, periodicity measures in equation (24) rely on the histogram of the DCT coefficients, h_{ab} and the quantization steps, a and b. The first quantization step, denoted with b, is unknown, and a measure which is independent of b is gotten by taking the maximum overall detectable values of b:

$$M\left(h_{ab;a}\right) = \max_{b \in B_a} M\left(h_{ab,a,b}\right), \tag{26}$$

where all the set of all detectable steps b is given by: $B_a = \{b | a/b / \notin \mathbb{N}, b \leq b_{max}\}$,

If the computed measure, $M(h,a)$ is above a predetermined threshold, it is presumed that double quantization periodic artifacts are present in the DCT coefficient histogram, then DCT coefficients, $C_k(x,y)$ are categorized as double quantized, otherwise, they are considered single quantized. The threshold $T(k,a(k))$ is empirically determined from single quantized DCT coefficients so that a desired false positive rate is obtained.

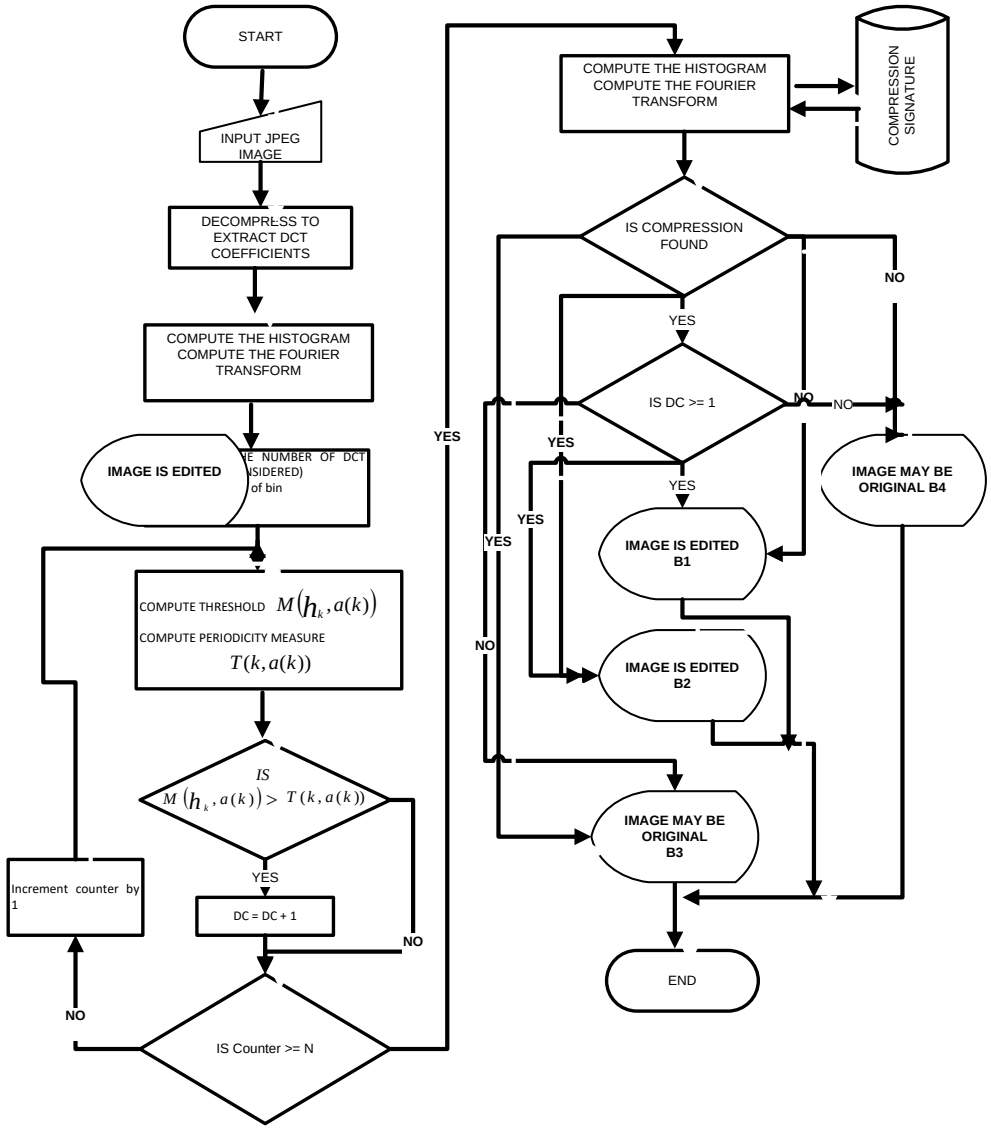


Figure 5: Flowchart for Detection of Double JPEG Compression

The result from this phase is also categorized into four:

- The image is edited B1
- The image is edited B2
- The image may be original B3 and
- The image is likely to be original B4

where B1, B2, B3, B4 denotes variable that stores the result of the analysis from this phase which are later used for further analysis.

C. Combined Analysis

The results collected from the two analyses as stated above are passed into logic circuit for final output.

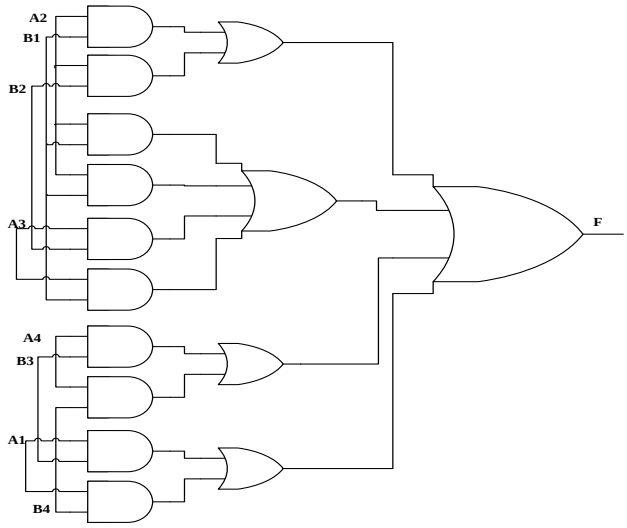


Figure: 6 Logic Circuit showing the hybrid approach

The output from the logic circuit is further grouped into four classes:
Class A: This Image has been tampered with
Class B: There is high probability that this image is tampered
Class C: There is high probability that this image Original
