

2016 INFORMATION SECURITY FOR SOUTH AFRICA

Proceedings of the
2016 ISSA Conference

17 – 18 August 2016
54 on Bath Hotel
Rosebank
Johannesburg
South Africa



information security
SOUTH AFRICA



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNISA | 
university of south africa



UNIVERSITY
OF JOHANNESBURG
JOHANNESBURG

*Edited by
HS Venter, M Loock, M Coetzee and MM Eloff*

ISBN 978-1-5090-2472-8
IEEE Catalog Number CFP1666I-USB

2016 Information Security for South Africa

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For reprint or republication permission, email to IEEE Copyrights Manager at pubs-permissions@ieee.org. All rights reserved. Copyright ©2016 by IEEE.

IEEE Catalog Number CFP1666I-USB

ISBN 978-1-5090-2472-8

Contents

Introduction.....	iv
Focus.....	v
Conference Committees.....	vii
Review Committee.....	viii
Review Process.....	xi
A framework towards governing “Bring Your Own Device in SMMEs” <i>Noluvuyo Fani, Rossouw von Solms and Mariana Gerber</i>	1
PoPI Act - opt-in and opt-out compliance from a data value chain perspective: A South African insurance industry experiment <i>Paulus Swartz and Adele Da Veiga</i>	9
An Interactive Visual Library Model to Improve Awareness in Handling of Business Information <i>Petrus Marthinus Jacobus Delpont, Mariana Gerber and Nader Sohrabi Safa</i>	18
Mobile Device Usage in Higher Education Institutions in South Africa: A Case Study <i>Ryan De Kock and Lynn Futcher</i>	27
SHA-1 and the Strict Avalanche Criterion <i>Yusuf Motara and Barry Irwin</i>	35
Specific Emitter Identification for Enhanced Access Control <i>Jeevan Samuel and Warren Du Plessis</i>	41
Review of Data Storage Protection Approaches for POPI Compliance <i>Nicholas Scharnick, Mariana Gerber and Lynn Futcher</i>	48
CDMA in signal encryption and information security <i>Olanrewaju Wojuola, Stanley Mneney and Viranjay M Srivastava</i>	56
Effect of Varying Mobility in the Analysis of Black Hole Attack on MANET Reactive Routing Protocols <i>Lineo Mejaele and Elisha Ochola</i>	62
The Pattern-richness of Graphical Passwords <i>Johannes Vorster, Renier van Heerden and Barry Irwin</i>	69
Dridex: analysis of the traffic and automatic generation of IOCs <i>Lauren Rudman and Barry Irwin</i>	77
Context-Aware Mobile Application for Mobile Devices <i>Mfundo Masango, Francois Mouton, Alastair Nottingham and Jabu Mtsweni</i>	85
Team Formation in Digital Forensics <i>Wynand van Staden and Etienne van der Poel</i>	91
Social Network Phishing: Becoming Habituated To Clicks and Ignorant To Threats? <i>Edwin Donald Frauenstein and Stephen Flowerday</i>	98
Identity Management for e-Government (Libya as a case study) <i>Othoman Elaswad and Christian D. Jensen</i>	106
Recognizing Surgically Altered Faces using Local Edge Gradient Gabor Magnitude Pattern <i>Chollette Olisah and Peter Ogedebe</i>	114
Adaptable Exploit Detection through Scalable NetFlow Analysis <i>Alan Herbert and Barry Irwin</i>	121
Unsupervised Learning for Robust Bitcoin Fraud Detection <i>Patrick Monamo, Vukosi Marivate and Bhekisipho Twala</i>	129

Introduction

ISSA2016 is the annual conference for the information security community that continues on the successful recipe established in 2001. The upcoming conference is held under the auspices of the University of Johannesburg Academy for Computer Science and Software Engineering, the University of South Africa School of Computing and the University of Pretoria Department of Computer Science.

The ISSA2016 Conference will run from the 17th (Wednesday) to the 18th of August (Thursday) 2016.

The conference has grown each year in various ways. Not only have delegate and presenter numbers been on the rise, but interest from industry has also grown and been displayed through sponsorship of the conference or aspects thereof. We believe that the quality and relevance of the information presented by industry practitioners and academics has also evolved over the years, as have the opportunities for senior research students to present their research to a critical and representative audience.

Conferences have become a major focus area - and often a money spinner - in many industries, so at any time you will see a number of conferences being advertised in fields such as information security. What sets the ISSA conference apart is that it is not intended to generate a profit for an organisation, and it does not encourage marketing of products and services through presentations. Instead, the proceeds from registration fees are reinvested to ensure that the conference grows each year. In exchange for their investment in the conference, sponsors are afforded an opportunity to present company-specific information that has a bearing on the conference themes, and presentations submitted by potential speakers are sent through a vigorous review process, managed by a team of respected international experts in information security.

We trust that the annual ISSA conference will continue to be recognised as an platform for professionals from industry as well as researchers to share their knowledge, experience and research results in the field of information security on a South African, but also on an international level.

To ensure ongoing improvement, we again encourage input from all those interested in the field of Information Security, particularly those who are actively seeking to progress the field, to take part and share their knowledge and experience.

We look forward to seeing old friends and new participants at ISSA.

Hein Venter, Marianne Loock, Marijke Coetzee, Mariki Eloff and Jan Eloff

Conference Co-organisers

www.infosecsa.co.za

Focus

Information security has evolved and in the last few years there has been renewed interest in the subject worldwide. This is evident from the many standards and certifications now available to guide security strategy. This has led to a more clear career path for security professionals.

The convergence of technologies together with advances in wireless communications, has meant new security challenges for the information security fraternity. As hotspots become more available, and more organisations attempt to rid their offices of "spaghetti" so the protection of data in these environments becomes a more important consideration.

It is this fraternity that organisations, governments and communities in general look to for guidance on best practice in this converging world.

Identity theft and phishing are ongoing concerns. What we are now finding is that security mechanisms have become so good and are generally implemented by companies wanting to adhere to good corporate governance, so attackers are now looking to the weak link in the chain, namely the individual user. It is far easier to attack them than attempt to penetrate sophisticated corporate systems. A spate of spyware is also doing the rounds, with waves of viruses still striking periodically. Software suppliers have started stepping up to protect their users and take some responsibility for security in general and not just for their own products.

The conference focuses on all aspects of information security and invites participation across the Information Security spectrum including but not being limited to functional, business, managerial, theoretical and technological issues.

Invited speakers will talk about the international trends in information security products, methodologies and management issues.

In the past ISSA has secured many highly acclaimed international speakers, including:

- Pieter Geldenhuys, Vice-chair of the Innovation Focus Group at the International Communications Union, Geneva, Switzerland. Topic: BUSINESS UNUSUAL: Strategic insight in creating the future. Leveraging the value of the Hyper-connected world.
- Wayne Kearney, Manager: Risk & Assurance at Water Corporation. Topic: Why are management shocked with all the "PHISH" caught? A case study in perspective.
- Prof. Dr. Sylvia Osborn, Associate Professor of Computer Science, The University of Western Ontario, Ontario, Canada. Topic: Role-based access control: is it still relevant?
- Prof. Dr. Steve Marsh, Associate Professor at University of Ontario, Institute of Technology. Topic: Trust and Security - Links, Relationships, and Family Feuds.
- Alice Sturgeon manages the area that is accountable for identifying and architecting horizontal requirements across the Government of Canada. Her topic made reference to An Identity Management Architecture for the Government of Canada
- Dr Alf Zugenmaier, DoCoMo Lab, Germany. His topic was based on Security and Privacy.
- William List, WM List and Co., UK. His topic was: Beyond the Seventh Layer live the users.

- Prof. Dennis Longley, Queensland University of Technology, Australia. His topic was: IS Governance: Will it be effective?
- Prof. TC Ting: University of Connecticut, and fellow of the Computing Research Association, United States.
- Prof. Dr. Stephanie Teufel: Director of the International Institute of Management in Telecommunications (iimt). Fribourg University, Switzerland.
- Rich Schiesser, Senior Technical Planner at Option One Mortgage, USA Rick Cudworth, Partner, KPMG LLP, International Service Leader, Security and Business Continuity - Europe, Middle East and Africa.
- Dario Forte - CISM, CFE, Founder, DFLabs Italy and Adj. Faculty University of Milano.
- Reijo Savola - Network and information security research coordinator, VTT Technical Research Centre of Finland.
- Mark Pollitt - Ex Special Agent of the Federal Bureau of Investigation (FBI) and professor at the Daytona State College, Daytona Beach, Florida, USA.
- Prof Joachim Biskup - Professor of Computer Science, Technische Universität Dortmund, Germany.
- Dr Andreas Schaad - Research Program Manager, SAP Research Security & Trust Group, Germany.
- Prof Steven Furnell - Head of School, School of Computing and Mathematics (Faculty of Science and Technology), University of Plymouth, UK.
- Prof Matt Warren - School of Information and Business Analytics, Deakin University, Australia.
- Christian Damsgaard Jensen - Associate Professor, Institute for Mathematics and Computer Science, Technical University of Denmark.

The purpose of the conference is to provide information security practitioners and researchers worldwide with the opportunity to share their knowledge and research results with their peers.

The objectives of the conference are defined as follows:

- Sharing of knowledge, experience and best practice
- Promoting networking and business opportunities
- Encouraging the research and study of information security
- Supporting the development of a professional information security community
- Assisting self development
- Providing a forum for education, knowledge transfer, professional development, and development of new skills
- Promoting best practice in information security and its application in Southern Africa
- Facilitating the meeting of diverse cultures to share and learn from each other in the quest for safer information system

Conference Committees

General Conference Chairs

Hein Venter (Department of Computer Science, University of Pretoria)
Marijke Coetzee (Academy for Computer Science and Software Engineering, University of Johannesburg)
Marianne Loock (School of Computing, University of South Africa)
Mariki Eloff (Institute for Corporate Citizenship, University of South Africa)
Jan Eloff (Department of Computer Science, University of Pretoria)

Organising Committee

Hein Venter (Department of Computer Science, University of Pretoria)
Marijke Coetzee (Academy for Computer Science and Software Engineering, University of Johannesburg)
Marianne Loock (School of Computing, University of South Africa)
Irene Venter (Department of Computer Science, University of Pretoria)
Mariki Eloff (Institute for Corporate Citizenship, University of South Africa)

Conference Programme Committee

Marijke Coetzee (Academy for Computer Science and Software Engineering, University of Johannesburg)
Marianne Loock (School of Computing, University of South Africa)
Mariki Eloff (Institute for Corporate Citizenship, University of South Africa)
Stephen Flowerday - 2016 guest editor for the SAIEE Africa Research Journal (University of Fort Hare)

Conference Honorary Committee




















The following members are honorary committee members of the ISSA conference. These committee members are honoured for their effort as founding members of the ISSA conference since 2000. Although they are not so actively involved in organising the conference, they are still performing an important advisory roll in the conference. The current conference committee feels obliged to honour them as such.



Jan Eloff (University of Pretoria)
Mariki Eloff (Institute for Corporate Citizenship, University of South Africa)
Les Labuschagne (School of Computing, University of South Africa)

On behalf of the general conference chairs, we would like to extend our heartfelt appreciation to all the conference committees for their hard work in organising ISSA 2016! Without your continuous hard work and efforts, ISSA 2016 would not have been possible. Again, we thank you!

Reviewers

A rigorous double-blind refereeing process was undertaken by an international panel of referees as shown below. The task of a reviewer is often a thankless task, however, without them this conference would not be possible. The ISSA Organising Committee would like to extend their heartfelt thanks to the following reviewers whom include leading information security experts from around the world:

Name	Company/Affiliation	Country	
Hanifa Abdullah	University of South Africa	South Africa	
Mary Adedayo	University of Pretoria	South Africa	
Atif Ahmad	University of Melbourne	Australia	
Sampson Asare	University of Botswana	Botswana	
Elmarie Biermann	Private	South Africa	
Hettie Booysen	Private	South Africa	
Reinhardt Botha	Nelson Mandela Metropolitan University	South Africa	
Rachelle Bosua	University of Melbourne	Australia	
KP Chow	University of Hong Kong	Hong Kong	
Nathan Clarke	University of Plymouth	UK	
Evan Dembsky	University of South Africa	South Africa	
Moses Dlamini	University of Pretoria	South Africa	
Paul Dowland	University of Plymouth	UK	
Lynette Drevin	North-West University	South Africa	
David Ellefsen	University of Johannesburg	South Africa	
Eduardo Fernandez	Florida Atlantic University	USA	
Stephen Flowerday	University of Fort Hare	South Africa	
Evangelos Frangopoulos	University of South Africa	Greece	
Steven Furnell	University of Plymouth	UK	

Lynn Fitcher	Nelson Mandela Metropolitan University	South Africa	
Virginia Greimann	Boston University	United States of America	
Stefanos Gritzalis	University of the Aegean	Greece	
Barry Irwin	Rhodes University	South Africa	
Christian Damsgaard Jensen	Technical University of Denmark	Denmark	
Jason Jordaan		South Africa	
Anne Kayem	University of Cape Town	South Africa	
Hennie Kruger	North-West University	South Africa	
Grace Leung	University of Johannesburg	South Africa	
Stefan Lindskog	Karlstad University	Sweden	
Buks Louwrens	Nedbank / University of Johannesburg	South Africa	
Sean Maynard	University of Melbourne	Australia	
Tayana Morkel	University of Pretoria	South Africa	
Francois Mouton	Council for Scientific and Industrial Research	South Africa	
Martin Olivier	University of Pretoria	South Africa	
Rolf Oppliger	eSECURITY Technologies	Switzerland	
Jacques Ophoff	University of Cape Town	South Africa	
Mauricio Papa	University of Tulsa	USA	
Guenther Pernul	University of Regensburg	Germany	
Indrajit Ray	Colorado State University	USA	
Rayne Reid	Nelson Mandela Metropolitan University	South Africa	
Karen Renaud	University of Glasgow	United Kingdom	

Reijo Savola	VTT Technical Research Centre Finland	Finland	
George Sibiya	Council for Scientific and Industrial Research	South Africa	
Paul Simon	Air Force Institute of Technology	United States of America	
Aelita Skarzauskiene	Mykolas Romeris University	Lithuania	
Bobby Tait	University of South Africa	South Africa	
Barend Taute	CSIR	South Africa	
Stephanie Teufel	University of Fribourg	Switzerland	
Kerry-Lynn Thomson	Nelson Mandela Metropolitan University	South Africa	
Dustin van der Haar	University of Johannesburg	South Africa	
Johan Van Niekerk	Nelson Mandela Metropolitan University	South Africa	
Brett van Niekerk	University of Kwazulu Natal	South Africa	
Wynand van Staden	University of South Africa	South Africa	
Basie von Solms	University of Johannesburg	South Africa	
Edgar Weippl	Secure Business Austria	Austria	
Stephen Wolthusen	Norwegian Information Security Lab	Norway	
Alf Zugenmaier	Munich University	Germany	

Review Process

ISSA uses a double blind peer-review process to ensure the quality of submissions before acceptance. Authors initially submit abstracts to determine if the paper meets the goals and fits into the theme of the conference. The ISSA Program Committee assesses each submission for relevance and fit. Authors are then notified whether their abstracts were accepted, and if so, invited to submit a full paper for peer review.

On the due date, authors submit full papers, anonymised by the authors for the double blind review process. Each paper goes through an administrative review and is assigned to at least three reviewers selected from an international panel of reviewers, in order to confirm that the paper conforms to the specifications and quality for the conference. If a paper does not meet the requirements, the author is asked to make the required changes as indicated by reviewers and asked to resubmit the paper, or to consider submitting the paper to another conference.

A Review Committee is invited to participate, consisting of both local and international experts in the field of Information Security. A process is followed by the Program Committee to allocate papers to reviewers based on their area of expertise. Reviewers are subject matter experts, of which over 50% are international. Reviewers usually have 5 or 6 categories that they are willing to review against. Each reviewer will establish the number of papers they can review in a specific time period and are allowed to bid on the papers they want to review. An automated process allocated papers to each reviewer according to their preferences.

Each paper is reviewed by a minimum of two reviewers in a double blind review process. Papers are reviewed and rated on a 10 point system with 1 being poor and 10 being excellent as follows:

- Originality (1-10)
- Significance (1-10)
- Technical quality (1-10)
- Relevance (1-10)
- Presentation (1-10)
- Overall Rating (1-10)

Reviewers' confidence in their own rating is also taken into account by the algorithm that calculates the final score. Reviewers are encouraged to make anonymous suggestions to the author(s) of the paper.

Based on the final score (1-10), a paper with 5 or below points can be recommended for a poster/research-in-progress session and a 9 to 10 point paper can be put in the "best paper" category. An acceptance rate of between 30% and 40% is expected for the conference.

Authors are notified of the outcome of the review process which includes the anonymous suggestions and recommendations of the reviewers. Authors then have to submit the final version of the paper that will then be included in the formal conference proceedings. A CD version of these proceedings was also published and distributed at the conference with ISBN 978-1-4799-3383-9. All CD proceedings from all previous ISSA conferences are also available at www.infosecsa.co.za/past.

A framework towards governing “Bring Your Own Device in SMMEs”

Noluvuyo Fani, Rossouw von Solms and Mariana Gerber

Center for Research in Information and Cyber Security

NMMU

NMMU, University Way, Port Elizabeth, 6001, South Africa.

s207068382@nmmu.ac.za, rossouw.vonsolms@nmmu.ac.za , mariana.gerber@nmmu.ac.za

Abstract — Information is a critically important asset that has been used for decades within organizations. Like any asset, there are threats to the information that impact processes such as; email retrieval and access to organizational system services. As a consequence of the threats, attention to the security of the information is important. Technology is utilized to secure information and the cost affiliated to the technology can be dire. As technology evolves with each transitory decade, there are different phenomenon’s that attempt to process and secure organizational information whilst reducing costs. The evolution of technology has developed a new phenomenon called “Bring Your Own Device” (BYOD). BYOD is a phenomenon that allows employees to use their own personal mobile device to complete organizational tasks. The adoption of BYOD expands from large organizations to small, medium and micro enterprises (SMMEs). With the adoption of BYOD there are benefits and more significantly risks associated to BYOD. Therefore, this paper will discuss the SMME context and its challenges towards the governance of BYOD. In addition, there will be a discussion on how organizations can govern BYOD in an SMME context by considering the existing BYOD approaches and provide an approach suitable for SMMEs. Furthermore, the suitable BYOD approach for an SMME context will further be evaluated and compared against the existing BYOD approaches that were identified. The research process of the study is conducted within the design-oriented research paradigm utilizing a cyclic approach.

Keywords- *BYOD, SMMEs, mobile devices, information*

Security.

I. INTRODUCTION

There are various assets that are composed within an organization. Assets such as; humans, information and capital are composed within the organization. An asset can be categorized as tangible or intangible. An intangible asset is an asset that *cannot* be seen or touched (e.g. patents) and a tangible asset *can* be seen and touched (e.g. computer) [1]. There is a value attached to each asset whether it is tangible or intangible, and therefore, assets are important within an organization and should be protected.

Information is a valuable intangible asset. It can be defined as “data with attributes of relevance and purpose. It is usually in the format of a document or visual and/or audible message.

Additionally, information should convey a message that must be understood” [2]. Organizations utilize information to complete their daily tasks as information is a universal form of communication. The communication of information will be through sources of; emails, telephonic, paper-based documentation etc. The information communicated will be specific to each organization and might contain some “secrets” of the organization [3]. Due to the uniqueness pertaining to the organizational information, organizations should implement security mechanisms that should safeguard the confidentiality, integrity and availability (CIA) of the information [4].

The security mechanisms implemented will reduce the likelihood of breaches to the CIA of information and the information remains intact [3]. The technological tools attained to process and secure information have changed and adapted to the changes and needs of organizations. Organizations prefer technology that allows an ease of use and accessibility while maintaining or reducing costs. Technology has developed to a rapid extent of bringing forth a phenomenon referred as “Bring Your Own Device” (BYOD) [5]. BYOD is an exciting development, which has caused an alteration in the way business is conducted and is affiliated with many benefits. However, with any technology, there are risks associated with BYOD.

A. BYOD phenomenon

BYOD is an acronym for “Bring Your Own Device” and can also be referred to as the *Consumerization of Information Technology*. BYOD can be defined as “the practice of allowing employees to bring to the workplace their own mobile devices that are capable of connecting to the organizational network.” [6]. The dual-use of a mobile device for personal and organizational purposes has offered the benefits of:

- **Accessibility** – Accessibility to organizational resources via the organizational network, allowing employees to work “anytime” and “anywhere”.
- **Increased Productivity and Innovation** – Minimal training is required due to the familiarity with the mobile device, thus, there is increased production and innovation.
- **Cost-Savings** - BYOD can assist in the reduction of costs towards organizational expenses as the device is purchased and owned by the employee [7].

The benefits affiliated with BYOD have allowed BYOD to gain momentum with both organizations and employees. Employee demand for the implementation of BYOD leaves the organization with minimal choice but to adapt to the changing environment [8]. With the benefits and adoption of BYOD from both large organizations and SMMEs, organizations should remain aware of the risks of implementing BYOD, as the confidential organizational information is accessed through the BYOD devices. The risks of implementing BYOD ranges from; data leakage, lost devices and hacking [9]. The next subsection discusses BYOD in an SMME environment.

B. BYOD in an SMME

SMMEs are encompassed by limitations in their budgets and resources. The benefits of the implementation of BYOD in an SMME could reduce the budgets and costs affiliated to the resources. This is due to circumstances such as the cost affiliated with the purchase of the BYOD devices is handled by the employee [10].

When an SMME implements BYOD limited budgets and resources should not be the only issue fixated on, but every aspect affiliated with BYOD must be taken into account. With this in mind, caution must be applied by the SMMEs as they can become easily susceptible to the risks associated with BYOD. There are BYOD initiatives such as; strategies, recommendations and frameworks outlined in literature. Before embracing these BYOD initiatives, SMMEs should understand their particular requirements and what is appropriate in their environment.

C. Requirements for BYOD in SMMEs

According to the National Small Business Amended Act No. 102 of 2004, a SMME definition is “a separate and distinct business entity, which is managed by one or more owner(s), which predominantly conducts its business in any sector and/or subsector of the national economy”. The SMME is all-encompassing of requirements such as; scalability, utility, efficacy and quality. Table 1 below provides a brief description of the allocated SMME categories [11]:

Table 1: Categories and descriptions of SMMEs[11]

Categories	Description
Survivalist enterprises	Operates in the informal sector of the economy. Minimal training or asset investments. Therefore, resulting in a lack of business growth.
Micro enterprises	One to five employees, usually the owner and family. An informal enterprise with no license, formal business infrastructure. Basic business skills and training.
Very small enterprise	Middle class economy. 10 paid employees or less Consists of self-employed artisans (electricians) and other pro.
Small enterprise	Approximately 100 employees. Registered, fixed business premises. Consists of complex management structure or managed by a single owner.
Medium enterprise	Owner managed and approximately 200 employees. Operates from fixed infrastructure with all formal necessary necessities for business.

The small stature and limited resources of SMMEs makes SMMEs vulnerable to weaknesses to their information. It is common that incidents of breaches to the SMMEs network and other resources develop. The pressure of the sustainability and the maintenance of existing SMME resources provides difficulties in monitoring other factors such as information security. The phenomenon of BYOD provides a competitive edge for any organization but with the strains and limitations found in SMMEs, the adoption of BYOD may be a hindrance instead of a competitive advantage. [12].

With the harsh reality of the limitations in SMMEs, there is a need for a solution that will cater for the desire of BYOD in an SMME environment. There are requirements that have to be taken into account when the BYOD solution is formulated. The requirements for BYOD solution in SMMEs should cater for the following:

- **Scalability** – Solution scalable for an SMME environment.
- **Utility** – Solution is usable in an SMME environment.
- **Efficacy** - Solution is efficient and developed with the SMME environment in mind.
- **Quality** – The solution formulated should provide value in an SMME environment.

This concludes that with the requirements for BYOD in SMMEs taken in context, an appropriate solution for the governance of BYOD can be devised. However, before formulation of the solution, it is vital to also consider the protection of the information within SMMEs as information is an asset in an organization regardless of organizational stature and limitations.

D. Information security characteristics of BYOD

Before the phenomenon of BYOD, organizations provided employees organizational mobile devices. With the phenomenon of BYOD, devices have the dual use of being used as a personal and organizational devices. As a consequence, employees have the advantage of accessibility to personal applications and services [13]. The security of the organizational information can be compromised when dealt with unknown applications and services entering the organizational network.

The IT department can only manage a certain degree of security on accessibility to personal applications and services. Therefore, employees within an organization should be made aware of their role in the security of organizational information. [6]. A foundation of characteristics for a suitable solution should be compiled as the initial phase for governing BYOD. Below is a list of eight BYOD characteristics identified from literature that an organization should follow:

BYOD Characteristics:

C1: There must be risk identification:

- “BYOD is an institutionalised security risk which small scale organizations need to assess and

evaluate before blindly embracing the practice” [6].

- “There are many potential risks and threats to confidential information resources and assets in organizations use BYOD devices” [7].

C2: There must be security requirements stipulated for BYOD:

- “The main goals of information security are confidentiality, integrity and availability” [3].
- “Legal and liability issues should be considered and stated in the BYOD policy” [14].

C3: The organizational context must be considered:

- “Uncontrolled environments present more dynamic risks within the specific context and circumstances of that environment” [15].
- “Organizations require accurate and reliable information because they communicate and manage substantial information resources” [7].

C4: There must be a BYOD device analysis:

- “BYOD consists of the use of personal devices. Only the definition does not state which devices it concerns” [16].
- “Devices should be registered for participation in the BYOD program, officially approved for use, and provisioned with required security settings” [5].

C5: The organization must take into context the employee role:

- “Users of mobile devices need to be aware of threats the mobile device threats and have competent skills to secure their devices” [17].
- “Users should be educated as they perform their daily activities, with frequent policy reminders that are non-intrusive and relevant to their current task” [19].

C6: There must be IT administration within the organization:

- “Organizations BYOD should realize the impact BYOD can have on technical support” [16].
- “It is crucial for organizations to employ a proper security model for mobile devices as security challenges will increase in organizations” [20].

C7: There must be a BYOD policy:

- “Policies are a good starting points for gaining control on an enterprise as they provide guidelines for BYOD adoption” [6].
- “The policy should provide clarity on how devices will be used and how IT can meet those needs” [21].

C8: An organization must have compliance:

- “A BYOD policy is likely to improve compliance by educating employees the risks associated with their devices” [6]. “Violation of the policy should have severe punishment” [22].
- “Companies must re-evaluate BYOD compliance” [23].

Once the BYOD characteristics are met, a solution can be formulated. In order for organizations to manage the demands of implementing BYOD, there are frameworks that have been formulated in literature. The upcoming section will present an analysis of some of the existing frameworks for managing BYOD.

II. EXISTING FRAMEWORKS FOR BYOD

There are many factors that dictate the approach in the formulation of a BYOD framework. As a result, the formulated frameworks for BYOD are specific to each environment or organization. In this paper, there are four frameworks that are considered for analysis as the existing formulated BYOD frameworks. The objective of the frameworks is the governance and management of BYOD. In this section, an outline of each framework will be provided and the distinction between each framework should be apparent. Subsequently, there will be a tabulated mapping of the BYOD characteristics mentioned earlier (Section I) and the identified frameworks. The objective of the mapping is to analyze whether the identified frameworks meet the requirements stipulated in the BYOD characteristics and if they cater for an SMME environment.

A. BYOD Security Framework

The first framework identified is the BYOD security framework [5]. This framework is divided into seven phases for managing BYOD. A brief description of each phase is as follows:

- **Plan:** Understand the context of the business. Identify the relevant users and resources they access.
- **Identify:** Devices are registered, approved, and provided with the appropriate security.
- **Protect:** The information held within the devices requires protection.
- **Detect:** The organization should prevent, or respond to and recover from, intentional or unintentional different threat events identified.
- **Respond:** The organization should respond to identified threats.
- **Recover:** The organization must be able to fully recover from the event.
- **Assess and Monitor:** An organization should assess and monitor the value and competence of the BYOD security program [5].

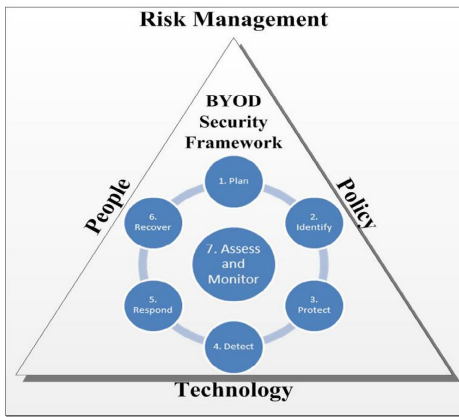


Figure 1: BYOD Security Framework [5]

Figure 1 illustrates the BYOD Security Framework. In illustration, the seven phases are encompassed by the three pillars of people, technology and policy. The purpose of the framework is to provide a foundation for a BYOD security program. The framework can be constantly amended. Furthermore, the BYOD Security Framework is formulated to form part of the risk management framework [5].

B. BYOD framework for a management system

The BYOD framework formulated governs BYOD by seeking assistance from the ISO/IEC 27000-series and strategic management. There are three steps that are specified in the proposed framework. The three steps are visualized in Figure 2 and are concisely defined as follows:

- **Analysis:** The organisation determines the relevant issues affect overall strategy and information security.
- **Design:** More analyses is conducted and there is the development of strategies. Existing policies are updated.
- **Action:** The organization should perform a risk assessment. When the risk assessment is completed, the strategy can be implemented.

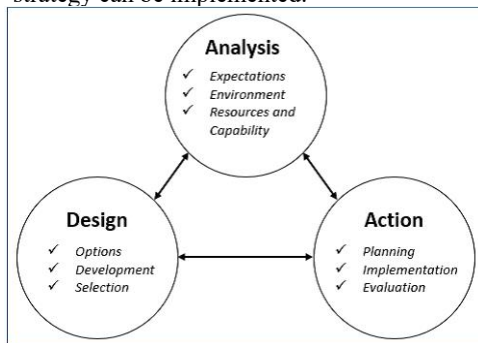


Figure 2: BYOD framework for a management system [24]

The framework provides a security and strategic way of thinking when an organization adopts BYOD [24].

C. BYOD privacy & culture governance framework

The third framework is the Bring Your Own Device implementation framework [25]. This framework maps the

organizational culture and privacy concerns within the organization. Once the mapping is complete, a policy is developed. The components prescribed in the framework are as follows:

- Determine the culture within the organization based on employee views.
- Delineate the characteristics that the organizational culture is based on.
- Identify the privacy concerns that would be applicable to the organization.
- Clearly define the individual concerns with regards to privacy.
- Conduct a privacy concern valuation based on employee's views. The assessment can assist in improving employee satisfaction.
- Develop a policy that takes account of the privacy concern assessment.
- Implement cloud management control, relate to the organizational culture.

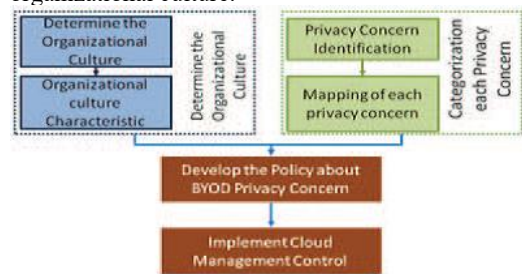


Figure 3: BYOD privacy & culture governance framework [25]

In Figure 3 the relationships of different components of the framework are diagramed. The purpose of the framework is to determine if organizations benefit in the implementation BYOD when organizational culture and cloud management control is adapted [25].

D. Enterprise and BYOD space BYOD Security Framework

The Enterprise and BYOD space BYOD Security Framework was formulated to protect the enterprise networks when BYOD is implemented. The represented framework is divided into two sides; the Enterprise side and the BYOD side. Below is brief description of each side:

- **Enterprise side:** includes the corporate resources and device management. The network access controls personal space and enterprise space.
- **BYOD side:** provides the functions that assist in separating corporate space, enforcing security policies, and the protection of corporate data [26].

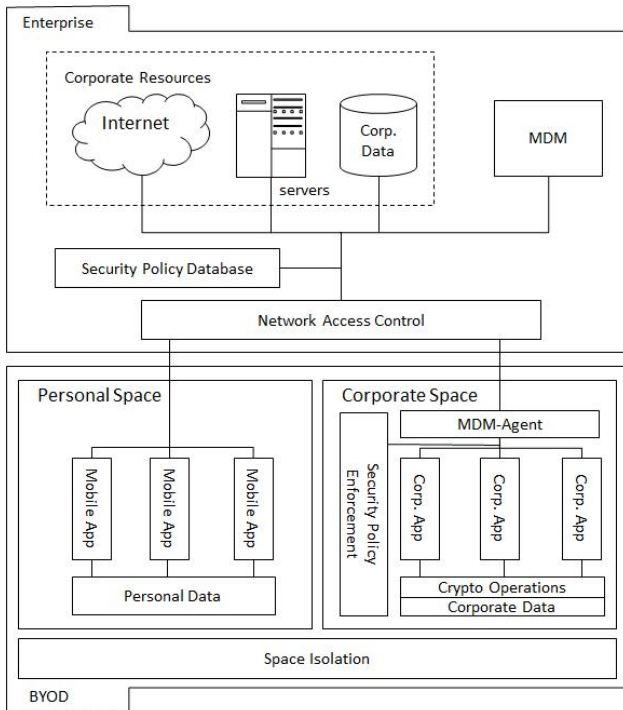


Figure 4: Enterprise and BYOD space BYOD Security Framework [26]

The enterprise and BYOD space BYOD Security Framework is presented in Figure 4. The framework provides protection to the organizational information by separating the network spaces that a BYOD user can access into enterprise space and BYOD space. This permits BYOD users to work in controlled and protected spaces. [26].

The four BYOD frameworks discussed above are similar in their intention of governing BYOD. Although, it is apparent that they are different in the way they are formulated and implemented. Eight characteristics were mentioned earlier and they provided a foundation for an appropriate BYOD solution. In Table 2 there is a mapping of the eight BYOD characteristics and the identified existing BYOD frameworks. The mapping analyses whether the identified frameworks meet the eight BYOD characteristics, and whether they cater for an SMME environment.

Table 2: Mapping of the BYOD characteristics and existing frameworks

Authors	C1	C2	C3	C4	C5	C6	C7	C8	SMME
[5]	✓	✓	✓	✓		✓			
[27]	✓	✓	✓		✓				
[25]					✓	✓	✓		
[26]		✓	✓	✓		✓	✓	✓	

From the analysis of the tabulated mapping, it is noticeable that although the frameworks meet some of the BYOD characteristics, but none of the identified BYOD frameworks cater for an SMME environment. Therefore, it would be difficult to assume that they would be appropriate to be

implemented in an SMME environment. BYOD high level management framework

The solution to the phenomenon of BYOD is not about developing numerous frameworks, but rather a framework that will allow the organization to reap the benefits of BYOD while taking into account the organizational environment and information protection. As small organizations, SMMEs are adopting and want to adopt BYOD. But in doing so, they encounter issues when it comes to a solution for the governance of BYOD. Thus, it is essential that a solution for governing BYOD in SMMEs is formulated.

III. BYOD MANAGEMENT SYSTEM (BYODMS)

The previous section demonstrated the four BYOD frameworks from literature. As evaluated, the frameworks lack in their diversity and alignment with the eight BYOD characteristics. Furthermore, they lack in addressing BYOD within an SMME environment. As a result of the challenges discussed, the proposed solution depicted in this section is that of the BYOD high level management framework. The BYOD high level framework was formulated through a rigorous research process within the design-oriented research paradigm utilizing a cyclic approach. The first phase is an analysis phase where the problem is analysed, in the second phase is a design phase where solution is developed. The third phase is an evaluation phase which consists of the validation of the artifact against the specified objectives, methods etc. The fourth phase is the diffusion phase where the solution is finalised [28].

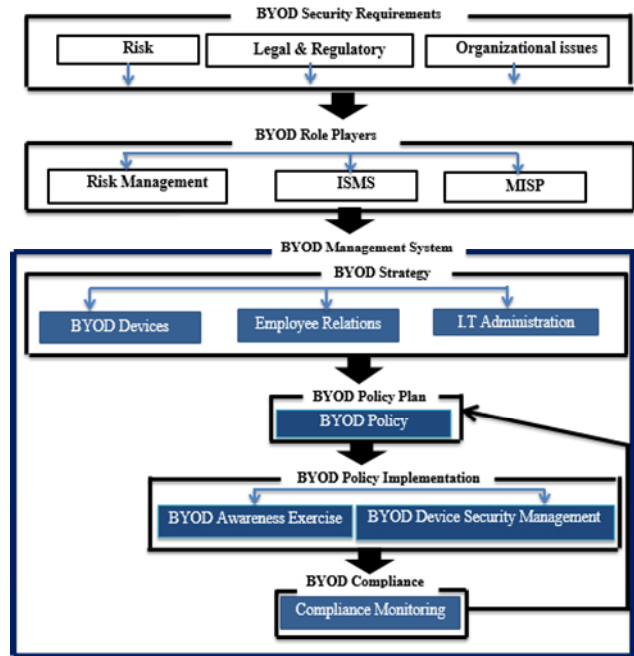


Figure 5: BYOD high level management framework

Figure 5 illustrates the BYOD high level management framework. The BYOD high level management framework is divided into six sections; the BYOD Security Requirements, Security Role Players, BYOD Strategy and the BYOD Policy

Plan, BYOD Policy Implementation and BYOD Compliance. The purpose of the BYOD strategy is for executive management to make all the decisions that are required for the governance of BYOD. The decisions to be made should take into account the following three components; BYOD Devices, Employee Relations and IT Section. The six components will be further divided into the following:

- **BYOD Security Requirements:**
 - **Risk:** Determine risks to the CIA of the information
 - **Legal & Regulatory issues:** Identify legal and regulatory issues
 - **Organizational issues:** Identify other security requirements
- **BYOD Role Players:**
 - **Risk Management:** Identify BYOD risks
 - **ISMS:** Secure organizational information
 - **MISP:** Determine what the MISP states about information security
- **BYOD Strategy:**
 - **BYOD Devices:**
 - **Type of device:** Decisions about the type of device to be incorporated into the municipal environment for BYOD.
 - **Device registration:** It is essential that the preferred devices are registered.
 - **Employee Relations:**
 - **Eligibility and Registration:** Decisions need to consider the eligibility of employees and the registration of the eligible employees.
 - **Awareness Programs:** Executive management must decide on the awareness programs.
 - **IT Section:**
 - **Compatibility testing:** The BYOD devices require compatibility testing.
 - **Authentication and Authorization:** BYOD users need to be authenticated and authorized.
 - **Information separation:** Information should be separated on the BYOD device.
 - **Device and Application Management / Security:** The information and applications within the BYOD device, require constant security and protection.

Once the decisions are concluded, a draft of a BYOD Policy should follow. The BYOD Policy will be inclusive but not limited to the policies, controls, education and control measures. The BYOD Policy can be divided into the following components:

- **BYOD Policy Plan:**
 - **BYOD Policy:** A BYOD Policy should be a documented guideline for BYOD.

- **BYOD Implementation:**
 - **BYOD Awareness Exercise:** This component will consist of the educational, awareness and training aspects of BYOD.
 - **BYOD Device Management:** BYOD device management will be addressed in this component.
- **BYOD Compliance:**
 - **Compliance Monitoring:** There needs to be constant monitoring of compliance for BYOD.

The BYOD high level management framework was formulated under the design-oriented research paradigm. The identified SMME environment and stakeholder for this study is local government, particularly at a District Municipality, situated in the Southern Cape. The District Municipality is applicable to this study because currently there is no BYOD management in place in local government and has aspects that pertain that it as an SMME.

The initial draft of the framework was based on a literature study, which was further justified through a process of cycles of refinement. The literature study and initial draft of the framework was presented during a visit to the District municipality. The literature study portrayed that there are various components that are composed within a BYOD framework. Therefore, a mind map of all the different components was illustrated.

A mind map also known as “brain map” or “mental map” was developed by Tony Buzan during the 1970s. It can be defined as an outline with ideas and pictures radiating out from a central concept (main idea). From the central concept key ideas radiate out, like the branches of a tree. The branches contain key words written in capitals over the line. [29].

Once the mind map has been drafted, a focus group was scheduled to substantiate the components on the mind map. A definition for a focus group is as follows; “*a group of interacting individuals having some common interest or characteristics, brought together by a moderator, who uses the group and its interaction as a way to gain information about a specific or focused issue*” [30]. Following the implementation of the focus group, a survey questionnaire was formulated which was inclusive of the proposed components to be contained in the framework. The questionnaire was constructed on an Excel spreadsheet and divided into the three components: BYOD Policy, BYOD Awareness Exercise and the BYOD Device Management interlaced in the BYOD Policy Plan hierarchy.

A second visit was scheduled to the District Municipality where formal semi-structured interviews were conducted with two representatives from the municipality. The semi-structured interview lasted approximately an hour and data was gathered through a survey questionnaire. The purpose of the semi-structured interviews was to further analyse a suitable solution for the municipal environment. An illustration of the process towards the implementation of the BYOD high level management framework discussed above is represented in Figure 6.

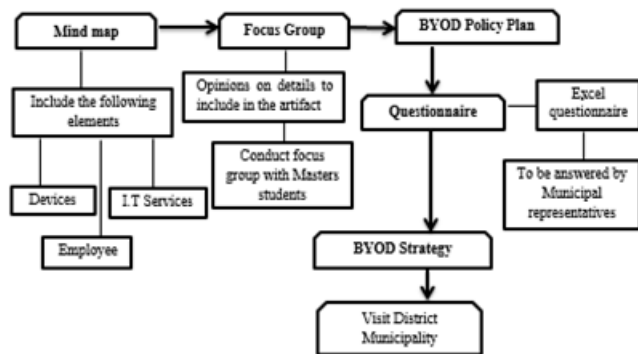


Figure 6: Process model for the BYOD high level management framework

The proposed BYOD high level management framework is a solution that wants to govern and manage BYOD within an SMME related environment. The previous section discussed four frameworks existing in literature that also aim to govern and manage BYOD. The upcoming section will provide an evaluation of the four existing frameworks in literature and the BYOD high level management framework.

IV. EVALUATING FRAMEWORKS FOR BYOD

BYOD is a phenomenon that warrants constant management and governance. The previous section, proposed a BYODMS framework and the preceding section provided four existing frameworks in literature. Consequently, when compared to the four existing frameworks in literature, the development of the BYOD high level management framework raises the question; is it a suitable solution compared to the solutions that currently exist? Therefore, this section will provide a critical evaluation of the BYOD high level management framework against the four existing frameworks discussed earlier in the paper.

Each of the four existing frameworks in literature and the BYOD high level management framework, have their benefits and when compared to each other, there are similarities that can be observed. Furthermore, the observation findings from the four existing frameworks provide elements that can be deemed as missing from the frameworks. Consequently, the BYOD high level management framework has been adapted to bridge the missing elements by fulfilling the eight BYOD characteristics an organization must follow for governing BYOD. Furthermore, when formulating the BYODMS, the SMME context was taken in account. Table 3 tabulates the similarities and missing elements between the BYOD high level management framework and the four existing frameworks from literature.

Table 3: Evaluation of frameworks

Frameworks	Similarities to BYOD high level management framework	Missing elements
BYOD security framework [5]	<ul style="list-style-type: none"> - Understands the business environment. - Registers BYOD devices - The organization has measures for device and information protection - Organizations provides continuous monitoring. 	<ul style="list-style-type: none"> - The framework is technical aspect of governing BYOD devices. - The SMME environment is not cited.
BYOD framework for a management system [24]	<ul style="list-style-type: none"> - Determine threats through a risk assessment - Develop strategies or policies for the governance of BYOD - Planning before policy implementation. An evaluation of the strategy. 	<ul style="list-style-type: none"> - The role of compliance isn't considered. - Adoption in an SMME environment is not cited.
BYOD privacy & culture governance framework [25]	<ul style="list-style-type: none"> - Determine the culture of the organization - Provide a clear definition for the respective privacy concerns - Develop a policy 	<ul style="list-style-type: none"> - The employee role is not considered. - The SMME environment is not cited.
Enterprise and BYOD space BYOD security framework [26]	<ul style="list-style-type: none"> - The organizational context must be considered. - There is a BYOD device analysis and IT administration. - There is a BYOD policy in place. - Compliance is incorporated. 	<ul style="list-style-type: none"> - There is a lack of adequate risk management. - The SMME environment is not cited.

The evaluation of the four existing frameworks in literature against the BYOD high level management framework tabulated in Table 3, indicate that they seemingly lack in addressing BYOD within an SMME environment. It could be hypothesized that the existing frameworks in literature address the governance of BYOD within large organizations. Thus, it can be determined that the BYOD high level management framework is the appropriate solution for the governance of BYOD in SMMEs. Furthermore, the BYOD high level management framework was developed with the eight BYOD characteristics in mind. Therefore, the BYOD high level management framework meets the eight BYOD characteristics that an organization should follow when implementing a governance oriented solution for BYOD in SMMEs.

V. CONCLUSION

BYOD is redefining how employees and organizations conduct daily business tasks. The adoption of BYOD in both large and small organizations governs an era where the filtration of personal and business is becoming blurry. The

risks associated to BYOD are undeniable. But, with proper governance, BYOD can be managed.

This paper studied and discussed the BYOD phenomenon and how BYOD is affecting SMMEs. It was derived that there is a need for a BYOD solution within an SMME environment and the solution should adhere to eight BYOD characteristics. As a result, four existing frameworks in literature were studied to determine if there is a solution that exists and meets the eight BYOD characteristics for an SMME BYOD solution. Once it was concluded that the four existing frameworks meet some of the characteristics but not all, the BYOD high level management framework was formulated. Following the formulation of the BYOD high level management framework, there was an evaluation of the frameworks for BYOD. Thus, it was determined that the BYOD high level management framework is an appropriate solution for BYOD. For future work, a suggestion of the formulation of a BYOD policy for SMMEs.

REFERENCES

- [1] D. Palacios-Marqués, P. Soto-Acosta, and J. M. Merigó, "Analyzing the effects of technological, organizational and competition factors on Web knowledge exchange in SMEs," *Telemat. Informatics*, vol. 32, no. 1, pp. 23–32, 2015.
- [2] L. A. Joia, "Measuring intangible corporate assets, linking business strategy with intellectual capital," *Intellect. Cap.*, vol. 1, pp. 68–84, 2000.
- [3] B. M. B. Suhail Qadir Mir, Mehraj-ud-din Dar, S M K Quadri, "Information availability: Components, Threats and Protection mechanisms," *J. Glob. Res. Comput. Sci.*, vol. 2, no. 3, 2011.
- [4] E. Fakhruddinova, J. Kolesnikova, O. Yurieva, and A. Kamasheva, "The Commercialization of Intangible Assets in the Information Society," *World Appl. Sci. J.*, vol. 27, pp. 82–86, 2013.
- [5] N. Zahadat, P. Blessner, T. Blackburn, and B. A. Olson, "BYOD security engineering: a framework & its analysis," *Comput. Secur.*, vol. 55, pp. 81–99, 2015.
- [6] K. Madzima, M. Moyo, and H. Abdullah, "Is Bring Your Own Device an institutional information security risk for small-scale business organisations?," 2014.
- [7] A. B. Garba, J. Armarego, D. Murray, and W. Kenworthy, "Review of the Information Security and Privacy Challenges in Bring Your Own Device (BYOD) Environments," *J. Inf. Priv. Secur.*, vol. 11, no. 1, pp. 38–54, 2015.
- [8] *The Role of IS Assurance & Security Management*, vol. 1. 2013.
- [9] A. A. Dedeche, F. Liu, M. Le, and S. Lajami, "Emergent BYOD Security Challenges and Mitigation Strategy Research Methodology," pp. 1–17, 2013.
- [10] B. Van Ommen, "IT Security in SMEs: Necessary or Irrelevant?," 2014.
- [11] *National Small Business Amendment Act*. 2004.
- [12] J. Devos, H. Van Landeghem, D. Deschoolmeester, and J. Devos, "Rethinking IT governance for SMEs," *Emerald*, 2012.
- [13] S. Kabanda and I. Brown, "Bring-Your-Own-Device (BYOD) practices in SMEs in Developing Countries – The Case of Tanzania," in *25th Australasian Conference on Information Systems*, 2014.
- [14] T. A. Yang, R. Vlas, A. Yang, and C. Vlas, "Risk management in the era of BYOD the quintet of technology adoption, controls, liabilities, user perception, and user behavior," *Proc. - Soc. 2013*, pp. 411–416, 2013.
- [15] S. Allam, S. V. Flowerday, and E. Flowerday, "Smartphone information security awareness: A victim of operational pressures," *Comput. Secur.*, vol. 42, pp. 55–65, 2014.
- [16] M. Hensema, "Acceptance of BYOD among Employees at Small to Medium-sized Organizations," *19th Twente Student Conf. IT*, pp. 1–8, 2013.
- [17] M. A. Harris, K. Patten, and E. Regan, "The Need for BYOD Mobile Device Security Awareness and Training," in *Proceedings of the Nineteenth Americas Conference on Information Systems*, 2013, no. January.
- [18] A. Weeger and H. Gewald, "Factors Influencing Future Employees Decision-Making to Participate in a BYOD Program: Does Risk Matter?," 2014, pp. 0–14.
- [19] S. Charbonneau, "The role of user-driven security in data loss prevention," *Comput. Fraud Secur.*, vol. 2011, no. 11, pp. 5–8, 2011.
- [20] Eslahi Meisam, Var Naseri Maryam, H. Hashim, N. M. Tahir, and E. H. M. Saad, "BYOD: Current State and Security Challenges," *IEEE Symp. Comput. Appl. Ind. Electron.*, pp. 189–192, 2014.
- [21] K. Dulaney and P. Debeasi, "Managing Employee-Owned Technology in the Enterprise," 2011.
- [22] A. C. Johnston, M. Warkentin, and M. Siponen, "AN ENHANCED FEAR APPEAL RHETORICAL FRAMEWORK : LEVERAGING THREATS TO THE HUMAN A SSET THROUGH SANCTIONING RHETORIC," vol. 39, no. 1, pp. 113–134, 2015.
- [23] A. M. French, C. Guo, and J. P. Shim, "Current Status , Issues , and Future of Bring Your Own Device (BYOD)," *Commun. Assoc. Inf. Syst.*, vol. 35, 2014.
- [24] M. Brodin, "Combining ISMS with strategic management : the case of BYOD COMBINING ISMS WITH STRATEGIC MANAGEMENT : THE CASE OF BYOD," no. August, 2015.
- [25] N. Selviandro, G. Wisudiawan, S. Puspitasari, and M. Adrian, "Preliminary study for determining bring your own device implementation framework based on organizational culture analysis enhanced by cloud management control," in *2015 3rd International Conference on Information and Communication Technology (IColCT)*, 2015, pp. 113–118.
- [26] Y. Wang, J. Wei, and K. Vangury, "Bring your own device security issues and challenges," *2014 IEEE 11th Consum. Commun. Netw. Conf.*, pp. 80–85, 2014.
- [27] M. Brodin, "Management issues for Bring Your Own Device," 2015.
- [28] H. Österle, J. Becker, U. Frank, T. Hess, D. Karagiannis, H. Kremer, P. Loos, P. Mertens, A. Oberweis, and E. J. Sinz, "Memorandum on design-oriented information systems research," *Eur. J. Inf. Syst.*, vol. 20, no. 1, pp. 7–10, 2011.
- [29] M. Davies, "Concept mapping, mind mapping and argument mapping: What are the differences and do they matter?," *High. Educ.*, vol. 62, pp. 279–301, 2011.
- [30] M. a Masadeh, "Focus Group : Reviews and Practices," *Int. J. Appl. Sci. Technol.*, vol. 2, no. 10, pp. 63–68, 2012.

PoPI Act - opt-in and opt-out compliance from a data value chain perspective: A South African insurance industry experiment

Paulus Swartz
School of Computing
University of South Africa
South Africa
paulus.swartz@absa.co.za

Adéle Da Veiga
School of Computing
University of South Africa
South Africa
dveiga@unisa.ac.za

Abstract—Personal information is collected and processed by various companies when individuals buy products and services, share their information on social media or enter their details in competitions and so on. This personal information, which could potentially also be shared with third party companies, is analysed to tailor services and products to consumer's preferences and online behavior, with the objective of creating a data value chain.