Class D: Unknown if this image has been tampered with or it is original

Conditions for Class A

Combining the results from the two phases as depicted by the logic circuit, the below is the condition for class 1 using control structure.

```
If (A2= 1 && B1 =1) OR (A2 =1 && B2 =1)
        {
                Print (This Image has been tampered with)
        }
```

Conditions for Class B

Also, the condition for Class 2 using control structure is stated below

```
If (((A2= 1 && B1 =1) OR (A2 =1 && B2 =1))
  OR (A3 = 1 && B2=1) OR (A3 =1 && B1=1 )))
        {
                Print (There is high probability that this image is tampered)
        }
```

Conditions for Class C

The condition for class 3 is given as:

```
If ((A4 = 1 && B3 =1) OR (A4 =1 && B4 = 1))
        {
                Print (There is high probability that this image Original)
        }
```

Conditions for Class D

```
If ((A1 = 1 && B3 =1) OR (A1 =1 && B4 = 1))
        {
                Print (Unknown if this image has been tampered with or it is original)
        }
```

V. RESULTS

We collected over 2000 images from five different image capture devices such as Nikon Digital Camera, Samsung Digital Camera, Infinix Mobile Phone, Techno Spark2 mobile phone and HP Scanner. Out of these, 1000 were selected randomly i.e. 200 images from each device. Out of the remaining 1000 images, 200 images were selected from each device and were grouped into 5different groups. Group one was labeled original image, group two was edited with photoshop software, group three was edited with Microsoft Paint, while group four was edited with CorelDraw and group five was edited with GIMP. As presented above, our first approach is to check the EXIF metadata using our EXIF analyser to detect doctored image. The results generated were compared with known forensic tools. The results are presented in table 1 below.