When the Protection of Personal Information (PoPI) Act (2013) comes into effect in South Africa, companies will have to comply with the conditions of PoPI and protect individuals' personal information accordingly. Companies will only be allowed to use personal information for the agreed purpose it was collected for and must obtain individuals' consent to share or further process their information.

This research sets out to monitor the flow of personal information through an experiment to establish if data value chains are shaped within the South African insurance industry, and to establish whether the consumer's personal information, which is part of the data value chain, is processed in line with certain conditions of PoPI.

The experiment highlighted that some of the insurance companies in the selected sample did not comply with the opt-in or opt-out preferences of the researcher. In addition some did not meet with the condition to obtain consent before sharing personal information with third parties for marketing purposes. No formal data value chains could be identified during the time frame of this experiment as it was found that the researcher was contacted randomly about generic marketing and communication offerings.

Keywords—Protection of Personal Information Act; PoPI; data value chain; direct marketing; opt-in; opt-out; personal information; privacy; e-mergent

I. INTRODUCTION

Privacy is not a new concept. It is the right of the individual to be free from secret observation and to determine with whom, how and whether or not to share personal information [1]. For most people "privacy" is a meaningful and valuable thing, but the term has different meanings in different contexts [2]. Privacy is an essential component of individual freedom, civil liberties, autonomy and dignity [3, 4]. The right to privacy

is the right to an individual's autonomy and personality, which is the individual's general right to immunity [3].

Individuals have a reasonable privacy expectation that companies, like telephone or internet providers, banks, government, medical practitioners and insurance companies would secure their personal information [3]. However, the right to privacy in the digital world is under attack as tracking surveillance is increasing and individuals' personal records are becoming more vulnerable while being stored digitally [3]. Contextual integrity is destroyed by selling or reusing digital information; even if users give their consent, they are not always aware of the purpose for which their information is later used [5]. Mismanagement of personal information processing, storage, use, collection and exchange can violate human rights. This could result in people losing trust in organisations, especially if the information is not secured and processed in accordance with regulatory requirements [6].

In South Africa the Protection of Personal Information Act (PoPI) (2013) was promulgated in November 2013 [7, 8]. This Act regulates the processing of personal information by public and private organisations domiciled in South Africa. PoPI includes a condition relating to unsolicited marketing, namely that consent is required in certain circumstances when existing or new customers are contacted. Organisations must comply with the conditions of PoPI and may only contact individuals in line with those conditions.

This research paper discusses research carried out to determine if customers' opt-out and opt-in preferences are honoured in the flow of personal information in the insurance industry, as required by PoPI. This research project forms part of a larger research project that honours students from the School of Computing at the University of South Africa participate in as part of their B.Sc. Honours degree.

The remainder of the research paper is structured as follows: Section II presents the research problem. This is followed by section III, which gives an overview of PoPI. Section IV discusses the research methodology and section V contains the results of the experiment. A discussion of the results and

limitations is presented in section VII, followed by the conclusion in section VIII.

II. RESEARCH QUESTIONS

The personal information processed by the insurance industry could be used to analyse individual preferences and to obtain competitive advantage with the aid of data value chains. These data value chains help organisations to make more effective strategic and operational decisions by directing services to specific customers or market segments [9]. However, data value chains could also pose a risk to the confidentiality and privacy of the information being shared with potential third parties or when used for purposes not agreed with the data subject (the person whose information it is).

Section 69 of PoPI prohibits unsolicited marketing unless the customer (data subject) consents to it. This means that new customers must opt-in for electronic communication, for example via cell phone text messages or e-mails. New customers may be contacted only once in order to obtain their consent (or opt-in) for marketing purposes.

Although it is thought that organisations have started the process of implementing the conditions of PoPI, many might not have. Once the provisions of PoPI come into effect, organisations will have one year to comply with the Act. Research [2] shows that only 12% of small and medium enterprises (SMEs) are in the process of complying, while 16% believe they are compliant, 56% are not aware of the conditions of PoPI and 16% are not compliant. Many organisations believe that it will require significant effort to become PoPI compliant, with some estimating that it could take in excess of 9 000 hours [10]. While some organisations have started with the implementation process, research indicates that it could take more than a year to become compliant, while many organisations believe that it could take up to three to five years to become compliant [30].

The following two research questions have therefore been formulated:

- What personal data value chains are there in the insurance industry in South Africa for the flow of personal information?
- Do South African insurance companies only contact customers if they have opted in for marketing and communication purposes as required by PoPI?

The answers to these research questions can indicate to organisations whether they comply with certain conditions for marketing in PoPI, and whether they are using clients' personal information to offer value services in the context of a value chain.

III. AN OVERVIEW OF PRIVACY LEGISLATION

A. International Privacy Regulation

Data protection laws have been adopted by over 100 countries and others are in the process of adopting privacy laws [8, 11]. The EU's Data Directive 95/48/EC is one of the

best-known privacy laws [12]; [13] and it has recently been updated to the General Data Protection Regulation (GDPR) [14]. The GDPR addresses new technological developments and harmonises national data protection laws across the EU member states [15].

According to the parliament text of the proposed GDPR, consent must be explicit and indicate affirmative agreement from the data subject, and is valid as long as personal information is processed for the purpose it has been collected for. Reference [12], argue that the objective of privacy legislation is to enable the individual to (i) manage or control the flow of personal information and (ii) to give the individual autonomous space.

Owing to the growth of modern computing, data protection laws have been implemented in many countries. In 1974 the United States of America drafted their privacy legislation. Germany followed in 1977 and France in 1978 [12]. South African citizens' personal information is also processed outside South Africa by multinational organisations and through the internet, which renders the information vulnerable [6]. The sensitivity of personal information changes as it flows through the economy, therefore the security and privacy requirements are dynamic [16] and should at all times be processed in line with the regulatory requirements of the relevant jurisdictions.

B. Protection of Personal Information Act (PoPI), 2 13

The purpose of PoPI is to provide a constitutional right to privacy by protecting the individual's personal information when processed by a responsible party. In this context the individual is referred to as a data subject, who is the "person to whom personal information relates" and who is an identifiable, living, natural or juristic person [17]. The responsible party is the, "public or private body or any other person which, alone or in conjunction with others, determines the purpose of and means for processing personal information" [17].

Personal information is information relating to the data subject, such as biographical information (e.g. race, gender, marital status, disability or religion), education, medical or financial information, e-mail and physical addresses, biometric information, and even information about personal opinions and views, including correspondence [17].

PoPI regulates the manner in which personal information may be processed, in line with international standards and established conditions, according to the prescription of the minimum threshold requirements for the lawful processing of personal information.

PoPI also provides for the rights of data subjects and remedies available to them, to protect their personal information from processing that is not in accordance with the Act. PoPI provides for the establishment of an information regulatory body with certain duties and powers in line with the conditions of PoPI and the Promotion of Access to Information Act (PAIA), 2000 [18]. Funds for the

Office of the Information Regulator have been approved by the Minister of Justice and Treasury [19] in support of the implementation of the sections relating to the Information Regulator.

PoPI has a significant impact on an organisation's policies, employees, information technology infrastructure, third party service providers and procedures if the organisation aims to comply with the provisions of the Act [20]. It impacts responsible parties that collect, process and store the personal information of customers, employees and third parties as part of their operational activities [7]. The next section gives some insight into the perceived positive and negative impacts of PoPI.

C. Positive impact of PoPI

PoPI will have a positive impact from an organisational and data subject perspective.

Preventive Measures: Responsible parties who collect personal information must be accountable and transparent, and should safeguard personal information according to condition 7 of PoPI [34]. According to [7], companies are now implementing proactive technical and organisational measures in the hope that these will prevent the leaking of personal information. These measures should ensure that companies' databases are secure to prevent data leakage and to protect their investments.

Transparency: Another advantage, according to reference[7], is that companies will be more transparent in terms of how, what and where personal information is stored within the company. Companies must notify data subjects when personal information is processed (section 18, [17]), and data subjects have the right to opt-in or out, free of charge, to receive marketing communication (section 69, [17]). Consent must be given before personal information is shared with third parties for marketing purpose (sections 11 and 20, [17]), therefore data subjects should not under normal circumstances receive unsolicited text messages or phone calls [7]. All businesses or parties responsible for big data and the analyses of an individual's habits, purchase behaviours or health status must treat the information as if it has been collected by means of questionnaires [21]. They must therefore be transparent in their use of the personal information, ultimately protecting the right of the individual while abiding by ethical principles.

Individuals' Rights: If data are inaccurate, misleading, excessive or incomplete, or if data have been obtained unlawfully, data subjects can rightfully request an update, deletion or correction of their personal information according to section 16 of PoPI [22]. Reference [21] argues that the laws protecting the privacy of personal data give individuals rights to all their data, irrespective of the source. PoPI also enables individuals to institute civil proceedings under certain circumstances if there has been interference with the protection of their personal information (sections 5 and 99 of [17]).

D. Negative Impact of PoPI

Many organisations believe that PoPI will have a negative impact on them.

Marketing Costs: The Consumer Protection Act [23] of South Africa only allows for an opt-out mechanism. Section 11(5) of the Consumer Protection Act, 2008, states that if a consumer opt-out to receive direct marketing, no person must charge the consumer a fee to effect it. PoPI stipulates that affirmative consent is required, which means that individuals have to opt-in to receive direct marketing messages (section 69, [17]). PoPI also requires that the customer be given reasonable time to object, at no cost to the data subject, which means that the business is responsible for all costs when the customer opts out at a later stage [24]. Companies must update their IT systems to flag the option to opt-in or opt-out of direct marketing (section 11, [17]). Company processes for responsible parties and third parties must also be updated according to section 13 of PoPI, with provision that personal information can only be shared if the purpose is specific, the quality of information is ensured (section 16, [17]) and the information is safeguarded (section 19, [17]). This has an impact on the system design and administration process, on contracts with third parties.

Infrastructure Cost to Company: Critics have warned that the PoPI regulatory scheme will discourage economic activity and put undue burdens on businesses [25], because many businesses will have to make supplementary investments in information technology systems or use third-parties vendors in order to comply with PoPI.

Compliance Time Frames: To be compliant within one year is impracticable, as shown by a survey conducted by South African businesses in 2013. It could take up to three years to become fully compliant [10, 25]. Organisations have to overcome huge challenges to become compliant. Companies that are already implementing measures to comply with PoPI requirements are concerned that they will not be compliant in time [10]. A study done by Cibecs in 2012 shows that 26% of South African companies are in the process of complying with the requirements of the PoPI Act, and are therefore upgrading their IT infrastructure measures [26]. Research [26] indicates that as many as 38% of the companies surveyed still have outdated compliance measures in place.

E. Data Value Chains

A data value chain is the management and coordination of data across the service continuum, where a collaborative partnership is formed, and where data collection is coordinated from various stakeholders while analysing the data to optimise service delivery and product development [27]. Data that is generated by companies facilitates re-use and value generation based on the data, over and over again [28]. The European Commission states that the data value chain is the, "centre of the future knowledge economy, bringing the opportunities of the digital developments to the more traditional sectors (e.g. transport, financial services, health, and manufacturing)" [28].

Personal information can be utilised in data value chains to provide an improved service offering through focussed marketing, which can allow organisations to channel identified services to specific customers. However, customers must willingly share their personal information and organisations must only use it in line with the consented preferences of the customers.

F. Use of Personal Information for marketing purposes

Direct marketing entails that the marketer communicates directly with a customer or client in the hope that the customer will respond positively to the marketer's request [29]. Any type of electronic communication, like a text message (SMS) or a video message (MMS) to a mobile telephone, e-mails, mobile device application advertising and social media marketing, is a tool used by the marketer to advertise a service or products. Section 11 of the Consumer Protection Act (CPA) 68 of 2008 stipulates that every person has the right to privacy and to refuse to accept any approach or communication if the purpose is for direct marketing. According section 69 of PoPI, the processing of personal information for the purpose of direct marketing is prohibited, unless the marketer has the consent of the data subject, is a customer of the responsible party, the responsible party has the customer's contact details and they market similar products or services of the responsible party to the data subject.

Subsection 4 of section 69 of [17] states that the identity and address or contact details of the sender must be known to enable the recipient to respond to the request if they wish to do so. Consumer preference information is used by direct marketers to combine groups of consumers with the same interest and taste, and this information is beneficial for businesses as well as for the consumer to receive communication messages according to their personal preferences [30], which relate to a data value chain.

According to section 69 of PoPI, the customer must grant permission for the processing of personal information and must also have the option to cease any communications. The consent option for processing personal information is referred to as "opt-in", and the rejection of future communications from the marketer is referred to as the "opt-out" option. Consumers are sometimes misled about their choice to opt-in or opt-out on the company's websites or application forms. For example, the default setting on most websites is to opt-out, or the questions that are asked ("Please send me newsletters" or "Please do not send me newsletters") are trivial and might influence consumer decisions [31]. Because of inattention, and cognitive and physical laziness, default answers are given. Often the opt-in option is ticked by default [32]. Marketers tend to set the "yes" option as default if they need the consumers to opt-in for the processing of personal information [32]. As such compliance with PoPI could impact negatively on organisations' freedom to use marketing and communication initiatives.

IV. THE INSURANCE INDUSTRY

To be competitive in the insurance industry, companies have to market their products. Cold-calling is a method used by insurance companies to market their products, and according to [33], the Financial Advisory and Intermediary Services Act (FIAS) 37 of 2002 [34] and CPA address the issue, but it is PoPI that will eliminate the cold-calling sales technique. According to a study done in the health services, 6% of data breaches are committed by insurance companies, the third highest out of 17 industries [35]. Cybersecurity insurance is expanding rapidly in the insurance market, with a forecast of \$7.5 billion in annual sales globally by 2020 by the global cyber insurance market [36, 37]. If an insurance company wants to provide cybersecurity insurance in South Africa, it must set an example and comply fully with the requirements of PoPI. The research reported in this paper has focused on the insurance industry, because it processes large quantities of personal information. The research results can provide the insurance industry with insight into possible gaps in compliance with PoPI.

V. RESEARCH METHODOLOGY

This section outlines the research methodology.

1) Positivism Paradigm

The positivism paradigm applies to this research. A postivism paradigm is based on realist ontology beliefs, where there is an object reality according to representational epistemology where symbols are used to explain and describe this objective reality accurately [38, 39]. Reference [39] states that positivism can reveal the causal relationship that exists within social life, such as the flow and use of personal information in the economy.

2) Experimental Design

An experimental design was used for this research project. This design allowed the researcher to have full control over the experiment and strengthens the internal validity [40]. Reference [41] suggest that if the experiment is carried out correctly, the testing effect, mortality, history and maturation, as possible pitfalls of internal validity, will not have an effect on the research outcome. Nevertheless, these pitfalls were avoided. Two groups were involved in the research, namely the experiment group and the control group; a stimulus was applied to the experiment group and no stimulus was applied to the control group [41].

3) Sample

This research focused on the insurance industry in South Africa. The insurance industry collects personal information through online applications, telephonic marketing, hard copy applications and also their claim process.

The geographical area was limited to South Africa. The head offices of the insurance companies included in the sample are mainly located in the metropolitan cities of each province.

Twenty insurance companies were included in the sample. The sampling method used for this research project was a convenience sample [36]. A prerequisite for inclusion in the sampling was that the insurance company had to have a website where online insurance applications could be requested.

B. Research Design

1) Experiment Preparation

To conduct the experiment, two new cell phone numbers and six new e-mail addresses were utilised, which allowed the researcher to supply his personal information when requesting online quotations from the sample of insurance companies.

Once the researcher had requested quotes from an insurance company via its website, the researcher’s personal information, including a cell phone number and/or e-mail address, was requested and processed by the insurance company. The information was therefore included in the customer records in the companies’ databases. In this way the researcher’s personal information was deposited in the economy, and the researcher was able to monitor the flow of his personal information (each new cell phone number and e-mail address). The researcher used an identical profile in his dealings with all the insurance companies selected for the research. While the research project was undertaken, the cell phone numbers and e-mail accounts were not used for any other purpose.

Table 1 shows the two new cell phone numbers, Cell phone A and Cell phone B, which were created in March 2015 for the purpose of this research. When the researcher purchased sim cards from the service provider, his information was verified in line with the Regulation of Interception of Communications and Provision of Communication-Related Information (RICA) Act, 2002 [43].

Company Name	Cell phone A / E-mail A	Cell phone B / E-mail B
Company A - J	Opt-In	Opt-Out

Table 1: Group 1 Companies

The six e-mail accounts were created in April 2015. Two of the six e-mail addresses were linked with the cell phone numbers in Table 1. The cell phone numbers and related e-mail addresses were disclosed to the first ten insurance companies (see Table 1). This experiment was conducted from a consumer perspective and hence to protect the confidentiality of the companies in the sample their names are withheld.

The remaining four e-mail addresses (see Table 2) were included in the personal information supplied to the next ten insurance companies in the sample. Combination of two email addresses was used in group 2 for opt-in and opt-out to determine the data value chains that will be created without

the cellphone numbers linked to the email addresses. It was found that no information could be submitted without a cell phone number. Cell phone A was therefore also submitted with e-mail addresses C and D, and cell phone B was submitted with e-mail addresses E and F.

Company	Cell phone A / E-mail C	Cell phone A / E-mail D	Cell phone B / E-mail E	Cell phone B / E-mail F
Company K - T	Opt-In	Opt-In	Opt-Out	Opt-Out

Table 2: Group 2 Companies

In order to monitor whether the opt-in and opt-out preferences were maintained, the researcher opted in for all communication when requesting online quotations using cell phone number A, and opted out for all online quotations using cell phone number B.

For the control group, the researchers purchased four cell phone sim cards. The researchers activated the numbers on the network by sending at least one SMS to another number. No stimuli were applied to these numbers, meaning that the researcher did not disclose the cell phone numbers to any company nor use it for phone calls or text messaging. This would eliminate any biased results during the experiment, because the cell phone numbers were not subjected to any experimental treatment.

2) Conducting the Experiment

Personal information was disclosed to the insurance market in May 2015. The method used to disclose personal information was to request life insurance or short-term policy quotations from insurance companies using the online application tools on the companies’ websites.

3) Data Collection

Data were collected by means of telephone calls, SMSs and e-mail messages received from companies that contacted the researcher on either of the two cell phone numbers or any of the six e-mail addresses created for this experiment. Information about each telephone call and SMS was recorded daily on a spreadsheet, and information about e-mail messages received was recorded twice per week.

The time frame for collection was restricted to the period March to October 2015 to accommodate students enrolled for the one-year Unisa honours degree module.

During this time the researcher recorded certain aspects, such as the origin of contact details; whether the researcher opted in for the communication; whether there was an option to opt-out of any future communication; whether the researcher was liable for any cost when opting out; and whether the researcher was contacted by an automated telephone. Questions such as “Where did you get my contact details?” and “Do you have my name and surname?” were asked to

telemarketers or call centre callers who telephoned the researcher. This provided the researcher with an indication of whether the researcher was known to the company and whether calls were made to random numbers. It also helped the researcher to establish whether they obtained his personal information as a result of the online insurance application process.

4) Findings

Table 3 outlines a summary per company of the number of times the researcher where contacted where the opt-in or opt-out preference was selected. In total the researcher was contacted 84 times during the data collection period of which 47 contacts were linked to the companies included in the sample and 37 were related to companies who contacted the researcher were not part of the sample 55% of all communications were received via SMSs and 28% via e-mail messages. Telephone calls only accounted for 17% of his contact with direct marketers. The 28% e-mail messages that were received were generated when quotations had been requested from insurance companies.

Thirty percent of the total communications received (84) related to contacts where the customer opted in for communication by the insurance companies. For 70% of the contacts the researcher had not opted in for communications by entities indicating that the opt-out preference had not been complied with.

Company Name	Opt – In Number of contacts	Opt-Out Number of Contacts
Company A	1	3
Company B	0	0
Company C	0	1
Company D	1	1
Company E	1	1
Company F	1	0
Company G	2	0
Company H	0	0
Company I	1	0
Company J	2	0
Company K	1	8
Company L	0	0
Company M	2	0
Company N	2	0
Company O	6	6
Company P	3	0
Company Q	2	0
Company R	0	0
Company S	2	0
Company T	0	0
Total	27	20

Table 3: Number of contacts received per company for opt-in and opt-outs

Phone calls were received from the insurance companies who called about the quotations requested. Only 22% (10 out of 46) SMSs sent by the insurance companies had the researcher’s personal information; 18% were sent by the cell phone service provider.

The remaining 60% of SMSs received came from entities who had no permission to contact the researcher and who had no information about the researcher.

Where the researcher opted out for marketing communications, 20 communications were received. Of these, 30% were received from cell phone B combined with e-mail B, which indicated non-compliance with the opt-out reference.

The other 70% were received from cell phone B combined with e-mails E and F. Similarly, where the researcher opted in for future communications, a total of 27 communications were received of which 33% were from e-mail A linked to cell phone A and 67% were from e-mails C and D.

Interestingly, in 37 of the instances the researcher was contacted by 18 different companies who were not part of the sample. This indicated that the information could have been shared with third parties who used it to contact the researcher.

These companies did not have permission to contact the researcher for marketing purposes via the cell phone numbers and e-mail addresses used in the research. Most of these companies only contacted the researcher once, but 2 companies contacted the researcher at least 8 times each during the experiment to offer financial services or to notify him that he had won a competition.

On cell phone B, where the researcher opted out for communication, the researcher received 9 calls, seven SMSs and four e-mail messages. The researcher received 5 calls, two SMSs and 20 e-mail messages on the cell phone number where opt-in was elected for communication.

The results indicate that half (six out of 12) of the contacts made by Company “O” were permitted and used e-mail address C or D. There was no consent for the other half of the communications received from Company “O”, as the researcher had opted out when using e-mail addresses E and F. There was no consent for 90% (8 out of 9) of contacts made by Company “K” – there was a privacy disclaimer on the website regarding the protection of personal information that said the researcher would only be contacted about a requested quotation.

Company “A” contacted the researcher 4 times. This company also had a privacy disclaimer on their website, indicating that they would protect the researcher’s personal information and would only contact the client about a quotation requested. The researcher elected to opt-out of communication from Company “A”, but no option was provided to opt-out during the application process. The websites of Company A and Company K did not offer opt-in or opt-out options on their application/quotation systems, but they did include privacy

disclaimers that promised to protect the customer's personal information.

38% of the entities that contacted the researcher did not have the researcher's personal details, and it was unknown how the researcher's contact details had been obtained for 35% of the communications received.

Only 43% of the SMSs received included the option to opt-out of communications. Most of the 43% of the SMSs that included the option to opt-out, indicated that standard rates would apply to opt-out. None of the phone calls received were from an automated calling machine.

The control group received a total of 70 communications, 9 missed calls and 61 SMSs. Three of the cell phone numbers (Cell Phone Provider I; Cell Phone Provider II; and Cell Phone Provider III) did not receive any communication except SMSs from organisations. Cell Phone Provider IV accounted for 9 missed calls, of which six were from different numbers, as well as six SMSs. These SMSs were messages from financial service providers or a message that a competition had been won. This cell phone number might have been owned and used by another individual in the past, which could explain these messages. This can be further investigated to determine the source, why the cellphone number is link to a marketing database.

VI. DISCUSSION

The researcher had only given consent for 33% of the 84 communications received (e-mail messages, SMSs and telephone calls). The opt-out group received more calls and sms's (9 calls, 7 sms's) than the opt-in group (5 calls, 2sms's), during the research. Seventy six per cent of the entities had not obtained the researcher's consent to contact him. Section 69(1) of PoPI stipulates that data subjects must give their consent to the responsible party to process their personal information and must opt-in for marketing purposes. 42% of the responsible parties did not have any consent to contact the researcher for marketing purposes. In addition, Companies 1 to 18 were not even part of the sample. This answers research question 2, indicating that customers are contacted even though they have not opted in for marketing and communication purposes as required by PoPI.

This indicates that insurance companies might not be fully compliant as yet, as some do not have the proper consent options (opt-in/opt-out mechanism) for the client (data subject) or do not abide by it, while others have disclaimers on their websites, but do not abide by them.

At the time when the data were submitted via the insurance companies' websites, only five out of 20 companies made provision for clients to opt-in or opt-out for any marketing communications. In future, organisations will have to give new customers the option to opt-in for marketing and communication, and allow existing customers to opt-out at

any time for marketing and communication purposes, as per section 69 of PoPI.

There were 2 companies that included a privacy disclaimer on their websites, stating that they valued the researcher's personal information, would protect it and would only contact the researcher about the product or service he is interested in. However, these companies did not comply with section 69 of PoPI, which states that a data subject must give his/her consent for the processing of personal information for marketing purposes.

Almost half of the SMSs received were sent by entities that did not give the researcher the option to opt-out of direct marketing communications. This means that the responsible persons or third party did not comply with the regulations of PoPI. Data subjects must be given the option to opt-out of or withdraw their consent for the processing of information and future marketing communications from third parties as per section 69(4b) of PoPI.

Direct marketing from the companies that were not part of the sample did not comply with PoPI regulations, as these companies contacted the researcher via SMS for marketing purposes without consent to do so.

The third party or organisation responsible for direct marketing must supply its address or contact details as per section 69(4b) of PoPI, to enable recipients to opt-out of any future communication. 43% of the companies that sent SMSs without an opt-out option did not comply with PoPI. Section 69(3) of PoPI, states that the customer must have the option to consent or cease communication at free will, at any future marketing communications.

Because SMSs were received from unknown senders as well as organisations who were not included in the sample, it was difficult to establish the origin of all messages or how personal information was leaked or shared to these entities, because the researcher was in no position to confirm how the entity got the information to make contact with the researcher. However, these messages indicated that data, specifically personal information, were shared in the economy with third parties as the cell phone numbers and e-mail addresses were only used when submitting the information on websites of insurance companies included in the sample.

In conclusion, it was found that at the time of the research there were no significant personal data value chains created by the insurance industry in South Africa for the flow of personal information, which answered research question 1. The researcher was contacted randomly and not for products or services tailored to his demographic profile. It was also concluded that the researcher had been contacted by companies that had no consent, which answered research question 2.

VII. LIMITATIONS

For the purpose of the experiment it was assumed that companies were in the process of becoming compliant with

PoPI as it has been promulgated for three years now and companies will only have one year to become compliant once in effect. A limitation was that the conditions of PoPI, apart from those relating to the establishment of the Information Regulator, were not yet enforced, which means that companies do not have to be compliant as yet unless they are a multinational organisation operating in other jurisdictions with data protection laws. This could be why the results of the research indicated non-compliance for certain sections and conditions of PoPI. Taking into consideration that it could take between three to five years to become compliant it is anticipated that companies should have started to implement measures to prepare for compliance.

Another limitation of the research project was the limited time frame available to monitor communication to the cell phones and e-mail addresses created. This could not be avoided, because the honours project had to be concluded within a year.

A further limitation was that some cell phone numbers had previously belonged to other people, therefore some of the communications received via sms during the research might have been meant for the previous owner of a cell phone number. Not all communications were therefore necessarily applicable to the research.

Another limitation to consider in the research project is that there was no control over the information processed in line with the RICA Act of 2002 [43] by the store and the service provider from whom the sim cards were purchased. Personal information could also have been leaked during this process.

VIII. CONCLUSION

The objective of this research was to establish if any personal information value chains were created in the insurance industry through the flow of personal information and secondly, whether certain conditions of PoPI were complied with from a marketing perspective. An experimental design was used within the insurance industry of South Africa as a sample population. This was investigated by establishing whether the customer (data subject) was contacted by the companies in the selected sample if they had not opted in for any communication.

The results indicated that 67% of the entities did not have the researcher's consent during the research to contact the researcher. In addition, the senders of 57% of SMSs had not given the researcher the option to opt-out. The researcher was contacted by companies who were not included in the sample, indicating that data could have been shared or leaked. It was also found that the researcher was contacted randomly for ad hoc marketing that was not tailored to the researcher's demographic traits. No significant personal data value chains could be identified.

Future research using a longer time frame is necessary to monitor the data flow, and to further investigate the establishment of data value chains and compliance with PoPI. Additional value will be added if the experiment is repeated once PoPI commences.

IX. REFERENCES

- [1] Chen, L.F. and Ismail, R., "Information Technology program students' awareness and perceptions towards personal data protection and privacy", 3rd International Conference on Research and Innovation in Information Systems, ICRIS, pp. 434–438, 2013.
- [2] Doyle, C. and Bagaric, M., "The right to privacy: appealing, but flawed", *The International Journal of Human Rights*, Vol. 9 No. 1, pp. 3–36, 2005
- [3] Hiranandani, V., "Privacy and security in the digital age: contemporary challenges and future directions", *The International Journal of Human Rights*, Vol. 15 No. 7, pp. 1091–1106, 2011.
- [4] Van der Sloot, B., "Do privacy and data protection rules apply to legal persons and should they? A proposal for a two-tiered system", *Computer Law & Security Review*, Elsevier Ltd, Vol. 31 No. 1, pp. 26–45, 2015.
- [5] Goodman, E., "Design and ethics in the era of big data", *Interactions*, Vol. 21 No. 3, pp. 22–24, 2014.
- [6] Borena, B., Belanger, F. and Ejigu, D., "Information Privacy Protection Practices in Africa: A Review Through the Lens of Critical Social Theory", 2015 48th Hawaii International Conference on System Sciences Information, pp. 3490–3497, 2015.
- [7] De Bruyn, M., "the Protection of Personal Information Act and Its Impact on Freedom of Information", *International Business & Economics Research Journal*, Vol. 13 No. 6, pp. 1315–1340, 2014CIBECs, "2012 State of business data protection in South Africa", Available from: <http://offers.cibecs.com/state-of-business-data-protection-in-sa>, (Accessed 23 March 2015), pp.14, 2012
- [8] Milo, D. and Ampofo-anti, O., "A not so private world", *Without Prejudice*, Vol. 14 No. 09, pp. 30–32, 2013.
- [9] Prinsloo, P., Archer, E., Barnes, G., Chetty, Y., & van Zyl, D., Big(ger) data as better data in open distance learning. *International Review of Research in Open and Distance Learning*, Vol. 16 No. 1, pp. 284–306, 2015.
- [10] PricewaterhouseCoopers (PwC), "The protection of personal information bill: The journey to implementation", Available from: <https://www.pwc.co.za/en/assets/pdf/pop-white-paper-2011.pdf> (Accessed 24 February 2016), 2011.
- [11] Greenleaf, G., "Sheherezade and the 101 Data Privacy Laws: Origins, Significance and Global Trajectories", *Journal of Law, Information Science*, Vol. 23 No. 1, pp 1-48, 2014.
- [12] Olinger, H.N., Britz, J.J. and Olivier, M.S., "Western privacy and/or Ubuntu? Some critical comments on the influences in the forthcoming data privacy bill in South Africa", *International Information and Library Review*, Vol. 39 No. 1, pp. 31–43, 2007.
- [13] Directive 95/46/EC of the European Parliament and of the Council of 1995. Available from : http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf (Accessed 24 February 2016)
- [14] General Data Protection Regulation (GDPR) of 2012. Available from: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012PC0011&from=EN> (Accessed 24 February 2016).
- [15] Hunton and Williams, "The proposed EU General Data Protection Regulation: A guide for in-house lawyers", Available from: https://www.huntonregulationtracker.com/files/Uploads/Documents/EU%20Data%20Protection%20Reg%20Tracker/Hunton_Guide_to_the_EU_General_Data_Protection_Regulation.pdf (Accessed 24 February 2016), 2015.
- [16] Diaz-Tellez, Y., Bodanese, E.L., Nair, S.K. and Dimitrakos, T., "An architecture for the enforcement of privacy and security requirements in internet-centric services", *Proc. of the 11th IEEE Int. Conference on Trust, Security and Privacy in Computing and Communications, TrustCom-2012 - 11th IEEE Int. Conference on Ubiquitous Computing and Communications, IUCC-2012*, pp. 1024–1031, 2012.
- [17] Protection of Personal Information Act (PoPI) 4 of 2013, South African Government, Available from: <http://www.acts.co.za/consumer-protection-act-2008/index.html> (Accessed 20 June 2015).
- [18] Promotion of Access to Information Act (PAIA) 2 of 2000, South African Government, Available from: <http://www.acts.co.za/promotion-of-access-to-information-act-2000/index.html> (Access 20 June 2015).

- [19] Heyink, M., Funds Approved for Establishment of Privacy Regulator, Privacy Online, Available from: http://www.privacyonline.co.za/news/2015/06/Funds_Approved_Establishment_Privacy_Regulator (Accessed 20 June 2015)
- [20] Pillay, L., “The partial commencement of the Protection of Personal Information Act, 2013”, Without Prejudice, Vol. 14 No. 8, p. 54, 2014.
- [21] Wilson, S., “Big data held to privacy laws, too”, Correspondence, Macmillan Publishers Limited., Vol. 519, p. 414, 2015.
- [22] Magolego, B.N., “Personal data on the Internet – can POPI protect you?”, De Rebus, No. 548, pp. 20–22, 2014.
- [23] Consumers Protection Act (CPA), 68 of 2008. South African Government, Available from: <http://www.acts.co.za/consumer-protection-act-2008/> (Accessed 16 October 2015).
- [24] Calaguas, M., “South African Parliament Enacts Comprehensive Data Protection Law: An Overview of the Protection of Personal Information Bill”, Africa Law Today, No. 3, pp. 1–6, 2013.
- [25] Swart, I.P., Grobler, M.M. and Irwin, B., “Visualization of a data leak”, 21st Conference on the Domestic Use of Energy, pp. 1–8, 2013.
- [26] Botha, J.G., Eloff, M.M. and Swart, I., The effects of the PoPI Act on small and medium enterprises in South Africa. In Information Security for South Africa (ISSA), 2015 (pp. 1-8). IEEE, 2015.
- [27] Miller, H.G. and Mork, P., 2013. From data to decisions: a value chain for big data. IT Professional, 15(1), pp.57-59.
- [28] European Commission. (2013) A European strategy on the data value chain. Retrieved from <https://ec.europa.eu/digital-agenda/en/news/elements-data-value-chain-strategy>. (Accessed 07 July 2013)
- [29] Hamann, B. and Papadopoulos, S., “Direct marketing and spam via electronic communications: An analysis of the regulatory framework in South Africa”, De Jure, Vol. 47 No. 1, pp. 42–62, 2013.
- [30] Dolnicar, S. and Jordaan, Y., “A Market-Oriented Approach to Responsibly Managing Information Privacy Concerns in Direct Marketing”, Journal of Advertising, Vol. 36 No. 2, pp. 123–149, 2007.
- [31] Lai, Y.-L. and Hui, K. L., “Internet Opt-In and Opt-Out: Investigating the Roles of Frames, Defaults and Privacy Concerns”, 2006 ACM SIGMIS CPR Conference on Computer Personnel Research, pp. 253–263, 2006.
- [32] Bellman, S., Johnson, E.J. and Lohse, G.L., “On site: to opt-in or opt-out?: it depends on the question”, Communications of the ACM, Vol. 44 No. 2, pp. 25–27, 2001.
- [33] Millard, D., “Hello, POPI? On cold calling, financial intermediaries and advisors and the Protection of Personal Information Bill”, Journal of Contemporary Roman-Dutch Law, Vol. 76, pp. 604-622, 2013.
- [34] Financial Advisory and Intermediary Services (FIAS) Act, 2002 (Act No. 37 of 2002). Available from: <http://www.acts.co.za/financial-advisory-and-intermediary-services-act-2002/> (Access 5 March 2016).
- [35] Widup, S., Bassett, G., Hylender, D., Rudis, B., Spittler, M., “2015 Protected Health Information Data Breach Report”, Available from: http://www.verizonenterprise.com/resources/reports/rp_2015-protected-health-information-data-breach-report_en_xg.pdf, (Accessed 5 March 2016), 2015.
- [36] PricewaterhouseCoopers (PwC), “Turnaround and transformation in cybersecurity”, Available from: <https://www.pwc.com/gx/en/consulting-services/information-security-survey/assets/pwc-gsiss-2016-financial-services.pdf> Accessed 5 March 2016), 2015.
- [37] CIBECS, “2012 State of business data protection in South Africa”, Available from: <http://offers.cibecs.com/state-of-business-data-protection-in-sa>, (Accessed 23 March 2015), pp.14, 2012
- [38] Brewer, J.D., “The A-Z of Social Research Positivism”, SAGE Research Methods, pp. 236–238, 2015.
- [39] Cohen, D. and Crabtree, B., The Positivist Paradigm, Available from: <http://www.qualres.org/HomePosi-3515.html>, (Accessed 27 June 2015), 2008.
- [40] Staller, K., “Encyclopedia of Research Design”, *Encyclopedia of Research Design: Qualitative Research*, pp 1159-1164, 2010
- [41] Miller, R.L. and Brewer, J.D., “The A-Z of Social Research Research design”, SAGE Research Methods, pp. 263–269, 2003.
- [42] Seltman, H.J., “Experimental Design and Analysis”, p. 35., 2013
- [43] Regulation of Interception of Communication and Provision of Communication –Related Information Act (RICA), Act 70 of 2002, South African Government, Available from: <http://www.acts.co.za/regulation-of-interception-of-communications-and-provision-of-communication-related-information-act-2002/> (Accessed 24 February 2016).

An Interactive Visual Library Model to Improve Awareness in Handling of Business Information

Petrus M.J. Delpoort, Mariana Gerber, Nader S. Safa

Centre for Research in Information and Cyber Security

NMMU

Port Elizabeth, South Africa

S211253502@nmmu.ac.za, Mariana.Gerber@nmmu.ac.za, sohrabisafa@yahoo.com

Abstract— Information technology has changed organisational processes significantly. However, information security is still a controversial issue among experts in this domain. Information security breaches lead to loss of reputation, competitive advantages, intellectual properties, productivity, and revenue and in the worst scenario leads to bankruptcy. In this regard, awareness plays a vital role to mitigate information security threats. This study aims to present different threats that effect confidentiality, integrity and availability of information, pertaining to administrative employees, in an integrated and informative design, based on a review of literature. In addition, a possible interactive visual library is proposed, through a proof of concept that contributes to administrative employees' information security awareness. The results shed some light on this information security awareness issue, and provides the means for further academic study.

Keywords— information security; administrative employees; business information; sensitive data; awareness; threats

I. INTRODUCTION

Information has become the lifeblood of modern organisations and core to most business processes [1]. Due to this, an organisation could seriously be harmed or even become bankrupt if proper information security is not implemented and maintained. According to the reputable company named Bitdefender, employees of an organisation remain the weakest link in the organisational sphere and may consequently pose a serious threat to information security [2]. Another study undertaken by Whitman, suggests that the third highest information security threat in an organisation is the act of human error or failure [3]. This clearly shows the importance of employees' behaviour in the domain of information security. One can clearly see that without proper awareness, an organisation's employees could compromise business information. The term business information is used as an amalgamation of valuable data, sensitive data and information within a business environment, essentially any information that is considered an asset to the organisation. In reference to an organisation's employees, this paper focusses purely on the role that administrative employees, sometimes called administrative assistants, play whilst handling business information. An administrative employee is defined as a person that is employed to assist with various clerical tasks in an office

setting such as correspondence, keeping records, making appointments, and carrying out similar tasks [4]. This paper classifies administrative employees as the main point of contact in most organisations, whether by phone, email or physical contact due to the role of liaising with customers, and even potential attackers, on a daily basis.

Consequently, without administrative employees having proper knowledge of existing threats, the organisation may undergo severe repercussions. Therefore, the objectives of this paper is to firstly, focus on the role of administrative employees and actions they perform, especially when encountering certain threat scenarios and how these threats can compromise business information. Secondly, it will also propose a possible solution, through a proof of concept, which can contribute in raising the administrative employees' awareness on secure handling of business information.

With the mentioned background in mind, this paper will continue to discuss the research approach. Thereafter the problem will be explored in more detail by contextualizing information security threats. The paper will then end off by providing a proof of concept to the proposed interactive visual library.

II. METHODS/APPROACH

The following methods were used to address the objectives of this paper.

Firstly, a literature review was done in order to identify top threats that exist within a typical organisation. After comparing the top threats from various sources, the top five threats that focus on organisational environments were extracted which provide the input into the survey.

Secondly, a survey in the form of a questionnaire was done. Typically a questionnaire is done to identify a specific pattern or behavioural existence within a certain group [5]. The questions were adopted from previous studies in this domain [3]. The focus of this survey is administrative employees within an organisation. In order to have participated in this survey, the participant had to adhere to the following three characteristics. These characteristics are as follows:

- Participant must be an administrative employee within any sector except the IT sector, as it might provide a bias result.
- Participant has to work with a computer on a daily basis.
- Participant must handle business information on a regular basis

Only when a participant adhered to these three characteristics, was the participant able to participate in the survey. Each participant, which belong to various institutions or organizations, received the same questions and a specific pattern or behavioural existence was identified. The data were analysed and reported on, which showed a sample of the current awareness levels regarding administrative employees within organisations.

Thirdly, argumentation was used in order to identify critical aspects in creating an interactive visual library, which aims to address the awareness levels of administrative employees within a typical organisation.

Lastly, a prototype was discussed which served as a proof of concept. Using this research approach, the following section will start contextualizing information security threats

III. ADMINISTRATIVE EMPLOYEES VS. THREATS

Information security techniques may lose their usefulness if misused, misunderstood, or not used by end-users. Due to this, information security awareness is a crucial activity in any organisation [6]. It is clear that without awareness or knowledge of information security threats, one might become a victim within an organisational environment and most likely compromise business information. As mentioned previously, information is the lifeblood of an organisation and therefore administrative employees must securely handle business information.

It is essential to have a clear understanding of what a threat is within an organisation; therefore, the following subsection will identify typical threats within an organisation.

A. Threats within an Organisation

In order to address the objective of this paper, the first step is to identify the top five threats administrative employees might encounter whilst handling business information within an organisation.

A threat, regarding information security, is a category of entities that present a danger to a current asset [7]. The asset is anything that the organisation finds vital to keep secure in order to continue normal business operation. The threat typically moves the asset from a secure state to an insecure state whereby compromising either confidentiality, integrity or the asset's availability. When an asset is compromised, the organisation could suffer tremendously. Many threats exist which could lead to compromised assets, however this paper focusses on the threats that fall under the category of "Act of human error, or failure" as identified by Whitman [3]. By combining literature [8] and various online threat reports, such as the annual Threat Horizon report [9], a cross examination,

based on highest occurrence, revealed the top threats that exist within a typical organisation [10][11]. Table 1 displays a list of the top five internal threats.

TABLE I. TOP FIVE THREATS INSIDE AN ORGANISATION

Threat:	Description/Example:
1)	<ul style="list-style-type: none"> - Due to internal employees with privileged system access performing deliberate attacks on company. - Disgruntled employees/ Outsider attacks.
2)	<ul style="list-style-type: none"> - The trusting nature of internal employees is exploited by extracting sensitive information. - E.g. a perpetrator pretending to be an IT technician, asking employees for personal passwords to fix underlying problems, however the perpetrator uses this information to breach security in.
3)	<ul style="list-style-type: none"> - Employees use the organisation's internet for personal use during working hours. - Employees might use the internet to watch a video clip, log onto social media sites or play games, hereby dramatically increasing chances of hidden malicious files entering organisation's network.
4)	<ul style="list-style-type: none"> - Technological advances enable one to carry data easily through portable devices such as USB flash-disk, MP3 player and cell phones. - Employees use these methods daily without thinking about consequences. - E.g. business information is stored on employee's flash-disk, and the flash-disk with sensitive business information is misplaced. The leakage of business information can have a devastating effect on the business. - Employees give out their passwords/credentials to other employees, not considering the consequences it may have on the organisation.
5)	<ul style="list-style-type: none"> - Piracy is a major issue. - Could have major reputational damage on organisation. - Employees often save personal files or programs on organisation's network; these include pirated software, movie files, or even pornography. - Effective policies should be in place to monitor for such files together with effective policy enforcement.

Administrative employees interact with business information and might subsequently confront various threats. Therefore, awareness about these threats are critical, and simultaneously, from an organisational perspective, to have effective policy control and enforcement in place. The following subsection will discuss the organisation's responsibility towards awareness.

B. Awareness within an Orgni ation

Security awareness is how well users understand the importance of information security and how well they exercise information security controls to protect the organisation's data and networks [13]. According to Ernst & Young, for the

effective protection of information, it is recommended that organisations invest more in training and awareness programs to help prevent information security breaches [14].

Without these awareness programs, administrative employees typically have one of two standings with regard to awareness: either the employee is ignorant or the employee is negligent. Whether ignorant or negligent, the administrative employees' insecure handling of data may be a vulnerability that could lead to the compromising of business information.

The impact of compromised business information may lead to a business losing reputation, competitive advantages, intellectual properties, and in the worst scenarios, lead to bankruptcy in the business sector. It is therefore imperative to eradicate any lack in knowledge that may exist with administrative employees about secure handling of business information. A fundamental point is firstly to be aware of threats. If one's awareness is deficient, one is more likely to be the cause of compromised business information.

To address this deficiency, it is of utmost importance to properly educate, train, and raise awareness on how to handle business information securely. This raises a question: how does one raise administrative employees' awareness? Different methods of effectively establishing awareness amongst administrative employees exist. One of the most efficient methods to raise awareness is to make use of the Information Security Awareness Training (ISAT) which provides awareness and training or workshops to educate administrative employees on issues related to information security [13].

Some guidelines in combination with an ISAT program from Cisco, a network technology company, suggest that an organisation should establish a security awareness and education practice. This is vital in gaining employee support, due to the fact that employees who believe that security programs are important, are more likely to follow specific procedures [12]. Cisco further suggests that a practice should:

- Educate and train employees about company expectations for protecting data.
- Include security awareness and practices in new-hire orientation events.
- Train employees about security considerations when answering the phone and connecting to the internet, social networking, and collaboration sites.
- Train employees about physical security concerns, such as allowing only employees with badges to enter buildings.

By using a combination of these suggested guidelines, it is possible to have a proper awareness program. As mentioned before, employees are seen as the weakest link inside an organisation, concerning information security. In agreement to this, the following section will determine the actions of administrative employees when facing the top five identified threats.

IV. ANALYSIS AND RESULTS

In further addressing the objective of this paper, the next step is to determine the level of knowledge and awareness administrative employees possess when encountering threats while handling business information. This is done using a survey in the form of a questionnaire.

A. Instrument Design and Categorization

For the design of the questionnaire, suggestions from Rowley [5] are followed which formulates the different questions. Not only the suggestions, but also the predominant threats, that were identified from literature, are integrated into the design of questions asked to participants. This will enable one to find a correlation with facts from literature and results from the survey.

The questionnaire consists of fifteen questions. Twelve of the fifteen questions are scenario-based questions. Each question will contain a scenario, allowing the participant to select one of four provided answers. Depending on the chosen answer of a participant, a different outcome is calculated.

The twelve scenario based questions are divided into three main categories, allowing an easy method of determining the level of awareness. The different categories are abstracted from the well-known McCumber Cube as seen in Figure 1. The three different categories are confidentiality, integrity, and availability (CIA) respectively. Also from Figure 1, the three different information states are important namely: transfer, storage, and processing state.

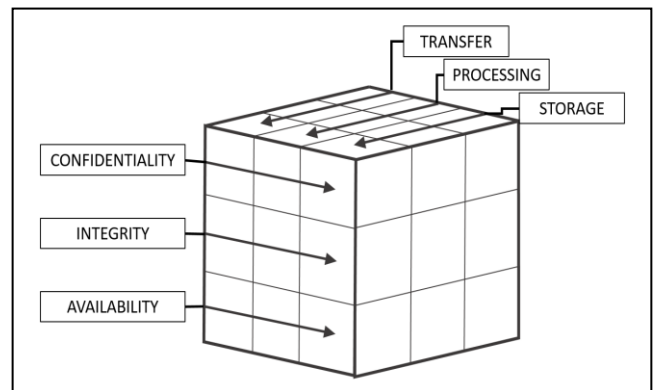


Figure 1. The Well-known McCumber Cube

For the exact explanation of the questionnaire, Figure 2 is used for reference. From Figure 2, one can identify the three categories on top. Underneath each category, any of the three different states may be used in a scenario. For instance, scenario 1, 4, 7, and 10 will be based on either the transfer, storage, or processing state. However, all four of these scenarios will focus only on the category of confidentiality.

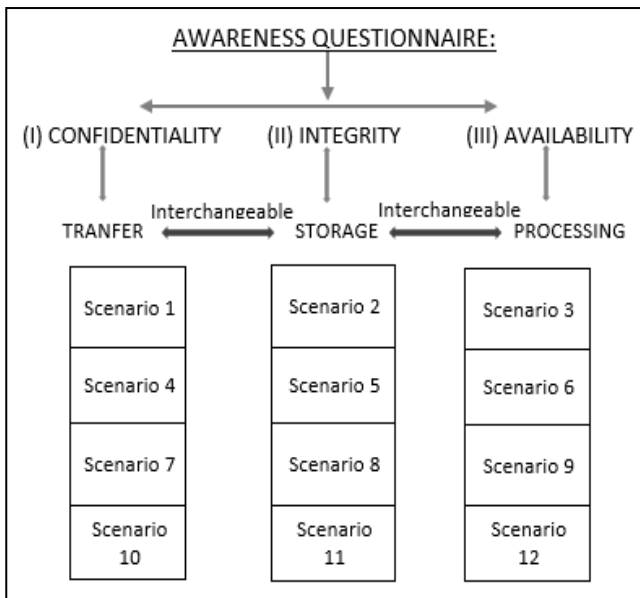


Figure 2. Visual Representation of Questionnaire Structure

Each scenario question was designed to test a specific category's awareness level. The following subsection highlights the method behind measuring the awareness level.

B. *eighted Ranking Scale*

In order to collect usable quantitative data from the survey, a weighted ranking scale was designed. This weighted ranking scale allows the participant's awareness level to be determined. For instance, depending on the answer that a participant selects for a particular scenario, the participant will be ranked on the weighted ranking scale. Figure 3 shows the weighted ranking scale together with the association of each number. This weighted ranking scale is used in the following subsection to explain the overall working of the questionnaire.

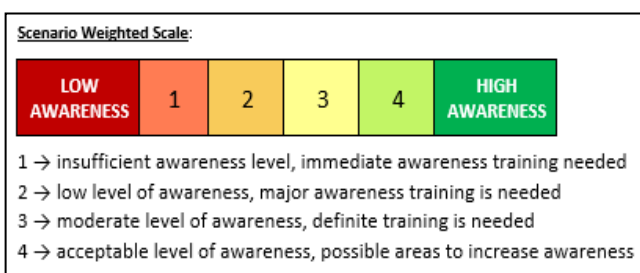


Figure 3. Weighted Ranking Scale

C. *E ample and E planation*

As mentioned earlier, each individual category (CIA) has four underlying scenarios. These four scenarios not only test the awareness level of its underlying category, but also the awareness level of the top identified threats respectively. Consequently, one is able to adequately test for awareness

amongst administrative employees on various threats and categories.

The first scenario of the questionnaire is now used as an example to explain the overall working of the questionnaire. The participant will be given the scenario as shown in Figure 4.

Figure 4. Sample Scenario Based Question from Questionnaire

After reading the scenario, the participant will then have to choose one of the four provided answers. The answers are not sorted in any particular order when presented to the participant. However, when the answers are reviewed, to determine the awareness level, the answers will be sorted according to most suitable answers. Below is how the answers are sorted when determining the level of awareness:

- 1) Make use of friend/co-worker's flash drive and print it on their computer.
- 2) Email the file to a friend/co-worker's computer and print.
- 3) Decide to print file at home and bring to work the following day.
- 4) Contact technician to first repair the printer, then you print.

The value next to the answer enables the researcher to determine the awareness level of the participant. For instance if the participant chose the answer "Decide to print file at home and bring to work the following day". This issues the participant a weight of three (3). The three (3) on the weighted scale shows that the participant has a moderate awareness level; definite awareness training is needed for this particular scenario. Some scenarios allow the participant to select a fifth option named "I don't know". The weight of this option holds a value of one (1). This symbolizes that the participant is unaware and/or unknowledgeable of the threat. The sum of all the weight values allows one to determine the participant's overall awareness of the three categories, and the top identified threats.

The example scenario given above for instance, tests the participant's knowledge regarding the threat of information leakage, under the confidentiality category while business information is in the transferring state. Table 2 lists all the scenario questions from the questionnaire, including a brief explanation of what is being tested.

TABLE II. EXPLANATION OF EACH SCENARIO IN QUESTIONNAIRE

Scenario No	Explanation
1	This scenario is determining the level of awareness concerning information leakage while in the transferring state. Transferring the sensitive file to another computer may result in information leakage.
2	The level of awareness is determined concerning integrity of business information while in a state of storage. If one changes the file, the integrity is compromised.
3	The level of awareness is determined concerning availability of business information while in a state of storage. If one removes the record, the availability of that information is compromised.
4	The level of awareness is determined concerning social engineering breaches. If the employee is not aware of these threats, it might compromise security and confidentiality of business information.
5	The level of awareness is determined concerning malicious attacks and integrity of business information while in a state of processing. If one use outdated logs, the business information might unreliable due to malicious activity
6	The level of awareness is determined concerning illegal activity and availability of business information while in a state of transferring. Surfing particular websites might severely affect the network availability and provide an open door into network
7	This scenario is determining the level of awareness concerning business information confidentiality and social engineering while in the processing state. If sensitive information is seen by unauthorized personnel, it might have a major effect on company.
8	The level of awareness is determined concerning integrity of business information while in a state of storage. Changing information without proof of validity can easily breach business information's integrity.
9	The level of awareness is determined concerning availability and illegal activity while in a state of storage. Storing personal files at work can compromise the availability for necessary space needed for business information also creating an unsafe environment.
10	This scenario is determining the level of awareness concerning business information confidentiality while in the storage state. Not locking up hardcopies of sensitive information could easily have information leakage as effect.
11	The level of awareness is determined concerning malicious attacks and integrity of business information while in a state of processing. Using special privileges to change business information may cause a breach of integrity.
12	The level of awareness is determined concerning availability of business information while in a state of processing. Turning off or restarting a server will most definitely reduce the availability of business information for all employees.

Each of the twelve scenarios was weighted using the weighted ranking scale. The following subsection will report on the results of the questionnaire.

D. Questionnaire Results

Seventy-five questionnaires were distributed, by using email, amongst eleven sectors. Fifty responses were received

which are categorized in the different organisational sectors as shown in Table 3.

TABLE III. DISTRIBUTION OF DEMOGRAPHIC SECTORS

Demographic Sector	Number of respondents
Academic Sector	4
Engineering Sector	4
Financial Sector	12
Human Resource Sector	2
Insurance Sector	1
Marketing Sector	5
Medical Sector	1
Production Sector	3
Public Service Sector	5
Safety & Security Sector	7
Secretarial Sector	6

After collecting all the responses, an average score was calculated, and mapped to a percentage value, for each category. A zero (0) percentage means no awareness exist compared to a hundred (100) meaning excellent awareness exist.

In order to have calculated an average percentage for each participant, each value of the weighted scale was given a certain percentage. The following list shows the percentage value for each weighted scale ranking.

- Weight of 1 → 0% to 25%
- Weight of 2 → 26% to 50%
- Weight of 3 → 51% to 75%
- Weight of 4 → 76% to 100%

By using this list, Figure 5 was created to represent the percentage for the overall average score of the three individual categories.

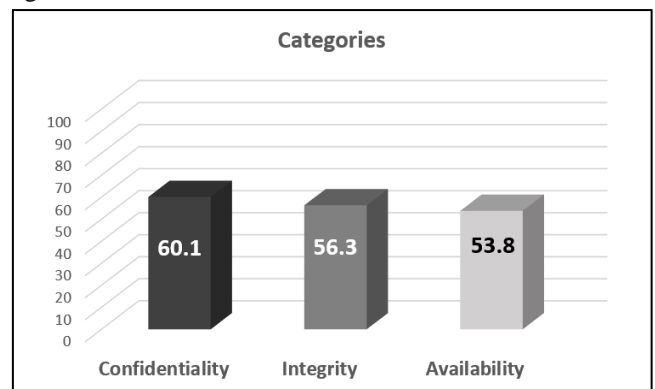


Figure 5. Overall Category Results

One can see that in the Confidentiality category, the participants scored 60.1%. This percentage is equal to a three (3) on the weighted ranking scale. The three (3) shows a moderate level of awareness, which is acceptable, however still suggesting that definite training is needed to increase awareness in this category. In the Integrity category, the participants scored 56.3%. This places them on a weight of three (3), just rising above a two (2). This also shows a moderate level of awareness exists, however it is in the early stages of awareness, and more awareness straining is

necessary. Lastly, in the Availability category, the participants scored 53.8%. The weighted ranking is a three (3) however; it is extremely close to a two (2), which emphasises a need for major awareness training in the availability category.

All three categories might seem acceptable at first, however the following three figures focus on the percentage value for each underlying scenario. This highlights specific threat areas that need attention within its respective category.

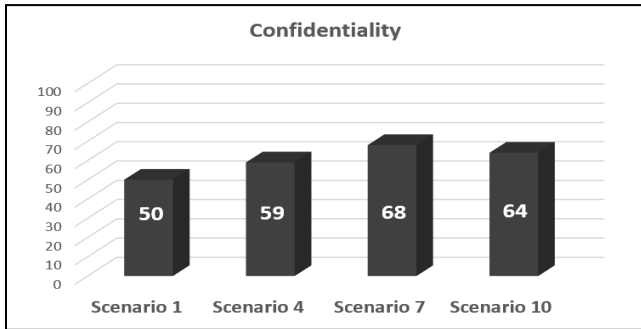


Figure 6. Results of Category: Confidentiality

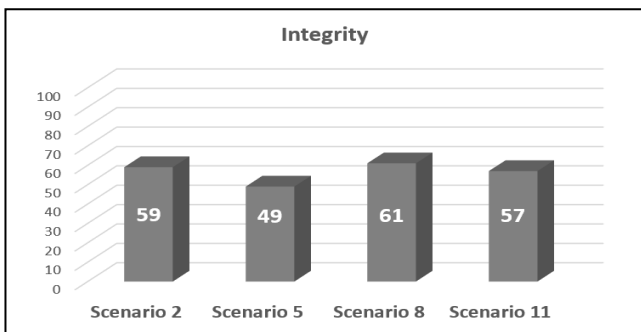


Figure 7. Results of Category: Integrity

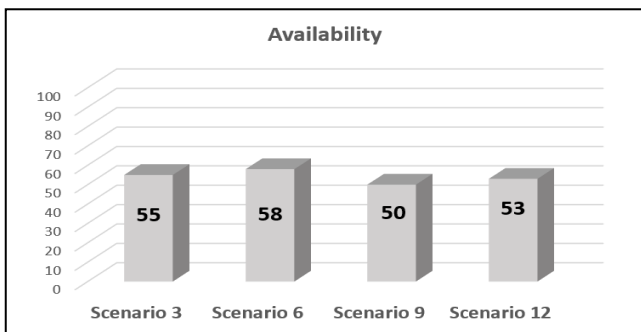


Figure 8. Results of Category: Availability

The previous three graphs highlight the three lowest scenarios, which is scenarios 1, 5 and 9. The following subsection will discuss the scenarios in more detail.

E. Questionnaire Findings

The three scenarios can be mapped to an individual threat area. Scenario 1 maps to threat of information leakage. Scenario 5 maps to the threat of malicious attacks and lastly scenario 9 maps to the threat of illegal activity. These three threats require more awareness training than the rest; however,

it is essential that all threats receive proper awareness training. Further findings collectively suggest that there is a need for proper awareness training especially under the category of availability. If one were to look at the overall percentage score of the survey (56.7%) one would realize that it is barely a moderate level of awareness. This implies that overall; there is a definite need for awareness training on certain areas whilst major awareness training is needed in other areas. It is essential that this is addressed accordingly, otherwise without proper awareness training in these areas; administrative employees can become vulnerable, in turn leading to compromising an organisation's business information.

In order to address the much-needed awareness training, the following section proposes a possible solution that could aid in raising administrative employees' awareness.

V. INTERACTIVE VISUAL LIBRARY

From the previous sections, it is clear that awareness has to be raised regarding administrative employees. A possible way of raising awareness is in the form of an interactive visual library. For the purpose of this paper, an interactive visual library can be defined as a dynamic method in which theory principles are provided. The theory principles are presented in a holistic manner by which knowledge is transferred in an engagement between a person and computer. The question however is, what aspects are important in creating an interactive visual library? To answer the question, the following section will highlight the aspects that are deemed critical in creating such a library.

A. Critical Aspects

The first critical aspect to consider is the learning style. Three different learning styles can be derived from Neil Fleming's VAK model. The three styles are visual, auditory and kinaesthetic learning styles [15]. The three different styles are individually explained as:

- **Visual** – Visual material is preferred and helps with better remembrance.
- **Auditory** – Learners use the form of auditory sound waves to study and remember.
- **kinaesthetic** – Prefer to work in groups together with the fact that a classroom affects the learner negatively.

According to Woda and Kubacki-Gorwecki [15] a staggering 65% of learners are visually orientated learners. The proposed solution therefore only focusses on the visual learning style. Providing a visual interactive library to promote awareness with administrative employees is suited considering the fact that a majority of administrative employees would better comprehend with a visual learning style. Ultimately, the interactive library promotes satisfaction, usability, and acceptance with the administrative employees.

The second critical aspect is to consider the fact that it is interactive. With regards to computers, Stevenson defines interactive as a two-way flow of information between a computer and a computer-user; the reaction of the computer

responding to a user's input [4]. From the definition, it is clear that it allows the user to interact with the computer. The fact that the solution is interactive, allows the user to control the library according to his or her needs. An example of the interactive requirement, for instance, would allow the user to use a swiping gesture to rotate the threats in order to select a desired threat.

The third and final critical aspect to the proposed solution is that the solution is in the form of a library. To clarify, the proposed solution holds a collection of threats in one single integrated library. This negates the need for a user to search online for threats. As it is difficult to search for a threat online if one is not aware that a particular threat exist. Therefore, this library would ease the searching of threats. Thus, the third aspect promotes ease of use with users.

A combination of these three critical aspects forms the basis of the interactive visual library. In the following subsection, an example of such an interactive visual library is given, which serves as a proof of concept.

B. Interactive Visual Library Prototype

As mentioned in the previous subsection, the first aspect of the prototype is for it to be of visual nature. Figure 9 serves as a demonstration on how the prototype is structured. As seen from Figure 9, the prototype is in the visual form of a three-dimensional rotating wheel. The wheel is visually stimulating to promote satisfaction, usability, and acceptance with the administrative employees.

Each three-dimensional block is interactive in the sense that one could click on it. Depending on the platform it operates on, one uses either a computer mouse or even a finger on a touch display such as a smartphone. Each three-dimensional block represents a threat in itself. Each threat listed on the blocks is not only abstracted from literature's top five threats, but also on the results of the survey. In this main screen, one is able to rotate by swiping the blocks like a wheel. A user can rotate through all the blocks and select a specific block.

Once an administrative employee clicks on a specific block, another screen will be displayed. Figure 10 shows a clear view of the block that was clicked on. This screen educates the administrative employee on the specific threat. Details are given regarding the threat, such as description of the threats and possible examples of the threat, amongst others. Tips on how to protect against the specific threat will also be highlighted.

Figure 10 serves as an example for each of the three-dimensional blocks. Each threat has the same layout of information as well as interactive in the same way. As mentioned previously, the fact that all the threats are together in one library negates the need for an administrative employee to search for threats on the internet. This adheres to the third aspect of being an integrated library. Functionality is also embedded to allow an educational video to be played.

Figure 11 is a view of the tips specifically concerning the threat of social engineering. Once the user clicks on the button called "Tips to keep safe", this screen will appear. Typically, the same happens when the user clicks on the "More examples" button. The proposed solution currently serves as a prototype for a proof of concept on raising awareness amongst administrative employees. This prototype should be modified to extend its functionality. However, the fundamental part is to incorporate the three critical aspects into the design of the library.

VI. CONCLUSION

The first objective of this research has been achieved by identification of different threats through a review of literature, pertaining to acts of human error, or failure. The second objective was attained by integrating threats with three main elements in security – confidentiality, integrity, and availability of information. Consequently, this led to an integrated and informative design. Finally, a possible interactive visual library, through a proof of concept, was proposed which can possibly contribute to employees' information security awareness.

Further research can be done by other researchers in this domain to reveal other approaches that can improve information security awareness in organizations. Further research will also be beneficial in expanding the scope of identified threats, to incorporate more areas coexisting with acts of human error, or failure. In addition, further research is required to assess the contribution of the interactive visual library in raising administrative employees' awareness. Although the research focusses on administrative employees, further research can be done by expanding the scope to encompass all employee groups within an organization, such as IT personnel, executive management as well as support personnel. It would be beneficial to compare the results from future research with the results of this study in order to identify possible patterns.



Figure 9. Demonstration of Visual Nature

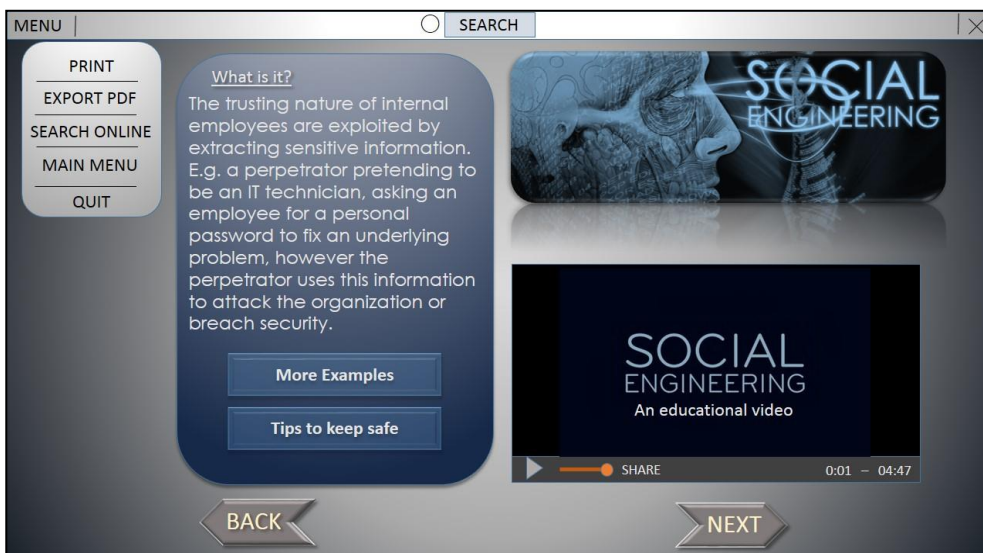


Figure 10. Block Containing Information of a Particular Threat

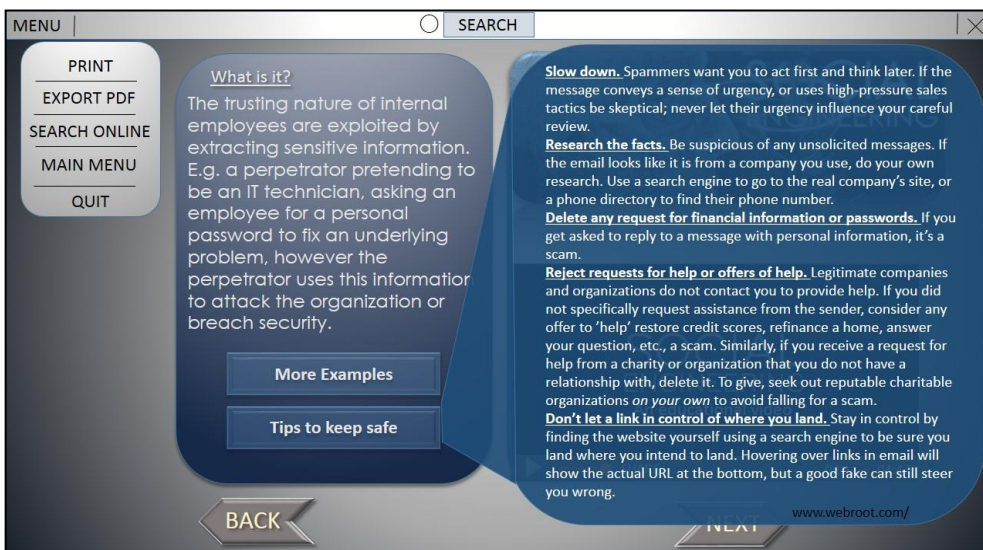


Figure 11. Tips to Keep Safe

ACKNOWLEDGMENT

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors, and are not necessarily to be attributed to the National Research Foundation.

REFERENCES

- [1] R. von Solms and S. H. (Basie) von Solms, "Information Security Governance: A model based on the Direct-Control Cycle," *Comput. Secur.*, vol. 25, no. 6, pp. 408–412, Sep. 2006.
- [2] DBIR, "Verizon, Data Breach Investigation Report." [Online]. Available: http://www.verizoneenterprise.com/resources/reports/rp_data-breach-investigations-report-2011_en_xg.pdf. [Accessed: 22-Apr-2014].
- [3] M. E. Whitman, "Enemy at the Gates: Threats to Information Security," *Commun. ACM*, vol. 46, no. 8, pp. 91–95, 2003.
- [4] Oxford Dictionaries, "Oxford Dictionaries," Oxford University Press, 2016. [Online]. Available: www.oxforddictionaries.com/definition/english/administrative-assistant. [Accessed: 10-Apr-2016].
- [5] J. Rowley, "Designing and using research questionnaires," *Manag. Res. Rev.*, vol. 37, no. 3, pp. 308–330, 2014.
- [6] M. T. Siponen, "A conceptual foundation for organizational information security awareness," *Inf. Manag. Comput. Secur.*, vol. 1, no. 8, pp. 31–41, 2000.
- [7] M. E. Whitman and H. J. Mattord, *Principles of Information Security*. Boston, USA: Cengage Learning, 2012.
- [8] M. E. Whitman, "In defense of the realm: understanding the threats to information security," *Int. J. Inf. Manage.*, vol. 24, no. 1, pp. 43–57, Feb. 2004.
- [9] Threat Horizon, "Threat Horizon 2013: Information security-related threats of the future," 2013. [Online]. Available: <https://www.securityforum.org/research/>. [Accessed: 24-May-2014].
- [10] C. Waxer, "The Top 5 Internal Security Threats - IT Security," *ITsecurity.com*. [Online]. Available: <http://www.itsecurity.com/features/the-top-5-internal-security-threats-041207/>. [Accessed: 24-May-2014].
- [11] Whittle, "The top five internal security threats." [Online]. Available: <http://www.zdnet.com/the-top-five-internal-security-threats-3039363097/>. [Accessed: 24-May-2014].
- [12] CSO Staff, "The Ten Habits of Highly Secure Employees." [Online]. Available: <http://www.csoonline.com/article/2123078/access-control/the-ten-habits-of-highly-secure-employees.html>. [Accessed: 25-May-2014].
- [13] E. B. Kim, "Recommendations for information security awareness training for college students," *Inf. Manag. Comput. Secur.*, vol. 22, no. 1, pp. 115–126, 2014.
- [14] Ernst and Young, "Ernst & Young's 2008 Global Information Security Survey," 2008. [Online]. Available: [http://www.ey.com/Publication/vwLUAssets/GISS2012/\\$FILE/EY_GIS_S_2012.pdf](http://www.ey.com/Publication/vwLUAssets/GISS2012/$FILE/EY_GIS_S_2012.pdf). [Accessed: 20-Apr-2014].
- [15] M. Woda and K. Kubacki-gorwecki, "Students Learning Styles Classification For e-Education," in *ICIT 2011 The 5th International Conference on Information Technology*, 2011.

Mobile Device Usage in Higher Education Institutions in South Africa

Ryan De Kock and Lynn A. Futcher

Center for Research in Information and Computer Security, School of ICT, Nelson Mandela Metropolitan University

Email: s211109940@nmmu.ac.za, Lynn.Futcher@nmmu.ac.za

Abstract— Cyber security threats are on the rise as the use of personally owned devices are increasing within higher education institutions. This is due to the rapid adoption of the Bring Your Own Device (BYOD) trend. In 2012, 20% of students used laptops globally for academic purposes, 30% used tablets, and 40% used smart phones. In addition, 60% of higher education institutions in the United States and United Kingdom allow students, faculty and non-academic staff to access their network using personally owned mobile devices.

A great concern is that although BYOD is widely accepted in higher education institutions, security is somewhat lacking. In addition, cyber-security threats have switched their focus to mobile devices. Therefore, the number of new mobile vulnerabilities reported each year has increased. Furthermore, in 2012, 10% of global cyber security breaches took place in the education sector with a total of 100 breaches resulting in the exposure of 10,000 identities. This placed the educational sector at the top of the list with the third most cyber-security breaches in 2012, behind the healthcare and retail sectors.

A literature survey, together with a single explanatory case study involving a higher education institution in South Africa were used to determine typical mobile device usage in an academic context. As a result of completing the study, it is clear that there is a high demand for the use of BYOD in higher education institutions in South Africa and that BYOD is vital to the academic success of its students. This paper discusses mobile device usage in higher education institutions in South Africa. In addition, it provides some key factors for higher education institutions to consider when dealing with the increased demand for BYOD usage.

Keywords— Mobile device usage; Bring Your Own Device; Higher Education Institutions

I. INTRODUCTION

Gartner [1] defines Bring Your Own Device (BYOD) as:

“An alternative strategy that allows employees, business partners and other users to use a personally selected and purchased client device to execute enterprise applications and access data. It typically spans smartphones and tablets, but the strategy may also be used for PCs. It may or may not include a subsidy.”

BYOD was first introduced in 2009 by Malcolm Harkins, Intel’s chief information security officer, after realizing that more and more employees wanted to use their own mobile devices in the workplace [2]. Intel’s leaders did not dismiss the possibility of this new trend due to the risks involved. Instead, they embraced the technology by setting up effective employee-owned device policies, resulting in increased connectivity to Intel’s network, greater employee productivity and improved security measures.

As the adoption of the BYOD trend is increasing in today’s organizations of different sectors, higher education institutions also encourage students and staff to use their own devices in exchange for the benefits offered by this trend. Furthermore, it is predicted that BYOD will become the leading practice for all educational environments by the year 2017 [3]. This highlights the overwhelming increase in the BYOD paradigm in the education sector.

The purpose of this paper is to determine mobile device usage in higher education institutions in South Africa. This is achieved through a case study of a South African higher education institution, implementing BYOD. The following section discusses the research design implemented in this study followed by background information on the use of BYOD in higher education institutions. Thereafter, the case study data is presented and discussed followed by key factors derived from the literature and the case study data.

II. RESEARCH DESIGN

In addition to the literature survey conducted to gain a better understanding of mobile device usage in higher education institutions, this study also makes use of a case study. The case study was used to gather a large amount of data and information required to determine the current state of mobile device usage in South African higher education institutions.

Yin [4] suggests that there are three different types of case studies. These include explanatory, exploratory and descriptive case studies. However, this research makes use of the descriptive case study. This type of case study is used when the researcher is seeking to describe a natural phenomenon which occurs within the data in question [4]. As for this research, a descriptive case study is used to describe how a South African higher education institution is implementing BYOD.

In addition, there is more than one type of case study design. In fact, Yin [4] proposes that there are two types of case study

designs, the single- and multi-case design (which involves cases within cases). The design used for this study makes use of the single case study design, as it focuses on a single case.

Therefore, a single descriptive case study involving a higher education institution in South Africa was used to determine typical mobile device usage in an academic context. The following section discusses mobile device usage in higher education institutions.

III. BACKGROUND

Although the concept of BYOD was only first introduced in 2009 [2], organizations and higher education institutions have shown an increasing interest in and tolerance for employees and students using their own mobile devices for work and academic purposes.

Liz Gosling, director of Information Technology (IT) services at Auckland University of Technology, states that the IT demands in higher education institutions differ from the technology requirements within an enterprise organization [5]. Therefore, to draw a comparison between higher education institutions and enterprise organizations, the BYOD users within each of these need to be compared to determine where they are similar and where they differ. Fig. 1 illustrates this comparison.

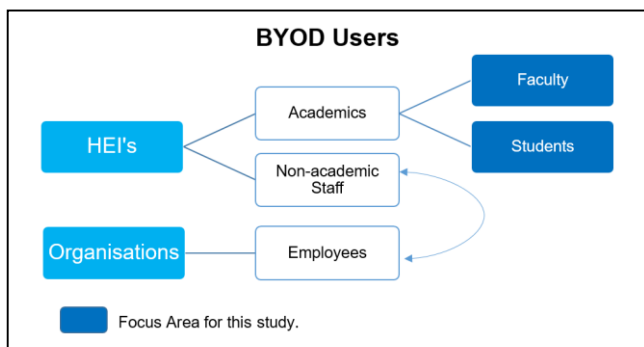


Figure 1: A comparison between BYOD users in organizations and HEI's.

As depicted in Fig. 1, the BYOD users in higher education institutions differ from organizations since they comprise students, non-academic staff and faculty, whereas organizations only include various employees. Furthermore, the employees within an organization are similar to the non-academic staff members within a higher education institution. These include human resources, marketing, accounting and finance, management, employees, etc. Throughout this paper, faculty refers to any academic staff such as lecturers, professors, etc.

Higher education institutions are realizing the importance of addressing the demand of BYOD within their institutions. This is supported by the findings in a survey conducted by Bradford Networks [6]. The survey questioned professionals representing over 500 higher education institutions in the United States and United Kingdom. It was found that 85% of higher education institutions allow students, faculty and non-academic staff to use their personal devices on their network, while 6% of the respondents reported that they have no plans to implement BYOD in the future. Furthermore, they found that 84% of the institutions that do not allow BYOD receive requests to use personal devices on their networks [6]. From Fig. 2, it is clear

that there is a high demand for mobile device usage in higher education institutions. Fig. 2 is based on the results of an international survey conducted by Educause in 2014 [7]. The survey was sent to 213 higher education institutions across 15 countries.

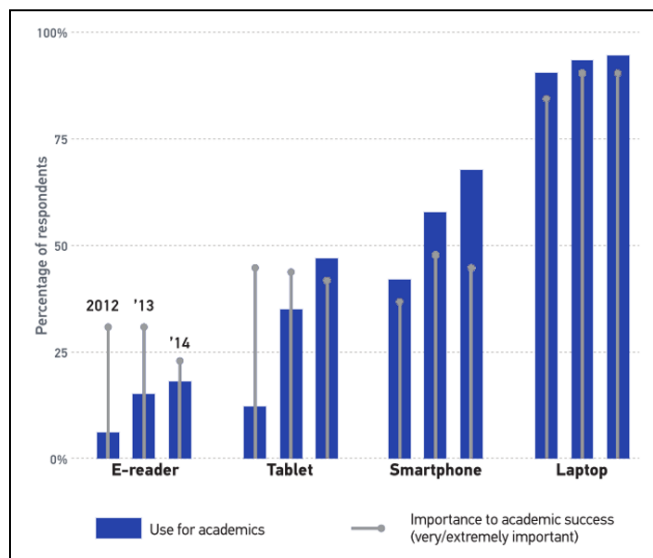


Figure 2: Use and importance of devices for academics [7].

Fig. 2 illustrates how important the use of BYOD is within the education sector as well as the percentage of students and staff that use personally owned devices for educational purposes. As illustrated in Fig. 2, 92% of students used laptops for academic purposes in 2014, 44% used tablets, 68% used smart phones, and 16% used e-readers [7]. An additional figure extracted from the international survey conducted by Educause in 2014, shows students' experiences with various types of technology for academic purposes. This is depicted in Fig. 3.

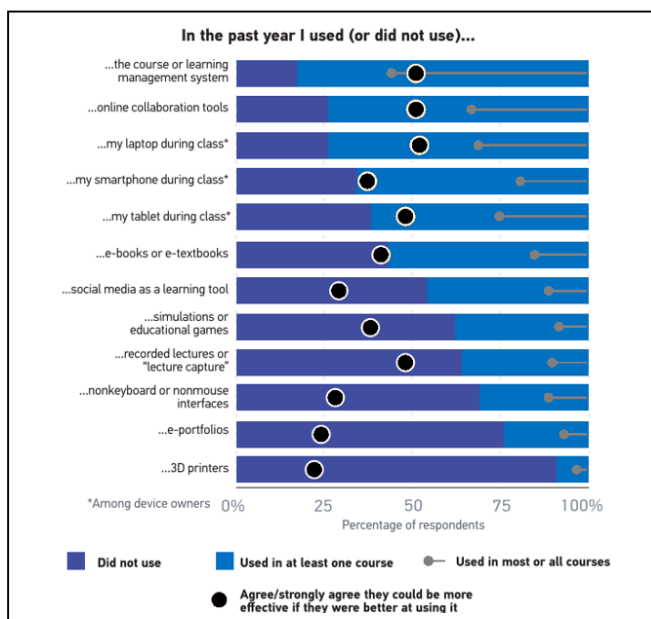


Figure 3: Use of technology for academic purposes [7].

Fig. 3 shows students' experiences with various types of technologies and their opinions about being more effective if they were better skilled at using certain technologies. Although students are skilled in most technologies, the use of e-books and recorded lectures should be considered. Furthermore, higher education institutions should provide enough online content to support their course content. Most students have used the learning management system (LMS) in at least one course (83%), but only just over half (56%) have used it in most or all of their courses, as depicted in Fig. 3. An LMS is a fundamental component in higher education. These systems function as digital learning environments, administrative systems for course management, and enterprise systems for institutional analytics and other purposes [8].

The above mentioned surveys clearly reflect a wide acceptance of BYOD in higher education institutions. Stavert [9] therefore, suggests three main reasons for why education institutions transition to BYOD. These include:

1. Financial pressure – Not all higher education situations can afford state of the art personal technology for all its students and staff. However, with the use of BYOD, students, faculty and non-academic staff can use their own mobile devices.
2. Pressure from students and staff – Higher education institutions are pressured by students, faculty and non-academic staff to use their own mobile devices for work and academic purposes.
3. Digital device ownership and use – Mobile devices have become more affordable over the last couple of years. These devices provide students, faculty and non-academic staff with 24/7 access to ideas, resources, people and communities. This has led to a large increase in ownership of mobile devices.

In addition, the reasons behind the great levels of acceptance in higher education institutions may be due to the fact that the purpose of an educational institution is to provide knowledge which is achieved by providing information regarding a particular subject. Today the internet is a major source of information on almost any subject. Higher education institutions may also have subscriptions to online journals and libraries which most of them provide for free to students. With the use of BYOD, students can easily access these sources of information from anywhere [10].

A great concern is that although BYOD is widely accepted in higher education institutions, security is somewhat lacking. Most higher education institutions have allowed some form of BYOD mostly via network access control (NAC) without implementing any BYOD policy [10]. This is very risky as higher education institutions are exposing their networks to various threats like unauthorized access, attacks of malware and viruses from student devices connected to the institution's network, loss of data, etc. This is also supported by an international survey conducted by the SANS Institute in 2014. They found that 60% of higher education institutions are concerned with the use of faculty and non-academic staff owned mobile devices while 30% are concerned with the use of student owned mobile devices on their networks [11].

The greater concern over faculty- and non-academic staff-owned mobile devices makes sense, since they handle large amounts of sensitive data, whereas students typically only handle their own. However, it was specifically the exposure of this type of data that landed Iowa State University in trouble in April 2014, when it was discovered that nearly 30,000 student records between 1992 and 2012 were exposed on 5 departmental servers [12]. While the servers were taken over by attackers wanting the computing power to create Bitcoins, the fact remains that privacy-protected data subject to regulatory compliance was inadvertently exposed on their servers.

It is therefore clear that there is a high global demand for mobile device usage in higher education institutions and that security is somewhat lacking. The following section discusses mobile device usage in South African higher education institutions.

IV. CASE STUDY

In accordance with the case study approach, a representation of any population was not intended, but rather a single case was chosen [13]. For this purpose, only South African higher education institutions implementing BYOD where eligible. According to the South African Higher Education Act 101 [14], a higher education institution can be defined as an institution that provides higher education on a full-time, part-time or distance basis which is:

- a) Merged, established or deemed to be established as a public higher education institution under this act;
- b) Declared as a public higher education institution under this act;
- c) Registered or provisionally registered as a private higher education institution under this act.

Given the above mentioned definition of what a higher education institution is, a prominent higher education institution within South Africa, was selected as the single case for this case study.

The Nelson Mandela Metropolitan University (NMMU) opened on 1 January 2005, due to the merging of three very different institutions as a result of the South African government's countrywide restructuring of higher education. Therefore, NMMU brings together the traditions of both technikon and university education, and draws on more than a century of quality higher education in an institution that offers a wide range of academic, professional and technological programs at varying entrance and exit levels. Furthermore, the NMMU has approximately 26 602 students and approximately 4 515 (1 702 faculty and 2 813 non-academic staff) permanent and contracted staff members, based on six campuses in the Nelson Mandela Metropole and George.

The mission statement of the NMMU is "*to offer a diverse range of quality educational opportunities that will make a critical and constructive contribution to regional, national and global sustainability*". This can only be achieved through the deployment and use of appropriate Information and Communication Technologies (ICT). The NMMU must furthermore also operate and be perceived as a safe and reliable

institution that ensures the security and proper use of its information assets.

The NMMU provides Wi-Fi access to students, faculty, non-academic staff and guests on their campuses. They also recognize the value of personal devices used for work and study purposes. In the past few years, NMMU has invested R7 000 000 on Wi-Fi across all seven campuses, and a further R750 000 to improve the quality of the Wi-Fi coverage. They have also upgraded 70 traditional lecture venues to enable faculty and students to use modern technology and provided support for NMMU’s Learning Management System (Moodle). In addition, the university handed over 250 computing devices using selection criteria that covered all campuses and all faculties but with a focus on off-campus students. Furthermore, they estimate that an additional R2 000 000 will be spent on modernizing the remaining venues in 2016. Due to this, mobile device usage at the NMMU has increased over the past few years.

The case study data was obtained from key ICT staff members from the NMMU. They were asked to supply BYOD related documents and archival records where available. Several freely available documents were also obtained from internal systems within the NMMU. The documents obtained include network logs, a list of suggested software, survey results, information security awareness and training initiatives, policies and procedures.

Fig. 4 illustrates mobile device usage among students, faculty and non-academic staff at the NMMU. These percentages refer to the number of users who used their smartphones, laptops and tablets to access the NMMU network 3 or more times per week in 2014.

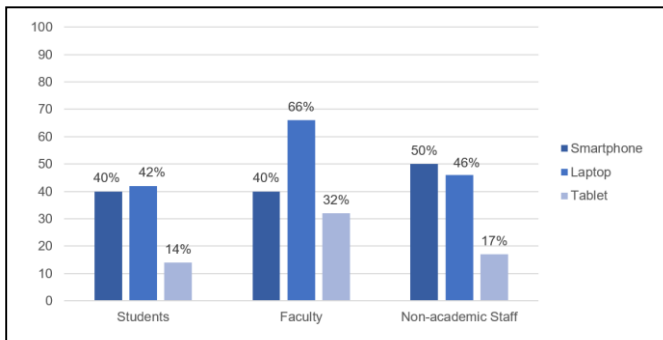


Figure 4: Mobile device usage for 3 or more times per week in 2014.

In 2014, faculty and non-academic staff accessed the NMMU network using mobile devices more frequently than students, as depicted in Fig. 4. Furthermore, students and faculty primarily used laptops when accessing the NMMU network, while non-academic staff primarily used smartphones. Tablets were not used very often by any of the user groups in 2014, as depicted in Fig. 4. However, Fig. 5 illustrates an enormous increase in tablet usage by students in 2015. In fact, tablet usage among students increased by approximately 55% from 2014 to 2015, as depicted in Fig. 5.

It can also be seen that student smartphone usage increased from approximately 40% in 2014 to approximately 85% in 2015. Furthermore, laptops which were primarily used by students in 2014, were surpassed by smartphones and tablets in 2015.

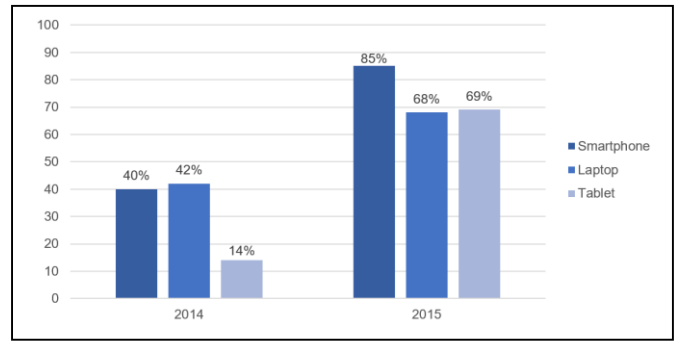


Figure 5: Student mobile device usage for 3 or more times per week.

Smartphone usage among faculty and non-academic staff also increased in 2015, as depicted in Fig. 6. However, the increase in smartphone usage resulted in a decrease in laptop and tablet usage among faculty and non-academic staff in 2015. Fig. 6 refers to both faculty and non-academic staff. However, the figure only depicts mobile device usage for 3 or more times per week, therefore only illustrating the frequent use of mobile devices accessing the NMMU network.

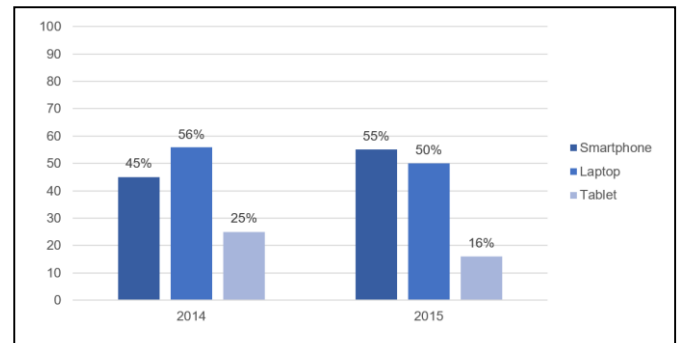


Figure 6: Staff mobile device usage for 3 or more times per week.

Smartphone usage among faculty and non-academic staff increased by approximately 10% from 2014 to 2015, while laptop usage decreased slightly by approximately 6% and tablet usage decreased by approximately 9%, as depicted in Fig. 6.

The frequent use of mobile devices at the NMMU increased significantly among students in 2015. However, the frequent use of mobile devices among faculty and non-academic staff has decreased slightly with the exception of a slight increase in smartphone usage. Figs. 7, 8 and 9 illustrate what the mobile devices depicted in Figs. 4, 5 and 6 were used for in 2015, therefore only depicting the frequent use (3 or more times per week) of mobile devices.

Fig. 7 illustrates what students, faculty and non-academic staff used their smartphones for in 2015.

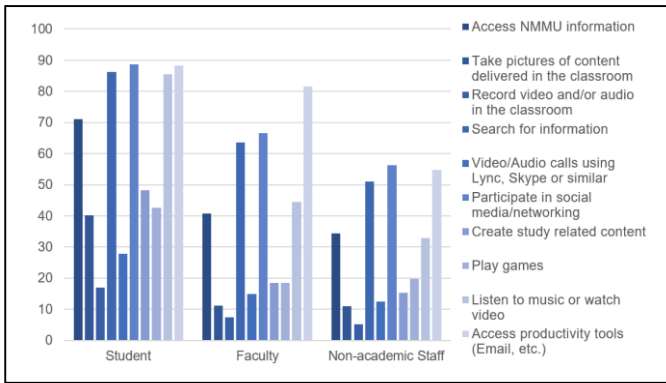


Figure 7: Student, faculty and non-academic staff smartphone usage for 3 or more times per week in 2015.

In 2015, students and non-academic staff primarily used their smartphones to participate in social networking followed by accessing productivity tools, such as emails, and to search for information. However, faculty primarily used their smartphones to access productivity tools followed by participating in social networking, and searching for information. Therefore, students and non-academic staff used their smartphones for similar purposes, while faculty shows a slight exception of accessing productivity tools more frequently than participating in social networking. The use of smartphones to access NMMU related information is also relatively popular among students, faculty and non-academic staff, as depicted in Fig. 7.

Fig. 8 illustrates what laptops were used for among students and staff in 2015. In Fig. 8 staff refers to both faculty and non-academic staff at the NMMU.

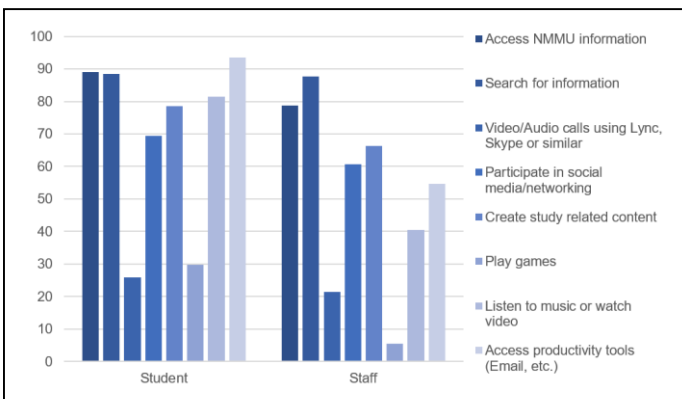


Figure 8: Student and staff laptop usage for 3 or more times per week in 2015.

In 2015, students primarily used their laptops to access productivity tools and NMMU related information as well as to search for information, as depicted in Fig. 8. Faculty and non-academic staff, however, mainly used their laptops to search for information, access NMMU related information and to create study related content such as presentation slides and worksheets, etc. Therefore, both students and staff used their laptops to frequently access NMMU related information and to search for information on the internet, as depicted in Fig. 8.

Fig. 9 illustrates what tablets were used for among students and staff in 2015, where staff refers to both faculty and non-academic staff.

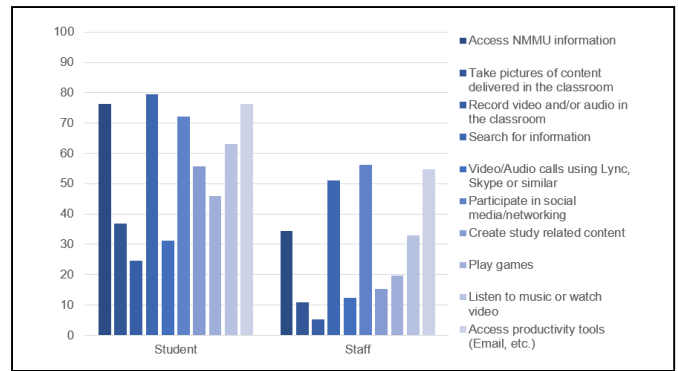


Figure 9: Student and staff tablet usage for 3 or more times per week in 2015.

In 2015, students primarily used their tablets to search for information followed by accessing productivity tools and NMMU related information, as well as participating in social networking. Whereas faculty and non-academic staff primarily used their tablets to participate in social networking followed by accessing productivity tools, searching for information, and accessing NMMU related information. Therefore, both students and staff primarily used tablets for similar reasons in 2015, as depicted in Fig. 9.

Fig. 10 illustrates which tools and technologies NMMU faculty are interested in using for academic purposes.

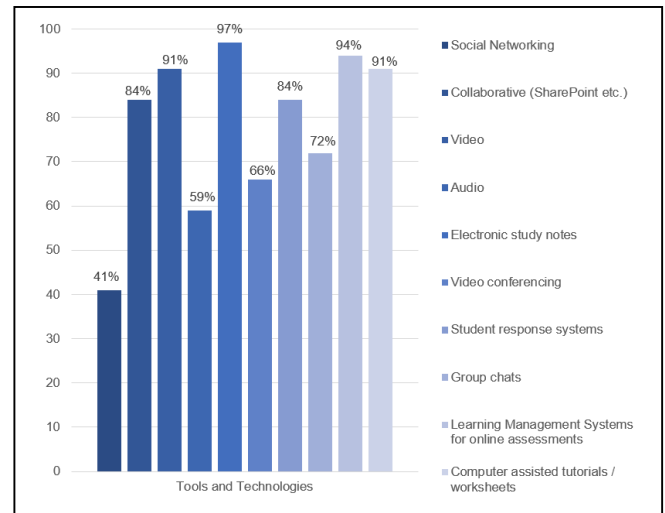


Figure 10: Faculty interest in technology usage for academic purposes in 2015.

This graph illustrates various teaching and learning assessment methods and the percentage of faculty interested in using them for academic purposes.

According to Fig. 10, faculty are mostly interested in using electronic study notes, learning management systems for online assessments, videos, and computer assisted tutorials and worksheets to aid teaching and learning methods at the NMMU. In 2014, 83% of students found that faculty are using technology to enhance their learning experience. This increased fractionally in 2015. Furthermore, students at the NMMU are currently receiving study material in various forms from faculty. Fig. 11

illustrates how NMMU students received their study material in 2015.

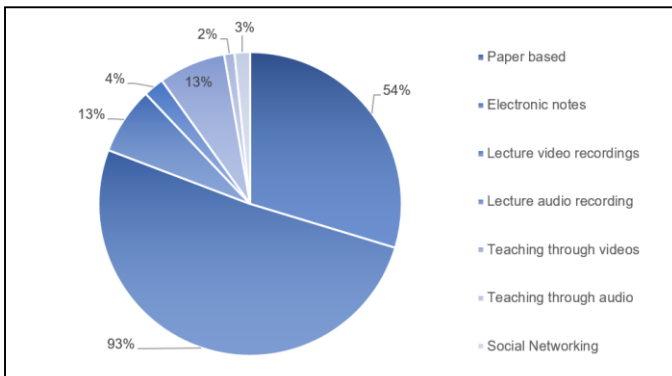


Figure 11: How students received study material from faculty in 2015.

In 2015, students primarily received their study material in the form of electronic notes, followed by paper-based study material. However, paper-based study material decreased by 7%, while electronic notes increased by 4% from 2014 to 2015. Teaching through videos and lecture video recordings were also relatively popular, but insignificant in comparison to electronic and paper based study materials.

From these results, it can be concluded that faculty are currently integrating technology into the curriculum to enhance students learning experience. In addition, there has been a significant increase in BYOD demand. The following section discusses key factors that higher education institutions should consider when dealing with this increased demand for BYOD usage.

V. KEY FACTORS

Given the predicted increase in mobile device usage at South African higher education institutions and the integration of technology into the curriculum, higher education institutions need to consider several key factors when dealing with the increased demand for BYOD usage as discussed in this section. These factors are derived from both literature and the case study data.

Mobile Device Management (MDM) – Some higher education institutions may consider adopting an MDM solution. Although not a new technology, MDM is only starting to gain in sophistication due to the invasion of employee-owned devices into the workplace [15] and because the number of confidential business information leakages via mobile devices has continued to rise [16]. An MDM solution can be seen as a partial system for the management of BYOD risks, such as data leakages, loss of organizational control and visibility, and ease of mobile device loss [15]. This is achieved through comprehensively managing mobile devices by monitoring their status and controlling their functions remotely using wireless communication technology such as Wi-Fi or Over-the-Air (OTA), as well as managing the required organizational resources [16]. Although relatively expensive, higher education institutions that can afford to implement an MDM solution should do so. However, it is essential for higher education institutions to realize that the implementation of an MDM

solution is not necessarily sufficient to cope with the proliferation of devices on their campuses. Therefore, higher education institutions need to make sure their technology and policies deliver the data security and management efficiency they seek [17].

Since there are no commercial off-the-shelf solutions for MDM that work on every platform [18], and that all MDM solutions offer the same basic capabilities, choosing an MDM solution should not be based on technical security needs alone. Instead, it should be supported by non-technical elements of information security such as policies and processes [15].

Develop a concise and inclusive acceptable use policy (AUP) – Higher education institutions face a unique set of challenges when implementing BYOD [19]. These challenges are differentiated according to student, faculty and non-academic staff. Each user group brings with it a unique set of demands. Before developing an AUP, higher education institutions first need to determine the intended goals and results of the policy document [20]. These include outlining authorized use, prohibited use, systems management, policy violation procedures, policy review and specifying limitations of liability [21]. In addition, higher education institutions need to determine what systems, services, and sensitive data students, faculty and non-academic staff need to access using their personal mobile devices [19]. Furthermore, the policy needs to accommodate the uncertainty of emerging technologies that will continue to end up on campuses [19]. Therefore, institutions need to find a way to draft a policy that is sufficiently broad to allow for future technologies yet sufficiently detailed to be enforceable.

Data security – Higher education institutions need to review and implement appropriate safety measures to protect their students, faculty, non-academic staff, and databases populated with sensitive information [22]. However, for higher education institutions to achieve this, they need to consider various threats [23]. These include unauthorized access to sensitive data stored on the mobile devices; unauthorized access to data stored on the institution's network; attacks from malicious software; and the ability to impersonate an authorized user. In addition, sensitive data should be classified and encrypted [11].

Network infrastructure – Opening a higher education institution's network to student, faculty and non-academic staff mobile devices increases the strain on the institution's network [20]. Therefore, institutions need to ensure that their network infrastructure is capable of meeting the BYOD demands. To achieve this, institutions need to determine how many mobile devices students, faculty and non-academic staff have and ensure sufficient bandwidth is available to accommodate these devices [9]. In addition, they need to ensure that their network is maintained by the IT department [11]. Ease of access and quality of service also plays a major role, since students, faculty and non-academic staff will most likely expect 24/7 network access [9]. Several higher education institutions use network segmentation to improve performance and increase security [20]. This allows them to provide a network for students and a separate network to be used by faculty and non-academic staff, thereby avoiding data and security conflicts and protecting student information.

Develop a software infrastructure – In a BYOD environment, students, faculty and non-academic staff will use a variety of mobile devices. A significant challenge for any higher education institution is to provide software tools that can be utilized by their users on any device [20]. This requires considerable planning. Therefore, institutions need to make use of platform-independent tools, cloud-based storage, and web-based applications.

Develop a portal – Higher education institutions need to create a central location that collects software tools and other resources [20]. This provides students, faculty and non-academic staff with a central location from where they can access web applications, general information, distinct-licensed software and other educational resources.

Build a curriculum – Higher education institutions need to find a way to incorporate technology into the curriculum [20]. This will enable students to learn and complete assignments anywhere, anytime. Furthermore, students will most likely be encouraged to bring their personal devices to campus if the curriculum supports their use. In addition, faculty should be able to grade assignments quicker and send feedback to students using the LMS.

Provide ongoing education and training – Higher education institutions should find ways to educate students, faculty and non-academic staff of the dangers associated with the use of BYOD [24]. They should be made aware of ways to access and use data safely, as well as how they can protect sensitive information. Education and training should also include social media usage, personally identifiable information, strong passwords and privacy settings [25]. Without training and education, users could inadvertently put personal data as well as the institutions' data at risk. Furthermore, students, faculty and non-academic staff should clearly understand the appropriate and inappropriate use of their personal devices [3].

Address equity – Higher education institutions need to maintain equity among students by ensuring that no student is disadvantaged through the lack of available technology [3]. Several higher education institutions allow students who cannot afford their own mobile devices to loan devices from them [9]. It is essential that all students have equal opportunities in this regard.

Plan financially for sustainability – Higher education institutions need to be well-prepared for the possible challenges introduced by BYOD. Financial sustainability allows higher education institutions to plan ahead for mobility [22]. This will allow them to add devices to their network without adding strain. In addition, the allocation of funds is essential to enabling higher education institutions to follow through on their BYOD projects, plans, and the integration of technology [22]. Sufficient investment in bandwidth, infrastructure, personnel, and new technology is needed to provide a robust and scalable network infrastructure to support the increasing number of devices [3].

Help desk – A well run help desk is central to the smooth operation of a BYOD program. The role of the help desk should be expanded to cater for multiple devices and operating systems [26]. Furthermore, higher education institutions should ensure that processes, procedures and systems are in place so that

technical support can be provided promptly and efficiently to students, faculty and non-academic staff [9].

Higher education institutions should consider these key factors when dealing with the increased demand for BYOD usage on campus.

VI. CONCLUSION

It is clear that higher education institutions in South Africa are observing a tendency in faculty and students that use their laptops, smart phones, tablets, e-readers and other mobile devices as a resource for enhancing their learning experience [10]. Furthermore, some higher education institutions may consider adopting an MDM solution to address the potential breach points associated with the implementation of BYOD. However, while MDM has some level of protection, the use of MDM alone is an insufficient resource for the implementation of BYOD [27]. Therefore, higher education institutions need to find more innovative and effective ways to safeguard valuable information and protect students, faculty and non-academic staff from security violations and data loss. The key factors discussed in this paper serves as a good starting point for higher education institutions. The ultimate goal should be for higher education institutions to safely provide enhanced learning resources to its students and to safeguard faculty and non-academic staff within their comfort zone. The explosion of mobile devices in higher education institutions is clearly cause for both celebration and concern.