Table 1: Accuracy of our approach in relation to other Forensic tools

	Photoshop	Coreldraw	Paint	GIMP
Photoforensic	100%	0%	0%	100%
Imageedited	100%	0%	50%	100%
Forensically	100%	0%	0%	100%
Our approach	100%	100%	0%	100%

From the result presented above, our approach was able to detect forgery performed using photoshop photo editing software accurately as well as those images edited with Coreldraw and GIMP. Meanwhile, it failed to detect manipulations performed using Microsoft Paint photo editing software. On investigation, we observed that while other image editing software alters the metadata of the image, Microsoft Paint does not. This was the reason why our approach failed in this regard. This established one of the reasons while we decided to use two approaches in detecting forged image. Relying only on the report of EXIF analysis is not enough evidence to conclude if an image is tampered with or not.

Table 2: Accuracy of our approach in Detecting Double JPEG Compression

Q1/Q2	50	55	60	65	70	75	80	85	90	95
50	-	91	92	96	93	95	99	97	99	99
55	89	-	94	96	95	97	93	99	95	98
60	87	89	-	93	94	93	95	97	94	99
65	83	84	87	-	90	92	99	97	98	97
70	87	90	85	89	-	97	94	96	99	96
75	91	89	88	90	93	-	93	97	94	96
80	86	91	92	89	90	93	-	95	99	95
85	85	87	93	90	92	95	93	-	96	97
90	89	86	92	89	91	95	93	96	-	99
95	85	87	90	87	93	93	97	94	98	-

As stated above, the remaining 1000 images left are used for the second phase of our approach. Out of the remaining images were divide it into two; the first 500 were duplicated into 10 different folders. Each of the folders was double compressed with different quality factors. We use 90 different quality factors (i.e 50/55, 50/60, 50/65, etc). The compressed images were tested with our algorithm and the results are presented in table 2.

Table 2 shows the accuracy of our second approach to detecting the existence of double compression in any given image. We did not test for Q1=Q2 since we stated earlier that an image is said to be double compressed if first compress with a quality factor Q_1 and recompressed with another quality factor Q_2 where $Q_1 \neq Q_2$.

From figure 7, it can be deduced that the highest accuracy is obtained when Q_1 and Q_2 are at the highest value i.e $Q_1 = 95$ and $Q_2 = 90$.

Also, we evaluate our second approach by comparing the results of the analysis with those results presented by Taimori, et al (2016) and Dong, et al (2011).

Table 3: Comparism of our approach with that of Taimori et al (2016) and Dong (2011)

Approach	Q1/Q2	50	55	60	65	70	75	80	85	90	95
Taimori	50	10.95	84.31	90.15	97.45	90.88	99.64	97.08	89.42	89.05	78.47
Dong		12.04	11.31	2.55	30.29	16.06	98.18	96.35	18.61	26.64	9.49
Proposed		-	91.00	92.00	96.00	93.00	95.00	99.00	97.00	99.00	99.00
Taimori	55	89.78	15.69	88.69	93.43	91.24	83.58	84.31	87.59	97.81	79.56
Dong		27.01	9.49	30.29	0	0.73	0	6.93	2.19	58.03	10.58
Proposed		89.00	-	94.00	96.00	95.00	97.00	93.00	99.00	95.00	98.00
Taimori	60	97.45	90.88	20.07	95.62	95.99	90.15	98.54	94.53	95.99	91.61
Dong		50.73	16.06	7.66	75.18	44.89	10.22	83.58	58.39	11.68	20.07
Proposed		87.00	89.00	-	93.00	94.00	93.00	95.00	97.00	94.00	99.00
Taimori	65	98.54	96.35	94.89	15.69	95.62	90.51	98.18	99.27	85.04	86.13
Dong		82.85	10.22	39.78	8.03	82.48	0	0	48.54	1.46	2.55
Proposed		83.00	84.00	87.00	-	90.00	92.00	99.00	97.00	98.00	97.00
Taimori	70	98.18	99.27	98.91	94.16	21.9	95.99	96.35	100	99.64	89.78
Dong		85.4	92.34	93.8	24.09	5.47	35.4	0	97.45	79.56	5.84
Proposed		87.00	90.00	85.00	89.00	-	97.00	94.00	96.00	99.00	96.00
Taimori	75	99.27	99.64	99.27	99.64	90.15	18.25	97.45	97.81	97.08	92.7
Dong		98.91	98.91	95.26	98.91	15.69	1.09	84.31	0.73	21.17	0.73
Proposed		91.00	89.00	88.00	90.00	93.00	-	93.00	97.00	94.00	96.00
Taimori	80	98.91	98.54	99.64	99.27	100	99.27	17.15	98.18	99.64	97.45
Dong		99.27	95.62	95.99	96.72	99.64	77.37	0	79.2	0	2.55
Proposed		86.00	91.00	92.00	89.00	90.00	93.00	-	95.00	99.00	95.00
Taimori	85	100	99.64	100	98.54	99.27	100	100	17.15	100	100
Dong		98.18	99.64	99.64	98.91	98.18	94.16	95.26	0	50.73	8.03
Proposed		85.00	87.00	93.00	90.00	92.00	95.00	93.00	-	96.00	97.00
Taimori	90	98.91	99.27	99.27	98.91	99.27	100	100	100	20.8	99.64
Dong		97.08	98.91	100	97.81	99.27	97.45	97.81	98.18	0	0
Proposed		89.00	86.00	92.00	89.00	91.00	95.00	93.00	96.00	-	99.00
Taimori	95	100	100	100	99.64	100	99.64	100	100	100	8.76
Dong		98.18	98.18	100	97.08	99.64	96.35	94.53	94.89	98.91	0
Proposed		85.00	87.00	90.00	87.00	93.00	93.00	97.00	94.00	98.00	-

From the result presented in table 3, it is observed that our approach performed beautifully in all circumstances. Table 3 reveals that where the other methods as presented in table 3 failed, our approach performed well and where their approached outperformed our approached, our results are very close to theirs.

VI. DISCUSSION

With this tool, we have shown that it is possible to distinguish between original image and doctored image. Our approach was based on examination of the EXIF parameters (hidden information in JPEG image) and detection of double quantization in any JPEG image. The combination of the two approaches gives a better result of image tampering detection.

From the analysis, it was observed that relying only on the result of EXIF metadata analysis cannot guarantee a good result as experience forger can tampered with the metadata also, we observed that some photo editing software sometimes does change the EXIF parameters even when the content of the image is tampered with. Hence, we deem

it fit to combine the result from EXIF metadata analysis with detection of effect of double quantization on image. This gives near a perfection result.