Since this study only includes a single case, further research could include performing such a case study on other higher education institutions in South Africa. Furthermore, future research will consider the development of a framework to aid South African higher education institutions with the implementation of BYOD.

VII. ACKNOWLEDGEMENTS

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.

REFERENCES

- [1] Gartner, Inc., "Gartner Predicts by 2017, Half of Employers will Require Employees to Supply Their Own Device for Work Purposes," 1 May 2013. [Online]. Available: <http://www.gartner.com/newsroom/id/2466615>. [Accessed 28 May 2015].
- [2] J. Roman, "BYOD: Get Ahead of the Risk," 1 November 2012. [Online]. Available: <http://www.bankinfosecurity.in/byod-get-ahead-risk-a-4394>. [Accessed 2 May 2015].
- [3] T. Probert, "BYOD – an educational revolution?," *Educational Technology*, pp. 72-73, 2012.
- [4] R. K. Yin, *Case Study Research: Design and Methods* 5th Edition, United States of America: SAGE Publications, Inc., 2013.
- [5] C. Sliep, "Bring Your Own Device and Information Technology Service Delivery: A Higher Education Intitution Case Study," p. 165, 2013.

- [6] Bradford Networks, "The impact of BYOD in education," May 2013. [Online]. Available: http://thebooks.s3.amazonaws.com/The_Impact_of_BYOD_in_Education.pdf. [Accessed 27 August 2015].
- [7] J. Bichsel and E. Dahlstrom, "ECAR Study of Undergraduate Students and Information Technology, 2014," Educause, Louisville, 2014.
- [8] M. Brown, J. Dehoney and N. Millichap, "What's Next for the LMS?," *Educause Review*, 22 June 2015.
- [9] B. Stavert, "Bring your own device (BYOD) in schools," *NS Department of Education and Communities*, 2013.
- [10] K. R. Afreen, "Bring Your Own Device (BYOD) in Higher Education: Opportunities and Challenges," *International Journal of Emerging Trends Technology in Computer Science (I ETCS)*, pp. 233-236, 2014.
- [11] R. Marchany, "Higher Education: Open and Secure?," *SANS Institute*, June 2014.
- [12] Iowa State University, "Iowa State IT staff discover unauthorized access to servers," 22 April 2014. [Online]. Available: <http://www.news.iastate.edu/news/2014/04/22/serverbreach>. [Accessed 23 November 2015].
- [13] K. M. Eisenhardt and M. E. Graebner, "Theory building from cases: Opportunities and challenges," *Academy of Management Journal*, vol. 50, no. 1, pp. 25-32, 2007.
- [14] Republic of South Africa, "Higher Education Act 101 of 2003," 2003.
- [15] A. Dedeche, S. Lajami, M. Le and F. Liu, "Emergent BYOD Security Challenges and Mitigation Strategy," *The University of Melbourne*, pp. 1-17, 2013.
- [16] W. Jeon, R. Keunwoo and D. Won, "Security requirements of a mobile device management system," *International Journal of Security and its Applications*, vol. 6, no. 2, pp. 353-358, 2012.
- [17] M. Davis, "BYOD: Why Mobile Device Management Isn't Enough," 20 11 2012. [Online]. Available: <http://www.informationweek.com/it-leadership/byod-why-mobile-device-management-isnt-enough/d/d-id/1107487?>. [Accessed 7 August 2015].
- [18] D. Baranwal, S. Ravindran and R. Sadana, "BYOD in the Enterprise — A Holistic Approach," *ISACA Journal*, pp. 1-8, 2013.
- [19] S. Difilipo, "The policy of BYOD: Considerations for higher education," *Educause Review*, pp. 60-61, 1 April 2013.
- [20] Intel Education, "BYOD Planning and Implementation Framework," 2012. [Online]. Available: <http://www.k12blueprint.com/sites/default/files/BYOD-Planning-Implementation.pdf>. [Accessed 30 November 2015].
- [21] A. Green, "Management of security policies for mobile devices," *Proceedings of the 4th annual conference on information security curriculum development*, 2007.
- [22] A. S. Ackerman and M. L. Krupp, "Five Components to Consider for BYOT/BYOD," *International Conference on Cognition and Exploratory Learning in Digital Age*, pp. 35-41, 2012.
- [23] I. Bernik and B. Markelj, "Mobile devices and corporate data security," *International Journal of Education and Information Technologies*, pp. 97-104, 2012.
- [24] N. Hockly, "Tech-savvy teaching : BYOD Technology Matters," *Journal of Modern English Teachers*, vol. 21, no. 4, pp. 44-45, October 2012.
- [25] S. Emery, "Factors for Consideration when Developing a Bring Your Own Device (BYOD) Strategy in Higher Education," *University of Oregon*, pp. 1 - 111, July 2012.
- [26] B. Dixon and S. Tierney, "Bring Your Own Device To School," *Report by Microsoft Corporation*, pp. 1-16, 2012.
- [27] E. B. Koh, J. Oh and C. Im, "A Study on Security Threats and Dynamic Access Control Technology for BYOD , Smart-work Environment," in *International Multi-conference of Engineers and Computer Scientists*, Hong Kong, 2014.

SHA-1 and the Strict Avalanche Criterion

Yusuf Moosa Motara

Department of Computer Science
Rhodes University
Grahamstown 6140, SOUTH AFRICA
Email: y.motara@ru.ac.za

Barry Irwin

Department of Computer Science
Rhodes University
Grahamstown 6140, SOUTH AFRICA
Email: b.irwin@ru.ac.za

Abstract—The Strict Avalanche Criterion (SAC) is a measure of both confusion and diffusion, which are key properties of a cryptographic hash function. This work provides a working definition of the SAC, describes an experimental methodology that can be used to statistically evaluate whether a cryptographic hash meets the SAC, and uses this to investigate the degree to which compression function of the SHA-1 hash meets the SAC. The results ($P < 0.01$) are heartening: SHA-1 closely tracks the SAC after the first 24 rounds, and demonstrates excellent properties of confusion and diffusion throughout.

I. INTRODUCTION

Many computer scientists know little about the inner workings of cryptographic hashes, though they may know something about their properties. One of these properties is the “avalanche effect”, by analogy with the idea of a small stone causing a large avalanche of changes. The “avalanche effect” explains how a small change in the input data can result in a large change in the output hash. However, many questions around the effect are unanswered. For example, how large is the effect? After how many “rounds” of a compression function can it be seen? Do all inputs result in such an effect? Little experimental work has been done to answer these questions for any hash function, and this paper contributes experimental results that help in this regard.

A boolean n -bit hash function H is the transform $\mathbb{Z}_2^m \rightarrow \mathbb{Z}_2^n$, where m is an arbitrary non-negative number. A cryptographic hash function attempts to obscure the relationship between the input and output of H , and the degree to which this is accomplished is directly related to the (second-)preimage resistance of the hash function. This implies that two similar inputs should have very different outputs.

The Strict Avalanche Criterion (SAC) ([1], [2]) formalizes this notion by measuring the amount of change introduced in the output by a small change in the input. It builds on the definition of *completeness*, which means that each bit of the output depends on all the bits of the input, in a way that is cryptographically relevant. Using the definition of H as above, an output $H(x) = y$ is obtained for an input x . The initial bit of x is now flipped, giving $H(x_0) = y_0$. This process is repeated for $x_{1..n}$, resulting in $y_{1..n}$. The SAC is met when the Hamming distance between y and $y_{0..n}$ is, on average, $\frac{n}{2}$.

There are three contributions that this paper makes to the existing body of research:

- 1) A clarified understanding, with justification, of what the SAC is;

- 2) Experimental SAC results for a particular cryptographic hash (SHA-1);
- 3) An exploration of intermediate results

Section 2 of this paper examines related work and argues that the SAC as proposed by Webster & Tavares [1] has been misunderstood in much of the contemporaneous critical literature. Section 3 introduces salient points of a well-known cryptographic hash (SHA-1) which is assumed to exhibit the SAC, and describes an experimental design to test its SAC-compliance; to the best of the authors’ knowledge, there are no other studies which have conducted similar experiments. Section 4 presents experimental results, and some discussion follows.

II. RELATED WORK

The original definition [1] of the SAC is:

Consider X and X_i , two n -bit, binary plaintext vectors, such that X and X_i differ only in bit i , $1 < i < n$. Let

$$V_i = Y \oplus Y_i$$

where $Y = f(X)$, $Y_i = f(X_i)$ and f is the cryptographic transformation, under consideration. If f is to meet the strict avalanche criterion, the probability that each bit in V_i is equal to 1 should be one half over the set of all possible plaintext vectors X and X_i . This should be true for all values of i .

Forré [2] expresses this as:

Let \underline{x} and \underline{x}_i denote two n -bit vectors, such that \underline{x} and \underline{x}_i differ only in bit i , $1 \leq i \leq n$. Z_2^n denotes the n -dimensional vector space over 0,1. The function $f(\underline{x}) = z, z \in \{0, 1\}$ fulfills the SAC if and only if

$$\sum_{\underline{x} \in Z_2^n} f(\underline{x}) \oplus f(\underline{x}_i) = 2^{n-1}, \text{ for all } i \text{ with } 1 \leq i \leq n.$$

Similarly, Lloyd [3] understands the SAC as:

Let $f : Z_2^n \mapsto Z_2^m$ be a cryptographic transformation. Then f satisfies the strict avalanche criterion if and only if

$$\sum_{\underline{x} \in \mathbb{Z}_2^n} f(\underline{x}) \oplus f(\underline{x} \oplus \underline{c}_i) = (2^{n-1}, \dots, 2^{n-1})$$

for all $i, 1 \leq i \leq n$.

where \oplus denotes bitwise exclusive or and \underline{c}_i is a vector of length n with a 1 in the i th position and 0 elsewhere.

Other works ([4], [5], [6], [7]) follow in the same vein. However, these definitions calculate the sum over *all* possible inputs as leading to the fulfillment of the SAC, which is contrary to the original definition. The original definition separates a *baseline* value from the *avalanche vectors*, and states that the SAC holds true when “the probability that each bit [in the avalanche vectors] is equal to 1 should be one half over the set of all possible plaintext vectors” [1]. Therefore, a better test of whether $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2$ fulfills the SAC would use a universal quantifier,

$$\forall \underline{x} \in \mathbb{Z}_2^n, P(f(\underline{x}) = f(\underline{x}_i)) = 0.5$$

for all \underline{x}_i which differ from \underline{x} in bit $i, 1 \leq i \leq n$

A simple example clarifies the difference. Babbage [6] uses Lloyd’s [3] definition of the SAC and defines a SAC-compliant function:

Define $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2$ by

$$\begin{cases} f(x_1, \dots, x_n) = 0 & \text{if } x_1 = 0 \\ f(x_1, \dots, x_n) = x_2 \oplus \dots \oplus x_n & \text{if } x_1 = 1 \end{cases}$$

The simplest function of this nature is $f(\underline{x}) = x_0 \wedge x_1$. Then, taking $g(\underline{x}) = f(\underline{x}) \oplus f(\underline{x} \oplus 01)$ and $h(\underline{x}) = f(\underline{x}) \oplus f(\underline{x} \oplus 10)$,

\underline{x}	$f(\underline{x})$	$g(\underline{x})$	$h(\underline{x})$	$P(f(\underline{x}) = f(\underline{x}_i))$
00	0	0	0	1.0
01	0	0	1	0.5
10	0	1	0	0.5
11	1	1	1	1.0
Sum:	2	2		

Note that the sum of each of the third and fourth columns is 2^{n-1} , as predicted, and that this function fulfills the summed definition of the SAC. However, the first and last rows do not fulfill the original definition of the SAC at all: the probability of change, given the baseline values 00 and 11, is 0.0 in each case. It is therefore more reasonable to regard the *row* probability as important. This understanding is also in accordance with the original text that defined the term. Under this definition, $x_0 \wedge x_1$ is not SAC-compliant.

It is worth noting that the original definition, as per Webster & Tavares [1], is slightly ambiguous. They state that “the probability that each bit in V_i is equal to 1 should be *one half* over the set of all possible plaintext vectors X and X_i ”; however, they also state that “to satisfy the strict avalanche criterion, every element must have a value *close to one half*”

(emphasis mine). Under Lloyd’s interpretation, the SAC is only satisfied when an element changes with a probability of precisely 0.5. This is an unnecessarily binary criterion, as it seems to be more useful (and more in line with the original definition) to understand how far a particular sample *diverges* from the SAC. Therefore, this paper regards the SAC as a continuum but takes Lloyd’s formulation as the definition of what it means to “meet” the SAC.

Preneel [4] suggests a generalisation of the SAC called the *propagation criterion* (PC), defined as

Let f be a Boolean function of n variables. Then f satisfies the **propagation criterion of degree k** , $PC(k)$, ($1 \leq k \leq n$), if $\hat{f}(\underline{x})$ changes with a probability of 1/2 whenever i ($1 \leq i \leq k$) bits of \underline{x} are complemented.

It can be seen that the SAC is equivalent to $PC(1)$. The same work defines an *extended propagation criterion* which regards the SAC as a continuum. Much of the subsequent work ([8], [9], [10], [11], [12], [13]) in this area has more closely examined the relationship between PC and nonlinearity characteristics. Many of these extend the PC in interesting ways and examine ways of constructing functions which satisfy $PC(n)$, but experimental research that targets existing algorithms is scarce.

Although there are proven theoretical ways to construct a function which satisfies the SAC [7], there is no way (apart from exhaustive testing) to verify that an existing function satisfies the SAC. By contrast, useful cryptographic properties such as non-degeneracy [14] or bentness [15] are verifiable without having to resort to exhaustive testing. However, the SAC metric is no worse in this regard than the correlation immunity [16] and balance [17] metrics which also require exhaustive testing.

III. EXPERIMENTAL DESIGN

The SHA-1 hash [18] is a well-known cryptographic hash function which generates a 160-bit hash value. It is the successor to the equally well-known MD5 cryptographic hash function which generated a 128-bit hash value. SHA-1 was designed by the National Security Agency of the United States of America and published in 1995 as National Institute of Standards and Technology (NIST) Federal Information Processing Standard 180-1.

A. Hash details

The SHA-1 hash is constructed using the Merkle-Damgård paradigm ([19], [20]), which means that it consists of padding, chunking, and compression stages. These stages are necessary for the hash algorithm to be able to handle inputs which are greater than 447 bits in length; however, they are unnecessary to consider in an examination of the compression function itself, since the strength of the Merkle-Damgård paradigm is predicated on the characteristics of the compression function. This paper examines only the compression function itself, and does not concern itself with padding, chunking, or Davies-Meyer strengthening [21].

The SHA-1 compression function makes use of addition, rotation, and logical functions (AND, OR, NOT, XOR), applied over the course of 80 rounds, to convert the 16 input words into a 5-word (160-bit) output. Each round affects the calculation of subsequent rounds, and the hashing process can therefore not be parallelized to any significant degree. A full description of the inner workings of SHA-1 is provided in FIPS 180-1 [18]. For the purposes of this work, it is sufficient to understand that each round generates a value that is used in subsequent rounds, and that there are 80 rounds in total.

B. Statistical approach

It is computationally infeasible to exhaustively test the degree to which SHA-1 meets the SAC since the input space (2^{64}) is too large. However, it is possible to use a sampling approach instead, where representative samples are drawn from a *population* and inferences are made based on an analysis of those samples. This approach relies on each input being statistically independent of other inputs. Generating such input can be an extraordinarily difficult task [22]; however, random.org¹ provides data which meets this requirement [23], [24]. A source which may be more random, but which has undergone far less scrutiny, is HotBits². Data for these experiments has therefore been obtained from random.org.

The inputs which make up the population should represent real-world usage, and the form of the input is therefore of concern. The inputs to the SHA-1 compression function are twofold: 16 32-bit words of input data and an initialization vector of 5 32-bit words, for a total of 21 32-bit words (or 672 bytes). The initial initialization vector is defined by the FIPS 180-1 specification, and the input data is padded and terminated such that the last two words processed by the algorithm encode the length of the input data. Subsequent initialization vectors are generated from the output of the previous application of the compression function. For any input which is larger than 1024 bytes, there is therefore at least one iteration of the compression function for which all 672 bytes are effectively "random" — if it is assumed that a previous iteration of the compression function can possibly result in the applicable initialization vector. To make this assumption, it is sufficient to assert that there are no values which *cannot* be generated as intermediate initialization vectors (given a pool of $\leq 2^{64}$ different bitstreams). Therefore, we can take independent 672-byte inputs as our population of concern.

The hypothesis to be tested is that SHA-1 meets the SAC. The desired margin of error is 1%, at a 99% confidence level. The required sample size is therefore determined by

$$n = \left(\frac{\text{erf}^{-1}(0.99)}{0.01\sqrt{2}} \right)^2 = 16587$$

where erf^{-1} is the inverse error function

Given the 2^{64} input space, this seems to be a very small number; however, "it is the absolute size of the sample

which determines accuracy, not the size relative to the population" [25]. Data collected during the experiment also indicates the degree to which SHA-1 does not meet the SAC, and the round at which the SAC comes into effect.

Each of the 16587 inputs is passed through a custom implementation of the SHA-1 compression function. This implementation has not been validated by NIST's Cryptographic Algorithm Validation Program³, but nevertheless passes all of the byte-oriented test vectors provided by NIST; in addition, source code for the compression function is available on request. When presented with a 672-byte input, the compression function outputs a list of 80 vectors, one for each round of the compression function. *Baseline* and *avalanche* vectors are generated for each input, and per-round compliance with the SAC is determined by these.

The primary question that this work seeks to answer is: to what degree do each of the output bits meet the SAC? To determine this, the per-input SAC value for each bit must be calculated, as described above. The geometric mean of the SAC values is representative of the central tendency. From the data that is generated to answer the primary question, two other questions may be fruitfully answered:

- **What is the distribution of SAC values per input?** The geometric mean provides a way to understand the degree to which an output bit meets the SAC, on average over a range of inputs. The distribution of SAC values quantifies how likely any particular input is to meet the SAC.
- **How quickly do the bits of the SHA-1 hash meet (or not meet) the SAC?**

For repeatability, it is disclosed that the data used to create inputs is the first $16587 \times 672 = 11,146,464$ bits generated by random.org from the 2nd to the 12th of January 2015. This data is available from <https://www.random.org/files/>.

IV. RESULTS

As shown by Figure 1, the SHA-1 hash diverges from the SAC by remarkably small amounts. The initial divergence is due entirely to the fact that the very last bits of a 672-bit input are found in rounds 15 and 16 and, when modified, have an exaggerated effect on subsequent rounds. This effect is largely due to the fact that the changes have not yet had time to diffuse through the rounds. Data which is most representative of the final hash output can therefore be seen in rounds ≥ 24 .

If sufficient time is provided for diffusion, a different picture emerges. Figure 2 shows the absolute divergence from round 24 onwards. Although the heatmap looks noisier, the most important thing to note is that the maximum divergence from the "ideal" SAC value of 0.5 is only 0.0009, which is within the margin of error for this sample size.

A 5-figure statistical summary (minimum, lower quantile, median, upper quantile, and maximum) of deviation from the SAC is plotted as Figure 5. In this graph, (round, bit) tuples have been converted to single value bits using the function $\text{bit}(r, i) = (r-1) \cdot 32 + (i-1)$. This was done to better illustrate

¹<https://www.random.org>

²<https://www.fourmilab.ch/hotbits/>

³<http://csrc.nist.gov/groups/STM/cavp/#03>

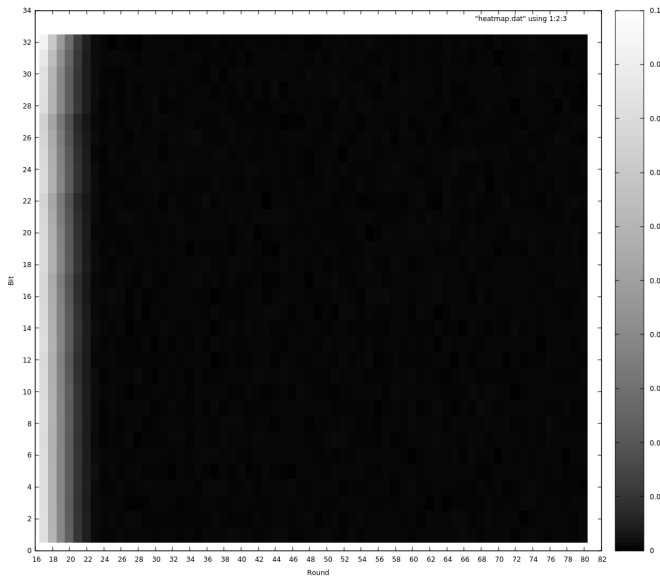


Figure 1. Divergence from SAC, round 17..80

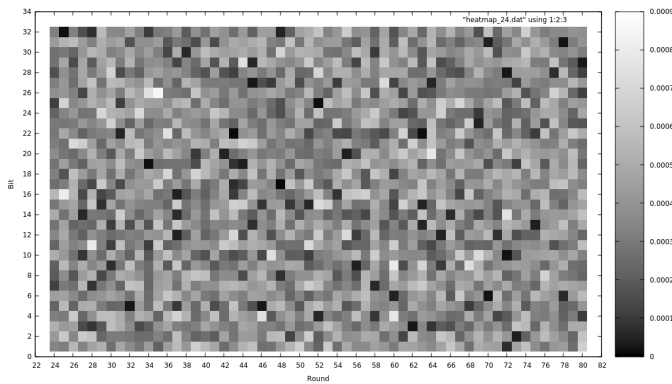


Figure 2. Divergence from SAC, rounds 24..80

noteworthy points, and because there is no round-specific pattern in the data. The median value is 0.0 throughout, and the lower and upper quartiles demonstrate remarkable consistency across the rounds despite minima and maxima which fluctuate significantly. It is interesting to note that rounds 24..44 show the same pattern as rounds 60..80, which are the final rounds of the hash. The distribution of values appears to remain constant from round 24 all the way up to round 80.

The distribution of SAC values for rounds ≥ 24 is shown in Figure 3, and there are few surprises here. It has a median, mean, and mode of 0.5, and appears to be a normal distribution. To verify whether the distribution is, in fact, normal, a quantile-quantile plot was generated. A quantile-quantile plot overlays points from a data-set on top of the theoretically-predicted distribution; if the actual points lie along the theoretically-predicted line, then the data fits the specified distribution.

Three possible distributions were plotted (see Figure 4):

- Normal ($\sigma = 0.019285397$, $\mu = 0.5$), using the standard deviation and mean of the data where round ≥ 24 .

- Log-normal ($\sigma = 0.059899039$, $\mu = 0.49855239$), estimated from the data.
- Weibull ($k = 9.6116811$, $\lambda = 0.52480750$), estimated from the data.

None of the distributions match the data exactly; in fact, the normal distribution is the worst fit, with log-normal and Weibull distributions being much closer fits. At present, the distribution that the data conforms to is unknown.

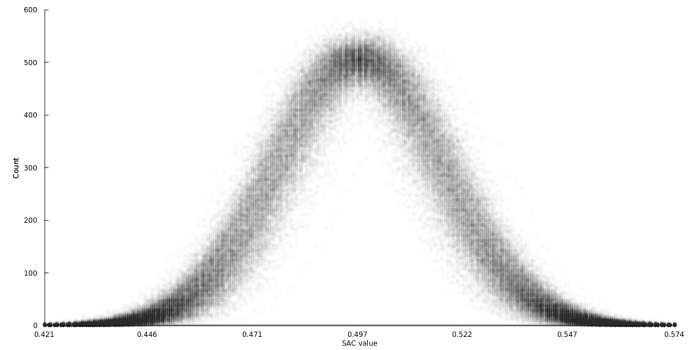


Figure 3. Distribution of SAC values

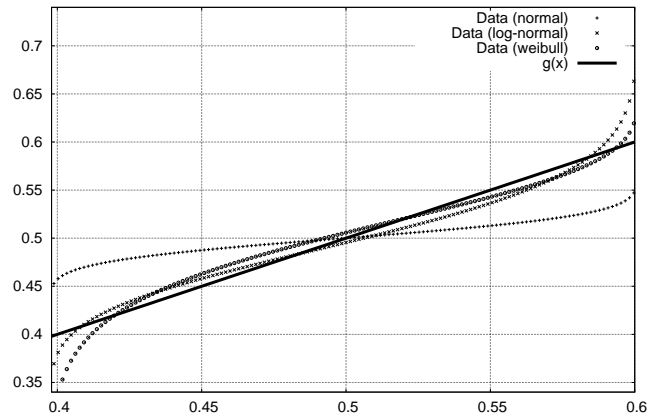


Figure 4. Quantile-quantile plot showing goodness of fit

Lastly, the averaged per-round distribution of SAC values is shown as Figure 6, which may be regarded as a 3D histogram where zero-buckets have been discarded. This graph attempts to show trends and changes in SAC values through rounds. The “spikiness” of the center is immediately noticeable, despite the SAC values being averaged. This reflects the fact that SAC values tend strongly towards 0.5. The right side of the graph is shorter than the left, and also higher; since zero-buckets have been discarded, this would indicate that SAC values tend to be distributed into more buckets as they tend towards zero, and conversely concentrated into fewer buckets as they tend towards 1. This tendency is present throughout the rounds.

The z-axis curve in the middle of the bits, which appears to be too pronounced to be an artifact of averaging, would seem to indicate that there are more non-zero buckets that are being filled as the bit-value increases. However, such an increase

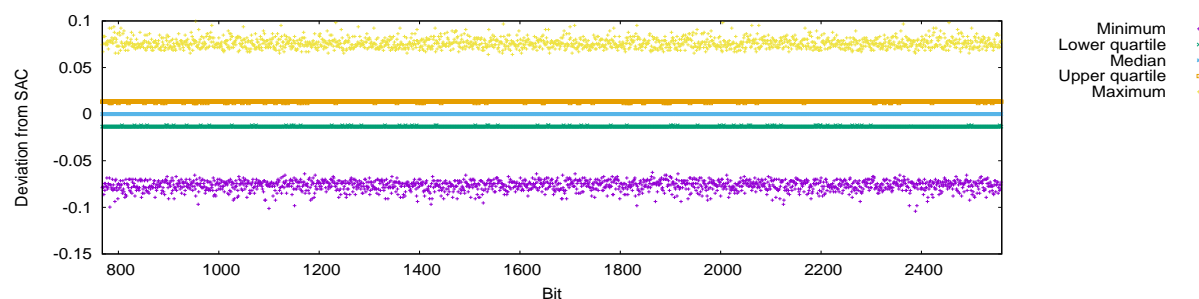


Figure 5. Summary statistics

should result in a decrease of SAC values at other points — and a corresponding dip at those points on the x-axis. There is no such dip, and other visualisations do not indicate such an increase. In the absence of any other explanation, it is believed that it is an artifact of the visualisation.

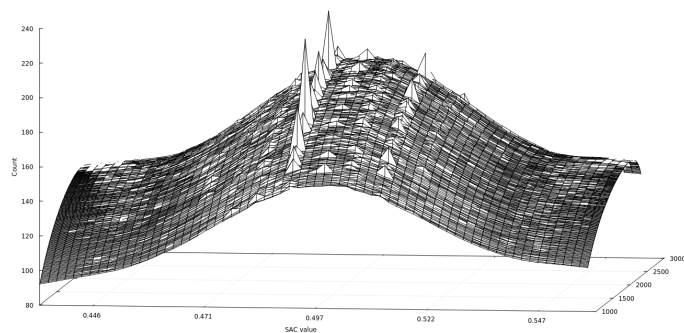


Figure 6. Distribution of SAC values, per-round

V. DISCUSSION

The questions posed may now be answered. As the experimental results show, each of the output bits meets the SAC by round 24, and it therefore takes only 8 rounds from the end of the input data for the SAC values to settle into a “stable” state. This stable state persists through all of the remaining rounds.

The distribution of SAC values does not fit any of the tested distributions exactly. However, the distribution displays a regularity that makes it quite possible that a less well-known distribution will fit. Further analysis of this could be worthwhile, since the distribution of SAC values may provide a different way to understand the behaviour of the hash.

One of the characteristics of a cryptographic hash is (second-)preimage resistance: the computational infeasibility of finding an input that results in a particular output. The SAC results obtained from these experiments highlight the difficulty of obtaining a specific preimage since, from round 24 onwards, the SAC is either met or very closely approximated. This makes it extraordinarily difficult to determine which input bit could contribute to a particular output change, since the answer is likely to be “any of them”!

The methodology that has been described above is not specific to the SHA-1 hash, and may be applied to any hash

function. It would be interesting to see it applied to other hash functions with a view to comparing their SAC values and distributions to the results above. Similarities and differences, and the possible reasons for them, would make for interesting research. For example, SHA-1’s spiritual predecessor, MD5, has also proven to be resistant to preimage attacks; could the reason be that it shares a similarly rapid achievement of close-to-SAC bits, followed by a similarly “stable” maintenance of the SAC through all of its rounds?

On an implementation note, it may be worthwhile to use a cloud computing platform (such as Google’s BigTable) for future experiments of this nature. The experiments have generated tens of gigabytes of data which take some time to query on a single machine. The scalable infrastructure of the cloud may allow queries, and hence experiments, to proceed more quickly.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

- [1] A. F. Webster and S. E. Tavares, “On the Design of S-Boxes,” in *Lecture Notes in Computer Science*. Springer Science + Business Media, 1986, pp. 523–534. [Online]. Available: http://dx.doi.org/10.1007/3-540-39799-x_41
- [2] R. Forré, “The Strict Avalanche Criterion: Spectral Properties of Boolean Functions and an Extended Definition,” in *Advances in Cryptology — CRYPTO’ 88*. Springer Science + Business Media, 1990, pp. 450–468. [Online]. Available: http://dx.doi.org/10.1007/0-387-34799-2_31
- [3] S. Lloyd, “Counting Functions Satisfying a Higher Order Strict Avalanche Criterion,” in *Lecture Notes in Computer Science*. Springer Science + Business Media, 1990, pp. 63–74. [Online]. Available: http://dx.doi.org/10.1007/3-540-46885-4_9
- [4] B. Preneel, “Analysis and design of cryptographic hash functions,” Ph.D. dissertation, PhD thesis, Katholieke Universiteit Leuven, 1993.
- [5] S. Lloyd, “Balance uncorrelatedness and the strict avalanche criterion,” *Discrete Applied Mathematics*, vol. 41, no. 3, pp. 223–233, feb 1993. [Online]. Available: [http://dx.doi.org/10.1016/0166-218x\(90\)90056-i](http://dx.doi.org/10.1016/0166-218x(90)90056-i)
- [6] S. Babbage, “On the relevance of the strict avalanche criterion,” *Electron. Lett.*, vol. 26, no. 7, p. 461, 1990. [Online]. Available: <http://dx.doi.org/10.1049/el:19900299>
- [7] K. Kim, T. Matsumoto, and H. Imai, “A Recursive Construction Method of S-boxes Satisfying Strict Avalanche Criterion,” in *Advances in Cryptology-CRYPTO’ 90*. Springer Science + Business Media, 1991, pp. 565–574. [Online]. Available: http://dx.doi.org/10.1007/3-540-38424-3_39

- [8] J. Seberry, X.-M. Zhang, and Y. Zheng, "Nonlinearly Balanced Boolean Functions and Their Propagation Characteristics," in *Advances in Cryptology — CRYPTO' 93*. Springer Science + Business Media, 1994, pp. 49–60. [Online]. Available: http://dx.doi.org/10.1007/3-540-48329-2_5
- [9] X.-M. Zhang and Y. Zheng, "Characterizing the structures of cryptographic functions satisfying the propagation criterion for almost all vectors," *Des Codes Crypt*, vol. 7, no. 1-2, pp. 111–134, jan 1996. [Online]. Available: <http://dx.doi.org/10.1007/bf00125079>
- [10] C. Carlet, "On the propagation criterion of degree 1 and order k," in *Lecture Notes in Computer Science*. Springer Science + Business Media, 1998, pp. 462–474. [Online]. Available: <http://dx.doi.org/10.1007/bfb0054146>
- [11] S. H. Sung, S. Chee, and C. Park, "Global avalanche characteristics and propagation criterion of balanced Boolean functions," *Information Processing Letters*, vol. 69, no. 1, pp. 21–24, jan 1999. [Online]. Available: [http://dx.doi.org/10.1016/s0020-0190\(98\)00184-7](http://dx.doi.org/10.1016/s0020-0190(98)00184-7)
- [12] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine, "Propagation Characteristics and Correlation-Immunity of Highly Nonlinear Boolean Functions," in *Advances in Cryptology — EUROCRYPT 2000*. Springer Science + Business Media, 2000, pp. 507–522. [Online]. Available: http://dx.doi.org/10.1007/3-540-45539-6_36
- [13] A. Gouget, "On the Propagation Criterion of Boolean Functions," in *Coding Cryptography and Combinatorics*. Springer Science + Business Media, 2004, pp. 153–168. [Online]. Available: http://dx.doi.org/10.1007/978-3-0348-7865-4_9
- [14] S. Dubuc, "Characterization of linear structures," *Designs, Codes and Cryptography*, vol. 22, no. 1, pp. 33–45, 2001.
- [15] O. Rothaus, "On "bent" functions," *Journal of Combinatorial Theory Series A*, vol. 20, no. 3, pp. 300–305, may 1976. [Online]. Available: [http://dx.doi.org/10.1016/0097-3165\(76\)90024-8](http://dx.doi.org/10.1016/0097-3165(76)90024-8)
- [16] T. Siegenthaler, "Correlation-immunity of nonlinear combining functions for cryptographic applications (corresp.)," *IEEE Transactions on Information Theory*, vol. 30, no. 5, pp. 776–780, 1984.
- [17] O. Staffelbach and W. Meier, "Cryptographic significance of the carry for ciphers based on integer addition," in *Advances in Cryptology-CRYPTO'90*. Springer, 1991, pp. 602–614.
- [18] P. FIPS, "180-1. Secure hash standard," *National Institute of Standards and Technology*, vol. 17, 1995.
- [19] R. C. Merkle, *Secrecy, authentication, and public key systems*. Stanford University, 1979.
- [20] P. Gauravaram, W. Millan, and J. G. Nieto, "Some thoughts on Collision Attacks in the Hash Functions MD5, SHA-0 and SHA-1." *IACR Cryptology ePrint Archive*, vol. 2005, p. 391, 2005.
- [21] R. S. Winternitz, "A Secure One-Way Hash Function Built from DES." in *IEEE Symposium on Security and Privacy*, 1984, pp. 88–90.
- [22] G. J. Chaitin, *Exploring RANDOMNESS*. Springer Science + Business Media, 2001. [Online]. Available: <http://dx.doi.org/10.1007/978-1-4471-0307-3>
- [23] C. Kenny, "Random number generators: An evaluation and comparison of random.org and some commonly used generators," *Departamento de Ciencias de la computación, Trinity College Dublin*, 2005.
- [24] L. Foley, "Analysis of an on-line random number generator. Project report, The Distributed Systems Group," *Computer Science Department, Trinity College Dublin*, 2001.
- [25] D. Freedman, R. Pisani, and R. Purves, *Statistics*, ser. International student edition. W.W. Norton & Company, 2007. [Online]. Available: <https://books.google.co.za/books?id=v0yxMwEACAAJ>

Specific Emitter Identification for Enhanced Access Control Security

J.N. Samuel¹ and W.P. du Plessis¹

Email: jeevanninansamuel@gmail.com, wduplessis@ieee.org

¹Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa

Abstract—This paper presents the application of specific emitter identification (SEI) to access control and points out the security caveats of current radio-based access remotes. Specifically, SEI is applied to radio frequency (RF) access remotes used to open and close motorised gates in residential housing complexes for the purposes of access control. A proof-of-concept SEI system was developed to investigate whether it is possible to distinguish between the RF signals produced by two nominally-identical access remotes. It was determined that it is possible to distinguish between the remotes with an accuracy of 98%.

I. INTRODUCTION

Access remotes are used to open gates to residential estates, houses and garages. On this basis they provide security as only people with the remote are able to gain access to these areas, akin to a key. However, the signal produced by these remotes can easily be read from low-cost software-defined radios (SDRs) and reproduced by another radio transmitter [1]. This allows for illegitimate access to residential estates, houses and garages. This motivates the need for making access remotes robust against replay attacks and cloning.

SDRs are radios whose hardware implementation is either replaced by corresponding software components or configurable via software [2]. Recently the concept of SDR has matured due to advancements in hardware and software technology to the point where anyone can purchase a SDR for \$10 to \$15 [3]. In particular the RTL2832U-based SDRs fit in this category of low-cost SDR. They consist of a number of models based on the tuning chip that they utilise. Three of the tuning chips used are the Raphael R820T, the Elonics E4000 and the Fitipower FC tuning chips. The R820T SDR can tune from 24 MHz to 1766 MHz [3], while the Elonics E4000 can tune between 52 MHz to 2200 MHz, though with a frequency gap between approximately 1100 MHz and 1250 MHz. Regardless of the type of tuner used, the RTL2832U receiver can sample data at up to 3.2 Msps and has an 8-bit analogue-to-digital converter (ADC) resolution. However, it has been found that the RTL2832U can only sample data reliably (i.e. without dropping samples) at sampling rates lower than 2.56 Msps [3]. While these low-cost SDRs have relatively low ADC res-

This work is based on the research supported in part by the National Research Foundation (NRF) (Grant specific unique reference number (UID) 85845). The NRF Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author(s), and that the NRF accepts no liability whatsoever in this regard.

olution and tuning range, they are sufficient for eavesdropping on radio communications.

Connected to a software application such as GNU Radio [4], an SDR can be used to listen to radio transmissions in its tuning range and to store these transmissions in a digital format. This makes it easier to perform replay attacks provided the assailant has a radio transmitter capable of transmitting on the same frequency as the captured communications. The HackRF is the cheapest SDR capable of receiving and transmitting radio communications at \$299 [5]. Access remotes are thus vulnerable to replay attacks since they perform access control by transmitting radio signals.

Specific emitter identification (SEI) is a technique used to uniquely identify radio transmitters, even those of the same make and model, using only their transmitted radio signals [6]. This means of identification is possible due to hardware tolerances in the radio frequency (RF) circuitry created during manufacturing [7]. SEI is also referred to as radio-frequency fingerprinting (RFF) or physical-layer identification. SEI aims to alleviate the mimicking or spoofing of the identities of radio devices as the identifying characteristics produced by SEI are inherently difficult to spoof [8], [9]. In this way, SEI is used to enhance the security of communication networks using wireless devices.

A typical approach used in SEI is to maintain a library of signal characteristics (or features) that uniquely identify a variety of emitters, and comparing the incoming signal of an emitter to the library of feature sets. The identifier (or label) of the feature set that best matches the incoming signal is assigned to it [6], [10]. An implicit requirement of this approach is that the feature sets cluster. This implies that feature values from a particular emitter are similar and repeatable for all signals produced by the emitter while appreciably distinct from feature values produced by a different emitter [6].

This paper demonstrates how conventional RF access remotes can be uniquely identified using low-cost SDR receivers and SEI. The success of this demonstration suggests that this is a viable approach to increasing the security which can be achieved using conventional RF access remotes.

Section II presents the design and implementation of a proof-of-concept software system that performs SEI to distinguish between two nominally-identical access remotes. Section III describes the results obtained from the study. Section IV concludes the paper.

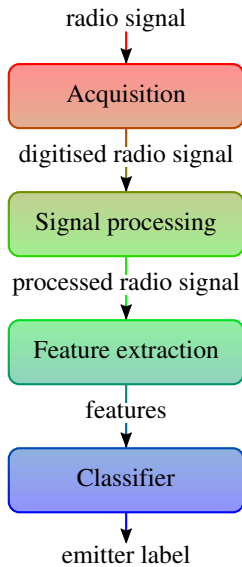


Fig. 1. SEI system overview.

II. SYSTEM DESCRIPTION

The overall SEI system depicted in Fig. 1 consists of the following elements which will be considered below.

- 1) The acquisition system acquires the RF signals produced by the access remotes. It then stores the data in a digital format for later processing.
- 2) Signal processing is then performed on the stored digital RF signals to remove any arbitrary variances in the signals that may distort the signals and affect signal classification.
- 3) The feature-extraction subsystem then extracts distinct features from the processed RF signals.
- 4) The classifier subsystem then takes the extracted features and builds an association between the radio signals and the transmitters from which they were produced.

A. Operating Characteristics of RF Access Remotes

RF access remotes operate in the industrial, scientific and medical (ISM) band at 433 MHz [11], [12]. This band is intended for the operation of equipment designed to use local RF energy for purposes other than telecommunications [13].

These access remotes transmit a modulated sequence of bits to the gate's receiver in order to open or close the gate. This usually takes the form of pulse width modulation (PWM) in which a logical 0 is represented by a short pulse, and a logical 1 is represented by a long pulse [1]. This simple form of modulation makes these access remotes susceptible to replay attacks allowing for illegitimate access to residential estates, houses and garages. This simple modulation scheme also allows for access remotes to be programmed by cloning the signal from another access remote [11].

For the development of this system, two RF access remotes that open the gate to a residential complex were considered. Each remote was distinctly labelled (A and B) as shown in Fig. 2.

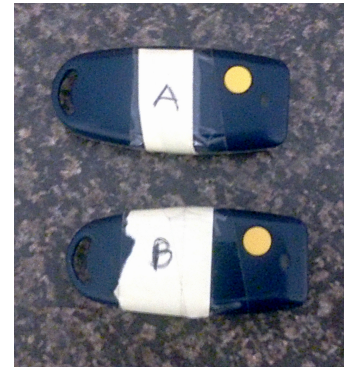


Fig. 2. Access remotes utilised for the development of the proof-of-concept SEI access control system.



Fig. 3. RTL2832U SDR with a Raphael R820T tuner.

The signal characteristics of the RF access remotes will be described in Section II-B during the elaboration of the signal acquisition setup.

B. Signal Acquisition

The signal acquisition system consists of two processes, namely the recording process and burst extraction process.

For signal recording, an RTL2832U SDR with an R820T tuner, shown in Fig. 3, was utilised and interfaced through a GNU Radio applet. The selected SDR is a relatively inexpensive SDR that can sample signals at up to 3.2 Msps and has 8-bit ADC resolution [3]. The SDR receiver was configured as shown in Table I.

The applet was run for 80 s while the button on the remote was continuously pressed for the duration of 80 s. After the 80 s, the applet stored the recorded samples in a binary file for later processing.

The recorded samples were then investigated in order to identify the characteristics of the signals produced by the access remotes. A single burst produced by an access remote is shown in Fig. 4. It is observed that the access remotes' signals consist of a 10.4-ms start pulse followed by twelve modulated pulses comprising a burst with a total duration of 13.6 ms. The start pulse is used for the detection of a signal produced by an access remote. The encoded burst pulses are seen to utilise PWM (as mentioned earlier) in which a short pulse corresponds to a 0 and a long pulse corresponds to 1. Based on this, each access remote transmitted the same bit sequence of 011001100001.

TABLE I
RECEIVER PARAMETERS.

Receiver parameter	Value
Low-noise amplifier (LNA) gain	5 dB
Center frequency	433.91 MHz
Sampling rate	1 Msp/s
Distance from receiver	20 cm

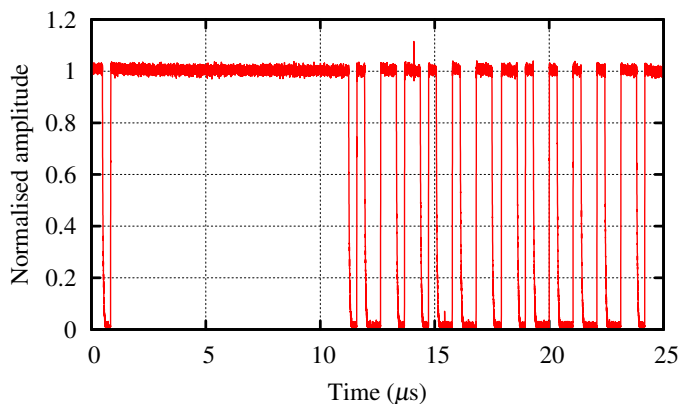


Fig. 4. Captured signal from access remote A.

In order to extract the individual access-remote bursts, burst extraction was performed on the digitally-stored RF signals produced by each access remote. A threshold algorithm was utilised to extract the individual bursts and is described in Algorithm 1. Essentially the algorithm parses the amplitude of the recorded samples and checks the number of consecutive samples above the threshold (the mean value of all the amplitude samples). If the number of consecutive samples is at least 90% of the preceding start pulse length (i.e. 10.4 ms as mentioned earlier), then a burst has been detected. The algorithm then extracts the complex in-phase and quadrature samples from the current search index to 13.6 ms later, which as mentioned earlier, is the approximate burst length. These extracted data constitute a single access remote burst. In this case, the search window is advanced by the sum of the start pulse length (i.e. 10.4 ms) and the burst length (i.e. 13.6 ms). If no burst is detected, the search index is advanced by 100 samples to continue parsing the data in a small window. For the development of this proof-of-concept system, more than a thousand bursts were extracted for each remote.

C. Signal processing

Following recording of the access-remote signals and extraction of bursts, the individual bursts are then further processed in order to remove any arbitrary variances in the bursts that are due to noise, amplitude variances and phase offsets.

The first step taken in processing is filtering out noise. This is typically done by first down-converting the recorded burst to its baseband frequency and then applying a low-pass filter to the signal [14]. In order to correctly filter the noise, the bandwidth of a burst had to be identified. This was done by taking the Fourier transform of a single burst and visually inspecting which frequency bins had the most energy, as shown

Data: Complex samples s

Result: Array of extracted bursts

$amp \leftarrow absolute(s)$;

$threshold \leftarrow mean(amp)$;

$searchIndex \leftarrow 0$;

$burstCount \leftarrow 0$;

while $searchIndex < total\ number\ of\ samples\ in\ amp$ **do**

$window \leftarrow [searchIndex\ to\ searchIndex + 10.4\ ms]$;

if $\Sigma(amp[window] > threshold) > 90\%$ of window

$length$ **then**

increment $burstCount$;

$extractionWindow \leftarrow$

$[searchIndex:searchIndex+window+13.6\ ms]$;

$burstArray[burstCount] \leftarrow s[extractionWindow]$;

increment $searchIndex$ by $(10.4\ ms + 13.6\ ms)$;

else

increment $searchIndex$ by 100 samples;

end

end

return $burstArray$;

Algorithm 1: Algorithm for burst extraction.

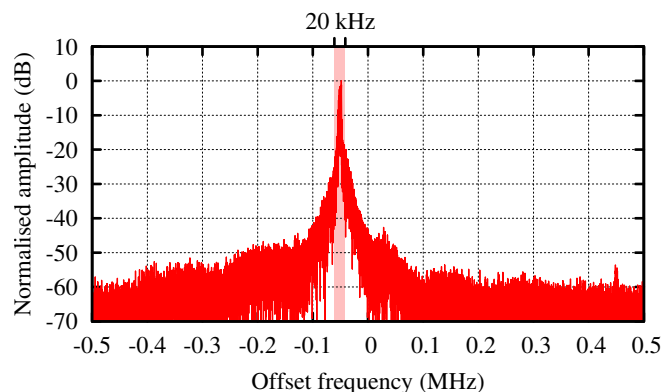


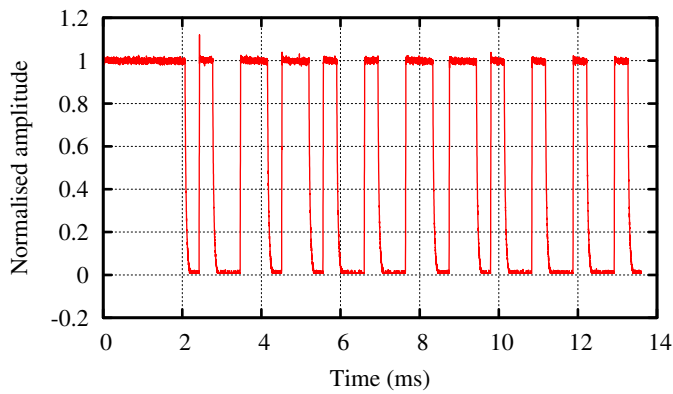
Fig. 5. Magnitude of frequency spectrum computed over a single burst.

in Fig. 5. In this way, the burst bandwidth was determined to be 20 kHz. Thus a finite impulse response (FIR) low-pass filter with a 3 dB cut-off frequency of 10 kHz was utilised in order to filter out noise. The FIR filter consisted of 10 000 coefficients and used a Hamming window.

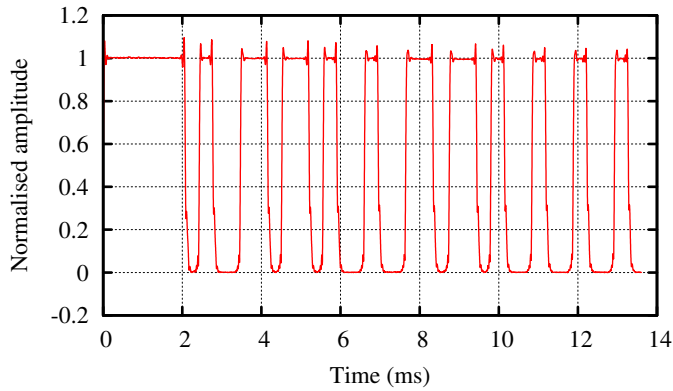
The effect of filtering is demonstrated in Fig. 6 which shows filtered and unfiltered bursts in Figs 6(a) and 6(b), respectively.

Following filtering, the amplitude representations of each burst need to be normalised between 0 and 1 so as to allow bursts recorded at different amplitudes to be compared. This prevents the feature extraction subsystem from producing feature vectors that differ due to amplitude variances between bursts. This would cause the misclassification of bursts even if they were produced from the same access remote.

Similarly, frequency offsets in the phase representations of each burst can cause misidentifications by the classifier. A



(a) Unfiltered.



(b) Filtered.

Fig. 6. Amplitude representation of an access remote burst.

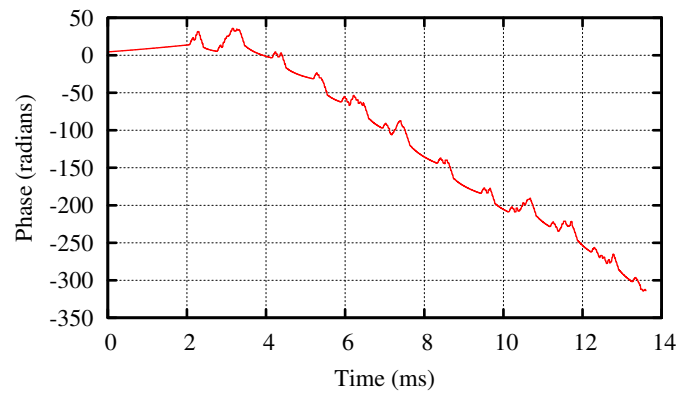
phase representation of a burst with and without a frequency offset is shown in Figs 7(a) and 7(b), respectively.

D. Signal difference inspection

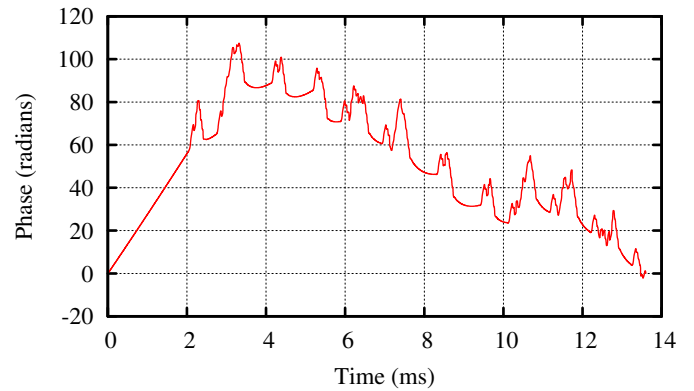
Once signal processing is complete, only then can the true differences between the signals produced by each access remote be determined.