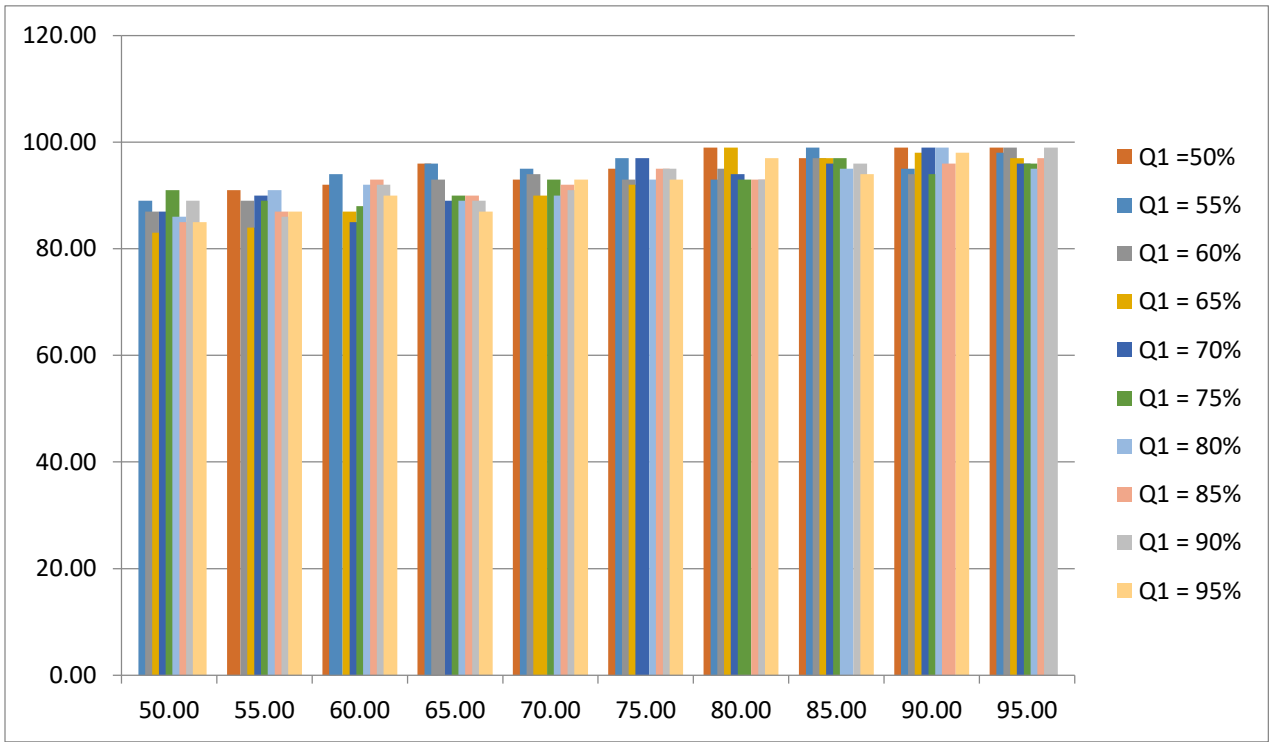


Figure 7: Bar chart showing Accuracy of the proposed approach.