Observing the differences between mean amplitude representations of 100 bursts produced by the individual access remotes, shown in Fig. 8(a), it is seen that there are distinct differences in the amplitude representations. However, these differences are not consistent as shown in Fig. 8(b). The average amplitude representation over the first 100 bursts for access remote A differs from the average amplitude representation for the next 100 bursts. The same holds true for access remote B. As mentioned earlier, for SEI to be successful it is imperative that the characteristics of the signal produced by a specific transmitter be consistent for all signals produced by that transmitter, while being appreciably distinct from the characteristics produced by another transmitter. Based on this observation, the amplitude representations of the access remotes are unlikely to achieve the ultimate goal of classifying the bursts emitted by them.

Observing the differences between the mean phase representations over 100 bursts of access remotes A and B alone (Fig. 9(a)), it is seen that the phase representations for each key



(a) Without frequency offset correction.



(b) With frequency offset correction.

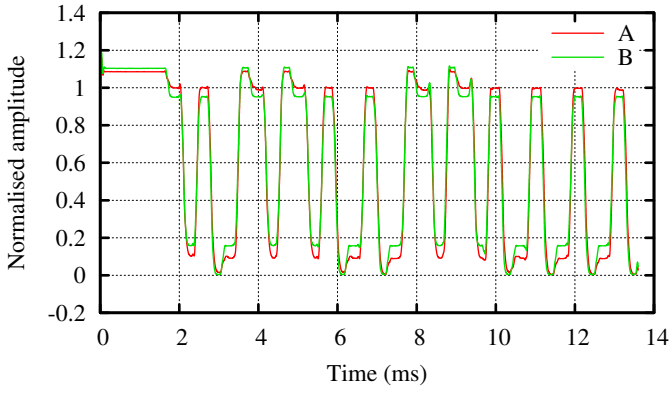
Fig. 7. Phase representation of an access remote burst.

differ significantly. As shown in Fig. 9(b), the post-processed mean phase representations do not exhibit the inconsistencies seen in the amplitude representations. As seen in Fig. 9(b), the mean phase representation for the first 100 bursts of access remote A is similar to mean phase representation for the next 100 bursts. The same holds true for access remote B. These phase differences are more distinct than the differences seen in the amplitude representation. On this basis, the phase representations of the access remotes would be better for the purposes of SEI.

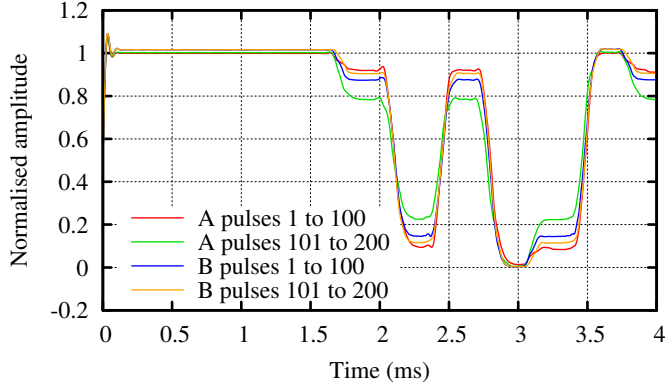
E. Feature Extraction

While it is possible to present the entire amplitude or phase representation to the classifier, this would be inefficient and may hinder the classification accuracy. This is because each sample in the phase and amplitude representations would be treated as a feature leading to an exorbitant number of features. Instead, a set of values that effectively summarises the shape of each representation are calculated. These values then serve as the features for each signal representation and the process is called feature extraction [15].

Statistical measures, namely variance, standard deviation, skewness and kurtosis, are typically used in the SEI of wireless devices such as Global System for Mobile Communications (GSM) cellular telephones [14]. For the development of this system, statistical feature extraction was utilised and is de-



(a) One set of averaged data.



(b) Comparison between different sets of averaged bursts.

Fig. 8. Mean amplitude representation of 100 bursts.

scribed in Algorithm 2. Each signal representation (the mean phase and amplitude representations over a certain number of bursts) is divided into a number of equally sized sub-regions (NR). For each sub-region, the variance, standard deviation, skewness and kurtosis are calculated. These statistical values are then standardised using [14]

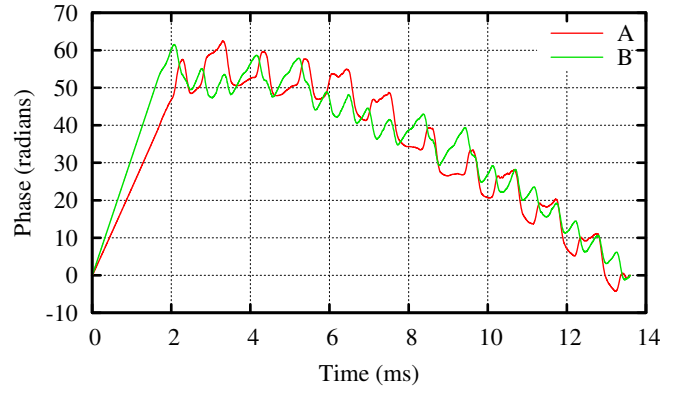
$$[\sigma, \sigma^2, \gamma, k] = \frac{\text{mean}([\sigma, \sigma^2, \gamma, k])}{\text{standard deviation}([\sigma, \sigma^2, \gamma, k])}. \quad (1)$$

Once statistical measures have been calculated for each sub-region, the variance, standard deviation, skewness and kurtosis are calculated over the whole signal representation. The number of sub-regions determined to work best for GSM cellphones was 5 [14], which leads to a total of 24 features per signal representation. Once the statistical features have been calculated for each signal representation, they are concatenated with the first 24 features corresponding to amplitude features and the latter 24 corresponding to phase features. The 48 features in total represent a single feature vector.

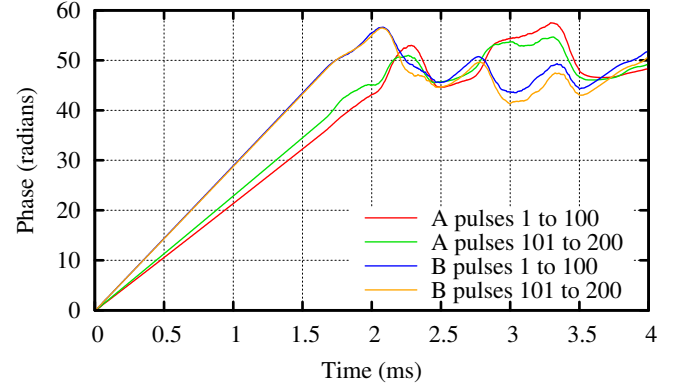
The average representation of the features for each access remote is shown in Fig. 10, with the amplitude and phase features in Figs 10(a) and 10(b), respectively.

F. Signal Classification

Once a set of feature vectors have been established, classification can take place. In order to perform classification,



(a) One set of averaged data.



(b) Comparison between different sets of averaged bursts.

Fig. 9. Mean phase representation of 100 bursts.

the feature vectors have to be segmented into training and test groups for each access remote. The training feature vectors serve to build an association between the feature vectors and the access remotes from which they were derived. This is done by presenting the classifier with a feature vector and an associated access remote label for all feature vectors in the training group. The test group of feature vectors is then used to evaluate the performance of the classifier. In this phase, each feature vector in the test group is presented to the classifier without a label, and the classifier returns the label of the access remote it deems most likely to correspond to the feature vector [16]. It is important to note that the training and test groups of feature vectors must be derived from different bursts. For the development of this system, training feature vectors were derived from the first 200 bursts for each remote. Test feature vectors were derived from bursts 200 to 1000 for each remote.

The classifier utilised was a k^{th} nearest neighbour (KNN) classifier. KNN computes a distance d between an m -dimensional input feature vector \mathbf{x} to a number of training feature vectors \mathbf{t}_r (with the same dimensionality). The label of \mathbf{x} is based on the most occurring label of its k nearest neighbours, where k is a positive integer [16]. The distance measure utilised in the implementation of KNN is the Man-

Data: Array of access remote bursts
Result: Feature vector of statistical features
amplitude representation of burst \leftarrow
 $\text{mean}(\text{absolute}(\text{low-pass filter}(\text{access remote bursts})));$
amplitude representations of burst \leftarrow
 $\text{normalise}(\text{amplitude representation of burst});$
phase representation of burst \leftarrow $\text{mean}(\text{angle}(\text{low-pass filter}(\text{access remote bursts})));$
phase representations of burst \leftarrow remove phase offset
from phase representation of burst;
NR \leftarrow 5;

for each representation **do**
N \leftarrow number of samples in representation;
s \leftarrow $\lfloor N/\text{NR} \rfloor$;
for $m = 1$ to NR **do**
g \leftarrow $m \times s$;
d \leftarrow $(g-s)+1$;
segment \leftarrow representation from samples d to g;
feature_vector(m) \leftarrow $\text{standardise}([\sigma \ \sigma^2 \ \gamma \ k \ \text{of} \ \text{segment}]);$
end
feature_vector(m+1) \leftarrow $\text{standardise}([\sigma \ \sigma^2 \ \gamma \ k \ \text{of} \ \text{entire representation}]);$
end
final_feature_vector \leftarrow concatenate feature_vector 1 to NR+1;
return final_feature_vector;

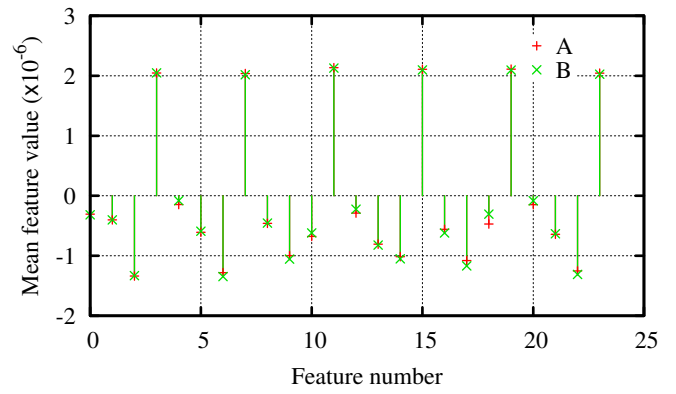
Algorithm 2: Algorithm for feature extraction.

hattan distance calculated by

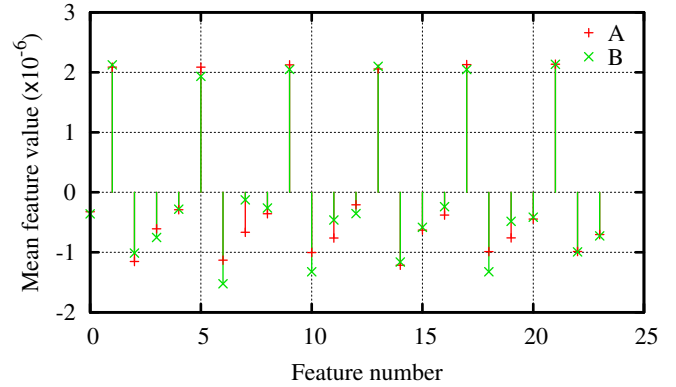
$$d = \sum_{i=1}^m |\mathbf{x}(\mathbf{i}) - \mathbf{t}_r(\mathbf{i})|. \quad (2)$$

Manhattan was the chosen distance measure due to the fact that Manhattan distance is best suited for features that measure dissimilar properties [16]. Given that feature extraction process considers amplitude and phase measures, Manhattan distance is apt for evaluating distance.

Once the distances for all feature vectors are computed, the labels of the k feature vectors corresponding to the lowest distance (nearest neighbours) are considered. The most occurring label among the k nearest neighbours is then assigned to \mathbf{x} . For this process, k is usually set to an odd number to avoid ties [16]. However, for the development of this SEI system, only the average feature vector per access remote was maintained in the memory of the KNN classifier. That is to say, only two feature vectors were presented to the KNN classifier for training, with each one corresponding to the mean of all training feature vectors produced by a particular access remote. Thus, k was set to 1. The processing detail for the KNN classifier is shown in Algorithm 3.



(a) Amplitude features.



(b) Phase features.

Fig. 10. Average representation of the features.

III. RESULTS

As mentioned earlier, feature vectors are derived by taking the mean representations of phase and amplitude over a certain number of bursts. In order to determine the effect that number of bursts utilised during feature extraction had on classification, the following two classification scenarios were considered.

- 1) Classification scenario 1 – Feature vectors were derived over 50 bursts.
- 2) Classification scenario 2 – Feature vectors were derived over 10 bursts.

The confusion matrices are shown in Table II. A summary of the classification accuracy (taken as the mean across the diagonal of the confusion matrix) for each of the scenarios and features used is shown in Table III.

The performance of the system is determined by how well the classifier is able to identify the bursts of the access remotes. Based on Table III, it is seen that the KNN classifier can identify bursts from access remotes A and B with an accuracy of at least 98% provided phase features are utilised. When amplitude features are used exclusively, the accuracy drops to between 53% and 55% depending on the number of bursts utilised to form a feature vector. Furthermore, the high accuracy for phase features in classification scenario 2 indicates that as few as ten bursts can be used to form a feature

Data: Feature vector (\mathbf{x}), training feature vectors (\mathbf{t}_r), class labels l and value for k

Result: Class labels and mean distance from nearest neighbours for each \mathbf{x}

$P \leftarrow$ number of feature vectors \mathbf{x} ;

$Q \leftarrow$ number of training feature vectors \mathbf{t}_r ;

for $p = 1$ to P **do**

for $q = 1$ to Q **do**

$\mathbf{x}' \leftarrow \mathbf{x}(\mathbf{p})$;

$\mathbf{t}_r' \leftarrow \mathbf{t}_r(\mathbf{q})$;

$d(q) \leftarrow \sum_{i=1}^m |\mathbf{x}'(i) - \mathbf{t}_r'(i)|$;

end

 Sort d in ascending order;

 Sort l based on sorted indices of d ;

if $k > 1$ **then**

 nearest_neighbours $\leftarrow l$ from 1 to k ;

 class_result(p) \leftarrow

 most_occurring_class(nearest_neighbours);

 distance_result(p) $\leftarrow \frac{1}{k} \sum_{i=1}^k d(i)$;

else

 class_result(p) $\leftarrow l(1)$;

 distance_result(p) $\leftarrow d(1)$;

end

end

return class_result and distance_result;

Algorithm 3: Algorithm for the KNN classifier.

TABLE II
CLASSIFICATION ACCURACY.

	Features used					
	Amplitude		Phase		Amplitude and phase	
	A	B	A	B	A	B
Classification scenario 1						
A	6.25%	93.75%	100%	0%	100%	0%
B	0%	100%	0%	100%	0%	100%
Classification scenario 2						
A	10%	90%	97.5%	2.5%	96.25%	3.75%
B	0%	100%	1.25%	98.75%	0%	100%

vector distinct enough to provide accurate classification. As a result, an access remote will only need to be sampled for less than a quarter of a second in order to produce the number of bursts required for accurate identification.

IV. CONCLUSION

In conclusion, the development of a proof-of-concept SEI access control system for RF access remotes proved successful. Offline classification was performed on RF bursts produced by two access remotes. When the phase features of the bursts were utilised, the bursts could be identified as belonging to a specific access remote with an accuracy in excess of 98%.

Furthermore, this classification can theoretically be performed

TABLE III
SUMMARY OF CLASSIFIER PERFORMANCE.

Classification scenario	Features used	Classification accuracy
1	Amplitude and phase	100%
1	Amplitude	53.13%
1	Phase	100%
2	Amplitude and phase	98.125%
2	Amplitude	55%
2	Phase	98.125%

by sampling bursts produced by an access remote for less than a quarter of a second. In light of these observations, SEI has been shown to hold tremendous potential to enhance the security of RF access remotes by providing physical-layer identification of the individual remotes and consequently makes these remotes less susceptible to replay attacks.

REFERENCES

- [1] T. Waterowski. (2016, July) H4ck33D – hacking a 433MHz remote control. <http://mightydevices.com/?p=300>.
- [2] M. Dillinger, K. Madani, and N. Alonistioti, "Introduction," in *Software Defined Radio: Architectures, Systems and Functions*, ser. Wiley Series in Software Radio. Chichester, England: Wiley, 2005, pp. xxxiii–xxxiv.
- [3] (2016, July) rtl-sdr – OsmoSDR. [Online]. Available: <http://sdr.osmocom.org/trac/wiki/rtl-sdr>
- [4] (2016, July) GNU Radio. [Online]. Available: <http://gnuradio.org>
- [5] (2016, July) NooElec – HackRF One software defined radio – SDR receivers – software defined radio. [Online]. Available: <http://www.nooelec.com/store/sdr/sdr-receivers/hackrf-one.html>
- [6] K. I. Talbot, P. R. Duley, and M. H. Hyatt, "Specific emitter identification and verification," *Technology Review Journal*, vol. 11, pp. 113–133, 2003.
- [7] B. Danev, D. Zanetti, and S. Capkun, "On physical-layer identification of wireless devices," *ACM Computing Surveys*, vol. 45, no. 1, pp. 6:1–6:29, Dec. 2012.
- [8] M. Williams, M. A. Temple, and D. Reising, "Augmenting bit-level network security using physical layer RF-DNA fingerprinting," in *Global Telecommunications Conference (GLOBECOM 2010)*, Miami, USA, Dec. 2010, pp. 1–6.
- [9] D. R. Reising, M. A. Temple, and M. J. Mendenhall, "Improving intracellular security using air monitoring with RF fingerprints," in *IEEE Wireless Communications and Networks Conference (WNC10)*, Sydney, Australia, Apr. 2010.
- [10] D. Zanetti, V. Lenders, and S. Capkun, "Exploring the physical-layer identification of GSM devices," Swiss Federal Institute of Technology Zurich, Department of Computer Science, Tech. Rep., 2012.
- [11] (2016, July) SENTRY learning 1/3/4 button 433MHz (binary, inverted trinary, SMART). http://www.martin-electronics.co.za/Learning_B_T_F_433Mhz.aspx.
- [12] (2016, July) SENTRY binary remote control transmitter three button (433Mhz) – SENTRY remote. <http://www.taskltd.com/task-online-store/remote-control-transmitters/binary-transmitters-433mhz/sentry-binary-transmitter-three-button-433mhz.html>.
- [13] ITU. (2016, July) Article 1: Terms and definitions. <http://life.itu.int/radioclub/rr/art01.htm>.
- [14] D. R. Reising, M. A. Temple, and M. J. Mendenhall, "Improved wireless security for GMSK based devices using RF fingerprinting," *International Journal of Electronic Security and Digital Forensics*, vol. 3, no. 1, pp. 41–59, Mar. 2010.
- [15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, USA: Wiley-Interscience, 2000.
- [16] S. Russell and P. Norvig, *Artificial Intelligence A Modern Approach*. New Jersey, USA: Pearson Education, 2010.

Review of Data Storage Protection Approaches for POPI Compliance

Nicholas Scharnick

Centre for Research in Information
and Cyber Security
Nelson Mandela Metropolitan
University
Port Elizabeth, South Africa
s209012789@nmmu.ac.za

Mariana Gerber

Centre for Research in Information
and Cyber Security
Nelson Mandela Metropolitan
University
Port Elizabeth, South Africa
Mariana.Gerber@nmmu.ac.za

Lynn Futcher

Centre for Research in Information
and Cyber Security
Nelson Mandela Metropolitan
University
Port Elizabeth, South Africa
Lynn.Futcher@nmmu.ac.za

Abstract—In business, information security has always been a debated topic amongst management and executives. Investing in something that is intangible is often not seen as priority expenditure as it brings no Return on Investment nor contributes to expanding the business. However, the newly enacted Protection of Personal Information (POPI) Act forces businesses to re-evaluate their stance on information security and data storage protection as POPI requires that “appropriate and reasonable security measures” be put in place to effectively protect all personal information that large organisations as well as smaller businesses process and more importantly store. However, the lack of comprehensive controls found within any one information security approach (information security standard, best practice or framework) to fully address the requirements of the POPI act, leaves businesses exposed to legislative action under POPI.

This paper, through the use of a detailed literature review and qualitative content analysis aims to analyze widely implemented information security approaches in the context of POPI compliance. Through identifying themes for data protection within various information security approaches, an evaluation of the comprehensiveness of these approaches and their proposed mechanisms for protecting data within businesses is conducted.

Keywords: Legislation; POPI; Information Security; Business; Data; Data Security; Storage Protection.

I. INTRODUCTION

Businesses are an important aspect of an economy in any country, without them minimal economic growth can occur. In South Africa, the largest business contributors are Small, Medium and Micro Enterprises (SMMEs), as these types of businesses make up 91% [1] of all business entities in South Africa, emphasizing the importance of assisting and supporting all business sizes in order to keep growing the economy.

In today’s competitive business landscape, technology can prove to be a major asset in conducting and growing an established or newly started business. Technologies such as Information Technology (IT) and IT systems have always had a place in any business; however, traditionally placed as a

mechanism to support the business operations. With the rapid evolution of modern technology, businesses now consider IT as an invaluable resource [2], thus IT has moved from a supporting role to a key driving force within the business [3], reaching a point where businesses rely on IT to conduct business. This newly found reliance on IT brings with it a fair number of concerns, including IT security, mismanagement of these systems as well as data protection [4], [5], [6].

This paper investigates how the issues of data storage and data protection within businesses, specifically SMMEs, can be addressed using various information security standards, best practices and frameworks in light of the recently enacted Protection of Personal Information (POPI) Act.

II. RESEARCH METHODOLOGY

For the purpose of this study, a background literature review was conducted in order to gain an understanding of the fundamental principles of information security and data protection. A further literature study was conducted on the recently enacted POPI Act in order to determine the current perception of the Act as well as what the act requires of a business in order to achieve compliance when protecting stored data. Upon determining the requirements for POPI compliance, specifically for protecting stored data, literature was consulted in order to determine which information security standards, best practices and frameworks are being widely adopted by businesses across the globe. Through the identification of such security approaches, the content thereof was analyzed using a qualitative content analysis. A qualitative content analysis can be defined as “A research technique for making replicable and valid inferences from texts to the contexts of their use” [7]. This analysis aimed to identify specific themes which relate to data, data security, protection of data and protecting Personally Identifiable Information (PII) found within the determined security approaches. The identified themes were then summarized within each analyzed information security approach along with the section in which they are located in the respective information security approach. Finally, all information security approaches and the themes identified are summarized to provide a holistic view of the comprehensiveness of addressing data protection.

III. BACKGROUND

The importance of IT in today's modern businesses coupled with the interconnectedness of society and the rapid expansion of the internet and internet based services, social media and networking leads businesses toward adopting IT based solutions in order to remain competitive in an e-commerce driven society. In business, an IT solution can provide a range of benefits, these include cultivating cost effective solutions to the businesses processes, reducing the overhead of business activities through efficient business practices and streamlining business processes through the incorporation of an IT solution [8]. However, should an IT solution be considered, the business needs to keep in mind that these advantages can be overshadowed by potential areas of concern that come with an IT solution. An organisation needs to consider the security implications of implementing an IT solution and ensure these security related concerns are addressed. Not only do these concerns need to be addressed, the organisation should consider the implementation of good information security practices in order to maintain the security of any information assets within the business as well as ensuring proper management of such security implementations.

Information security is defined as “protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction” [9]. Further, the three traditional characteristics at the core of information security, namely Confidentiality, Integrity and Availability (CIA) also need to be upheld and preserved [10]. In order for a business to address any information security concerns that arise from implementing an IT solution, the information security requirements of the business need to be established. Information security requirements are seen as “The amount of security needed to provide the required level of information security” [11]. To determine the appropriate “amount” of security that the business needs, the three originating sources that comprise information security requirements need to be considered. These three sources are stated as [10]:

- Assessing risks to the business, while considering the business goals and objectives.
- Legal, regulatory and contractual requirements that the business has to satisfy.
- Principles, objectives and business requirements that support the business's operations.

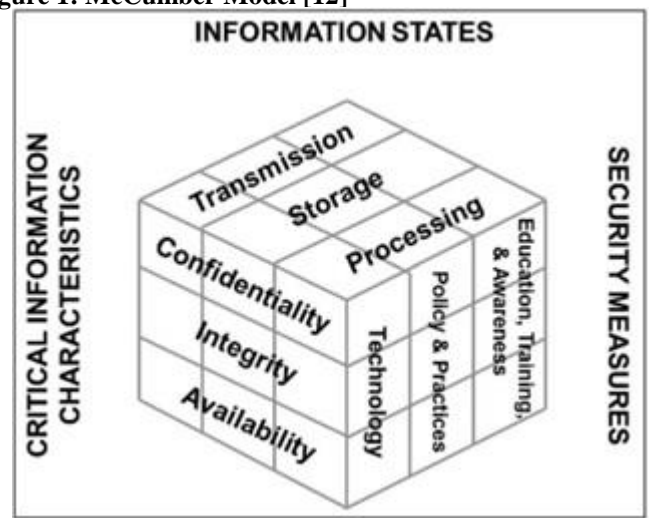
The definition of information security and information security requirements highlights that security is an ever important aspect that has to be considered in the modern and fully digital era that we live in. Thus the protection of a business's information assets becomes a key factor that needs to be addressed when considering how critical these information assets are in modern day business.

When considering “data” and “information” these terms are sometimes used interchangeably. However, these terms bring separate meanings within the context of an information asset. Information assets need to be protected as they originate from data which is then processed into a meaningful context. However, not all types of data require some form of processing (producing information from data) in order to become an

information asset. These types of data include: PII such as names, addresses and identification numbers, credit card details and banking details. These types of data are not contextualized through processing (thus becoming information) yet still hold value to a business, thus becoming an asset.

These information assets can be in a variety of forms within a business as there are three primary “states” that digital information can be in at any time. These states are: storage transmission and processing and are also known as data at rest, in motion and in use. The McCumber model [12], illustrated in Figure 1, depicts the three states of information in the context of the characteristics of information security as well as information security measures to maintain good information security practices within the business.

Figure 1: McCumber Model [12]



Focusing on the state of “Storage”, the business needs to consider upholding the three core characteristics of information security through the implementation of good information security practices. Such practices can be a combination of technological information security measures, sound information security policies and ensuring all employees are trained and educated on good information security practices and are aware of any information security concerns. During the process of determining the business security requirements for stored information, the business needs to address any legal requirements that are identified. Should the business store any PII, it needs to address its legislative requirements by complying to the recently enacted POPI Act [1.].

IV. PROTECTION OF PERSONAL INFORMATION

In order for the business to remain competitive and leverage all resources available, IT needs to be adopted [2]. As mentioned, IT provides a wide variety of advantages to the business, thus making the adoption and implementation of these systems more compelling to managers and business owners alike. However, the business has to consider the implications of implementing IT, addressing any concerns that there might be as well as ensuring that any compliance

requirements are met. This highlights a new challenge for businesses, addressing the legislative compliance requirements set out by the POPI Act. The POPI Act was signed into effect on 27 November 2013 and provides legislation on the protection of personal information in South Africa. While the Act has not yet been fully implemented or compliance enforced due to the act requiring the establishment of an information regulator, a large majority of the Act has not yet commenced. This will only occur once the regulator has been put in place and an official commencement date has been announced by the South African President. However, certain sections of the Act such as those pertaining to the establishment of the information regulator have commenced.

Currently, there are a limited number of sections within the POPI Act that have commenced, these being the following:

- Chapter 1: Definitions
- Chapter 5: Supervision: Part A – Information Regulator
- Chapter 12: General Provisions: Section 112 - Regulations
- Chapter 12: General Provisions: Section 113 – Procedure for making regulations
- Chapter 12: General Provisions: Section 113 – Procedure for making regulations

Within the POPI Act, there are two major sections that deal with the security aspect of information. The first is Chapter 3, titled “Conditions for lawful processing of personal information” which is divided into three parts. Part A contains several “conditions” numbered 1 to 8 as subsections and focuses on processing personal information from a general perspective. Part’s B and C focus on processing of special personal information and personal information of children respectively. The second being Chapter 9: “Trans-border informational flows” which outlines the requirements of a business should personal information be transferred to a third party residing in a foreign country.

Approaching Chapter 3 of the Act from the perspective of information security, an important “condition” within Part A of the Act, being Condition 7: “Security safeguards” is of significance. This condition provides legislative guidance and security requirements for businesses when processing or storing personal information. The condition has 4 subsections that relate to security. When considering data storage, Subsection 19 of the condition addresses security measures that pertain to maintaining the integrity as well as confidentiality of any personal information within an organisation. The requirements outlined by Subsection 19 ensure that confidentiality and integrity of information is maintained while at the same time the business takes “appropriate, reasonable, technical and organizational measures” [13] to prevent loss, damage, unauthorized destruction and unlawful access to any personal information held by the business.

Looking at the above statement taken from the Act, what is required for compliance is unclear and can be interpreted in many ways. The Act does not define what is seen as “appropriate, reasonable, technical and organizational measures” and could lead to uncertainty and confusion for business owners and managers alike. In business this can easily

occur as having insufficient knowledge or expertise on the security domain can hinder the understanding of what is required in order to implement “appropriate, reasonable, technical and organizational” measures for protecting personal information within the business.

Subsection 19 does provide limited guidance on addressing such concerns, stating that all foreseeable risks to the information held by the business are identified and that appropriate safeguards are put in place to mitigate such risk to the personal information. Further, the Act states that the business should ensure that such safeguards are effective in preventing or mitigating the associated risk and ensure that they are continually updated for such purpose.

Finally, the Act states that an organisation “must have due regard to generally accepted information security practices and procedures” [13] and ensure that any industry or profession related requirements are complied with when regarding good security practices.

Achieving compliance to POPI can be a demanding task as once again, there is a definitive lack of detail within the “guidance” for businesses to follow in order to become compliant with the requirements of the POPI Act. Considering that all businesses are required to comply with what is set out in the POPI Act, coupled with the legislative action that could be taken against businesses for non-compliance; places more strain on these businesses.

V. APPROACHES FOR POPI COMPLIANCE

Addressing the legislative compliance requirements of POPI can be done using one of the many well established information security approaches found within the information security domain. Some of the better recognized and more mature approaches will be analyzed and discussed in detail, placing focus on various themes found within these approaches that aim to address information security from a data protection perspective. The approaches discussed include key role players namely ISO, NIST, COBIT and ITIL.

A. ISO 27000 series of standards

The International Organisation for Standardization (ISO) 27000 standards are a series of best practice standards aimed at providing recommendations for the management of various information security topics within businesses [14]. Notable standards within the 27000 series are ISO27001, ISO27002 and ISO 27040. These standards address the implementation of an Information Security Management System (ISO 27001) [15], providing a code of practice for information security within a business (ISO 27002) [10] and providing guidance for storage security within the business (ISO 27040) [16].

Within these three ISO standards, a variety of data protection related themes can be found. Such themes provide guidance to the business on better protecting any business related data.

TABLE I. ISO 27000 SERIES THEMES

Themes	Source
Access Control	27002
Awareness and Training	27001/27002
Backup	27002
Data Confidentiality and Integrity	27040
Data Reliability and Availability	27040
Data Retention	27040
Information Protection	27002/27040
Media Handling	27002
PII	27002
Storage Management	27040

Referring to Table I, the themes drawn from the ISO 27000 series addresses the three main aims of information security, those being to ensure the confidentiality, integrity and availability of information. From the legislative aspect and specifically POPI, ISO provides guidance on PII. Considering the themes found in Table I, the ISO 27000 series addresses data protection from various angles, ensuring access control measures are in place, providing for retention and backup procedures as well as ensuring the business is aware of security concerns and trained to practice sound information security.

B. NIST

The National Institute of Standards and Technology (NIST) is a United States institution that aims to develop documentation and guidance on various security concerns. While there are a range of publications by NIST that address a wide variety of security related topic areas, Special Publication (SP) 800-122 and 800-171 focus on PII. PII is defined as “any information about an individual maintained by an organisation, which includes any information that can be used to distinguish or trace an individual’s identity” [17].

TABLE II. NIST THEMES

Themes	Source
Access Control	SP800-171
Awareness and Training	SP800-171/122
Breach Protection	SP800-122
Identification and Authentication	SP800-171
Information Confidentiality	SP800-122
Information Integrity	SP00-171
Media Protection	SP800-171
Personnel Security	SP800-171
PII Protection	SP800-122/171
Physical Protection	SP800-171
Privacy	SP800-122

Similar to the themes identified within the ISO series of standards, the NIST publications address aspects that are core to good information security practices. Protecting the confidentiality and integrity of any information within the business is of the utmost importance. Table II highlights an array of themes identified within the NIST publications which provide guidance for protecting data in the business. Key themes to note are breach protection and physical protection, ensuring physical hardware is secure and protected as well as virtual access is secured with the appropriate technical measures such as firewalls, encryption and other mechanisms.

C. COBIT

The Control Objectives for Information and Related Technology framework, or better known as COBIT, is a framework developed in 1996 by the Information Systems Audit and Control Association (ISACA). COBIT aims to provide a comprehensive and holistic approach to governing and maintaining IT within the organisation [18]. The most recent iteration of the framework, COBIT 5 contains five main principles. These principles include:

- Meeting stakeholder needs;
- Covering the enterprise end-to-end;
- Applying a single integrated framework;
- Enabling a holistic approach;
- Separating governance from management.

The COBIT framework contains control objectives for various topic areas and scenarios within the business. The major domains that these control objectives cover include: planning and organisation, acquisition and implementation, delivery and support; and monitoring.

TABLE III. COBIT THEMES

Themes	Source
Access Control	DS 11.6
Backup	DS 11.25
Data Integrity	DS 11.30
Storage Management	DS 11.19

The five main principles of COBIT suggest that the framework is more business oriented, focusing on aspects of the business such as ensuring full coverage of the enterprise as well as addressing its stakeholder’s needs. However, the COBIT framework does provide some guidance on information security within the business, including some basic aspects of data protection.

Table III summarizes the identified themes within COBIT, which include ensuring that the integrity of data is upheld, managing storage and backup of data as well as implementing access control measures within the organisation.

D. ITIL

The Information Technology Integrated Library (ITIL) is an IT service management framework that is widely adopted worldwide and aims to provide businesses with a practical approach to the identification, planning, delivery and support of IT services within businesses [19]. The ITIL framework is comprised of five volumes, covering major IT related areas such as: service design, strategy, operation and transition, as well as providing continual service improvement. When considering the ITIL framework from an information security perspective, each of the five volumes contributes to this topic in some form. However, service design and specifically Section 4.6 of this volume contributes to this in a major way. This includes guidance for an Information Security Management System (ISMS) within the organisation as well as an information security policy [20].

TABLE IV. ITIL THEMES

Themes	Source
Access Control	Section 4.5
Breach Protection	Section 4.6.5.2
Policy/Procedures/Controls	Section 4.6.4/4.6.5.1

The ITIL framework focuses on the topic of IT in the business, with the aim of ensuring the IT systems and services function effectively and are managed in such a way that supports this aim. Considering the identified themes; Table IV highlights that within ITIL there is little focus on information security, taking a similar approach as with COBIT by providing the basic guidance on security measures within the organisation. Thus ensuring appropriate access control and breach protection measures are put in place through the effective use of policies and controls.

TABLE V. SUMMARY OF THEMES

Themes	ISO	NIST	COBIT	ITIL
Access Control	27002	SP800-171	11.6	4.5
Awareness and Training	27001/27002	SP800-122/171		
Backup	27002		11.25	
Breach Protection		SP800-122		4.6.5.2
Data Confidentiality and Integrity	27040	SP800-122/171	11.30	
Data Reliability and Availability	27040			
Data Retention	27040			
Identification and Authentication		SP800-171		
Information Protection	27002/27040			
Media Handling	27002	SP800-171		
Personnel Security		SP800-171		
Physical Protection		SP800-171		
PII	27002	SP800-122/171		
Privacy		SP800-122		
Storage Management	27040		11.19	

VI. SUMMARY

Table V summarizes the themes found within each of the four reviewed information security approaches, providing a complete picture on the extent of guidance provided for data protection within each approach.

As previously mentioned, ISO and NIST have published a wider variety of documents for a range of security topics. Thus as seen in Table V, the approaches provided by these two bodies are more comprehensive in addressing security concerns specific to data protection over their counterparts COBIT and ITIL. However, this does not make any one standard, a superior choice over any other as an important point to note is “Every business has distinct data protection, backup and recovery needs and there is no one-size-fits-all solution” [21].

Table V provides an overview of the themes found within each of the four information security approaches. However, each information security approach addresses the identified themes in different ways. To follow, a comparative summary describes how each security approach addresses a specific theme in relation to each of the other approaches.

A. Access Control

The ISO series of standards addresses access control mechanisms for data protection within ISO27002. Providing generalized guidance on implementing controls for restricting access to information through the controlling of access rights of users as well as other applications. The NIST publication SP800-171 details access control requirements for protecting controlled unclassified information and takes a more detailed and technical approach over its ISO27002 counterpart.

SP800-171 states 22 requirements when addressing access control, these include:

- Monitor and control remote access sessions;
- Limit unsuccessful logon attempts;
- Encrypt controlled unclassified information on mobile devices.

However, COBIT and ITIL both address access control from a broader more business orientated perspective. The COBIT framework provides some basic requirements for a business to follow for addressing access control. While the approach of ITIL extends the businesses information security policies through access management, the ITIL approach does not set policies for access control but merely executes the information security policies already defined within the business.

B. Awareness and Training

The theme of awareness and training are covered fairly broadly within ISO and NIST. ISO27001 and ISO27002 both provide guidance for training and awareness within the business. ISO27001 focuses on this from the perspective of the Information Security Management System (ISMS), stating that all employees with responsibilities defined within the ISMS are competent to perform such duties.

ISO27002 approaches awareness and training by first ensuring employees are aware of the businesses security policies and procedures before being granted any access to information. ISO27002 also states that ongoing training should occur, ensuring employees are aware of any business related controls in place as well as its legal responsibilities. This especially holds true for the POPI Act, as it is the businesses legal responsibility under the POPI Act to protect any personal information stored within the business.

NIST SP800-122 states that “Awareness, training, and education are distinct activities, each critical to the success of privacy and security programs” [17]. From the viewpoint of POPI, it is imperative that the businesses employees are sufficiently trained to carry out and adhere to their responsibilities to protect any personally identifiable information or risk legislative action.

C. Backup

ISO27002 and specifically Section 10.5 of the standard, aims to address the topic of backup, providing guidance to the organisation with the objective of maintaining the integrity and availability of the information. In order to accomplish this, the guidelines put forward ensure that the frequency of performing backup’s falls in line with the business requirements. These guidelines also emphasise the importance of implementing an appropriate amount of physical and virtual security, while ensuring that the reliability of any backed up data is regularly tested.

Within the COBIT control objectives, DS11.25 addresses “Back-up Storage”, highlighting that back-up procedures should include the proper storage of data files and any related documentation. Further, DS11.25 states that the storage sites

for any backup data should be periodically reviewed to ensure the effectiveness of all security measures.

D. Breach Protection

Breach protection from the context of data protection is not addressed comprehensively within any of the four security approaches that were analyzed. ITIL makes mention of the management of security breaches within Section 4.6.5.2. Similarly, NIST approaches breach protection from a management perspective, stating that management of such breaches would require “coordination among personnel from across the organization” to ensure effective management of such an incident. However, NIST does state that breaches which involve PII need to be handled in an alternative manner. This is due to the potential sensitivity of the information as well as the possible impact such an incident might have on the business. From the perspective of POPI, should such a breach occur, notification to the information regulator as well as the data subject(s) involved needs to be made in accordance to Section 22 of the Act.

E. Data Confidentiality and Integrity

Within the NIST Special Publication 800-122, a section entitled “PII Confidentiality Safeguards” discusses two major options that organisations can select from in order to address PII concerns and maintain the confidentiality of any PII that it holds. The first option available to organisations is to develop various policies and procedures to address the concerns they have regarding the PII within the business. This would then serve as the “rules” to follow in order to protect the confidentiality of any PII.

The second approach that organisations can opt for is education, awareness and training programs for their employees. Such programs should focus on methods to ensure confidentiality as well as making employees aware of any potential security threats that could impact the confidentiality of the PII within the organisation.

COBIT once again takes a business and management approach to providing guidance, stating that the integrity of data be checked periodically.

F. Media Handling

ISO27002 addresses important aspects of media handling which includes the management of any removable media within the business as well as the disposal thereof when it is no longer needed. Guidance provided by the standard for managing any removable media within the business includes ensuring that accurate records are kept for auditing purposes when media is removed from the business by authorized personnel. Media management also includes the security of the physical storage media, storing it in an appropriate manner while keeping it safe and secure.

The NIST Special Publication SP800-171 provides more detailed guidance to the business for protecting its storage media. Stating various technical measures for securing such media while ensuring only authorized personnel have access as well as ensuring that the confidentiality of the information is upheld.

G. Personally Identifiable Information

The ISO27002 standard provides guidance for the development of an organizational privacy policy. This policy can be seen as an important step towards securing the data and more importantly, PII stored within the business. However, a privacy policy needs to be communicated to all employees within the organisation that handle any private information to ensure the effectiveness of the policy and to better protect sensitive personal information. While an organisational privacy policy is important, compliance with all relevant data protection legislation and regulation is crucial.

PII from the NIST perspective “should be protected through a combination of measures”. NIST provides guidance for the implementation of operational safeguards as well as detail on policy and procedure creation for protecting PII within the business.

H. Storage Management

Considering the importance of storage in today’s modern society, COBIT provides minimal guidance to businesses for addressing this topic. COBIT states that procedures should be developed to address any data storage concerns and ensure effective management thereof.

The ISO 27040 standard provides supporting controls for storage management and aims to address the security requirements of the organisation from a data protection and security standpoint. Once management and administrators are aware of the risks associated with storing information within the business, ISO 27040 strives to provide security guidance for protecting this information.

The summary of these themes provides an overview of each theme as addressed by its corresponding sources, as well as how each source approaches said theme. In order to provide an understanding and detail as to how each of the four security approaches address the topic of data protection, themes that were identified within more than a single security approach were discussed and compared in order to reveal the differences in the guidance provided.

VII. CONCLUSION

Within this paper, various information security standards, best practices and frameworks were discussed. These best practices aimed at providing guidance to businesses for implementing and maintaining sound information security controls and procedures within the business. This includes guidance on developing and implementing an Information Security Management System as well as implementing various information security controls and procedures to manage and reduce the information security risk of the organisation.

The ISO series of standards also includes ISO 27040. While being a fairly new standard, ISO 27040 sets out to guide businesses in better protecting their stored data by building on and extending the controls found in ISO 27001/2. The topic of protecting data and PII within businesses is of the utmost importance in today’s modern and technology centric era. This highlights the need for specific guidance on the topic of data protection. Other major and popular security best

practice frameworks were discussed, including the COBIT governance and control framework as well as the ITIL IT service management best practice framework.

Looking at the summarized findings (Table V) one can see that there is “no one size fits all” solution to addressing the protection of data and more specifically, when data is in storage. Should a business organisation want to comprehensively protect any stored data, a combination of the security approaches which have been analyzed within this paper can be applied to address a large majority of the potential areas for security concern.

VIII. FUTURE RESEARCH

Future research will focus on data storage and protection within SMMEs in South Africa. Making use of a survey aimed at identifying the types and scales of IT infrastructure within SMMEs and how they address the topics of information security and data protection within the business. The survey will also assist in determining the current understanding of the POPI Act as well as which security approaches SMMEs currently implement with respect to data storage and protection. From the results of the survey, coupled with the themes identified within this research paper, guidelines for SMMEs will be developed which will assist these businesses to better protect any stored data within the organisation, while at the same time also assisting in positioning themselves for the POPI Act and compliance thereto.

IX. ACKNOWLEDGEMENTS

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.

REFERENCES

- [1] Abor, J., & Quartey, P. (2010). Issues in SME development in Ghana and South Africa
- [2] Nieto, M. J., & Fernández, Z. (2005). The role of information technology in corporate strategy of small and medium enterprises. *Journal of International Entrepreneurship*, 3(4), 251-262.
- [3] Lester, D. L., & Tran, T. T. (2008). Information technology capabilities: Suggestions for SME growth. *Institute of Behavioral and Applied Management*, 72-88.
- [4] Lucchetti, R., & Sterlacchini, A. (2004). The adoption of ICT among SMEs: evidence from an Italian survey. *Small Business Economics*, 23(2), 151-168.
- [5] Hashim, J. (2007). Information communication technology (ICT) adoption among SME owners in Malaysia. *International Journal of Business and Information*, 2(2), 22
- [6] Blackburn, R., & Athayde, R. (2000). Making the connection: the effectiveness of Internet training in small businesses. *Education+ Training*, 42(4/5), 289-299.
- [7] Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. Sage.
- [8] Bruque, S., & Moyano, J. (2007). Organisational determinants of information technology adoption and implementation in SMEs: The case of family and cooperative firms. *Technovation*, 27(5), 241-253.

- [9] Andress, J. (2014). The basics of information security: understanding the fundamentals of InfoSec in theory and practice. Syngress.
- [10] ISO/IEC (2005). ISO 27002: 2005. Information Technology-Security Techniques-Code of Practice for Information Security Management. ISO.
- [11] Gerber, M., von Solms, R., & Overbeek, P. (2001). Formalizing information security requirements. *Information Management & Computer Security*, 9(1), 32-37.
- [12] NSTISSI No. 4011 (20 June 1994) National training standard for information systems security (InfoSec) Professionals
- [13] POPI (2013) "Protection of Personal Information Act", South African Government Gazette (2013) Retrieved from <http://www.justice.gov.za/legislation/acts/2013-004.pdf> Date retrieved: 12 August 2015
- [14] ISO/IEC (2014) ISO 27000: 2014. Information technology — Security techniques — Information security management systems — Overview and vocabulary.
- [15] ISO/IEC (2005). ISO 27001: 2005. Information technology — Security techniques — Information security management systems — Requirements
- [16] ISO/IEC (2015). ISO 27040: 2015. Information technology -- Security techniques -- Storage security
- [17] McCallister, E., Grance, T., & Scarfone, K. A. (2010). SP 800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII), National Institute of Standards & Technology, Gaithersburg, MD.
- [18] COBIT 5 A Business Framework for the Governance and Management of Enterprise IT (2013).
- [19] Arraj, V (2013). ITIL: The basics
- [20] Clinch, J. (2009). ITIL V3 and Information Security. Best Management Practice.
- [21] "Why SMEs should back up their data to the cloud" (2013, March 11) Retrieved from <http://www.theguardian.com/small-business-network/2013/mar/11/back-data-to-cloud-small-business>. Date retrieved: 22 March 2015.

CDMA in Signal Encryption and Information Security

Olanrewaju B. Wojuola, Stanley H. Mneney and Viranjay M. Srivastava

School of Engineering
University of KwaZulu-Natal
Durban - 4041, South Africa.

wojuolao@ukzn.ac.za, mneney@ukzn.ac.za, viranjay@ieee.org

Abstract— Code-division multiple-access (CDMA) is a communication technique that was developed originally for the military because of its jam-resistant properties. It is one of the early forms of jam-resistant, signal encryption techniques used in military applications for the purpose of wireless signal transmission and information-hiding from adversaries. In recent years, CDMA has also played a key role in mobile telephony as a multiple-access technique because of certain properties that make it suitable for commercial and civilian applications. This paper gives a brief exposition on CDMA as a signal encryption technique, and the position that it occupies in future wireless technology. This paper also compares CDMA technology with a relatively recent technique, interleave-division multiple-access (IDMA) that has been attracting significant attention in wireless circles.

Keywords— Code-division multiple-access, Interleave-division multiple-access, Signal encryption, Information security, Wireless communication.

I. INTRODUCTION

The advent of code-division multiple-access (CDMA) dates back to the 1940s. It was developed originally for the military as a means of establishing secure, jam-resistant communications [1-6]. During transmission, the existence of a CDMA signal can hardly be detected as it appears as noise spread out over a wideband channel, unlike amplitude or frequency modulated carriers that have concentrated energy over a narrowband. CDMA's large bandwidth makes it difficult to jam. In addition, CDMA signal energy is well below noise level, meaning that the signal is buried (hidden) in noise. It is for these reasons that CDMA can be used for covert transmissions.

CDMA relies on coding for user-separation. It involves the use of spreading codes (also known as spread-spectrum codes, pseudo-noise (PN) codes, or pseudo-random noise (PRN) codes or sequences) as user identification element. Examples of such codes include maximal linear code sequences, Gold codes, Walsh-Hadamard codes and Kasami codes [7-11]. In a CDMA system, each user is assigned a unique spreading code, and uses this code to encode (encrypt) the user's signal into a wideband signal. The receiver requires a knowledge of this unique code before the transmitted information can be detected and decoded. For good performance in multiple-access

applications, spreading codes are required to have minimum cross-correlation between them.

CDMA can be classified into four protocol types: direct sequence CDMA (DS-CDMA), frequency-hopping CDMA (FH-CDMA), time-hopping CDMA (TH-CDMA) and hybrid CDMA [1, 12, 13]. The last group (hybrid CDMA) is obtained from any combination of the first three, or CDMA with any other technique.

This paper gives a brief exposition on the DS-CDMA from encryption point of view. This paper also briefly considers interleave-division multiple-access (IDMA) and the position that the techniques occupy in future wireless technologies. Other advanced forms of CDMA and IDMA systems exists (e.g. MIMO CDMA systems, space-time coded multicarrier CDMA systems, multicarrier IDMA systems, etc.) [14-16], but these are not the focus of this paper.

In literature, IDMA is usually presented as a better alternative to CDMA. As we shall see later in this paper, this common view does not represent the true picture, particularly from information security point of view.

The rest of this paper is organised as follows. Basic principles of operation of CDMA systems are presented in section II. By appealing to the basic theory, the use of CDMA (or spread spectrum) in signal encryption and information security has been explained in section III, followed by an illustrative example in section IV. The system performance curves are used in section V to further explain the principles behind the use of spread spectrum techniques, its application is in Section VI. In section VII, IDMA has been introduced. This is followed by a critical look at the position of IDMA and CDMA in information security and future wireless systems in Section VIII. Finally, the section IX concludes the work and recommend the future aspects.

II. BASIC THEORY OF CDMA SYSTEMS

Consider a DS-CDMA system. Let $b_n(t)$ (with a bit time T) be the data for the n^{th} user and $C_n(t) = \sum_{i=1}^N c_n(t - iT_c)$ be the unique code for the user. If we assume that there are M users, then $0 \leq n \leq M$, and there are M unique codes. The coded output for each user is given by

$$y_n(t) = b_n(t) \cdot C_n(t) = b_n(t) \sum_{i=1}^N c_n(t - iT_c), \quad (1)$$

where T_c the chip time, is much less than the bit time T . This multiplication has the implication that the spectrum of the bit which is proportional to $1/T$ is now much larger and is proportional to $1/T_c$. Thus the encoding in (1) spreads (enlarges) the spectrum of the signal and it is for this reason that CDMA is sometimes referred as spread-spectrum multiple access (SSMA). The spread factor is given by the ratio T_c/T .

Assume the presence of other users in the communication channel. At the receiving side, the signals from all users reach the receiver simultaneously. For a Gaussian channel, received signal $r(t)$ is:

$$r(t) = \sum_{n=1}^M b_n(t) \sum_{i=1}^N c_n(t - iT_c) + n(t), \quad (2)$$

where $n(t)$ is additive white Gaussian noise with a double-sided power spectral density $N_o/2$. The signals from other users constitute interference. In order to recover the data from a specific user (selecting user 1), the composite signal is multiplied by the specific user code as in equation (3):

$$r_{n=1}(t) = b_1(t) \sum_{i=1}^N [c_1(t - iT_c)]^2 + \sum_{n=2}^M b_n(t) \sum_{i=1}^N [c_1(t - iT_c)] c_n(t - iT_c) + n(t) \sum_{i=1}^N [c_1(t - iT_c)], \quad (3)$$

Since $[c_1(t - iT_c)] \in \mp 1$, $[c_1(t - iT_c)]^2 = 1$, equation (3) thus reduces to:

$$r_1(t) = b_1(t) + \sum_{n=2}^M b_n(t) \sum_{i=1}^N [c_1(t - iT_c)] c_n(t - iT_c) + n(t) \sum_{i=1}^N [c_1(t - iT_c)]. \quad (4)$$

In equation (4), $b_1(t)$ is the recovered data for user 1; the second term represents multiple-access interference (MAI) from other users and the third term is noise which is spread out further.

In digital DS-CDMA represented by equation (1) to (4), the message signal is, in principle, multiplied directly by the code signal and the resulting signal modulates a carrier for onward transmission through a communication channel. The receiver correlates the received signal with the code of the user. Because each user's unique code has low cross-correlation with the other codes, the receiver is able to distinguish between users. Correlating the received signal with a code for a certain user de-spreads (decodes) the signal for the user.

III. CDMA AND INFORMATION SECURITY

The possibility of using CDMA in information-hiding centres around signal spreading as in fig. 1. In this figure, the message signal is multiplied by the spreading code to give the spread spectrum (SS) signal, spreading out the signal energy over a wideband. By spreading the spectrum of the signal, its energy or power density can be reduced to a level much lower than that of channel noise. Furthermore, the spreading process makes the signal itself to look like noise. Thus the signal is hidden inside the channel noise. An adversary cannot perceive the existence of the communication because the signal is buried in noise. A receiver can detect and decode the signal if and only if the receiver knows the spreading code with which the signal was encoded originally. Thus the code serves as the key for recovering the original information.

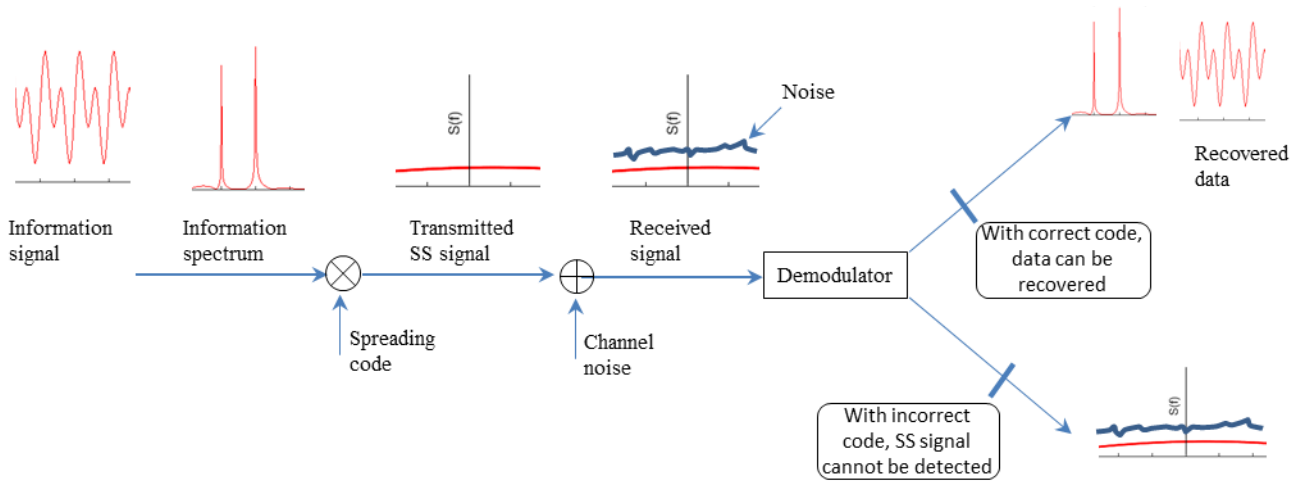


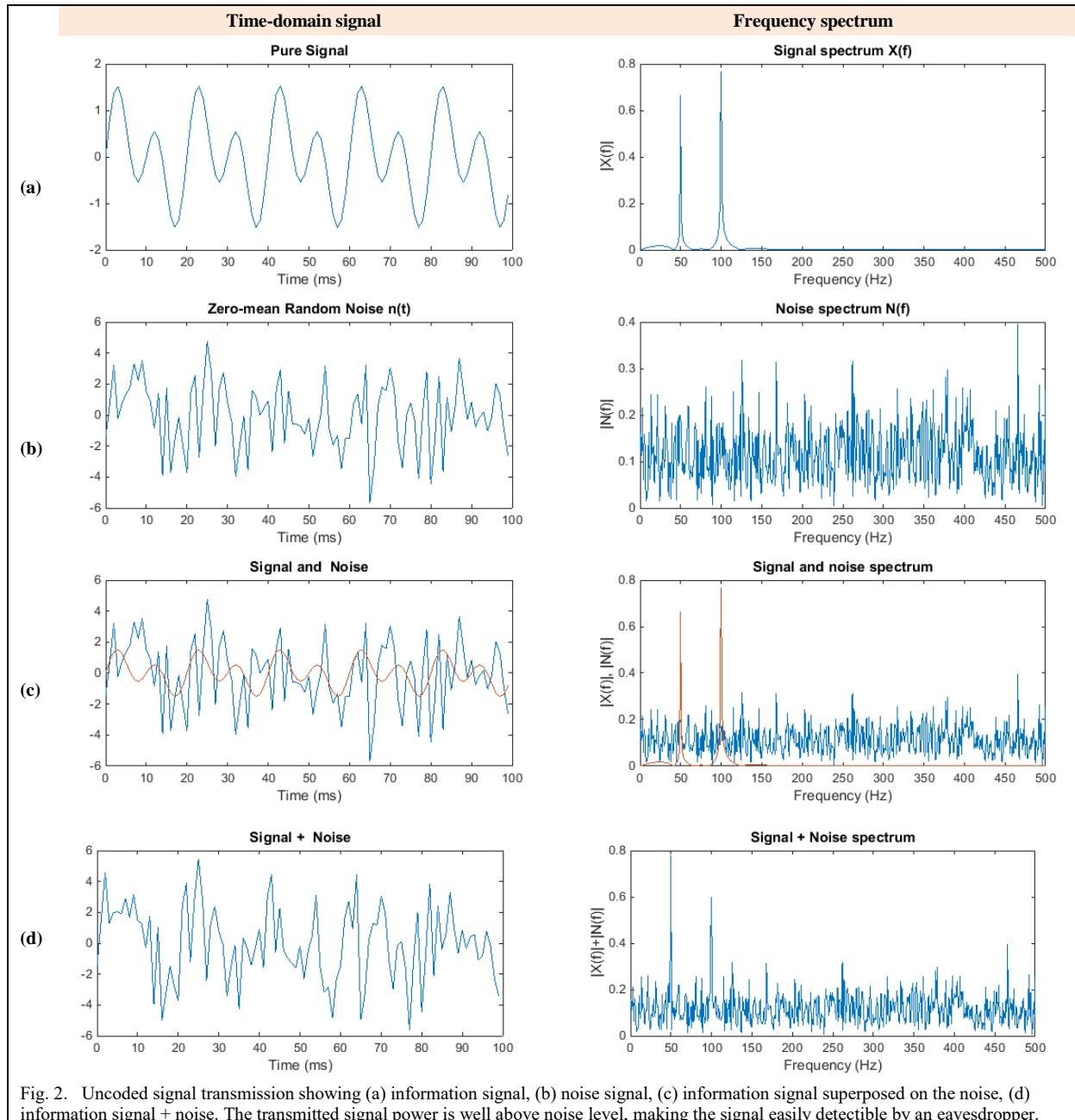
Fig. 1. The use of spread-spectrum techniques in information-hiding. (An adversary cannot perceive the existence of the communication because the signal is buried in noise).

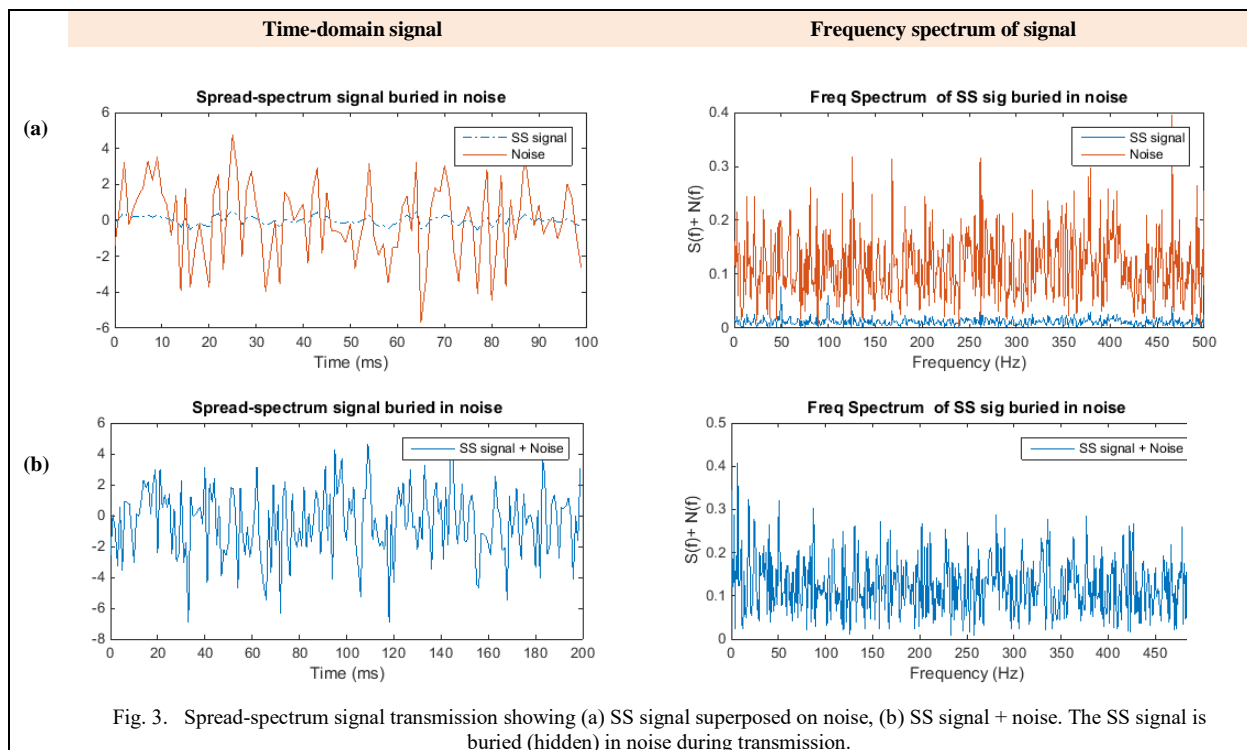
IV. AN ILLUSTRATIVE EXAMPLE

An illustrative example shall now be used to further exemplify the principles behind the use of spread-spectrum techniques in information security. We shall first consider plain, *uncoded* signal transmission. By this we mean transmission not involving the use of spreading codes. Let the information signal $x(t) = 0.7 \sin(2\pi(50)t) + \sin(2\pi(100)t)$. Fig 2(a) shows the signal both in time and frequency domain. Its frequency spectrum clearly shows the signal's component frequencies (50 and 100 Hz), in agreement with the analytic expression $x(t)$ for the signal. Now consider a Gaussian noise having a mean of zero and unit variance. Fig. 2(b) shows the noise and its spectrum. Fig. 2(c) shows the results of

superimposing the signal on the noise. Clearly, the signal power is much stronger than the noise, so that the signal can be picked up easily by an observer, making the signal vulnerable to attack. Fig 2(d) is a sum of the signal and the noise. The frequency spectrum of this figure shows that the presence of the signal is still visible even when mixed with noise. This shows the vulnerability of uncoded signal transmission.

Now consider the use of spreading codes. Let the signal energy be spread out over a wide bandwidth, to give a spread-spectrum (SS) signal. Fig 3(a) shows the SS signal, superimposed on the channel noise (the SS signal and the noise are superimposed in one plot both in time and frequency domain).





Clearly, the SS signal power is much less than that of the noise. Frequency spectrum (right graph of Fig. 3(a)) shows that the SS power spectral density is much below that of noise

Fig. 3(b) shows a plot of the sum of the signal and the noise. Clearly, this resulting sum-signal looks entirely like noise, and it is difficult to recognise the presence of the actual signal. That is to say, the spread spectrum signal is buried (hidden) in the noise. Therefore during transmission, it is difficult to detect the presence of the actual signal, making it hidden from eavesdroppers.

V. CDMA PERFORMANCE CURVES

Performance curves [17, 18] for a CDMA system gives another way of viewing the information-hiding capability of a CDMA system. Fig. 4 shows the system performance for Gold codes of different lengths N in terms of bit-error-rate (BER) as a function of signal-to-noise ratio (SNR). Here, zero decibel (0 dB) represents the point where signal strength is the same as that of noise. In order words, 0 dB represents the noise level.

Communication systems are normally operated at low BER. Therefore as we look at this performance curves, we shall be focussing on the behaviour at low BER. The right-most curve on the figure is that of uncoded signal transmission. By *uncoded* we mean transmission not involving the use of power is much above noise level. Therefore the signal is easily detectable by eavesdroppers spreading codes. At a BER of 10^{-5} , the uncoded system has an SNR of about 12 dB, which is equivalent to 15.85. That is, the signal power is 15.85 times bigger than the noise power. This implies that for the uncoded signal transmission, the signal power is much above noise level.

Now in Fig. 4, consider the performance curve for the shortest code length ($N = 31$). For this curve, at a BER of 10^{-5} the system SNR is about -2 dB, which is equivalent to about 0.63. That is, the signal power is about 0.63 times the value of the noise power. This implies that the 31-chip spreading code transmits the signal slightly below noise level. That is to say, the signal is slightly buried in noise.

Extending this treatment to the other code lengths gives a SNR of -8 dB, -14 dB and -21 dB for code length $N = 127, 511$ and 2047 respectively when BER is 10^{-5} . This implies that for the code length $N = 127, 511$ and 2047 respectively, the spread spectrum signal is 0.158, 0.040 and 0.008 times the value of

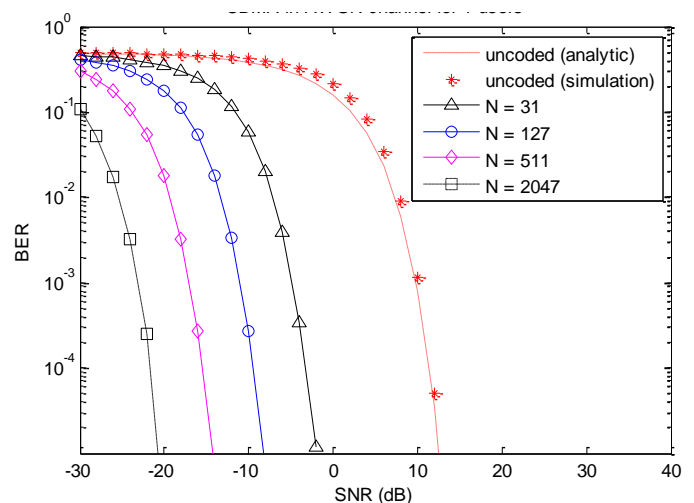


Fig. 4. CDMA performance curves showing the technique's information-hiding capability.

noise power. This clearly shows that the longer the code, the deeper the signal is buried inside noise. For the longest code ($N = 2047$), the spread spectrum signal is about 0.8% the noise power, meaning that the signal is very much below noise level, thus implying that the signal is deeply buried in noise. Therefore an eavesdropper will not perceive the communication because it is deeply buried in noise. Furthermore, whereas the original information is a narrowband signal, the encrypted version is a wideband signal having a bandwidth much larger than that of the original signal. These make it difficult for a casual observer to detect or jam the signal.

VI. CDMA IN MULTIPLE-ACCESS APPLICATIONS

Though developed originally for the military, CDMA has become an important worldwide technique in wireless communication because of certain properties that makes it attractive for commercial and civilian applications, These properties include: multiple-access capability, enhanced spectral efficiency, frequency diversity, unity cluster size and simplified frequency planning [1, 2, 12, 13, 19, 20]. Statistics [21] show that the number of CDMA subscribers grew from about 7.8 millions in 1997, to about 577 millions in 2010. Viterbi [4] indicates that as at 2002, over one hundred million consumers use devices that employ CDMA technology to provide wireless personal communication or position-location or both. CDMA is the mode of communication in the global positioning system (GPS) [22].

The use of CDMA in mobile telephony and other multiple-access applications is based on properties of spreading sequences. Among other things, orthogonality of spreading sequences is central to the system performance. Members of a set of functions $f(x)$ on a closed interval $[a, b]$ are said to be orthogonal if;

$$\langle f_i, f_j \rangle = \int_a^b f_i(x)f_j(x)dx = \delta_{ij}, \quad (5)$$

$$\text{where } \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Therefore for an orthogonal set of spreading codes $C(t)$ over a period $[0, T]$,

$$\langle C_i, C_j \rangle = \int_0^T C_i(t)C_j(t)dt = \delta_{ij} \quad (6)$$

Using this relationship reduces equation (4) to:

$$r_1(t) = b_1(t) + n(t) \sum_{i=1}^N [c_1(t - iT_c)] \quad (7)$$

This equation shows for an orthogonal set of codes, multiplying a received signal with the spreading code for a particular user isolates the user's signal from all others. All other signals are suppressed. This equation also shows at the receiving end, the noise term becomes spread out by the user

spreading code. By implication, the average noise power density becomes reduced by a factor of the process gain. For example, if the process gain is 1000, the average noise power density becomes reduced by this factor. Thus the noise power is suppressed. That is, the same signal spreading that enhances the desired signal simultaneously suppresses multi-user interference and channel noise.

VII. INTERLEAVE-DIVISION MULTIPLE-ACCESS

We shall now be considering interleave division multiple access (IDMA). IDMA is a relatively recent technique that was first proposed around the turn of the 21st century [23-26] as an alternative to CDMA. As the name implies, IDMA involves the use of interleavers as user-separating element. In IDMA, users are distinguished by user-specific interleavers instead of spreading codes used in CDMA.

In IDMA, signal spreading is avoided. This results in certain benefits which include avoidance of computationally intensive matrix multiplications and matrix inversion, and low-cost iterative multi-user detection [26-30]. However, these attractive features of IDMA are not without a price.

VIII. CDMA VERSUS IDMA IN INFORMATION SECURITY

IDMA is generally believed to be a promising candidate for future wireless technology. In literature, IDMA is usually presented as a better alternative to CDMA. However, it is useful to note that although IDMA has important benefits, it has its drawbacks. The limitations of IDMA are usually ignored in literature.

As stated earlier, CDMA involves signal spreading. The signal spreading requires matrix multiplication and inversion, and these are computationally intensive processes. In IDMA, signal spreading and hence matrix multiplication and inversion are avoided. This has important advantages because it minimises transmission bandwidth and computational requirements [26-30]. Because of these important benefits, IDMA is usually presented in literature as a better alternative to CDMA. For the same reasons, CDMA is also sometimes considered as being outdated and irrelevant to future wireless communication. A careful consideration shows that this is not the case. This is briefly explained as follows.

Signal spreading in CDMA has important advantages, some of which has been highlighted previously in this paper. Signal spreading produces low-level signals, spread out over a wideband. This makes it possible for CDMA systems to co-exist over the same bandwidth alongside with other transmission technologies like the frequency-division multiple-access (FDMA) whose energy is concentrated over a narrowband. This is important because it is a potential means of maximizing the use of the scarce electromagnetic spectrum. In contrast, IDMA does not possess this important benefit simply because of the absence of signal spreading in IDMA.

Apart from this, signal spreading is the secret behind CDMA's capability for covert transmission. This is important from information security's point of view. Signal spreading

results in very weak, low-level, noise-like signals which are difficult to detect. These characteristics can be used for keeping a spread-spectrum signal protected to maintain privacy of transmitted information. Furthermore, because CDMA signals are spread out over a wide frequency band, they are difficult to jam: jamming them requires excessive signal energy. IDMA systems lack these important benefits due to the avoidance of signal spreading in IDMA. Although IDMA has important potential benefits, the benefits are not without a price.

IX. CONCLUSION

Starting from basic principles, this work highlighted the inherent properties of CDMA that enables its use in signal encryption and information security.

This research work also considered the relevance of CDMA and IDMA techniques to information security and future wireless systems. While IDMA have certain desirable properties, it lacks the security features that are inherent in CDMA. Although IDMA has some potential benefits, it is not likely to replace CDMA in future communication systems.

REFERENCES

- [1] D. L. Nicholson, *Spread spectrum signal design. LPE and AJ systems*. Rockville, MD, USA: Computer Science Press, 1988.
- [2] A. J. Viterbi, *CDMA: principles of spread spectrum communication*: Addison-Wesley Pub. Co., 1995.
- [3] A. Viterbi, "Spread spectrum communications--Myths and realities," *IEEE Communications Magazine*, vol. 17, pp. 11-18, 1979.
- [4] A. J. Viterbi, "Spread spectrum communications: myths and realities," *IEEE Communications Magazine*, vol. 40, pp. 34-41, 2002.
- [5] R. C. Dixon, *Spread Spectrum Techniques*. Canada: John Wiley & Sons Ltd, 1976.
- [6] S. H. Mneney, "Wireless CDMA for rural application," in *AFRICON*, Stellenbosch, South Africa, 1996, pp. 408-13.
- [7] M. B. Mollah and M. R. Islam, "Comparative analysis of Gold Codes with PN codes using correlation property in CDMA technology," in *International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2012, pp. 1-6.
- [8] P. Samundiswary and P. Viswa Kalyan, "Performance Analysis of WCDMA using Different Spreading Codes," *International Journal of Computer Applications*, vol. 38, pp. 8-11, 2012.
- [9] G. Suchitra and M. L. Valarmathi, "Performance of Concatenated Complete Complementary code in CDMA systems," in *First UK-India International Workshop on Cognitive Wireless Systems (UKIWCWS)*, 2009, pp. 1-5.
- [10] A. Ziani and A. Medouri, "Analysis of different Pseudo-Random and orthogonal spreading sequences in DS-SS-CDMA," in *Multimedia Computing and Systems (ICMCS)*, Tangier, 2012, pp. 558-564.
- [11] D. Muirhead and M. A. Imran, "Alamouti Transmit Diversity for Energy Efficient Femtocells," in *73rd IEEE Vehicular Technology Conference (VTC Spring)*, Piscataway, NJ, USA, 2011, p. 5.
- [12] R. C. Dixon, *Spread spectrum systems*, 2nd ed. Chichester, Sussex, UK: Wiley, 1984.
- [13] R. Prasad, *CDMA for Wireless Personal Communications* Artech House, 1996.
- [14] K. Jyostna, F. Fadhil, N. M. Gouri, and B. N. Bhandari, "Performance analysis of SC MIMO-CDMA system using STBC codes," in *11th International Conference on Wireless and Optical Communications Networks (WOCN)*, 2014, pp. 1-5.
- [15] V. N. Mohammed, A. Kabra, P. S. Mallick, and L. Nithyanandan, "Multi access interference reduction in STBC MC-CDMA using binary orthogonal complementary sequence," in *International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, Ngercoil, India, 2015, pp. 1-4.
- [16] C. Bui Quang, Q. Zhang Tian, and J. Wu Wang, "A simplified advantage ACE and PSO algorithm for PAR reduction in STBC MC-CDMA systems," in *12th International Conference on Signal Processing (ICSP)*, 2014, pp. 1637-1642.
- [17] O. B. Wojuola and S. H. Mneney, "Multiple-access interference of Gold codes in a DS-SS-CDMA system," *SAIEE African Research Journal*, vol. 106, pp. 4-10, 2015.
- [18] O. B. Wojuola and S. H. Mneney, "Performance of even- and odd-degree Gold codes in a multi-user spread-spectrum system," in *4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE)*, Alborg, Denmark, 2014, pp. 1-5.
- [19] N. Yee and J. P. Linnartz, "BER of multi-carrier CDMA in an indoor Rician fading channel," in *Proceedings of the 27th Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, USA, 1993, pp. 426-430.
- [20] N. Yee, J. P. Linnartz, and G. Fettweis, "Multi-carrier CDMA in indoor wireless radio networks," *IEICE Transactions on Communications*, vol. E77-B, pp. 900-904, 1994.
- [21] The CDMA Development Group, "4Q 2010 CDMA Subscribers," Report Dec 2010.
- [22] Los Angeles Air Force Base (2011, 27 October). *Fact Sheet: Pseudorandom Noise (PRN) Code Assignments*, Available on <http://www.losangeles.af.mil/library/factsheets/factsheet.asp?id=8618>
- [23] W. K. Leung, L. Lihai, and P. Li, "Interleaving-based multiple access and iterative chip-by-chip multiuser detection," *IEICE transactions on communications*, vol. 86, pp. 3634-3637, 2003.
- [24] L. Ping, "Interleave-division multiple access and chip-by-chip iterative multi-user detection," *IEEE Communications Magazine*, vol. 43, pp. S19-S23, 2005.
- [25] P. A. Hoeher and H. Schoeneich, "Interleave-division multiple access from a multiuser theory point of view," in *4th International Symposium on Turbo Codes & Related Topics; 6th International ITG-Conference on Source and Channel Coding (TURBO-CODING)*, Munich, Germany, 2006, pp. 1-5.
- [26] P. Li, L. Lihai, W. Keying, and W. K. Leung, "Interleave division multiple-access," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 938-947, 2006.
- [27] M. K. Shukla, A. Gupta, and R. Bhatia, "A Survey on Various Interleavers in Iterative IDMA Communication System," in *International Conference on Special Functions and their Applications in Science and Engineering*, 2011.
- [28] P. Li, G. Qinghua, and T. Jun, "The OFDM-IDMA approach to wireless communication systems," *IEEE Wireless Communications*, vol. 14, pp. 18-24, 2007.
- [29] R. Gupta, B. Kanaujia, R. Chauhan, and M. Shukla, "Prime number based interleaver for multiuser iterative IDMA systems," in *International Conference on Computational Intelligence and Communication Networks (CICN)*, Bhopal, India, 2010, pp. 603-607.
- [30] K. Kusume, G. Bauch, and W. Utschick, "IDMA vs. CDMA: Analysis and Comparison of Two Multiple Access Schemes," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 78-87, 2012.

Effect of Varying Node Mobility in the Analysis of Black Hole Attack on MANET Reactive Routing Protocols

Lineo Mejaele*^ψ and Elisha Oketch Ochola*

*School of Computing, University of South Africa, Pretoria, South Africa

^ψMathematics and Computer Science Department, National University of Lesotho, Roma, Lesotho

Email: 48082236@mylife.unisa.ac.za, Ocholeo@unisa.ac.za

Abstract Mobile Ad-hoc Networks (MANETs) features such as open medium, dynamic topology, lack of centralised management and lack of infrastructure expose them to a number of security attacks. Black hole attack is one type of attack that is more common in MANET reactive routing protocols such as Ad-hoc On-demand Distance Vector (AODV) and Dynamic Source Routing (DSR). Black hole attack takes advantage of route discovery process in reactive routing protocols. In this type of attack, a malicious node misleads other nodes in the network by pretending to have the shortest and updated route to a target node whose packets it wants to interrupt. It then redirects all packets destined to a target node to itself and discards them instead of forwarding. This paper analyses the performance of AODV and DSR when attacked by black hole, by varying the mobility of the nodes in the network. The analysis is carried out by simulating scenarios of AODV based MANET and DSR based MANET using Network Simulator 2 (NS-2) and introducing the black hole attack in each of the scenarios. The different scenarios are generated by changing the mobility of the nodes. The performance metrics that are used to do the analysis are throughput, packet delivery ratio and end-to-end delay. The simulation results show that the performance of both AODV and DSR degrades in the presence of black hole attack. Throughput and packet delivery ratio decrease when the network is attacked by black hole because the malicious node absorbs or discards some of the packets. End-to-end delay is also reduced in the presence of a black hole attack because a malicious node pretends to have a valid route to destination without checking the routing table, and therefore shortens the route discovery process. The results also show that throughput decreases slightly when mobility of the nodes is increased in the network. The increase in the speed of the nodes decreases both end-to-end delay and packet delivery ratio.

Keywords-MANET; Reactive Routing Protocols; Black Hole Attack; Mobility

I. INTRODUCTION

The success of any kind of a network is intensely determined by the confidence people have in its security, it is therefore very crucial for both wired and wireless networks to be secured so as to offer protected communication [1]. Mobile Ad-hoc Network (MANET) is a group of mobile devices that can spontaneously interconnect and share resources via wireless communication channels, with no fixed network infrastructure or central management. MANETs can be assembled quickly with little cost because they do not require central monitoring or fixed network infrastructure. Mobile nodes in MANET do not necessarily have to be of the same

type. They can be PDAs, laptops, mobile phones, routers and printers, as illustrated by Fig.1. The nodes are equipped with antennas which operate as wireless transmitters and receivers. The antennas may be omnidirectional, highly directional, or a combination. The mobile nodes are resource constraint in terms of bandwidth and battery power [2, 3].

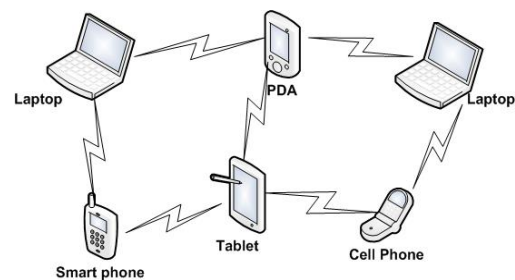


Figure 1. Mobile Ad-hoc Network

MANET is suitable to provide communications in many applications, particularly in cases where it is not possible to setup a network infrastructure. For instance, in a military operation, where there may be geographical barriers between participants, MANET can be setup to facilitate communication. Also because it is easy to set up, it may be of assistance to replace the damaged network infrastructure in disaster recovery operations where temporary network infrastructure is immediately needed [4, 5].

The features of MANETs expose them to many security attacks compared to other traditional networks. The high mobility and dynamic topology of MANETs make routing to be very challenging, that is why early research on MANET mostly concentrated on developing routing mechanisms that are efficient for a dynamic and resource constraint MANET. The security of protocols was given less attention when MANET routing protocols were defined. Black hole attack aims to disrupt the routing process of MANETs [1].

This paper aims to analyse the performance of MANET reactive routing protocols when attacked by the black hole. The two reactive routing protocols that are compared in the analysis are Ad-hoc On-demand Distance Vector (AODV) and Dynamic Source Routing (DSR). The mobility of the nodes in the network is varied during the analysis to determine the impact that mobility has on MANETs performance and to discover the protocol that is more preferable in a high mobility network. The effect of black hole attack is tested on reactive

routing protocols because black hole attack targets route discovery process and can easily attack reactive protocols since they discover the routes frequently.

The rest of the paper is structured as follows: Section II discusses the vulnerabilities of MANETs that expose them to attacks. Section III describes routing in MANETs and discusses the different categories of routing protocols, focusing more on reactive routing protocols. Section IV explains black hole attack, and describes some of the solutions that have been suggested to lessen the impact of the attack. Section V gives the simulation structure used to perform the analysis, presents the results obtained from the simulations and gives the analysis of the results. Section VI concludes the paper.

II. VULNERABILITIES OF MANETS

It is quite challenging to maintain security in MANETs because they have far more vulnerabilities than wired networks [6]. Any weakness in security system is vulnerability. Some MANETs vulnerabilities are presented as follows:

A. Lack of Secure Boundaries

The nodes in MANET are at liberty to move inside the network, join and leave the network any time. This makes it challenging to establish a security wall as compared to traditional wired networks that have a clear line of defense. In order to attack wired networks the adversaries must physically enter into the network medium, pass through firewalls and gateways before they have access to practice malicious behaviour to the target nodes in the network. However, in MANET the adversary can communicate with nodes within its transmission range, and become part of the network without any physical access to the network. The absence of secure boundaries causes MANET to be attacked at any time by any malicious node that is within the transmission range of any node in the network [7].

B. Lack of Centralised Management Facility

There is no central equipment such as a server for monitoring the nodes in the network and this increases the vulnerability problems of MANETs. Firstly, it becomes very difficult to detect the attacks in the absence of central control because the traffic in an ad-hoc network is very dynamic [8]. Secondly, lack of centralised management delays the nodes trust management. It becomes difficult to prior classify the nodes as trustworthy or untrustworthy because the security of the nodes cannot be presumed. Consequently, the nodes cannot be distinguished as trusted or non-trusted. Thirdly, lack of centralised authority can sometimes lead to decentralised decision-making. In MANETs, important algorithms depend on all nodes participating cooperatively, so the attacker can take advantage of this vulnerability and execute attacks that can ensure that the nodes are not cooperative [9].

C. Threats from Compromised Nodes in the Network

Each mobile node operates independently, which means it is free to join or leave the network at any time. It therefore becomes difficult for the nodes to set rules and strategies that can prevent malicious behaviour of other nodes in the mobile

network. Also, due to freedom of movement of the nodes, a compromised node can target different nodes in the network. Hence, it becomes quite challenging to identify malicious actions of a compromised node in the network, particularly in a huge network. As a result, internal attacks from nodes that have been compromised are more severe than external attacks because they are not easily identified due to the fact that a compromised node operated normally before it could be compromised [7].

D. Restricted Power Supply

Mobile devices in MANET get energy from batteries or other exhaustible means, so their energy is limited. This energy restriction can cause denial of service by the attacker; since the attacker is aware of the battery restriction, it can endlessly forward packets to the target node or make the target node to be involved in some time consuming activities. This will cause battery power to be exhausted and the target node will not operate anymore. Again, the limited power supply may cause a node in MANET to behave selfishly by not participating cooperatively in the network activities as a way to save its limited battery. This becomes a problem particularly when it is essential for the node to cooperate with other nodes [10].

III. ROUTING IN MANETS

The topology of MANETs keeps changing rapidly due to free movement of nodes joining and leaving the network any time. Routing is important in order to discover the recent topology so that an updated route to a certain node can be established and a message relayed to the correct destination [3, 11]. The traditional routing protocols such as distance vector and link state protocols that have been structured for hard wired networks cannot be directly applied to MANETs. This is because of mobility and dynamic topology, which are the fundamental characteristics of MANETs [12]. In order to overcome routing challenges in MANETs and attain effective routing, a number of routing protocols are defined specifically for MANETs. These protocols can be categorised into proactive, reactive and hybrid protocols based on the way paths are established and maintained by the nodes [13]. The hierarchy of the protocols is shown in Fig. 2. The reactive routing protocols discussed in this paper are shown in red in Fig. 2.

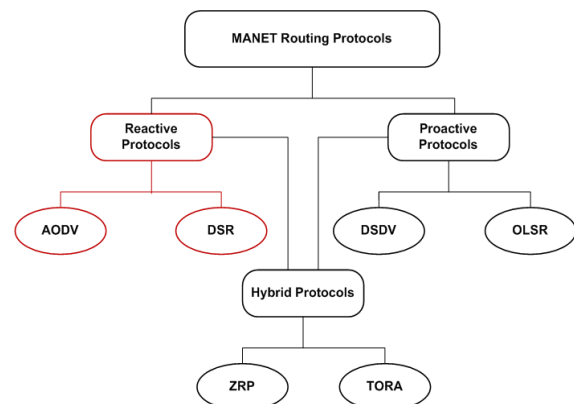


Figure 2. MANET Protocols Hierarchy

A. Proactive Protocols

These are table-driven routing protocols that try to keep a record of fresh and updated network routes. All the nodes in the network have a table to store the routing information [8]. The nodes exchange topology information so that they can all have the same view of the network. The exchanged information helps to reflect any changes in the topology. Whenever a node needs to send messages, it just searches the routing table for the path to the destination. The sending of the message is not delayed by the remote route discovery [11]. Maintaining an up-to-date topology in the routing tables causes a high control overhead.

B. Hybrid Protocols

Hybrid protocols are a mixture of proactive and reactive protocols. Their design merges the benefits of both proactive and reactive protocols to yield better results [14]. The hierarchical network model is used to structure majority of hybrid routing protocols. Firstly, all the routing information that is unknown is acquired by using proactive routing. Then reactive routing mechanisms are used to maintain the routing information when the topology changes [15].

C. Reactive Protocols

Reactive protocols are on demand routing protocols. As the name suggests, the routes to destination nodes are established only when the nodes must send data to destination whose route is unknown. This implies that the source node initiates the searching of routing paths only when needed. When a node wants to send data to a destination node, it starts a route discovery process within the network. Comparative to proactive protocols, the control overhead in reactive protocols is reduced; however the route searching process that occurs before data packets can be forwarded may cause source node to suffer long delays [16]. Reactive protocols use route discovery and route maintenance processes as explained below:

Route Discovery: Route discovery process is a cycle that involves a broadcast route request and a unicast reply that consists of paths that have been discovered [17]. All the nodes in the network keep a record in a routing table. This record consists of information about neighbouring nodes that can forward the packets so that they reach the destination. When a source node wants to send data packets to a destination node, and there is no routing information regarding the destination node in the routing table, the source node initiates a route discovery process [18]. In discovering the route, a source node broadcasts route request (RREQ) packet [19].

When the RREQ packet reaches any node in the network, the node compares the destination IP address to its IP address to determine whether it is the destination node. The node sends back a route reply (RREP) packet if it is the destination, but if it is not, it searches for a route to the destination in its routing table. If there is no route, it broadcasts the RREQ packet to nearby nodes. If there is a route to destination in its routing table, a node compares a RREQ packet sequence number with the destination sequence number in the table to find if the route is updated. The route in the routing table is considered fresh and updated if the destination sequence number in the table is

higher than the sequence number attached to the RREQ packet. The intermediate node with an updated route uses the opposite route to send a unicast RREP packet to the source node, and once the source node has received a RREP packet, it begins to send messages through this route. If the route in the table is not fresh enough, the node further sends the RREQ packet to its neighbours [18, 20].

Route Maintenance: During operation, any node that notices a damaged link sends a route error (RERR) packet. A RERR packet is relayed to every node that utilises the affected link for their communication to other nodes [20].

IV. BLACK HOLE ATTACK

The proper functioning of MANETs depends on the mutual agreement and understanding between the nodes in the network; however some nodes may become malicious and misbehave. Black hole attack is one of the harmful attacks caused by a malicious node that misbehaves in a network [21]. A malicious node exploits the process of discovering routes in reactive routing protocols. When a source node broadcasts a route request, a malicious node misleads other nodes by claiming to have the best path to the destination. The best path is determined by the shortness and freshness of the route. It achieves this by unicasting false route replies, directing data packets to be routed through it and just discarding them instead of forwarding [22]. A malicious node can work independently to launch the attack, and this is referred to as single black hole attack, or malicious nodes can work as a group and the attack is referred to as cooperative black hole attack [15].

A. Black Hole Attack Mitigations

There has been various research carried out to discover and mitigate the black hole attack in MANETs. The techniques were tested on AODV-based MANET. Some of the mitigation techniques are discussed below:

1) Detection, Prevention and Reactive AODV(DPRAODV)

In [23], DPRAODV is proposed. In this scheme, AODV protocol is modified to have a new control packet called ALARM and a threshold value. A threshold value is the average of the difference of destination sequence number in the routing table and sequence number in the RREP packet. In the usual operation of AODV, the node that gets a RREP packet checks the value of sequence number in its routing table. The sequence number of RREP packet has to be higher than the sequence number value in the routing table in order for RREP to be accepted. In DPRAODV, there is an extra threshold value that is matched to RREP sequence number, and if RREP sequence number is greater than the threshold value, then the sender is considered malicious and added to the black list. The neighbouring nodes are notified using an ALARM packet so that the RREP packet from the malicious node is not processed and gets blocked. Automatically, the threshold value gets updated using the data collected in the time interval. This updating of the threshold value helps to detect and stop black hole attacks. The ALARM packet contains the black list that has a malicious node. This list assists the neighbouring nodes not accept any RREP packet sent by a malicious node. Any node that gets a RREP packet looks into the black list and if the

reply comes from a node that has been blacklisted, it is ignored and further replies from that node will be discarded. Thus the ALARM packet isolates a malicious node from the network.

2) *Intrusion Detection System AODV (IDSAODV)*

IDSAODV is proposed in [24] in order to decrease the impact of black hole. This is achieved by altering the way normal AODV updates the routing process. The routing update process is modified by adding a procedure to disregard the route that is established first. The tactic applied in this method is that the network that is attacked has many RREP packets from various paths, so is assumed that the first RREP packet is generated by a malicious node. The assumption is based on the fact that a black hole node just sends a fake RREP packet, without searching through the routing table. Therefore, to avoid updating routing table with wrong route entry, the first RREP is ignored. This method improves packet delivery but it has limitations that; the first RREP can be received from an intermediate node that has an updated route to the destination node, or if RREP message from a malicious node can arrive second at the source node, the method is not able to detect the attack.

3) *Enhanced AODV (EAODV)*

In [25], the authors proposed EAODV. Similar to IDSAODV, EAODV allows numerous RREPs from various paths to lighten the effect of black hole attack. This method makes an assumption that eventually the actual destination node will unicast a RREP packet, so the source node overlooks all previous RREP entries, including the ones from malicious node and takes the latest RREP packet. The source node keeps on updating its routing table with RREPs being received until a RREP from the actual destination is received. Then all RREPs get analysed and suspicious nodes are discovered and isolated from the network. The limitation to this method is that it adds two processes that increase delay and exhaust energy of the nodes.

4) *Secure AODV (SAODV)*

The authors in [26] proposed a secure routing protocol, SAODV that addresses black hole attack in AODV. The difference between AODV and SAODV is that in SAODV, there are random numbers that are used to verify the destination node. An extra verification packet is introduced in the route discovery process. After getting a RREP packet, the source node stores it in the routing table, then sends an instant verification packet using reverse route of received RREP. The verification packet consists of a random number created by the source node. When two or more verification packets from the source node are received at the destination node, coming from different routes, the destination node stores them in its routing table and checks whether the contents contain the same random numbers. If the verification packets contain same random numbers along different paths, the verification confirm packet is sent by the destination node to the source node. The confirm packet consists of random number generated by destination node. If the verification confirm packet contains different random numbers, the source node will wait until at least two or more verification confirm packets contain same random numbers. When two or more verification confirm packets with

the same random numbers are received by the source node, it will use the shortest route to send data to the destination node. The security mechanism in this protocol is that malicious node pretending to be the destination node will not send the correct verification confirm packet to the source node.

5) *Trust-based Approach*

The authors in [27] suggested a trust based approach to mitigate the black hole attack. In this approach, every node keeps a trust value on all its neighbours. The trust value is computed as the proportion of discarded packets to forwarded packets. Each node ensures that the neighbouring node forwards the packets sent to it, unless the packet is destined to the neighboring node. As a way to ensure that the packets are forwarded, each node implements a caching mechanism by storing the packet being forwarded to the neighbouring node in its cache, and then promiscuously monitoring the neighbouring node to check whether it forwards the packet. If the neighbouring node forwards the packet, it compares it with the packet stored in its cache, and the node assumes the packet has been forwarded if they match. Else, after a set time the node assumes the packet has been discarded by its neighbour and the neighbouring node is suspected to be malicious. All the nodes in the network will get to know the behaviour of the neighbouring nodes, and can therefore periodically assign trust values that represent the trustworthiness of the neighbouring nodes. All RREP packets from a node that has been recognised as malicious are ignored, and the routes will only be selected through trusted nodes.

6) *Solution Using Packet Sequence Number*

In the regular operation of AODV, the source node compares the value of RREP sequence number with sequence number in its routing table. The RREP packet is accepted only if its sequence number has a value higher than the sequence number in source's routing table. A solution that requires the use of two additional small tables in every node is proposed in [5]. The sequence number for the last packet sent by a node is to be recorded in one table and another table should record the sequence number for last packet received from every node. Every time a packet is received or sent by a node, the tables are updated. During route discovery process, the source node broadcasts a RREQ packet to nearby nodes. The destination node or the intermediate node that has a fresh route to the destination will reply to the sender with RREP packet that contains the last packet sequence number received from the source node. The source node will therefore verify that the sequence number of RREP received matches the record it has in the table, and if it does not, the RREP packet is suspected to be from a malicious node. Since the sequence number is already part of communication in the base protocol, this solution does not increase overhead to the transmission channel. It makes it easy to recognise a suspicious reply.

V. SIMULATIONS AND RESULTS

The results are obtained from simulations implemented on Network Simulator 2 (NS-2) and are presented using graphs. NS-2 is distributed freely and is an open source environment which allows the creation of new protocols, and modification

of existing ones, so it is possible to introduce a black hole attack in NS-2 by modifying its source code [28]. A typical simulation with NS-2 consists of creating a scenario file that defines the position and movement patterns of the nodes, and a communication file that defines connection and traffic in the network. Each run of simulation produces a detailed trace file that shows events (such as number of packets delivered successfully) happening during simulation. Fig. 3 illustrates NS-2 simulation process.

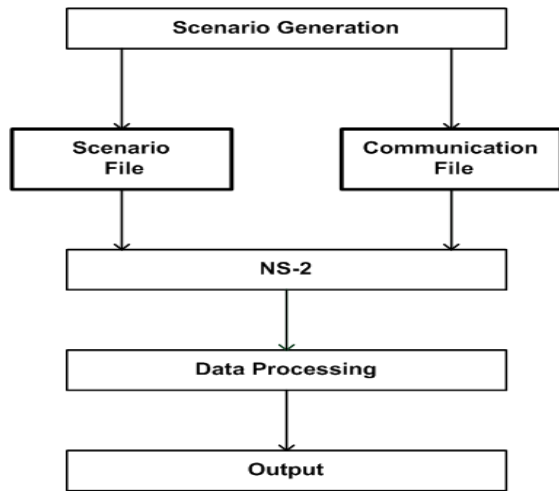


Figure 3. NS-2 Simulation Process

The simulation parameters used in this study are shown in Table I.

TABLE I. SIMULATION PARAMETERS

Parameter	Values
Simulator	NS-2.35
Mobility Model	Random Waypoint
Simulation Time	500 seconds
Terrain Area	670m x 670m
Number of nodes	20
Number of malicious nodes	1
Traffic Type	CBR (UDP)
Packet Size	512 bytes
Routing Protocols	AODV, DSR
Transmission Rate	4 packets/sec
Maximum Speed	20 ó 80 m/s
Pause Time	0 seconds
Transmission Range	250m

The performance metrics used are throughput, packet delivery ratio and end-to-end delay. In order to analyse the effect of mobility, the speed at which the nodes move was varied from 20m/s to 80m/s to create different scenarios. The total number of nodes and maximum number of connections were kept constant at 20 and 10 respectively. The results show the effect of mobility for both AODV and DSR protocols when the network is under a black hole attack and when there is no black hole attack.

A. Throughput

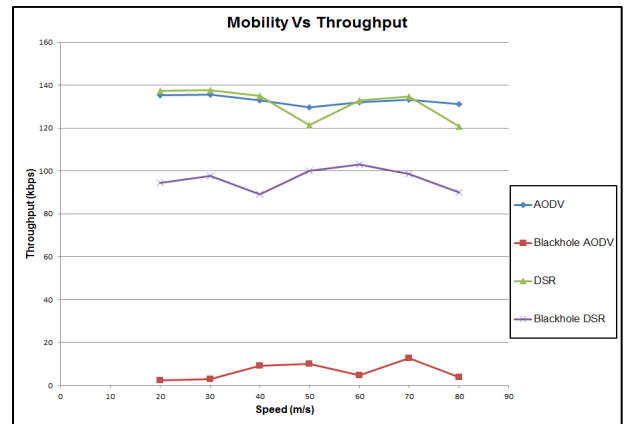


Figure 4. Throughput AODV vs. DSR

The simulation results of Fig. 4 show that increasing the speed of the nodes in the network does not bring significant change in throughput. For both protocols, throughput decreases slightly. This is caused by the rapid change of positions of the nodes, which may cause the path to the destination to change while some packets have been transmitted from the source node using the old route. Therefore the transmitted packets get lost on the way. Throughput of the network under black hole attack decreases because the malicious node discards some of the packets. AODV's throughput drops drastically compared to DSR's throughput.

B. Packet Delivery Ratio

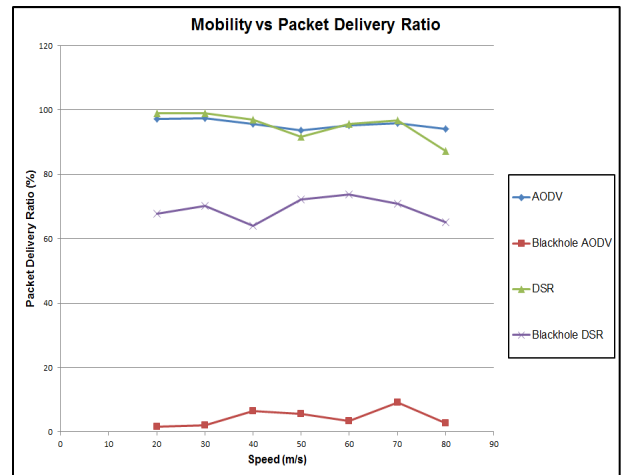


Figure 5. Packet Delivery Ratio AODV vs. DSR

When the mobility of the nodes is increased packet delivery ratio decreases a little. This is because some of the packets may get lost along the way to the destination when the path from the source node to the destination node changes due to rapid change of intermediate nodes' positions. The packet delivery ratio of AODV is very low compared to that of DSR when the black hole attack has been launched against the network.

C. End-to-end Delay

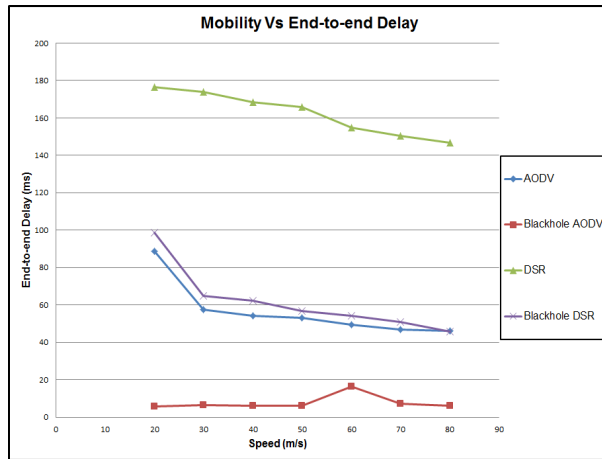


Figure 6. End-to-end Delay AODV vs. DSR

Fig. 6 shows that end-to-end delay decreases with increase in speed because the nodes' movement gets more frequent and the routing updates are regularly exchanged. When there is a black hole attack, end-to-end delay gets even lower because the malicious node pretends to have a valid route to the destination without checking in the routing table, so the route discovery process takes a shorter time.

VI. CONCLUSION

This paper has analysed the black hole attack on MANET reactive routing protocols (AODV and DSR). The analysis is done by varying the mobility of the nodes to determine the effect that mobility has on the way the protocols perform. The results obtained from simulations indicate that the performance of DSR degrades more than the performance of AODV when the speed of the nodes is increased, so it can be concluded that AODV is more preferred in a high mobility network. Furthermore, the results show that the black hole attack degrades the performance of both AODV-based MANET and DSR-based MANET, but the impact is more severe on AODV than DSR. It can therefore be concluded that DSR is more preferred in a network that is frequently attacked by the black hole.

ACKNOWLEDGMENT

We appreciate everyone who supported and encouraged us throughout this study. Most importantly, we thank the University of South Africa and the National university of Lesotho for providing necessary resources that supported the research.

REFERENCES

[1] C. Yu, T. K. Wu, R. Cheng and S. Chang, "A distributed and cooperative black hole node detection and elimination mechanism for ad hoc networks," *Emerging Technologies in Knowledge Discovery and Data Mining*, pp. 538-549, 2007.

[2] K. Osathanunkul and N. Zhang, "A countermeasure to black hole attacks in mobile ad hoc networks," in *Networking, Sensing and Control (ICNSC)*, 2011 IEEE International Conference On, 2011, pp. 508-513.

[3] B. Wu, J. Chen, J. Wu and M. Cardei, "A survey of attacks and countermeasures in mobile ad hoc networks," in *Wireless Network Security* Springer, 2007, pp. 103-135.

[4] C. Rajabhushanam and A. Kathirvel, "Survey of wireless MANET application in battlefield operations," (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. 2, pp. 50-58, 2011.

[5] R. Mishra, S. Sharma and R. Agrawal, "Vulnerabilities and security for ad-hoc networks," in *Networking and Information Technology (ICNIT)*, 2010 International Conference On, 2010, pp. 192-196.

[6] N. Sharma and A. Sharma, "The black-hole node attack in MANET," in *Advanced Computing & Communication Technologies (ACCT)*, 2012 Second International Conference On, 2012, pp. 546-550.

[7] Y. Rajesh and S. Anil, "Secure AODV protocol to mitigate black hole attack in Mobile Ad hoc Networks," *ICCNT 2012 International Conference On*, 2012, pp. 1-4.

[8] I. Zaiba, "Security issues, challenges and solution in MANET," vol. 2, pp. 108-109-112, 2011.