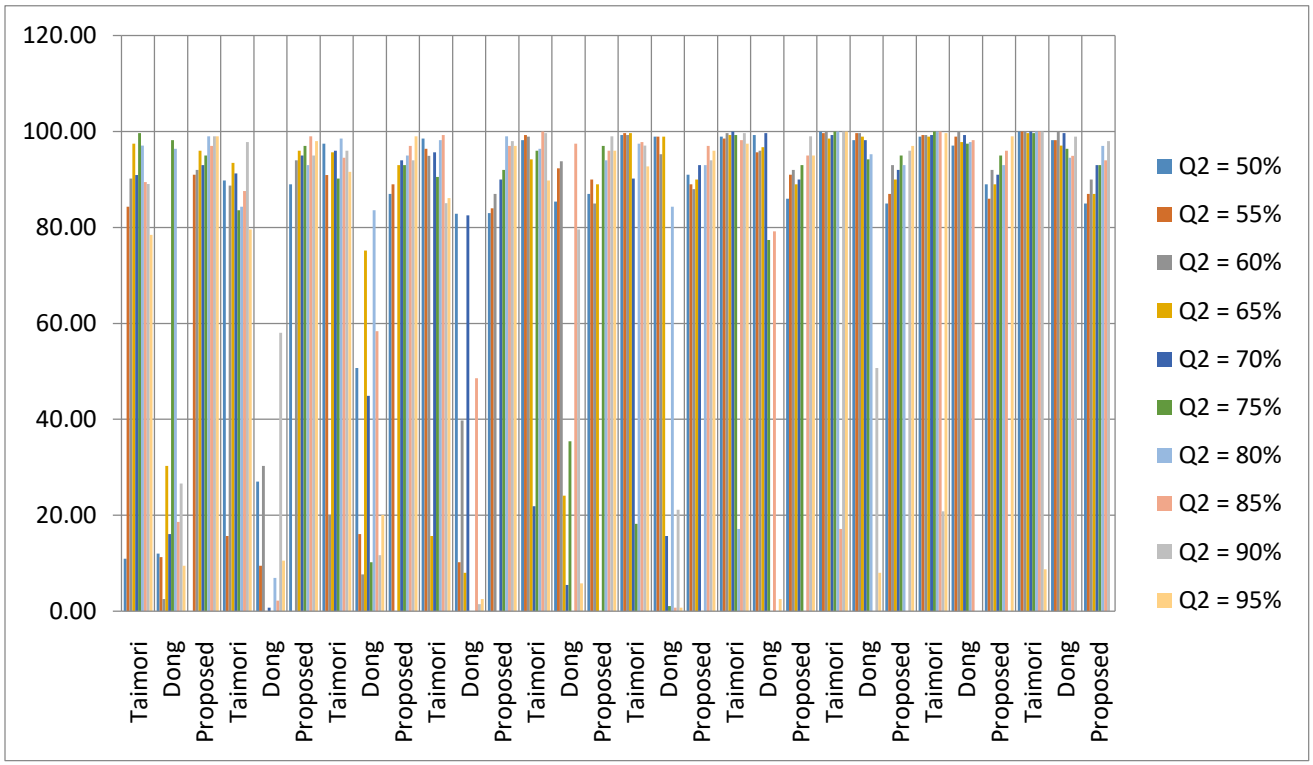


Figure 8: Bar chart showing comparism of our approach with that of Taimori et al (2016) and Dong et al (2011)

REFERENCES

Alin C. P. (2004). Statistical Tools for Digital Image Forensics. PhD thesis, at Dartmouth College Hanover, New Hampshire.

Dong L., Kong X., Wang B., & You X. (2011) Double compression detection based on Markov model of the first digits of DCT coefficients. *In: IEEE 6Th international conference on image and graphics (ICIG), pp 234–237*

Hitesh, C. P. & Mohit, M. P. (2015). An Improvement of Forgery Video Detection Technique using Error Level Analysis. *International Journal of Computer Applications (0975 – 8887) Volume 111 – No 15, February 2015*

Junfeng, H., Zhouchen L., LifengW., & Xiaoou, T. (2006). Detecting Doctored JPEG Images Via DCT Coefficient Analysis. *Springer-Verlag Berlin Heidelberg 2006*

Lin, Z., He, J., Tang, X., & Tang, C. K. (2009) Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. *Pattern Recognition 42(11), 2492–2501 (2009)*

Qing W. & Zhang R. (2016). Double JPEG compression forensics based on a convolutional neural network. *EURASIP Journal on Information Security (2016) 2016:23.*

Simone M., Marco T. & Stefano T. (2015). Discriminating multiple JPEG compressions using first digit feature. *SIP (2015), vol. 3, e19.*

Taimori, A., Farbod, R., Alireza, B., Ali A., & Babaie-Zadeh, M. (2016) A novel forensic image analysis tool for discovering double JPEG compression clues. *Springer Science+Business Media New York 2016.*

Tiziano B. & Alessandro P. (2012). *IEEE Transactions on Information Forensics and Security, vol. 7, No. 2, April 2012.*

Yujin Z. & Shilin W. (2011). Detection of Shifted Double JPEG Compression using Markovian Transition Probability Matrix. *APSIPA ASC 2011 Xi'an.*

Yi-Lei C. & Chiou-Ting H. (2011). Detecting Recompression of JPEG Images via Periodicity Analysis of Compression Artifacts for Tampering Detection. *IEEE Transactions on Information Forensics and Security, vol. 6, No. 2, June 2011.*

CSEAN

CYBER SECURITY EXPERTS
ASSOCIATION OF NIGERIA