[9] P. Goyal, V. Parmar and R. Rishi, "Manet: vulnerabilities, challenges, attacks, application," *IJCEM International Journal of Computational Engineering & Management*, vol. 11, pp. 32-37, 2011.

[10] U. K. Singh, S. S. KailashPhuleria and D. Goswami, "An analysis of Security Attacks found in Mobile Ad-hoc Network," *International Journal of Scientific & Engineering Research*, vol. 5, pp. 43-46, 2014.

[11] W. Li and A. Joshi, "Security issues in mobile ad hoc networks-a survey," *Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County*, pp. 1-23, 2008.

[12] B. Kannhavong, H. Nakayama, Y. Nemoto, N. Kato and A. Jamalipour, "A survey of routing attacks in mobile ad hoc networks," *Wireless Communications, IEEE*, vol. 14, pp. 85-91, 2007.

[13] N. Sharma and A. Sharma, "The black-hole node attack in MANET," in *Advanced Computing & Communication Technologies (ACCT)*, 2012 Second International Conference On, 2012, pp. 546-550.

[14] V. C. Giruka and M. Singhal, "Secure Routing in Wireless Ad-Hoc Networks," in *Signals and Communication Technology*, pp. 137-158, 2007.

[15] P. K. Singh and G. Sharma, "An efficient prevention of black hole problem in AODV routing protocol in MANET," in *Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2012 IEEE 11th International Conference On, 2012, pp. 902-906.

[16] F. Tseng, L. Chou and H. Chao, "A survey of black hole attacks in wireless mobile ad hoc networks," *Human-Centric Computing and Information Sciences*, vol. 1, pp. 1-16, 2011.

[17] A. N. Thakare and M. Joshi, "Performance Analysis of AODV & DSR Routing Protocol in Mobile Ad hoc Networks," *IJCA Special Issue on Mobile Adhoc Networks, MANETS*, pp. 211-218, 2010.

[18] R. Agrawal, R. Tripathi and S. Tiwari, "Performance evaluation and comparison of AODV and DSR under adversarial environment," in *Computational Intelligence and Communication Networks (CICN)*, 2011 International Conference On, 2011, pp. 596-600.

[19] R. H. Jhaveri, A. D. Patel, J. D. Parmar and B. I. Shah, "MANET routing protocols and wormhole attack against AODV," *International Journal of Computer Science and Network Security*, vol. 10, pp. 12-18, 2010.

[20] N. Purohit, R. Sinha and K. Maurya, "Simulation study of black hole and jellyfish attack on MANET using NS3," in *Engineering (NUiCONE)*, 2011 Nirma University International Conference On, 2011, pp. 1-5.

[21] M. Medadian, A. Mebadi and E. Shahri, "Combat with black hole attack in AODV routing protocol," in *Communications (MICC)*, 2009 IEEE 9th Malaysia International Conference On, 2009, pp. 530-535.

[22] A. Vani and D. S. Rao, "Removal of black hole attack in ad hoc wireless networks to provide confidentiality security service," *Int. J. Eng. Sci.*, vol. 3, pp.2377-2384, 2011.

- [23] P. N. Raj and P. B. Swadas, "Dpraodv: A dyanamic learning system against blackhole attack in aodv based manet," *IJCSI*, vol.3, pp.54-59, 2009.
- [24] R. Suryawanshi and S. Tamhankar, "Performance Analysis and Minimization of Blackhole Attack in MANET," *IJERA*, vol.2, pp.1430-1437, July-August, 2012.
- [25] Z. Ahmad, K. A. Jalil and J. Manan, "Black hole effect mitigation method in AODV routing protocol," in *Information Assurance and Security (IAS)*, 2011 7th International Conference On, 2011, pp. 151-155.
- [26] S. Lu, L. Li, K. Lam and L. Jia, "SAODV: A MANET routing protocol that can withstand black hole attack," in *Computational Intelligence and Security*, 2009. CIS'09. International Conference On, 2009, pp. 421-425.
- [27] J. Pan and R.Jain, "A survey of network simulation tools: Current status and future development,"
Internet:[http:// www1.cse.wustl.edu/~jain/cse567-08/ftp/simtools.pdf](http://www1.cse.wustl.edu/~jain/cse567-08/ftp/simtools.pdf),
Nov. 24, 2008 [May 5, 2016].

The Pattern-richness of Graphical Passwords

Johannes S. Vorster
Rhodes University
email: JSVorster@gmail.com
Barclays Africa
email: Jo.Vorster@absa.co.za

Renier P. van Heerden
Council for Scientific
and Industrial Research
email: renier@sanren.ac.za
Nelson Mandela Metropolitan University

Barry Irwin
Rhodes University
email: B.Irwin@ru.ac.za

Abstract—Conventional (text-based) passwords have shown patterns such as variations on the username, or known passwords such as "password", "admin" or "12345". Patterns may similarly be detected in the use of Graphical passwords (GPs). The most significant such pattern – reported by many researchers – is hotspot clustering.

This paper qualitatively analyses more than 200 graphical passwords for patterns other than the classically reported hotspots. The qualitative analysis finds that a significant percentage of passwords fall into a small set of patterns; patterns that can be used to form attack models against GPs. In counter action, these patterns can also be used to educate users so that future password selection is more secure.

It is the hope that the outcome from this research will lead to improved behaviour and an enhancement in graphical password security.

Index Terms—Information security, graphical passwords, password patterns, user authentication, user study.

I. INTRODUCTION

A. Background

Historical and current research into Graphical Passwords (GPs) cover a rich topic; see [1] for a review, we present only a small overview. GPs were first explored as an alternative to text-based passwords in the early 1990s. The first patent on the topic was registered to G. Blonder in 1995 [2], based on the idea of sequentially selecting points on an image – see Figure 1 (a). In this schema the user enrolls by selecting a number of points on an image. Authentication is then done by re-selecting the same points in the same order. Obviously the user cannot select the same point down to a pixel level, so the schema must inherently have some error margin. The size of the error region effectively defines a theoretical limit on the number of different passwords per image.

The initial idea from Blonder was soon followed by a variety of schemes that avoided the initial patent

by using other mechanisms of schemes. Draw-a-Secret (DAS), abstractly proposed by Syukri et.al. [3] and later implemented by Jermyn et.al. [4], uses a blank grid canvas and records a password as a sequence of strokes. In this scheme, each stroke travels through a number of grid-elements, and these are recorded to form the password – see Figure 1 (b).

In the early 2000s a number of alternative schemes were proposed and implemented. PassFaces, proposed by Brostoff and Sasse [5], used facial images – see Figure 1 (d). In this scheme, the user enrolls by selecting a number of faces from a large database of faces. During authentication one of the enrolment images are shown together with 8 other faces in a 3x3 grid. The user must go through a number of rounds, selecting the correct face from the 9 options during each round. Déjà Vu is a similar scheme proposed by Dhamija and Perrig [6]. It uses abstract – see Figure 1 (c). However it was shown during their study that enrolment rates for abstract images took twice as long as for face-based images.

Wiedenbeck et.al. [7] proposed a scheme called Pass-Points that is similar to that of Blonder, but it makes use of photos and well-defined tolerance circles. As was pointed out, this scheme needs to define an effective area around the enrolment points to ensure successful authentication. In a study by Van Oosterchot & Thorpe [8], the effective grid must be 19x19 around the point of enrolment to minimize login failures but maximize key-space.

Tao [9] proposed a recall-based scheme based on the board game Go. Users connect points placed on the intersections of grid-lines. This schema is perhaps the grandfather of the Android pattern unlock mechanism used on smartphones.

Background DAS (BDAS) [10] puts a background image on the DAS grid, allowing users to have a cued recollection of their password. Jansen's Picture Password scheme [11] is perhaps the most usable cue-recall based

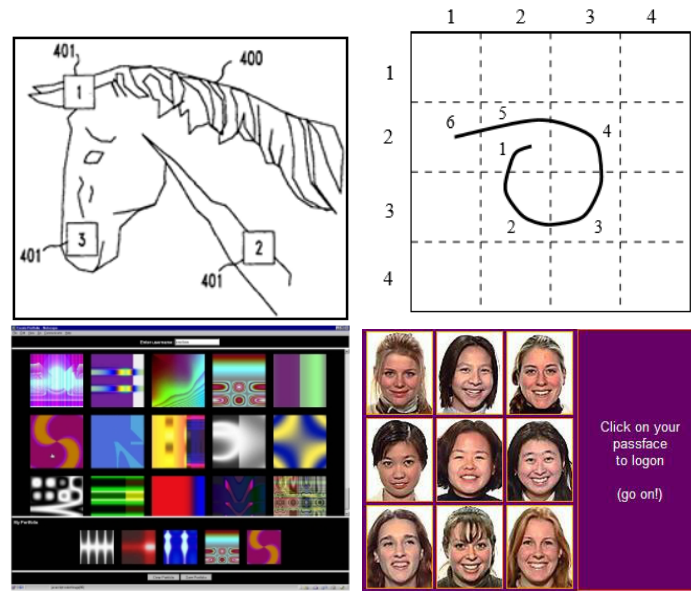


Fig. 1. Different graphical password schemes: (a: top-left) Blonder’s original patent image; (b: top-right) Draw-a-secret example; (c: bottom-left) Deja-vu; (d: bottom-right) Pass-faces.



Fig. 2. Graphical password schemes: (a: left) Pass-Go; (b: middle) Android pattern unlock; (c: right) Captcha-based GP.

scheme; using a grid overlay on an image and numbering the grid elements rather than using the actual image coordinates. This schema is similar to BDAS, but instead of drawing an image on top of the background, the user selects points on the image similar to the original Blonder scheme.

GP schemes have been implemented on mobile devices, but due to physical constraints entering such passwords is error-prone, less secure and less effective than with physical keyboards [12]. Popular GP schemes used on smartphones – see Figure 2 (b) is nothing more than a replacement of a numeric password scheme. It has been shown that the Android pattern unlock is equivalent to a 5 digit numeric password [13]. In a 2015 study of the Android pattern unlock it was found that most

passwords consist of two or three strokes and that there are directional biases (more left-to-right), but that the introduction of an over-layer image can improve this bias [14]. Another study found that 38% of passwords start in the top-left corner and that the combined percentage for starting in the bottom-center, bottom-right or right-middle, is only 8% [15].

More recently, alternatives have been proposed using more dynamic generation of images based on text-based passwords, such as the Captcha-based schema proposed by Gao, Wang and Dai [16]. The aim of such a GP scheme is to use hard Artificial Intelligence (AI) problems as a security primitive [17]. One obvious problem with such schemas is that it is strongly dependent on current technology and the entire scheme can be invalidated

by new technology.

Other authors have proposed using GPs not for direct authentication but for secondary security processes, such as password recovery. For example, Almuairfi et.al. [18] proposes the use of GPs as a substitute to the security questions used during recovery of lost passwords.

Many conventional password proposals have been mapped to GP equivalents. An interesting such case is that of "honeywords"; false passwords that are hashed and transmitted as part of normal password authentication. If such a password is used in an attempt to authenticate there is a high probability that the user's account has been compromised [19]. A similar scheme for GPs has been proposed [20].

B. Graphical Password Categories and Security

GPs can be categorized based on the mechanism that the users use to recall the correct password. There are three categories: recall-based, recognition-based and cue-based.

Recall-based schemes rely on the user remembering the password without any assisting framework. From the examples in the previous sections, DAS, Pass-Go and the Android pattern unlock are examples that fit into this category. These schemas have been extensively studied for security vulnerabilities.

Recognition-based passwords rely on the user recognizing an image from the password set from a larger set of images. Typical schemes in this group are PassFaces and Déjà Vu. One of the problems with such schemes are the number of rounds needed to enter passwords [10].

Cue-based passwords present the user with a cue, such as a background image on which a password is selected by clicking on the image. Blonder's original patent, Passpoints, BDAS and Jenson's scheme fall into this category. One of the most significant considerations when using such a scheme is that of hotspots, discussed below.

To use conventional passwords as a benchmark, studies [21] have found that conventional passwords have lengths between 6 and 13 characters with an average bit strength of 37.8 bits. Graphical passwords – for some schemes – have proven to give slightly stronger security [22]. PassPoints [7], for example, shows a significantly higher key-space size compared to conventional passwords. Using a background image for DAS – called BDAS – improves user password length [10].

One of the prominent critiques of GPs has been the threat of shoulder-surfing – an attacker observing the user during password entry. A significant number of

proposals have been generated to counter this threat [23], [24], [25], [26], [27].

C. Graphical Password Patterns

In the Biddle et.al. review of graphical passwords, it is noted that the size of the graphical password key space may be significantly smaller than the theoretical calculations due to password patterns, as is the case with conventional passwords. User-biases during password selection has been reported in numerous studies [28], [10], [29], [30], [31].

One of the first patterns that was recognized as part of graphical password selection is hotspots. In cued-recall schemes the background image plays an important role in the strength of the password; images with a low number of features tend to create password hotspots, that is, a large subset of the user population selects the same points on the image as part of their password [31]. This can be seen as the analogue of conventional passwords that often are selected from a small subset of characters. For example, in the RockYou dataset of 32 million passwords, 20.5% of passwords are number-only passwords [32].

Dirik et.al. [33] constructed an image processing algorithm that uses heuristics to attempt to identify potential hotspots in images that are then used to guess PassPoints passwords. Using the heuristics, they generate a dictionary of size 2^{32} entries and test against the PassPoints password set with a theoretical 40-bit size and report an 8% success rate in cracking the user passwords. This is a low success rate and one explanation proposed is that the the implemented heuristics did not match the patterns that humans would pick [34].

Van Oorschot & Thorpe [8] used heuristics to crack passwords from a database of click-based passwords. They identified four patterns, and found that 56% of passwords contain patterns from their 4-set of patterns. The patterns are: horizontal (15%), vertical (15%), diagonal (11%) and clock (5%); where clock is a clockwise pattern – circular clockwise or circular anti-clockwise.

II. METHODOLOGY AND STUDY EXECUTION

A. Overview and Aim

In this study, we wanted to understand what the types of patterns are that humans use when selecting GPs. The earlier studies did not involve the participants directly, that is, they never asked the participants for the method used. Therefore, we opted to use a qualitative study rather than a quantitative approach.

The aims of this study are: identify the types of patterns that users employ for selecting GPs. How do these patterns change if users are made aware of obvious security considerations for GPs, such as hotspots.

The first section of the study involved the selection of images and schemes to use. Since significant studies have already been conducted on DAS passwords, we focus on click-based passwords through various image-selection schemes and employ a Jankens’s type model [11].

The images used are presented in Figure 3. The Kitten and Hedgehog image, used by [35], [34], [36], was selected because of its low feature set. It is expected that these images should have a high hotspot incidence. The question we are interested in for these images are how the pattern changes after users are made aware of the hotspots they have selected as passwords. The Paperclip image was used by [8] and has a high number of features that should lead to few hotspots. The Company Logo image was generated using Interbrand and consists of some well-known brand icons. Such an image is similar to those used as alternatives to PassFaces and Déjà Vu. A critique often mentioned against PassFaces is that the passwords selected have patterns, such a all beautiful people or all women. The Faces image was included to investigate the password patterns in such images and to question users on why they had selected the passwords in these images.

B. Methodology

Interviews where held with 21 participants. During the interviews the five images were presented to the participants and they were asked to select passwords. Participants were also asked what the reason was for their password selection for each image. Once that was completed the participants were informed that graphical passwords are known for having hotspots or other patterns. Users were not shown any examples but told that in an image such as Kitten most users would select predictable points such as ears, eyes, nose and paws. No other images were referred to, nor were any other patterns pointed out, other than to mention that there are other patterns.

Participants were then asked if they wanted to revise their password selection and again had the opportunity to select passwords for the five images. Again participants were given the opportunity to identify the method that was used for the selection of the passwords.

The password patterns were analysed manually and classified. Password patterns were analysed and identi-

fied and also correlated with the reasons participants gave for their password patterns.

The participants were selected at random from a group of professionals that included project managers, business analysts, software developers and administrative staff.

III. RESULTS AND DISCUSSION

Analysis of the passwords during the first enrolment support previous results that report a significant number of passwords that use hotspots. We find 22% of initial enrolment passwords conform to hotspot clusters – see Fig. 4 for the Kitten hotspots. If compared to the patterns reported by [8] we find the same patterns, but with smaller percentages. For example we find only 2% of patterns conform to a vertical lines pattern, and 3% to a diagonal lines pattern – see Table I. Some of the non-hotspot patterns that show up in the first enrolment data set are shown in Figure 5.

Through inspection and the participant interviews we identify patterns that are independent of the image itself; that is, we find patterns such as zig-zag, or border-based patterns.

To further extend the analysis we set up a new classification scheme. First we define a category called Lines, which consists of all three of the line categories from [8]; vertical, horizontal and diagonal lines. We find that 27/105, or 26%, of the passwords fall into the Lines category. We then define a Border pattern, consisting of all squares or graphical password points that are on the border of the image. From the data we find that 6/105 passwords follow this pattern. When we define a Zig-zag category as a one row or column zig-zag pattern, then only 2% of first enrolment passwords conform to this pattern.

For the Paperclip image we can define only one pattern, and that is a Colour pattern. We define a password as conforming to the Colour pattern if all the points are selected to be paper clips with the same colour. Our initial expectation was that Paperclip would be the most secure, since it has such a high number of features

TABLE I
COMPARITIVE PATTERNS: FIRST ENROLMENT COMPARED WITH VAN OORSCHOT & THORPE [8] PATTERNS

Pattern	Instances and percentage	Comparitive from [8]
Hotspots	24/105 = 22%	n/a
Vertical lines	2/105 = 2%	15%
Horizontal lines	8/105 = 8	15%
Diagonal lines	3/105 = 3%	11%
Clock	0/105 = 0%	5%



Fig. 3. Images used in the study: (a: top-left) kitten; (b: top-middle) hedgehog; (c: top-right) paperclip; (d: bottom-left) logos; (e: bottom-right) faces.

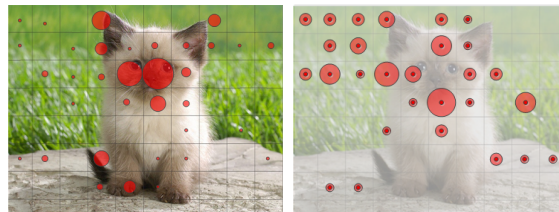


Fig. 4. Hotspot patterns between first (left) and second (right) enrolment.

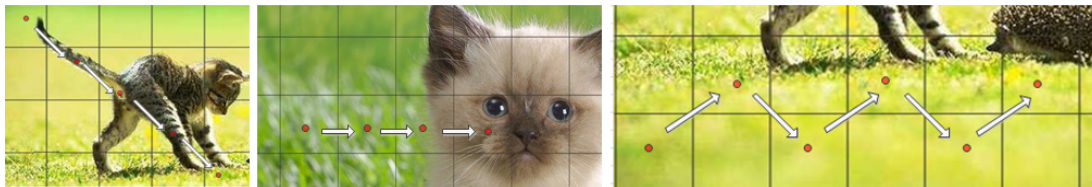


Fig. 5. Non-hotspot patterns found during 1st enrolment.

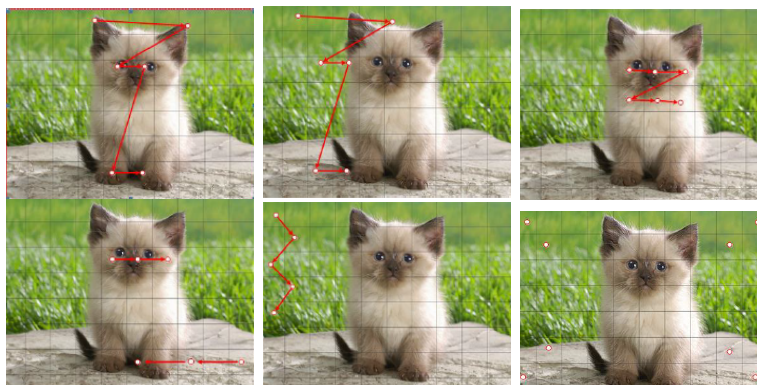


Fig. 6. Example-patterns found during enrolments. (top-left):hotspots; (top-middle):shifted hotspot; (top-right):independent, with hotspot offset; (bottom-left):combination dependent and independent; (bottom-middle):independent of picture; (bottom-right):corner-based pattern.

for the participants to select from. However, it turns out that 48% of passwords consists of a single or two-colour selection for the first enrolment. If the second enrolment's data is also included, then overall we find 33% of passwords conform to a one- or two colour pattern.

Lastly some passwords are clustered on a small number of points that located close to each other. If a 3x3 pattern is defined as a password that consists of only grid positions that fall within a 3x3 grid, we find 7% of all passwords captured to fall under this category.

In summary, the statistics for the five identified patterns are given in Table II. The statistics shows that the number of hotspots significantly decreases from first enrolment (28%) to second enrolment (10%) – see Fig. 4. A similar significant drop is seen in the statistics for the Colour pattern where the pattern is observed for 48% of password selections for the first enrolment and only 19% for the second enrolment. However, for other patterns, such as Lines, 3x3 and Border patterns there is no significant decrease in the observation of the patterns between first and second enrolment. This seem to signify that users change their behaviour only on the patterns that was specifically pointed during the education session between the two enrolments. That is, users tend not to generalize the existence of patterns and only try to avoid hotspots because that pattern was explicitly pointed out to them.

Overall 61% of first enrolment passwords fall into one of the five patterns identified. Even after user education the second enrolment still contain 46% passwords conforming to the identified patterns.

The remainder of yjr passwords, not fitted to the already mentioned patterns are *not* random. There are other patterns that were identified, but with much lower frequencies. For example about 4% of the passwords consist of line segments similar to Fig. 6 (bottom-left), consisting of 2 or more linear parts. The pattern represented in Fig. 6(top-right) consists of a pattern such as zig-zag or double-lines, but started at a hotspot; 2.4%

TABLE II
PASSWORD PATTERN STATISTICS

Pattern	1st Enrolment	2nd Enrolment	Combined
Lines	26%	27%	26%
Hotspot	28%	10%	19%
3x3	7	8%	7%
Border	6%	6%	6%
Colour	48%	19%	33%

of passwords have this pattern.

IV. LIMITATIONS

The study itself was focused on the gathering of qualitative information on password patterns. This is a relatively rare study type for GPs. Most researchers select quantitative studies, typically involving student subjects. Here we attempted to understand GP pattern selection not only by carefully investigating the passwords, but also by interviewing subjects as to what informed their password selection choices.

V. FURTHER WORK

This publication is the third in a series of publications that investigate GPs from different angles using qualitative methods. In the first study [37], the characteristics of GPs were investigated in the context of length and strength. It was shown that in conventional passwords there is character re-use, but in GPs, the re-use of symbols or positions on the image is significantly lower than what is statistically expected. The second study [36] investigated user perceptions regarding graphical passwords, concluding that users in general are still apprehensive to use such technologies for enterprise-level security, such as for authentication during financial transactions. This paper investigated the pattern-richness of GPs.

The use of GPs is now main stream in the sense that they are used widely in device security such as Android pattern unlock. There are, however, significant gaps in understanding what is required to make GPs operational in an enterprise environment.

In addition, since we have shown in this paper that user education has a appreciative effect on behaviour such as hotspot selection, we know that there is still a significant gap between user awareness of security in both conventional and graphical passwords.

VI. CONCLUSION

In this paper we set out to investigate user patterns in graphical passwords by using qualitative methods. We interviewed participants and asked them to enrol with five different images. After asking users for the the reasoning behind their selections and educating the participants on the dangers of hotspots, the users were asked to re-enrol with the same five images.

We find that although there is a significant drop in the number of hotspot passwords, there is still a appreciable pattern-based bias within the second enrolment password set. In particular we find that even after user education,

46% of the second enrolment passwords conform to the five identified categories.

REFERENCES

- [1] R. Biddle, S. Chiasson, and P. C. Van Oorschot, "Graphical passwords: Learning from the first twelve years," *ACM Computing Surveys (CSUR)*, vol. 44, no. 4, p. 19, 2012.
- [2] G. Blonder, "Graphical password. us patent 5559961, lucent technologies," *NJ: Murray Hill*, 1995.
- [3] A. F. Syukri, E. Okamoto, and M. Mambo, "A user identification system using signature written with mouse," in *Information Security and Privacy*. Springer, 1998, pp. 403–414.
- [4] I. Jermyn, A. J. Mayer, F. Monrose, M. K. Reiter, A. D. Rubin *et al.*, "The design and analysis of graphical passwords." in *Usenix Security*, 1999.
- [5] S. Brostoff and M. A. Sasse, "Are passfaces more usable than passwords? a field trial investigation," in *People and Computers XIV Usability or Else!* Springer, 2000, pp. 405–424.
- [6] R. Dhamija and A. Perrig, "Deja vu-a user study: Using images for authentication." in *USENIX Security Symposium*, vol. 9, 2000, pp. 4–4.
- [7] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon, "Passpoints: Design and longitudinal evaluation of a graphical password system," *International Journal of Human-Computer Studies*, vol. 63, no. 1, pp. 102–127, 2005.
- [8] P. C. van Oorschot and J. Thorpe, "Exploiting predictability in click-based graphical passwords," *Journal of Computer Security*, vol. 19, no. 4, pp. 669–702, 2011.
- [9] H. Tao, "Pass-go, a new graphical password scheme," Master's thesis, University of Ottawa, 2006.
- [10] P. Dunphy and J. Yan, "Do background images improve draw a secret graphical passwords?" in *Proceedings of the 14th ACM conference on Computer and communications security*. ACM, 2007, pp. 36–47.
- [11] W. Jansen, S. Gavrilva, V. Korolev, R. Ayers, and R. Swanstrom, "Picture password: a visual login technique for mobile devices," *NIST Report: NISTIR 7030*, 2003.
- [12] P. Bao, J. Pierce, S. Whittaker, and S. Zhai, "Smart phone use by non-mobile business users," in *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*. ACM, 2011, pp. 445–454.
- [13] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens." *WOOT*, vol. 10, pp. 1–7, 2010.
- [14] F. Alt, S. Schneegass, A. S. Shirazi, M. Hassib, and A. Bulling, "Graphical passwords in the wild: Understanding how users choose pictures and passwords in image-based authentication schemes," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 2015, pp. 316–322.
- [15] S. Uellenbeck, M. Dürmuth, C. Wolf, and T. Holz, "Quantifying the security of graphical passwords: The case of android unlock patterns," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 161–172.
- [16] H. Gao, X. Liu, S. Wang, and R. Dai, "A new graphical password scheme against spyware by using captcha." in *SOUPS*, 2009.
- [17] B. B. Zhu, J. Yan, G. Bao, M. Yang, and N. Xu, "Captcha as graphical passwords a new security primitive based on hard ai problems," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 6, pp. 891–904, 2014.
- [18] S. Almuairfi, P. Veeraraghavan, and N. Chilamkurti, "A novel image-based implicit password authentication system (ipas) for mobile and non-mobile devices," *Mathematical and Computer Modelling*, vol. 58, no. 1, pp. 108–116, 2013.
- [19] A. Juels and R. L. Rivest, "Honeywords: Making password-cracking detectable," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 145–160.
- [20] B. B. Zhu, J. Yan, D. Wei, and M. Yang, "Security analyses of click-based graphical passwords via image point memorability," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 1217–1231.
- [21] D. Florêncio and C. Herley, "A large-scale study of web password habits," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 657–666.
- [22] X. Suo, Y. Zhu, and G. S. Owen, "Graphical passwords: A survey," in *Computer security applications conference, 21st annual*. IEEE, 2005, pp. 10–pp.
- [23] S. Man, D. Hong, and M. M. Matthews, "A shoulder-surfing resistant graphical password scheme-wiw." in *Security and Management*. Citeseer, 2003, pp. 105–111.
- [24] S. Wiedenbeck, J. Waters, L. Sobrado, and J.-C. Birget, "Design and evaluation of a shoulder-surfing resistant graphical password scheme," in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2006, pp. 177–184.
- [25] H. Gao, Z. Ren, X. Chang, X. Liu, and U. Aickelin, "A new graphical password scheme resistant to shoulder-surfing," in *Cyberworlds (CW), 2010 International Conference on*. IEEE, 2010, pp. 194–199.
- [26] Y.-L. Chen, W.-C. Ku, Y.-C. Yeh, and D.-M. Liao, "A simple text-based shoulder surfing resistant graphical password scheme," in *Next-Generation Electronics (ISNE), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 161–164.
- [27] A. S. Gokhale and V. S. Waghmare, "The shoulder surfing resistant graphical password authentication technique," *Procedia Computer Science*, vol. 79, pp. 875–884, 2016.
- [28] D. Davis, F. Monrose, and M. K. Reiter, "On user choice in graphical password schemes." in *USENIX Security Symposium*, vol. 13, 2004, pp. 11–11.
- [29] S. Chiasson, R. Biddle, and P. C. van Oorschot, "A second look at the usability of click-based graphical passwords," in *Proceedings of the 3rd symposium on Usable privacy and security*. ACM, 2007, pp. 1–12.
- [30] P. C. van Oorschot and J. Thorpe, "On predictive models and user-drawn graphical passwords," *ACM Transactions on Information and System Security (TISSEC)*, vol. 10, no. 4, p. 5, 2008.
- [31] S. Chiasson, A. Forget, R. Biddle, and P. C. van Oorschot, "User interface design affects security: Patterns in click-based graphical passwords," *International Journal of Information Security*, vol. 8, no. 6, pp. 387–398, 2009.
- [32] M. Weir, S. Aggarwal, M. Collins, and H. Stern, "Testing metrics for password creation policies by attacking large sets of revealed passwords," in *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 2010, pp. 162–175.
- [33] A. E. Dirik, N. Memon, and J.-C. Birget, "Modeling user choice in the passpoints graphical password scheme," in *Proceedings of the 3rd symposium on Usable privacy and security*. ACM, 2007, pp. 20–28.
- [34] J. S. Vorster, "A Framework for the Implementation of Graph-

ical Passwords,” Master’s thesis, University of Liverpool, 12 2014.

- [35] S. Peach, J. Voster, and R. Heerden, “Heuristic attacks against graphical password generators,” in *Proceedings of the South African Information Security Multi-Conference (SAISMC 2010)*, Port Elizabeth, South Africa, 2010, pp. 272–284.
- [36] J. S. Vorster and R. van Heerden, “A study of perceptions of graphical passwords,” *Journal of Information Warfare*, vol. 14, no. 3, 10 2015.
- [37] —, “Graphical passwords: A qualitative study of password patterns,” in *The Proceedings of the 10th International Conference on Cyber Warfare and Security (ICCWS 2015)*, L. Armitstead, Ed. Academic Conferences Limited, February 2015, pp. 375–383.

Dridex: analysis of the traffic and automatic generation of IOCs

Lauren Rudman
Security and Networks Research Group
Department of Computer Science
Rhodes University
Grahamstown, South Africa
Email: g11r0252@campus.ru.ac.za

Barry Irwin
Security and Networks Research Group
Department of Computer Science
Rhodes University
Grahamstown, South Africa
Email: b.irwin@ru.ac.za

Abstract—In this paper we present a framework that generates network Indicators of Compromise (IOC) automatically from a malware sample after dynamic runtime analysis. The framework addresses the limitations of manual Indicator of Compromise generation and utilises sandbox environment to perform the malware analysis in. We focus on the generation of network based IOCs from captured traffic files (PCAPs) generated by the dynamic malware analysis. The Cuckoo Sandbox environment is used for the analysis and the setup is described in detail. Accordingly, we discuss the concept of IOCs and the popular formats used as there is currently no standard. As an example of how the proof-of-concept framework can be used, we chose 100 Dridex malware samples and evaluated the traffic and showed what can be used for the generation of network-based IOCs. Results of our system confirm that we can create IOCs from dynamic malware analysis and avoid the legitimate background traffic originating from the sandbox system. We also briefly discuss the sharing of, and application of the generated IOCs and the number of systems that can be used to share them. Lastly we discuss how they can be useful in combating cyber threats.

Index Terms—network security; malware; dridex; indicators of compromise

I. INTRODUCTION

Many security breaches or intrusions on computer systems are not reported, never made public or even detected [1]. This allows attackers to have free reign of victims' computers, which may have negative effects on organisations, if their employees' computers are compromised. When an organisation finds out about a compromised system or threat and responds accordingly, the information gathered may be valuable to others who experience a similar threat. This makes the sharing of information relating to the detection and identification of threats on an organisation network an important step in dealing with cyber-attacks [2]. The more that is known about a threat, the easier it is to understand, track and counter it.

An Indicator of Compromise (IOC) is defined by Harrington [3] as “a piece of information that can be used to identify a potentially compromised system. It could include suspicious IP addresses, domain names, email addresses, file hashes or a file mutex. This paper focuses on the automated generation of network related IOCs using samples of the Dridex malware [4] strain as test input. The paper will also discuss the process and

analysis used to find the information used in the generation of indicators.

The described system, takes as input a malware sample and outputs IOCs, using a collection of 100 Dridex malware binaries. The systems goal is to focus on the IOC artefacts which can be observed on a network connection – particularly DNS, HTTP, TCP, UDP, ICMP, FTP, SSH and target addresses. These indicators will be created from the PCAP file containing network traffic from automated dynamic malware analysis.

The remainder of the paper is structured as follows. Background information is presented in Section II, followed by an introduction to the common descriptive languages used for constructing IOCs in Section III. The generation system and data collection environment are described in Section IV. An overview of the observed network traffic, and the processing thereof is in Section V, followed, in Section VI, by the process of the generation of IOC's from the captured traffic. Section VII concludes the research, while proposed future enhancements are presented in Section VII.

II. BACKGROUND

There are two types of malware analysis, static and dynamic. Static analysis entails analysing the source code of the malware, never executing it. Dynamic analysis, on the other hand, is all about executing the malware and observing its behaviour on a system. Dynamic analysis is usually performed using a sandbox environment instead of an everyday computer. This is in case the malware potentially deletes files, changes the registry or even steals information. A sandbox is a restricted execution environment, that is run on a system, which allows the safe execution of malware without effecting the host system [5]. There are quite a few online malware analysis sites, such as Anubis, Comodo and Malwr, but these do not scale as they sometimes limit submission speed and the time results are given. It was decided to use a sandbox that runs locally on a system, instead of online.

The current system utilises the Cuckoo Sandbox as the analysis environment. Cuckoo is an open source automated malware analysis system that provides fast and complete analysis results [6]. It takes inputs such as Windows executables, DLL files, PDF documents, Microsoft Office Documents, URLs

and PHP scripts. Some malware does have anti-virtualization techniques and does not execute in a sandbox virtual machine environment [7]. The framework in this paper, will therefore not be able to successfully generate network IOCs from them as they will not generate network traffic. However after successful dynamic analysis of a sample that does execute, Cuckoo generates a PCAP file of captured packets and a report which includes screen shots, static analysis results, dropped files, DNS and HTTP requests and a behaviour summary.

No accepted standard format for the IOCs exists yet. There are however a few systems that have their own formats, such as OpenIOC¹, Cyber Observable Expression (CyBOX)² and Structured Threat Information Expression (STIX)³. These are discussed in Section III. The current systems typically require the manual input or tagging of information to generate IOCs, which is not scalable and would take a lot of time to generate many IOCs. One of the recent developments in the sharing of cyber threats is OASIS Cyber Threat Intelligence (CTI)⁴. The OASIS CTI Technical Committee, which includes the U.S Department of Homeland Security and other organisations have come together to develop standards to enable the analysis and sharing of threats and treat information. They are intending for the cyber threat information to be shared among trusted partners and communities [8]. It would be useful to have a standard format, so that IOCs can be easily shared without having to convert between formats. This would also allow for a greater distribution of IOCs and help security teams in tackling cyber threats.

There are other new solutions which allow for the uploading of IOCs in multiple formats, such as the AlienVault Open Threat Exchange (OTX)⁵. OTX is an online platform for sharing cyber threat information about malware or fraud campaigns and more. Another solution is the Malware Information Sharing Platform (MISP)⁶, which is a platform for sharing IOCs of targeted attacks.

When conducting a search for tools that automatically generate IOCs, a few simple scripts such as IOC_Creator⁷ and IOCAware⁸ were found. IOC_Creator generates OpenIOC formatted IOCs from unstructured data, although it lacks detail and is not comprehensive in terms of network Indicators. IOCAware uses a Cuckoo report generated after a file is analysed and only generates an Indicator for an IP address contacted and no other network IOCs.

Dridex is a type of malware, with the primary goal of infecting computers, stealing credentials, and obtaining money from victims bank accounts [4]. It was first observed in the wild in November 2014 [9]. When the malware is installed, the computer becomes part of a botnet [4], which can be used to

send phishing emails. In mid October 2015 many command-and-control servers used by the Dridex botnet were taken down by the Federal Bureau of Investigation (FBI) with the help of the National Crime Agency (NCA) [10]. However in late October, security researchers found signs that the botnet might still be functioning [11]. According to [12], Dridex was barely seen from 24 December 2015, but resumed its operations again in early January 2016. In February of 2016, it was found that part of the Dridex botnet may have been hacked as part of its distribution channel was changed by replacing malicious links with an installer for the Avira antivirus [13]. In March 2016, the Dridex botnet started to send SPAM emails with JavaScript attachments that eventually install Locky ransomware [14]. According to [15], in May 2016, the botnet was compromised again to distribute a “dummy file” instead of the Dridex binary.

The actual Dridex malware is spread through multiple types of spam email attacks with a Microsoft Word or Excel document attached, which includes a payload that downloads the malware [9]. Macros must be enabled in Microsoft Word for the payload to work [9]. Once installed, Dridex uses HTML injections to retrieve banking details [9] and can even steal user credentials through keystroke logging, form grabbing, stealing cookies and screenshots [4] [10] [12]. It is able to steal banking details of nearly 300 financial institutions of generally English speaking countries [16].

Dridex was chosen as it is a topical malware strain and since the framework can take any malware binary as an input, we thought that Dridex would be a useful demonstrative family. For the purposes of this research, 100 binaries identified to contain variants of the Dridex strain were analysed to show how the system operates. These were dated within the last twelve months.

III. STIX

STIX stands for Standard Threat Information Expression [8] and is used to describe information about cyber threats. It is an XML based format and was created to have a language that allows threat information to be easily stored, analysed and shared in a consistent manner [17]. We chosen STIX as the IOC language of choice because is able to represent a wide number of network level indicators. The level of detail of a STIX object can vary from one single property of an object to multiple properties of an object and even the logical (AND/OR) combination of objects [18]. The allowance for multiple indicators to be logically combined allows for the creation of IOCs to be flexible and have a high level of detail when needed.

STIX allows for the creation of many types of cyber threats, such as observables, indicators, incidents, exploit targets and more. We will be focusing on the creation of STIX Indicators in this paper. A STIX Indicator is made up of CyBOX objects, which contain a number of cyber observables. A STIX Indicator gives the CyBOX objects context by adding a title and description. A set of related STIX Indicators is grouped by a STIX Report and and lastly the Indicators and Reports are grouped using a STIX Package. CyBOX is a language used

¹<http://openioc.org/>

²<https://cybox.mitre.org/>

³<https://stixproject.github.io/about/>

⁴<https://www.oasis-open.org/>

⁵<https://otx.alienvault.com/>

⁶<http://www.misp-project.org/>

⁷https://github.com/tkllane/openiocscripts/blob/master/ioc_creator.py

⁸<https://goo.gl/ipBjZL>

to describe "events of stateful properties that are observable in a cyber domain" [17]. CybOX's data model uses an XML schema as does STIX.

It was decided to create our own reporting module that uses a filtered PCAP file to generate detailed network-based STIX Indicators. STIX is also one of the formats chosen by the OASIS CTI team to be a standard in the future of cyber threat sharing [8].

A. CybOX Objects

There are a number of different CybOX objects that can be used to create network related IOCs. The objects we have used for our system are listed below:

- Address: can be used to store addresses which include e-mail, MAC and IP addresses.
- Port: stores a port value.
- URI: stores a URI.
- Socket Address: stores an IPv4 address and Port.
- DNS Query: stores properties of a DNS query.
- HTTP Session: can store properties of a HTTP request and the response.
- Network Connection object, which is used to store information regarding any type of network connection.

The Port, URI, Socket Address and DNS Query objects were used as described in the above list. The Address object was used to store an IPv4 address only and the HTTP Session was used to store properties of an HTTP GET/POST request, without the response. The Network Connection object was used to represent TCP, UDP, ICMP, SSH and FTP connections.

IV. DATA COLLECTION

A. Cuckoo Sandbox

We implemented the framework on top of the Cuckoo Sandbox [5] which we used to execute and perform a first pass analysis on each Dridex malware sample. Cuckoo was chosen as the analysis tool because it is written in Python and is fully customizable and extendable [6]. It takes a suspicious file as an input and performs dynamic malware analysis on it, then generates reports, screen shots and a PCAP file. Our framework will focus on taking these reports and PCAP file to generate network related IOCs.

The latest version of Cuckoo was downloaded and installed on a Debian 8 server. In order for the sandbox to function, it needs a VM, a snapshot of the VM and a few variables in the configuration files changed to suit the setup. Cuckoo was configured to allow the use of VirtualBox as its virtual machine manager and each sample was set to run for 30 minutes each. We also had to configure more specific settings for VirtualBox in Cuckoo, such as making sure Cuckoo ran the VM in headless mode, the IP address of the virtual machine and the name of the virtual machine together with its snapshot.

In order for Cuckoo to capture network traffic, the configuration file, `auxiliary.conf` had to be modified to enable the packet sniffer, give the path to the local installation of the `tcpdump` utility and the name of the network adapter to

capture traffic from. If these are wrong Cuckoo will not be able to capture traffic, or will record from the wrong adapter.

B. Virtual Machine Setup

As previously stated, VirtualBox was chosen to create and manage the virtual machines. Windows 7 Ultimate SP1 was installed on a VM to replicate an everyday user's computer. Windows XP may be used to test the in future work, but may not be too relevant these days as it is outdated and no longer supported by Microsoft⁹.

The VM was setup to have 2GB of RAM together with 20 GB of storage. A few outdated versions of programs were installed such as Mozilla Firefox, Adobe, Microsoft Office 2007, Java, Google Chrome, Flash player, Opera, Adobe air, iTunes and Mozilla Thunderbird. These were chosen because they are common everyday programs and Microsoft Office was chosen because it is one of the programs Dridex uses to run its payload. However, the samples may be secondary binaries that do not even utilise this. To ensure minimal security, the firewall, Windows defender, and Windows updates were turned off along with not installing an antivirus.

The virtual machine was allowed access to the internet through a bridged adapter. The IP address of the VM and other network settings were statically set because Cuckoo needs to know the IP of the VM.

C. Extracting useful information

A toolchain of Python scripts were used to extract and analyse information generated by Cuckoo. First a script was created to filter out (as best as possible) non malware related traffic from each PCAP file. A non-malicious image file was submitted to the Cuckoo Sandbox five times to observe traffic created by the VM's snapshot. By using the baseline PCAPs, all of the IP addresses that the VM communicated with were extracted (excluding the pre-configured DNS server and the VM's IP address). These addresses were added to a list of clean IP addresses to filter from the PCAPs after the malicious files are analysed by Cuckoo.

Next the DNS queries and responses were extracted, which allowed us to create a list of clean domain names and the resolved IP addresses (if the domain was resolved). Taking a look at the domain names, a second filter list of clean domains was created. This list included domains such as 'microsoft.com', 'google.com', 'bing.com', 'windowsupdate.com', 'apple.com', 'sun.com' and others. These domain names, of course depend on the specific operating system and program versions installed on the VM. The IP addresses that the domains resolved to were added to the first IP filter list, if they were not previously added.

These two filter lists were used to create a `tshark` filter that reads the Cuckoo generated PCAP and creates a new PCAP. The filtered PCAP has packets of the type TCP, UDP, DNS, HTTP and ICMP and does not have packets to or from the clean IP addresses and also does not have the DNS requests

⁹<http://windows.microsoft.com/en-us/windows/end-support-help>

and responses for domains in the domain filter list. Because of dynamic IP addresses these two filter lists would have to be regenerated, before running a number of samples through Cuckoo.

The `tshark` filter was saved to a bash script so it could be utilised by a custom Cuckoo processing module that was developed. A Cuckoo processing module¹⁰ is a python script that lets you analyse the raw output from Cuckoo and append some information to a global container that can be used by the reporting modules. After Cuckoo has executed a sample in a Virtual Machine the processing modules are called with the reporting modules following. The processing module that was created calls the `tshark` filter script after Cuckoo has executed a sample for 30 minutes and has generated a PCAP. The custom processing module worked well for the system and managed to filter out most unwanted baseline traffic. However, some samples would very rarely be found contacting local university IP addresses, such as IT management servers and printers, so these IPs were added manually to the clean filter list.

The next step was to work with these filtered PCAPs assuming they only contain malicious traffic and gather information that may be useful in the creation of IOCs. After all the enabled Cuckoo processing modules are finished executing, the reporting modules are run. A custom reporting module was created for the purpose of retrieving information from the filtered PCAP and using that information for the creation of STIX Indicators. Secondary filtering had to be implemented in the reporting module because of dynamic IP addressing where some Microsoft domains would resolve to a different address as found in the baseline PCAP files.

V. NETWORK TRAFFIC OVERVIEW

From the 100 Dridex samples that ran for 30 minutes, Cuckoo was able to execute 100% of them, with 50 samples generating network traffic. The traffic only added up to 31,45 MB. We are unsure as to why some of the samples did not show network activity. According to Rossouw et al. [19], this may be because the samples were invalid, or only active when there is user activity or they detected the sandbox and stopped working. We suspect it may also be because the malware needed more time to run (longer than 30 minutes) in order to generate traffic. Another possibility is that some of the malware was designed to activate only within a certain time period (which had expired). The take down of much of the Dridex infrastructure in October may also have played a role, in the reduced volumes of observed traffic. The following subsections will discuss the results found when analysing the filtered PCAPs which we assume contain malicious traffic. We will show information about TCP connections, DNS requests and responses, HTTP requests and any hard-coded IP addresses.

¹⁰<http://docs.cuckoosandbox.org/en/latest/customization/processing/>

A. DNS

Of the 50 samples that generated traffic within 30 minutes of running, 40 of the samples used the DNS protocol (port 53 UDP). Of the 40 samples, all of them used the pre-configured DNS server. If some of the samples had used their own resolver, the information could have been used in the creation of an IOC. Some malware strains use their own iterative/recursive resolvers to avoid leaving traces in logs or caches of preconfigured resolvers on a victim network [19].

Table I shows the 14 domain names that were requested by some of the 40 samples. Ten of the domains were resolved, which could mean that the other four are no longer in use or they could be blocked by the preconfigured resolver set by the university.

Looking at the TTL values of the ten resolved domains, seen in Table I the most popular domain, `icanhazip.com` has a TTL of 5 minutes. There are three very small values seen such as 3, 10 and 20 seconds, which are generally related to Content Delivery Networks (CDNs) [20]. In this case the three domains are related to online certificates, which could explain why the values are low. A TTL of zero, which is not seen in these results, can indicate the use of fast flux domains which are used to “provide flexibility among the command and control infrastructure of bots” [21]. However, many domains using a TTL of zero could be included as part of an extended IOC in future work.

The most popular domain `icanhazip.com` was requests by 33 of the 100 samples and 66% of all samples that generated network traffic. This site is non malicious and is used to determine the IP address of the host that loaded the page. According to [22], [23] and [24] malware authors use this domain and similar sites to obtain the IP of an infected computer, as part of the environmental determination used prior to contacting the Command and Control (C&C) node(s). This domain is often suggested to be used as an IOC according to [24]. In this case, we think that using this domain as an IOC is a good idea as it appeared many times from the samples. The other domain, `api.ipify.org`, is also a non malicious site and is similarly used to check the IP address of a client computer.

As seen in Table I, three distinct samples were seen using the following domains: `th.symcb.com`¹¹, `th.symcd.com`¹² and `ocsp.thawte.com`¹³ [25] and two of the three samples also queried `crl.thawte.com`¹⁴. These domains are non malicious in themselves, but have been identified to be requested by malware in some cases as referenced above on VirusTotal. These domains can be used together to create an IOC to represent the three samples.

The domains `malwagroup.org`, `thedirtydelicious.com`, `nerdmeetsgirl.com`, `tanhadhidown.ru`, `herssofhaprih.ru`, `nohissandbo.ru` were requested by the same sample and according to [26], these domains are used to download the

¹¹<https://www.virustotal.com/en/domain/th.symcb.com/information/>

¹²<https://www.virustotal.com/en/domain/th.symcd.com/information/>

¹³<https://www.virustotal.com/en/domain/ocsp.thawte.com/information/>

¹⁴<https://www.virustotal.com/en/domain/crl.thawte.com/information/>

TABLE I: Domain names requested from 40 of the samples

Domain name	Number of samples	IPs resolved	DNS TTL
icanhazip.com	33	64.182.208.185 64.182.208.184	300
th.symcb.com	3	23.42.5.163	20
th.symcd.com	3	23.42.11.27	3
ocsp.thawte.com	3	23.42.11.27	10
crl.thawte.com	2	23.42.5.163	900
ho7rcj6wucosa5bu.tor2web.org	1	194.150.168.70 38.229.70.4 65.112.221.20	3600
api.ipify.org	1	50.17.192.14 107.20.229.58 54.243.252.101	60
malwagroup.org	1	182.50.130.67	3600
thedirtydelicious.com	1	184.168.27.45	600
nerdmeetsgirl.com	1	184.168.47.225	600
tanhadhidown.ru	1		
herssofhaprih.ru	1		
nohissandbo.ru	1		
mcreport.org	1		

Pony and Dyre banking malware. These six domains can be combined used to create one IOC as they are only seen in one sample.

One sample used the ho7rcj6wucosa5bu.tor2web.org and api.ipify.org domains. As stated above, api.ipify.org, is used to retrieve the IP address of an infected host and ho7rcj6wucosa5bu.tor2web.org is a known malicious IP address [27] [28] and is using the Tor2web network gateway¹⁵. These two domains can be used to to create two indicators or one combined indicator. The last domain, mcreport.org, was not resolved and is only mentioned as the result of the analysis of a malicious file on [29]. This domain seems to be out of use at the moment, but can still be used to create an IOC.

Other than the domains themselves, the IP addresses that were resolved can also be used in the generation of IOCs. The resolved IP addresses may change, so this may not be as effective as an IOC because it may not be relevant for long.

B. HTTP

Only three samples from the 50 that generated network traffic utilised the HTTP protocol. This is a significantly small amount of of samples and does not give too much to work with in terms of IOC generation. Between the three samples, there were nine HTTP requests and four "200 OK" replies and four "404 Not Found" replies and one did not receive a reply.

When looking at some of the HTTP header fields, it was found that six out of the seven User-Agent fields did not correspond to the programs and operating system. For example one of the requests specified the Opera browser (not installed on the VM) and another specified that it was running on Ubuntu. The User-Agents are shown in Table II along with the ID of the sample. Sample 4 used three different User-Agents and so did Sample 48. It is interesting that every HTTP request had a different User-Agent value. Sample 99 was the only sample to use a correct User-Agent field. In

[19], which also found that malware forges their own User-Agents, it was suggested that one sample can use different User-Agent because of the modular nature of malware. Each different module has its own way of forging an HTTP request.

Sample 4, mentioned in Table II, HTTP GET requested three URIs. These were 70.127.18.124/online.htm, 178.137.58.176/main/htm and 94.139.196.46/home.htm. The first URI did not receive a reply and the last two received a 404 Not Found reply. Sample 4 did not send any DNS requests, so these addresses were not resolved from a domain name.

Sample 48, sent three unique HTTP GET requests to IP addresses that had not been resolved from DNS requests. The three requests were 79.119.76.125/online.htm , 79.119.76.12/welcome.htm and 122.118.192.8/index.htm, the last of which brings up a login page to a router. There is most likely a compromised computer behind the router that is part of the Dridex botnet. Since the previously mentioned IP addresses had not been resolved from DNS requests it makes them a good property to add to part of an indicator. Each of the HTTP requests can be made into an HTTP Session CyBOX object, that can be wrapped by a STIX Indicator.

Sample 99 sent three HTTP GET requests seen in Figure 1. These headers are very similar and show that the sample was attempting to download a file called 'k1.exe' from the three domains. The headers have identical accept-language, accept-encoding, accept, and user-agent values in the fields and two of the headers have the same GET parameter. The malware most likely sends three requests for the same file (assuming it is the same file) in case one or more of the domains is down. In the case of the sample run, the 'nerdmeetsgirl.com' request (shown in Figure 1a) received a HTTP/200 response. For HTTP requests for 'thedirtydelicious.com' and 'malwagroup.org', shown in Figures 1b and 1c, response codes were received indicating that the files were no-longer present for download.

From the four different HTTP requests all of them used the GET request method. Rossow et al [19], analysed the network

¹⁵<https://tor2web.org/>

TABLE II: HTTP User Agents and sample IDs

User Agent	Sample ID
Mozilla/5.0 (Windows NT 5.1; rv:21.0) Gecko/20130401 Firefox/21.0	4
Mozilla/4.0 (compatible; MSIE 10.0; Windows NT 6.1; Trident/5.0)	4
Mozilla/5.0 (compatible; MSIE 9.0; AOL 9.7; AOLBuild 4343.19; Windows NT 6.1; WOW64; Trident/5.0; FunWebProducts)	4
Mozilla/5.0 (Windows NT 6.2) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/28.0.1464.0 Safari/537.36	48
Opera/9.80 (Windows NT 6.1; U; es-ES) Presto/2.9.181 Version/12.00	48
Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:21.0) Gecko/20130331 Firefox/21.0	48
Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E; GWX:QUALIFIED)	99

output of multiple types of malware samples, they also found that GET request was the most popular request method over POST.

```
GET /wp-content/plugins/cached_data/k1.exe HTTP/1.0
accept-language: en-US
accept-encoding: identity, *,q=0
host: nerdmeetsgirl.com
accept: */*
user-agent: Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E; GWX:QUALIFIED)
connection: close
```

(a) HTTP request to 'nerdmeetsgirl.com'

```
GET /wp-includes/simplepie/net/k1.exe HTTP/1.0
accept-language: en-US
accept-encoding: identity, *,q=0
host: thedirtydelicious.com
accept: */*
user-agent: Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E; GWX:QUALIFIED)
connection: close
```

(b) HTTP request to 'thedirtydelicious.com'

```
GET /wp-content/plugins/cached_data/k1.exe HTTP/1.0
accept-language: en-US
accept-encoding: identity, *,q=0
host: malwagroup.org
accept: */*
user-agent: Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E; GWX:QUALIFIED)
connection: close
```

(c) HTTP request to 'malwagroup.org'

Fig. 1: Three HTTP GET requests generated by one of the samples

C. Other Protocols

In terms of TCP requests, 42 out of the 50 samples that generated TCP traffic with 108 connections established and 869 connections failing. The samples did not show any malicious ICMP, UDP, FTP or SSH traffic.

VI. IOC GENERATION AND RESULTS

In this section we discuss how the IOCs were generated and show an example of an IOC that was created. As stated above, we used STIX for the creation of indicators and we generated seven types of indicators, ICMP, TCP, UDP, SSH, FTP, DNS and HTTP. Figure 2 shows the resulting flow of the system, with IOCs and a final product and Intrusion Detection System or firewall rules as a potential final product. As seen in the image and stated previously, a malware sample is submitted to the Cuckoo sandbox, which dynamically analyses its behaviour and records the network traffic to a PCAP file. The

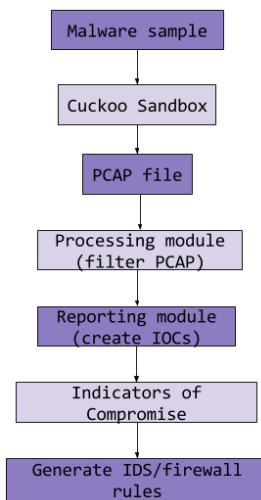


Fig. 2: IOC analysis and creation flow diagram

TABLE III: Cybox fields used for each IOC

IOC type	Network Properties
ICMP	IPv4 address, Type
TCP	IPv4 address, Port, TCP state
UDP	IPv4 address, Port
HTTP	Method, URI, Version, Host, Port, Accept, Accept language, Accept encoding, Authorization, Cache control, Connection, Cookie, Content length, Content type, Date, Proxy authorization
DNS	IPv4 address, Port, Domain name, Type
FTP	IPv4 address, Port, Request argument, Response argument
SSH	IPv4 address, Port, Public key

PCAP file is used by a custom processing module, which trims it, as described in Section IV-C. A custom reporting module then reads the new PCAP file and extracts useful values. Those values are then used to generate STIX Indicators for the specific protocols. The generation of the IOCs is discussed in more detail below.

A. IOC generation

In order to generate each IOC we need the correct values for each one. Table III shows the eight types of IOCs that were created and the network properties that were used to create them.

In order to generate the objects, we identified all the necessary packets for each malware sample and extracted the

```

-<cybox:Properties xsi:type="NetworkConnectionObj:NetworkConnectionObjectType">
  <NetworkConnectionObj:Layer3_Protocol>IPv4</NetworkConnectionObj:Layer3_Protocol>
  <NetworkConnectionObj:Layer4_Protocol>TCP</NetworkConnectionObj:Layer4_Protocol>
  <NetworkConnectionObj:Layer7_Protocol>HTTP</NetworkConnectionObj:Layer7_Protocol>
-<NetworkConnectionObj:Layer7_Connections>
  -<NetworkConnectionObj:HTTP_Session xsi:type="HTTPSessionObj:HTTPSessionObjectType">
    -<HTTPSessionObj:HTTP_Request_Response>
      -<HTTPSessionObj:HTTP_Client_Request>
        -<HTTPSessionObj:HTTP_Request_Line>
          <HTTPSessionObj:HTTP_Method>GET</HTTPSessionObj:HTTP_Method>
          <HTTPSessionObj:Value>/wp-content/plugins/cached_data/k1.exe</HTTPSessionObj:Value>
          <HTTPSessionObj:Version>HTTP/1.0</HTTPSessionObj:Version>
        </HTTPSessionObj:HTTP_Request_Line>
      -<HTTPSessionObj:HTTP_Request_Header>
        -<HTTPSessionObj:Parsed_Header>
          <HTTPSessionObj:Accept>*/*</HTTPSessionObj:Accept>
          <HTTPSessionObj:Accept_Language>en-US</HTTPSessionObj:Accept_Language>
          <HTTPSessionObj:Accept-Encoding>identity, *,q=0</HTTPSessionObj:Accept-Encoding>
          <HTTPSessionObj:Connection>close</HTTPSessionObj:Connection>
        -<HTTPSessionObj:Host>
          -<HTTPSessionObj:Domain_Name xsi:type="URIObj:URIObjectType">
            <URIObj:Value>nerdmeetsgirl.com</URIObj:Value>
          </HTTPSessionObj:Domain_Name>
          -<HTTPSessionObj:Port xsi:type="PortObj:PortObjectType">
            <PortObj:Port_Value>80</PortObj:Port_Value>

```

Fig. 3: STIX Indicator for a HTTP GET Request

properties shown in Table III. Next the python-cybox¹⁶ library was used to create CybOX objects out of the extracted values. These objects were then placed into STIX Indicators using the python-stix¹⁷ library. The ID numbers of the STIX Indicators were placed into a STIX Report and the Indicators and the Report was finally wrapped by a STIX Package.

Part of a STIX Indicator that was created to represent a HTTP GET request is shown in Figure 3. A STIX Indicator can include a Title and Description which we used to describe the IOC (not shown in the Figure). The layout of a CybOX HTTP Session object is shown using the information from the before mentioned, 'nerdmeetsgirl.com' request shown in Figure 1a. The Host, URI, Port, Protocols and more are represented in Figure 3. CybOX objects have the useful trait of including the creation date of an IOC. This IOC layout is advantageous because it is simplistic as it does not contain too much information. It also contains meta data which is useful for sharing, so people can understand what the IOC is about.

Each malware sample ended up with an XML file containing all the indicators that were created, which is the final product of the system. These XML files can easily be shared manually using the AlienVault OTX. AlienVault requires a STIX file to be uploaded with the extension changed to '.ioc' from '.xml' before it is uploaded. MISP also allows for the upload and sharing of STIX data. MISP is also useful because it allows for the export of IOCs in different formats including Intrusion Detection Systems (IDS) rules, OpenIOC, plain text, Snort rules and Suricata rules.

¹⁶<https://github.com/CybOXProject/python-cybox>

¹⁷<https://github.com/STIXProject/python-stix>

VII. CONCLUSION

In this work, we presented a framework for the automatic generation of Indicators of Compromise from a malware sample. The Dridex malware strain was used as an example set of malware for analysis and the samples generated PCAP files during dynamic analysis. An overview of the network traffic for DNS and HTTP protocols was shown, which resulted in some suspicious domain names, and HTTP request packets. The information gathered from these suspicious packets was used to generate the IOCs.

We can confirm that useful network-based IOCs can be generated from dynamic malware analysis while avoiding the legitimate background traffic originating from the sandbox system. An example of one of the IOCs can be seen in Section VI, and shows that the framework can create comprehensive STIX Indicators. Since the system can take any malware as an input, and uses PCAP files for the generation of Indicators, any malware that generates traffic during dynamic analysis in the sandbox used, will have a STIX file of IOCs generated.

The one downside of the system at the moment is that a baseline network traffic test has to be run every so often, but the method of filtering legitimate traffic from Cuckoo's PCAP file was very effective and lead us to create more accurate IOCs instead of creating IOCs from legitimate traffic, which would be troublesome. We believe that the framework, when expanded, will be a useful and scalable tool for the creation of all types of IOCs and could be used effectively in sharing cyber threats. This will help in combating cyber treats, by allowing the efficient generation of IOCs.

VIII. FUTURE WORK

Future research will be done with the aim of evaluating an optimal (and possibly flexible) means of sharing this IOC data in a way that it can be meaningfully utilised by others. This information will be used to expand the system to automatically share IOCs. Another useful expansion would be the creation of Intrusion Detection System and firewall rules from the STIX Indicators. This would allow for the data to be used as a defence mechanism against malware.

REFERENCES

- [1] D. W. Chris Johnson, Lee Badger, "Nist special publication 800-150 (draft) guide to cyber threat information sharing (draft)," October 2014. [Online]. Available: http://csrc.nist.gov/publications/drafts/800-150/sp800_150_draft.pdf
- [2] J. A. L. Denise E. Zheng. (2015, March) Cyber threat information sharing recommendations for congress and the administration. CSIS. [Online]. Available: http://csis.org/files/publication/150310_cyberthreatinfosharing.pdf
- [3] C. Harrington, "Sharing indicators of compromise: An overview of standards and formats," Conference Presentation, November 2013. [Online]. Available: https://www.rsaconference.com/writable/presentations/file_upload/dsp-w25a.pdf
- [4] US-CERT. (2015, October) Alert (TA15-286A) Dridex P2P Malware. Online Article. US-CERT. [Accessed on: 23 October 2015]. [Online]. Available: <https://www.us-cert.gov/ncas/alerts/TA15-286A>
- [5] D. Oktavianto and I. Muhardianto, *Cuckoo Malware Analysis*. Packt Publishing Ltd, 2013.
- [6] A. Provataki and V. Katos, "Differential malware forensics," *Digital Investigation*, vol. 10, no. 4, pp. 311–322, 2013.
- [7] X. Chen, J. Andersen, Z. M. Mao, M. Bailey, and J. Nazario, "Towards an understanding of anti-virtualization and anti-debugging behavior in modern malware," in *2008 IEEE International Conference on Dependable Systems and Networks With FTCS and DCC (DSN)*. IEEE, 2008, pp. 177–186.
- [8] C. Geyer. (2015, July) Oasis advances automated cyber threat intelligence sharing with stix, taxii, cybox. Blog Post. OASIS. [Accessed on: 29 November 2015]. [Online]. Available: <https://www.oasis-open.org/news/pr/oasis-advances-automated-cyber-threat-intelligence-sharing-with-stix-taxii-cybox/>
- [9] M. Sanghavi. (2015, March) DRIDEX and how to overcome it. Blog Post. Symantec. [Accessed on: 23 October 2015]. [Online]. Available: <http://www.symantec.com/connect/blogs/dridex-and-how-overcome-it>
- [10] Trend Micro. (2015, October) FBI, Security Vendors Partner for DRIDEX Takedown. Blog Post. Trend Micro. [Accessed on: 23 October 2015]. [Online]. Available: <http://blog.trendmicro.com/trendlabs-security-intelligence/us-law-enforcement-takedown-dridex-botnet/>
- [11] D. Bisson. (2015, October) The Dridex botnet ain't done yet, say researchers. News Article. Graham Cluley. [Accessed on: 23 October 2015]. [Online]. Available: <https://grahamcluley.com/2015/10/dridex-botnet-dead/>
- [12] D. O'Brien, "Dridex: Tidal waves of spam pushing dangerous financial trojan," Symantec, White Paper, February 2016. [Online]. Available: http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/dridex-financial-trojan.pdf
- [13] L. Frink. (2016, February) Dridex botnet distributor now serves avira. Blog post. Avira. [Accessed on: 29 April 2016]. [Online]. Available: http://blog.avira.com/dridex_serves_avira/
- [14] S. News. (2016, March) Dridex botnet spreading locky ransomware via javascript attachments. News Article. Security Week. [Accessed on: 29 April 2016]. [Online]. Available: <http://www.securityweek.com/dridex-botnet-spreading-locky-ransomware-javascript-attachments>
- [15] Z. Zorz. (2016, May) Dridex botnet hacked, delivers dummy file. Online Article. Help Net Security. [Accessed on: 6 May 2016]. [Online]. Available: <https://www.helpnetsecurity.com/2016/05/05/dridex-botnet-hacked/>
- [16] ——. (2016, February) Dridex botnet alive and well, now also spreading ransomware. Online Article. Help Net Security. [Accessed on: 29 April 2016]. [Online]. Available: <https://www.helpnetsecurity.com/2016/02/17/dridex-botnet-alive-and-well-now-also-spreading-ransomware/>
- [17] MITRE. About STIX. The MITRE Corporation. [Online]. Available: <http://stixproject.github.io/about/>
- [18] (2015) ObservableTypeCYBOX CORE SCHEMA. MITRE. [Accessed on: 1 November 2015]. [Online]. Available: <http://stixproject.github.io/data-model/1.2/cybox/ObservableType/>
- [19] C. Rossow, C. J. Dietrich, H. Bos, L. Cavallaro, M. Van Steen, F. C. Freiling, and N. Pohlmann, "Sandnet: Network traffic analysis of malicious software," in *Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*. ACM, 2011, pp. 78–88.
- [20] K. Fujiwara, A. Sato, and K. Yoshida, "Dns traffic analysiscdn and the world ipv6 launch," *Information and Media Technologies*, vol. 8, no. 3, pp. 833–842, 2013.
- [21] T. Holz, C. Gorecki, K. Rieck, and F. C. Freiling, "Measuring and detecting fast-flux service networks," in *NDSS*, 2008.
- [22] L. Teo, "Learning from the Dridex Malware - Adopting a Effective Strategy," SANS Institute, White Paper, October 2015. [Online]. Available: <https://www.sans.org/reading-room/whitepapers/detection/learning-dridex-malware-adopting-effective-strategy-36397>
- [23] AplusWebMaster. (2015, July) 'Changed Identification Numbers', 'Hilton Hotel' SPAM, 'Zombie 'Orkut' Phish ... Forum Post. Spybot. [Accessed on: 1 November 2015]. [Online]. Available: <https://forums.spybot.info/showthread.php?23632-SPAM-frauds-fakes-and-other-MALWARE-deliveries/page75>
- [24] B. Duncan. (2015) Upatre/Dyre - the daily grind of botnet-based malspam. Forum Post. SANS ISC InfoSec. [Accessed on: 1 November 2015]. [Online]. Available: <https://isc.sans.edu/forums/diary/UpatreDyre%20the%20daily%20grind%20of%20botnetbased%20malspam/19657/>
- [25] [Online]. Available: <https://www.virustotal.com/en/domain/ocsp.thawte.com/information/>
- [26] [Online]. Available: <https://www.virustotal.com/en/file/facc9a5f02e8d18c9cbac9ee760ffa38b2854e5d5c89a529e368be8857bc55a9f/analysis/>
- [27] B. Duncan. (2015, February) 2015-02-02 - malspam run pushes chanitor - subject: Logmein promo code - get 50MALWARE-TRAFFIC-ANALYSIS.NET. [Accessed on: 1 November 2015]. [Online]. Available: <http://www.malware-traffic-analysis.net/2015/02/02/index.html>
- [28] [Online]. Available: <https://www.virustotal.com/en/domain/ho7rcj6wucosa5bu.tor2web.org/information/>
- [29] [Online]. Available: <https://malwr.com/analysis/JzJkMmJkNtK3YmUyNDliZWfKMDNiZmQ3MmQ1YjJkZGU/>

Context Aware Mobile Application for Mobile Devices

Mfundo Masango*, Francois Mouton†, Alastair Nottingham‡ and Jabu Mtsweni§

Command, Control and Information Warfare
Defence, Peace, Safety and Security
Council for Scientific and Industrial Research
Pretoria, South Africa

*Email: gmasango@csir.co.za

†Email: moutonf@gmail.com

‡Email: anottingham@csir.co.za

§Email: jmtsweni@csir.co.za

Abstract—Android smart devices have become an integral part of peoples lives, having evolved beyond the capability of just sending a text message or making a call. Currently, smart devices have applications that can restrict access to other applications on the same device, implemented through user authentication. Android smart devices offer the capability of Android Smart Lock, which uses different authentication methods for unlocking the device based on the users location. However, Android Smart Lock does not allow locking for individual applications. A possible solution to this limitation is an application that performs user authentication using a context-aware approach. This paper proposes a context-aware application, which provides different user authentication methods that are set up according to the auto-detection of areas designated as safe zones by the user. This application aims to improve the overall security of the content of a given device by securing individual applications.

Index Terms—Android; Geofence; Google location services; Lock screen; Pattern lock;

I. INTRODUCTION

Android smart devices such as smart phones and tablets have evolved far beyond the capability of sending text messages or making calls, and are now arcades, personal navigators, storage devices, and social hubs. Another important innovation in modern Android devices is the development of context awareness with respect to the device's given surroundings [1]. Devices are now able to alert users when they are entering or exiting a particular area of interest, and may alert the user by means of a notification or alarm sound [2].

Applications that are context aware are able to track a user's current location, activating different authentication methods based on where the user is situated. These applications make use of geofences; virtual perimeters created around a real-world geographic areas which are dynamically generated [3]. For example, a user may want to be alerted when entering an area that has a McDonald's store, or when they are within the range of a house that they are viewing on a real estate website. In this case, when a user passes the area where the house is located, a notification can be sent to their mobile device that will inform the user of the house's proximity and

show the user the house's location via the Global Positioning System (GPS) [2]. The user may then use this information to find the house through a navigation app, such as Google Maps.

GPS receivers are highly accurate in determining a device's exact current location, achieved through the use of satellites [4], [5]. Geofencing is the practice of using GPS or Radio Frequency Identification (RFID) to define a geographic boundary [3]. Once this 'virtual barrier' is established, the administrator can set up triggers that send a text message, email alert, or app notification when a mobile device enters (or exits) the specified area. The location Application Programming Interface (API) available in Google Play services is used for making an application location aware, enabling an application to provide current location tracking and activity recognition [6]. The application is also able to track the current and last location of the user, as well as spoof false geo-locations [6], [7].

Currently, smart devices have applications that restrict access to other applications on the same device. For example, Smart Lock aims to protect a user's device by employing a variety of authentication methods commonly used to unlock a smart device [8]. Some of these methods for authenticating a user include pin, pattern lock, fingerprint, and facial recognition [9]. In general, however, support for protecting applications on an individual basis is limited. The research conducted in this paper proposes adapting some of Android Smart Lock features to protect applications on an individual basis by employing a separate set of user authentication methods. The proposed application has features that allow it to detect a device's current location while also detecting the user's current motion. The application then creates a geofence around a 'safe' location as defined by the user. Geofence transitions will be triggered upon entrance or exit of the virtual perimeter, and uses different authentication methods for unlocking specific applications based on a user's current location and the device's orientation and motion [10].

Securing personal information on a smart device was initially a lengthy process, but has reduced in complexity over

time and is now much more straight forward. A user may create a less secure password that may be easy to crack, in which case they may be given a chance to create a more 'secure' password. This 'secure' password may include a combination of characters, special characters and numerical values, but at the expense of authentication simplicity. Smart Lock was released as a possible solution to this problem at the device level. The application introduced in this paper attempts to solve this problem on a per application basis, providing finer granularity with respect to application security.

The paper is structured as follows. Section 2 gives background on devices adapting to user behaviour, Android Smart Lock and the Fingerprint authentication feature offered in Android 6.0 (Marshmallow). Section 3 describes the Context Aware Mobile Application. Section 4 identifies the application's features and describes the basic usage of the Context Aware Mobile Application. Section 5 discusses the advantages and limitations of the applications and proposes some scenarios which illustrate the use of the application. Section 6 concludes the paper and discusses potential future work.

II. BACKGROUND

The following subsections provide a background on: locking options available on Android operating system, Android Smart Lock, Fingerprint authentication — a feature that is introduced on devices with Android 6.0 (Marshmallow) or later, and background on devices adapting to user's behaviour. This background is provided in order for the reader to familiarise themselves with the current trends within the field of authentication mechanisms within the smart mobile device sphere.

A. Locking on an Android device

Smart devices offer different methods for locking a device, but the most commonly used lockscreen methods are none, swipe, pin, pattern and password [9]. However, some of these methods are not available on all smart devices, with some offering only none or swipe, which does not provide the same level of security as, for example, a pin or password method. Using swipe authentication for example, a user presses the unlock button on the device and with a simple swipe across the screen, the device is unlocked [9].

Setting a lockscreen on an Android device is a relatively simple process. A user needs to access security settings on the device and select a lockscreen method. The different methods available are slide, face unlock, pattern, pin and password. It is important to note however that when none of these options are selected, the default lock screen does not require authentication to unlock the device [11]. A pin is a numerical password of 4-to-17 digits, while the password is an alphanumeric string that may contain upper and lower case characters, integers, and special characters. The lock pattern method is a 3-by-3 grid of dots where one can draw straight lines to form a pattern, without visiting a single node more than once [12], [13]. Finally, the face unlock method requires a user to look

into the front camera and align within the marked area to perform facial recognition [14].

B. Android Smart Lock

Smart Lock is an Android feature introduced on devices with Android 5.0 (Lollipop) or later [15]. This is a security feature that was designed to allow a user to bypass secure lock screens when a user is near a trusted location. Smart lock allows a user to add a trusted location, expected to be a secure environment such as the user's home, where authentication is disabled to improve ease of access to the device. Figure 1 shows a user making use of Android Smart Lock.

Depending on a user's trusted locations, a user may prefer to have quick access to the device, bypassing user authentication. For example, a user may desire a less complicated unlocking mechanism to simplify device access while jogging along a predefined morning route than they typically use in untrusted public locations. Smart Lock will detect the current orientation and current location of the device. When the current location of the device is within a known geofence perimeter, user authentication can be adapted to provide a context appropriate access method to the device.

Facial recognition (or face unlocking) is another new feature that allows a user to unlock their device through the device's onboard camera [8]. This process is not guaranteed to work perfectly in all instances, and utilises a backup password or pattern lock in case a user's face is not recognised. One more new feature is body detection, which enables a device to detect when it is being held in the hands of the user or placed in a user's pocket [16]. The device remains unlocked while in the hands of a previously authenticated user, but resets to standard authentication when the smart device is set down [12].

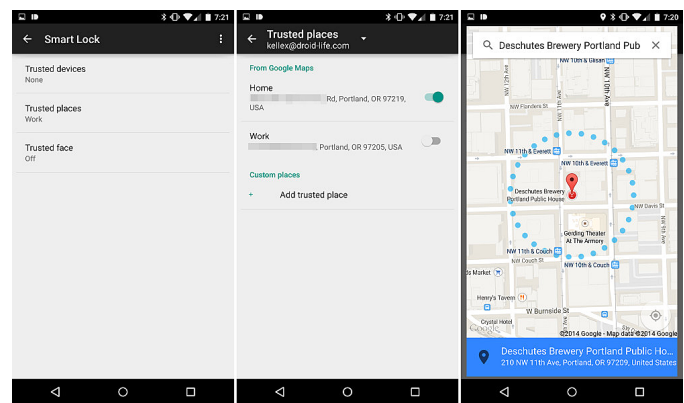


Fig. 1. Android Smart Lock

C. Fingerprint Authentication

Currently smart devices that operate using Android 6.0 (Marshmallow) have a fingerprint authentication method. The method serves as an alternative for entering a password, using a user's fingerprint as a means of authentication for gaining access to the device. A user has to create a backup password that will be used to gain access to the device if the user's

fingerprint is not recognised. The user's fingerprint is also used for identification when making purchases in some stores [17], [18].

D. Devices adapting to user's behaviour

User authentication on a device, whether a laptop, mobile device or smart watch, may be done differently depending on the device. A proposed conceptual model [19] allows devices to be part of a centralised authenticity key distributor that is used for verification of a user. Stephen Marsh has proposed a device comfort zone, where an enhanced notion of trust is enabled on a given personal device in order to better regulate the state of interaction between the device, its owner, and the environment [20]. This technology allows devices to gain each other's confidence, allowing them to share the process of user authentication with each other. The devices also secure the smart device's information, neglecting the applications on the smart devices which could be vulnerable to being accessed when all the devices are in proximity of each other. In these cases, a user's behaviour can also be used as a tool for security; for instance, information on the average times that a user accesses certain devices, as well as the current activity of the user on a given device, can contribute to the device better understanding of its user's patterns.

Applications are being developed to learn and adapt to a user's behaviour patterns, tracking the user's current location, heart rate, motion and so forth [21]. These applications are also developed to try and improve the security information on a smart device by providing different authentication methods for accessing the device [22]. Today, many users will own and use a variety of digital devices. For instance, a user may have a smart device, tablet, laptop and a smart watch. These different devices likely have different authentication methods for authenticating a user, and the smart device and tablet may even use the smart watch for user authentication, as these devices may be connected to the smart watch. This allows the devices to communicate with each other, sharing the same identity with each other until the smart watch is not within the area [23]. The devices can still authenticate a user via PIN, password, facial or voice recognition, and a laptop may also have a inbuilt fingerprint scanner for authenticating the current user.

III. CAMA APPLICATION

Context Aware Mobile Application (CAMA) is a mobile application that provides different user authentication methods based on device context. The different user authentication methods are triggered by the auto-detection of safe zones. A safe zone indicates a circular area which is parametrised by a geofence, which stores the name, longitude and latitude of the safe zone in a localised database [7], [3]. The first authentication method on the application is a 4 digit password, but different authentication methods can be customised based on the device's context (location, orientation and motion) on a per-application basis, such that the devices current context can

be used to select the most appropriate authentication model for gaining access to the application.

The authentication methods are triggered by using Google location services, which allows GPS to track the current location of the device, whether it be via mobile network, GPS or Wi-Fi [24]. When the device is in a safe zone, CAMA will not require any form of authentication to access the requested application. Authentication methods are activated once a user's current location is registered as outside of a safe zone. A saved location's latitude and longitude are used to calculate the distance between the current location of the device and any saved geofences. Locations that are not safe zones will require user authentication via a 4 digit password. When the user is outside fo the perimeter of the geofence, stronger authentication will be required to gain access to the application [25].

CAMA allows a user to register an account, create a user name, a 4 digit password, and a security question to be used for recovery of a forgotten password. Upon the successful creation of a user account, the application will check to ensure that GPS settings are turned on, so as to enable tracking of the user's current location [5]. With the use of Google location services, the current location of a user will be displayed on the provided map. The latitude and longitude of the current location are also displayed. Google time-line is also used by the application to identify areas a user visits regularly [26]. These locations are then suggested to a user as potential safe zones, allowing the user to save them as safe locations. Safe locations that have been identified by the user then have a geofence created around them with a radius of 0.1, as seen in Figure 2.

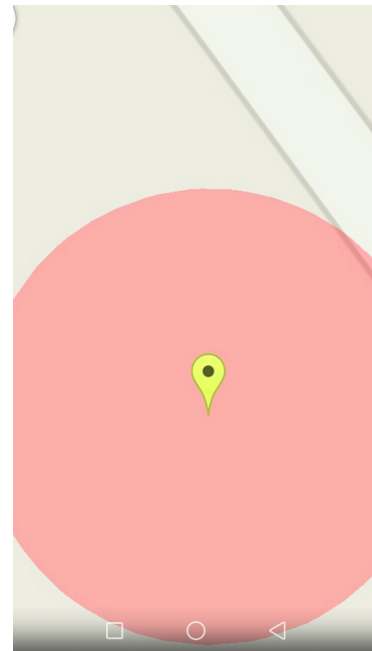


Fig. 2. Geofence with radius of 0.1

IV. CAMA USAGE AND FEATURES

A. Registration

A new user will be able to create an account. Upon successful creation of the account a user will be required to create a 4 digit password and select a question, which will be used for user authentication on the mobile application.

B. 4 digit password

A user successfully creates a 4 digit password, as seen in Figure 3. This password will be used for authenticating a user before saving a location, and will also be used to access the application when the authentication method is triggered.

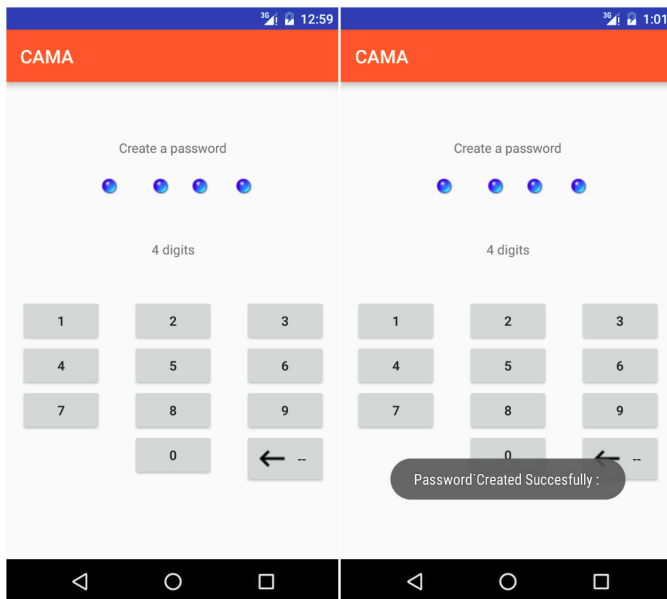


Fig. 3. 4 digit password successfully created

C. Security question

A user will be given default security questions such as:

- “Where was I born?”
- “What is my pet’s name?”
- “What is my favourite fruit?”
- “What did I meet my first love?”
- “Where do I attend church?”

The user will select one and enter an answer for the selected question. The question and answer will be used for changing a user’s forgotten password.

D. Adding a geofence

A user creates a geofence by pressing the ‘save location’ button, at which point the latitude and longitude of the location are displayed for the user. The user will then enter a location name and press the ‘save location’ button, which will create a geofence around that area. The device will then monitor the entering and exiting of the geofence, as it is now registered as a safe zone.

E. Removing a geofence

A list of saved geofences is created and displayed when a user presses the ‘list location’ button; on pressing and holding on a selected location on the list the user will be prompted to select whether the location should be removed as a safe zone.

A user’s current location is viewed and tracked on the map, with longitude and latitude of the exact current location displayed. A user may choose to save their current location as a safe zone via this application. The longitude, latitude and name of the current location are then stored. Locations that have been saved by the user are viewable on a map, as seen in Figure 4.

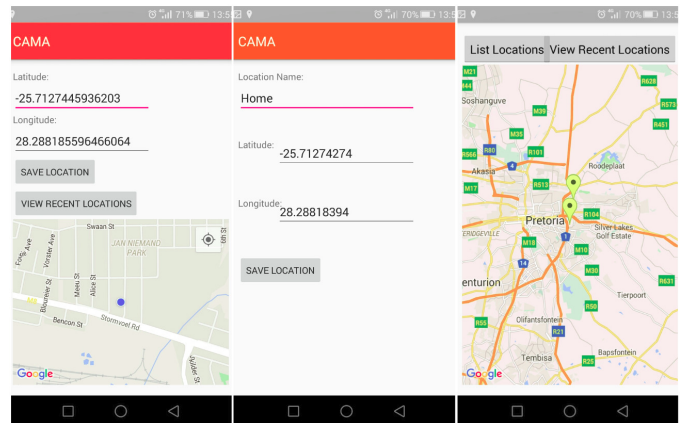


Fig. 4. Geofence creation

Figure 5 displays different use-cases of the application. The exact current location of the device is tracked and displayed with the current motion of the device. When the exact location of the current device is not within a designated safe zone, user authentication via a 4 digit password is required for accessing the application. The application will keep tracking the device’s current location against any safe zone defined by a user. The following subsections introduce the basic usage of the application.

F. Forgotten password

The question a user selected during registration will be shown, and the pre-determined answer as set by the user must then be entered. Upon the successful submission of an answer, the user will be allowed to create a new 4 digit password.

G. Within a safe zone

Upon entrance of a safe zone a user will receive a notification, indicating to the user that a geofence has been entered and no form of authentication will be required. A user will have ease of access to the application while the device remains within a safe zone.

H. Exit of a safe zone

Upon exit of a safe zone a user will receive a notification, indicating to the user that a geofence has been exited. This will trigger the 4 digit authentication method, requiring the

user to enter the 4 digit password before gaining access to the application.

I. Further distance from the safe zone

Based on a user's current location, the application will calculate the distance between the previously entered geofence, and based on that another authentication method will be triggered. The user will need to enter a user name and a password to gain access to the application.

J. Weak GPS Signal

The application will fallback to Wi-Fi and mobile network methods for determining a user's estimated current location. The security level will be a level higher.

K. Not able to determine a location

The application will act as if it is outside of a safe zone and the highest level of security will be immediately activated. This might be the case in buildings.

L. Usage of Google time-line

The CAMA application accesses a user's Google time-line and determines locations that are visited continuously by a user, and suggests these locations to a user as potential safe zones that a user may register in order to have ease of access to the application while in those frequently visited areas. Saving of these locations will be suggested in a timely manner to allow the user to make efficient decisions as to whether the suggested locations can be used as safe zones.

can therefore provide different user authentication methods for accessing device applications.

The application has some limitations. For example, notifications for entering and exiting of safe zones are not displayed in real-time, and the user must select from a list of default security questions rather than creating their own. The current authentication methods are designed for basic usage of the application, allowing for future development with a front-end user interface. The application's battery consumption is also very high, as the application currently uses fine and coarse locations for tracking the user's current location. This drains the battery as it uses Wi-Fi, GPS and mobile network communications to estimate the exact current location. The application's battery consumption could be reduced by using coarse locations to determine the location, which will allow the application to use Wi-Fi and the mobile network to estimate the user's exact location. With the application having the tracking of transitions as a service running continuously in the background, the entering and exiting of safe zones are monitored continuously. The strength in the application lies in that the application is able to successfully track, in real-time, the location of a user which allows for accurate tracking of entering and exiting of safe zones. This allows the application to switch between the different authentication methods in real-time, thus triggering the different authentication methods.

The proposed application in this paper could be further developed to provide more functionality to the user while still offering ease of access to the application. The following subsections propose scenarios which illustrate real-life use cases for the application.

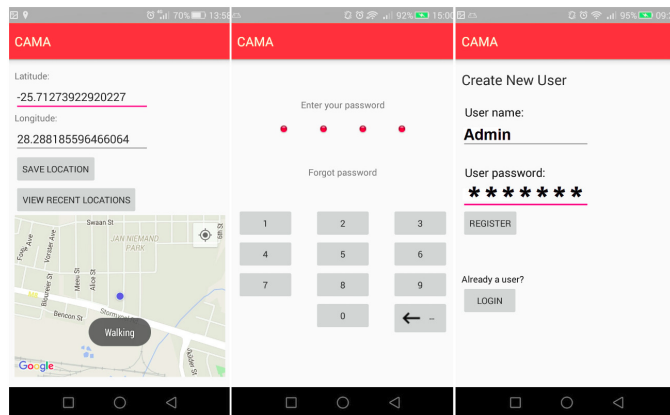


Fig. 5. Different uses cases within the CAMA application

V. DISCUSSION

Context aware mobile applications have been explored before, with some being developed to assist users with everyday tasks or with the usage of the mobile device. It may be that a small number of these applications are developed to enhance the security on a smart device, adding an extra security layer beyond those that are pre-installed on the smart device. The CAMA application proposes that an individual application can learn and understand user surroundings, and

A. Phone stolen

A user at the office may feel it necessary to secure an individual third party application for security reasons, as the application may have confidential information. If they have chosen to save the office as a safe location, in order to have ease of access to the application, they are still vulnerable to theft, and their device may be taken and accessed by an unauthorised person within the designated safe zone. However, once the device has been taken outside of the designated safe zone, in this case the office, the application will trigger different user authentication methods, locking access to the device.

B. Home

A user is at home and would like to lock his banking application for security purposes, as the user may not want his banking information to be viewed by any unauthorised person. The user may have a specific room in the house saved as a safe zone where the user would like to have ease of access to the banking application. Other rooms in the house will require user authentication before being able to access the application, as they are not safe zones.

C. Gym

A user goes for an afternoon gym session, and may wish to lock his email application for security purposes; the user will

then choose not to set the gym's location as a safe zone, and the application will require user authentication. The user may however want quick access to their email application at work and home, and so may choose to save these locations as safe zones, which will then require no user authentication.

These scenarios show that the application being proposed has capabilities beyond tracking a user's current location, providing notifications of entering and exiting safe zones, and saving locations as safe zones.

VI. CONCLUSION

Smart devices offer different user authentication methods such as slide, face unlock, pattern, pin and password. Today's Android smart devices have implemented more features to improve the security of smart devices and information it contains. User authentication methods have been implemented to ensure restriction of unauthorised users gaining access to the device. CAMA provides different authentication methods on an individual application and uses procedures that differ from Smart Lock, allowing the application to be self aware and responsible for its own user authentication methods.

The CAMA application is a context aware application that will track the user's current location, current motion and allow a user to save locations. The application makes use of the user's Google time-line and auto save locations that are recognised as consistently visited locations, and suggests locations for the user to save as safe zones. Locations that have been created have geofences around them which are used to provide different user authentication methods for gaining access to device applications. Specific applications can be individually secured and provide different user authentications methods depending on device context.

Future work will be done to store more information on the application such that a user may securely store emergency contact information, media files and documentation. Further development of the application will allow the application to provide more user authentication methods. This will allow a user to enter their own security question, as well as the implementation of other actions beyond saving locations and detecting the current motion of a user's device. For future avenues of research, it would be useful for a user to be able to integrate the authentication features of CAMA with other applications.

REFERENCES

- [1] C. W. Thompson, "Smart devices and soft controllers," *IEEE Internet Computing*, vol. 9, no. 1, pp. 82–85, Jan 2005.
- [2] L. Chamberlain. (2016, March) What is geofencing? GeoMarketing. [Online]. Available: <http://www.geomarketing.com/geomarketing-101-what-is-geofencing>
- [3] S. Rodriguez Garzon and B. Deva, "Geofencing 2.0: taking location-based notifications to the next level," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY, USA: ACM, 2014, pp. 921–932.
- [4] Garmin. (2016, April) What is gps? Garmin. [Online]. Available: <http://www8.garmin.com/aboutGPS/>
- [5] M. Avdyushkin and M. Rahman, "Secure location validation with wi-fi geo-fencing and nfc," in *Trustcom/BigDataSE/ISPA, 2015 IEEE*, vol. 1. IEEE, Aug 2015, pp. 890–896.

- [6] Google Developers. (2016, April) Using google api. Google. [Online]. Available: <http://www.developer.android.com/>
- [7] A. Popescu. (2013, October) Geolocation api specification. Google, Inc. [Online]. Available: <https://www.w3.org/TR/geolocation-API/>
- [8] Kellex. (2014, Nov) Android 5.0 feature: Google updates smart lock on lollipop to include trusted places. Droidlife. [Online]. Available: <http://www.droid-life.com/2014/11/18/android-5-0-feature-google-updates-smart-lock-o>
- [9] T. Sixta, "Gesture recognition for mobile phone unlocking," Master's thesis, Czech Technical University in Prague, Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, May 2014.
- [10] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, Mar. 2011. [Online]. Available: <http://doi.acm.org/10.1145/1964897.1964918>
- [11] R. B. Malhotra, A. A. Fulzele, and R. N. Verma, "A novel approach for android security system," *International Journal of Computer Engineering and Applications*, vol. 1, no. 1, January 2016.
- [12] G. Mazo. (2012, july) How to set up face unlock on your android phone. Androidcentral. [Online]. Available: <http://www.androidcentral.com/how-set-face-unlock-your-htc-one-x-or-evo-4g-lte>
- [13] Q. Kennemer. (2014, March) How to setup a lock-screen pattern, pin or password on your android device [android 101]. Phandroid. [Online]. Available: <http://phandroid.com/2014/03/20/android-101-lock-screen/>
- [14] S. A. A. K. Oka, P. I. K. G. Darma, and A. Arismandika, "Face recognition system on android using eigenface method," *Journal of Theoretical & Applied Information Technology*, vol. 61, no. 1, pp. 128 – 134, 2014.
- [15] N. Elenkov. (2014, dec) Dissecting lollipops smart lock. Google. [Online]. Available: <http://www.developer.android.com/>
- [16] J. Duino. (2015, aug) On-body detection explained. Androidcentral. [Online]. Available: <http://www.androidcentral.com/body-detection-explained>
- [17] B. Cha, K. Kim, and H. Na, "Random password generation of otp system using changed location and angle of fingerprint features," in *Computer and Information Technology, 2008. CIT 2008. 8th IEEE International Conference on*, July 2008, pp. 420–425.
- [18] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, "Touch me once and i know it's you!: Implicit authentication based on touch screen patterns," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, pp. 987–996. [Online]. Available: <http://doi.acm.org/10.1145/2207676.2208544>
- [19] A. Al Abdulwahid, N. Clarke, S. Furnell, and I. Stengel, "A conceptual model for federated authentication in the cloud," in *Proceedings of the 11th Australian Information Security Management Conference*. SRI Security Research Institute, Edith Cowan University, Perth, Western Australia, 2013, pp. 1 – 11.
- [20] S. Marsh, Y. Wang, S. Noël, L. Robart, and J. Stewart, "Device comfort for mobile health information accessibility," in *Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on*. IEEE, 2013, pp. 377–380.
- [21] M. Mitchell, "Context and bio-aware mobile applications," Master's thesis, Florida State University, 2011.
- [22] S. Marsh, Y. Wang, S. Nol, L. Robart, and J. Stewart, "Device comfort for mobile health information accessibility," in *Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on*, July 2013, pp. 377–380.
- [23] S. Davidson, D. Smith, C. Yang, and S. Cheah, *Smartwatch User Identification as a Means of Authentication*, Department of Computer Science and Engineering Std., 2016.
- [24] P. J. Brown, J. D. Bovey, and X. Chen, "Context-aware applications: from the laboratory to the marketplace," *IEEE Personal Communications*, vol. 4, no. 5, pp. 58–64, Oct 1997.
- [25] M. Heinel, "Android security," Master's thesis, Offenburg University of Applied Sciences, June 2015.
- [26] M. Knoll. (2016, April) Can't remember last night? google's location history can tell where you were. Trendblog. [Online]. Available: <http://trendblog.net/cant-remember-last-night-google-location-history-can-help-you/>

Team Formation in Digital Forensics

Wynand JC van Staden
UNISA School of Computing
UNISA Science Campus
Florida Park
wvs@wvs.za.net

Etienne van der Poel
UNISA School of Computing
UNISA Science Campus
Florida Park
evdpoel@unisa.ac.za

Abstract—A major problem in Digital Forensics (DF) is the often huge volumes of data that has to be searched, filtered, and indexed to discover patterns that could lead to forensic evidence. The nature of, and the process by which the data was collected also means that the data contain information about persons that are not involved, or only incidentally involved in the crime under investigation. Privacy is therefore another potential issue that needs to be dealt with in a DF investigation. This paper shows that techniques of the Team Formation (TF) task can be used to address both of these problems.

The TF task can be formulated to fit the DF arena: to commit a crime, the culprit(s) may require the assistance of several other individuals, which implies that a team of some sort gets established. During a post-mortem DF analysis, an investigator may only have one, or a few names to start with. One of the key challenges is finding possible co-conspirators. From a TF point of view, the culprit is trying to find the best team to commit the crime. This paper proposes that automated techniques could be used to discover potential teams from the data.

The TF task in DF requires the recording of skill-sets, and the generation and/or discovery of a graph depicting interaction between candidates. If the data consist of an email corpus and peoples' roles in an organisation (such as in the Enron data), both of these are available.

We consider the TF problem in general and extend it to the DF arena by considering the information that an investigator may have access to during the investigation.

Index Terms—Digital Forensics, Digital Forensic Investigation, Cyber-crime, Team-formation, Social Network Analysis, Expert Finding

I. INTRODUCTION

The post-mortem forensics analysis of communications data, such as an email corpus can be an extremely difficult and time-consuming task due to the volume and weakly structured nature of the data. The analysis process usually involves a traditional brute-force search for specific patterns, filtering to reduce the search space, and indexing of the resulting documents or parts of documents. The patterns, filters and indexing mechanisms are often hand-crafted by the investigator, usually specific to the potential crime being investigated.

Proposals use machine learning [1] and data-mining techniques [2] to guide the investigator's efforts by highlighting 'low-hanging fruit'. These techniques and tools save time and allow the investigator to more quickly find results that could lead to evidence.

The creation and formation of teams has been studied in operations research and the management and social sciences. In operations research the Team Formation (TF) problem

consists of assigning people with certain skills to a task to be accomplished, for example building a software development team. In the social sciences the TF techniques often used to do a post-hoc discovery of teams, by using individuals' communication patterns.

Crimes often involve the creation of teams, where a team would not be as rigid and designed as in the case of a software development team. Such a team is likely to be sub-optimal from a skills perspective, as there would be the additional constraint that the potential team member would have to be willing, or be able to be coerced to commit acts that would assist in the crime. There may even be unwitting team members, who participates in the crime through the simple act of doing their jobs. The TF task in the planning and execution of a crime therefore has possible additional dimensions.

This paper shows that techniques used in TF discovery can be applied to the DF task to automatically discover potential teams involved in the crime. This means that the investigator has a much smaller set of potential culprits to start investigating, using more traditional investigation techniques. It also has the benefit that the investigator does not need to look at the data of potentially innocent persons whose data happens to form part of the corpus. This has positive implications for privacy.

The TF problem is therefore considered from the perspective of the culprit(s): if they wanted to commit a crime, who would the best team be to accomplish this? The word 'team' should be considered a loose term, as the team may involve people who are simply doing their normal jobs, or may involve people, who has information required to accomplish aspects of the crime, and may or may not know that they are providing the information to aid in the commission of a crime.

Applying TF techniques can be viewed as intelligent automated filters that aim to (hopefully substantially) reduce the list of potential suspects. As in any investigation, these persons should remain 'just' suspects until further corroborating evidence is found.

To illustrate the concepts of applying TF discovery in DF, the Enron email corpus¹ was used as the data under investigation. Since the Enron data-set has undergone several releases in which data has been removed (at the request of persons whose data was within the data-set) the data provided

¹The Enron corpus was downloaded from <http://tinyurl.com/myjmcjl>

can no longer be used to identify those who were indicted, implicated, or sentenced – hence, for the moment, we cannot provide error rates or accuracy (recall and precision), however, it is important to understand that the purpose of the proposed techniques is not to provide an automated system for solving cyber-crime – the purpose is to provide tools and techniques that can guide an investigator through the investigation, and importantly, potentially protect the privacy of parties that may not be involved in the crime.

A. Contribution

This paper contributes to the field of Digital Forensics (DF) by applying techniques of the Team Formation (TF) task from a digital forensic perspective. It is argued that the TF task can be applied during a post-mortem analysis of seized data to guide the investigator, by narrowing down the list of suspects, focusing on persons of immediate interest, and avoiding investigating potentially innocent persons. To facilitate the use of TF, however, the team formation task has to be placed in the correct context.

In general, TF considers social network graphs and potential team members' skills and expertise to build a team to complete a specific task. The important difference between this work and others is that the team formation problem is framed in the DF paradigm, specifically with the focus on guiding the investigator during the analysis.

It is shown that standard Information Retrieval (IR) techniques can be employed to extract information from an email corpus, that can lead to identifying teams. The formulation of the TF task in the DF paradigm will allow further research into automation of the guidance provided to the investigator. A formal notation for the TF task is also proposed. This notation can be used when reasoning about the team formation problem in this and future research.

Additionally, by allowing the investigator to focus specifically on persons of interest (i.e those in the team), the privacy of others whose data forms part of the seized data may be protected.

B. Structure of the paper

The rest of the paper is structured as follows:

- Section II provides background information on DF, the TF task and related work.
- Section III frames the TF task in the DF paradigm, and provides formal definitions for ranking individuals.
- Section IV provides some examples of the application of the ideas presented in the paper to the Enron mail corpus.
- Finally, section V provides concluding remarks.

II. BACKGROUND AND RELATED WORK

Reformulation of the team formation problem concerns itself with two important pieces of work. Firstly, DF provides the paradigm within which the problem is contextualised, secondly, the team formation problem provides the concepts and tools needed to reformulate and understand the problem. Each of these is discussed in turn in the following sections.

A. Digital Forensics

Digital Forensics (DF) is defined as the “...preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of evidence in a digital context [3].” Using sound forensic techniques and proper controls digital data that could potentially be evidence is gathered, analysed and presented in context as part of the cyber-crime investigation. Politt [4] calls this the creation of a narrative.

This paper is concerned with digital evidence in the form of data. In particular, the post-mortem analysis (as opposed to live analysis) of de-obfuscated data. Since data can be hidden, a lot of DF research goes into the finding and identification of data. These techniques involve file-carving to find deleted data [5], [6], similarity hashes to identify files or parts of files [7], [8], to name but two². Once data has been de-obfuscated, that is, their meaning can be readily inferred, an analysis on the content can be done which will contribute to the narrative.

The analysis of the data can also be seen as a de-obfuscating effort (since data is now added to the narrative, and therefore its meaning in the narrative becomes clear). However, this paper will stick to the term analysis in order to avoid confusion.

Sifting through large volumes of data is typically accomplished through brute force approaches in which strings of data are matched against search queries, or where meta-data is matched against search queries. Such meta data consists of file-types, time-stamps, file-ownership and so on. Fei et al. [1] propose the use of Self-Organising Maps (SOMs) [9] to guide the investigator. Their technique uses meta-data to detect anomalies in the data, and the investigator is thus guided by focussing analysis on those pieces of data.

Beebe has proposed the use of text-mining to achieve better retrieval rates [10] and as a way to search through large corpora [2], and Pollitt has shown that Natural Language Processing (NLP) techniques such as Named Entity Extraction (NEE) can be useful during the creation of the narrative [4].

The use of automated guidance during a forensic investigation is therefore well established, and this paper builds on those ideas.

B. Expert finding and Team Formation

Finding experts is the problem of identifying individuals who may hold knowledge. This particular problem dates back as far as the 1990s [11], and the particular challenge set by the text-retrieval conference (TREC) in 2005 set the scene for renewed research in the field [12].

The particular problem in expert finding is estimating the expertise of an individual. Most notable approaches [11], [13] use a probability distribution model in order to estimate the expertise level. Zhang et al. [14] proposes a propagation based approach to finding an expert within a social network.

The use of social graphs to find criminal associations has been studied by Xu et al [15]. They use shortest-path algorithms to identify associations in criminal networks. However,

²The decryption of data is also, of course, part of the de-obfuscation problem.

their evaluation is run purely on the associativity of the links in the network.

Once an expert is found, a social graph is typically used to establish a team of experts within the graph. Team formation is a well researched problem outside digital forensics. Lappas et al [16] make use of minimum-span trees to build a team of experts on topics within a social graph. They show that constructing such a structure is NP-Hard.

Rangapuram et al. [17] extend team formation as presented by Lappas et al to include budget and location constraints. They also allow an upper bound on the team size, and well as a constraint to indicate the minimum level of expertise required to complete the task the team is identified for.

Rahman [18] considers the team formation problem from an economic perspective, and the concept of opaque and translucent teams are introduced. An opaque team shares knowledge within the team in order to maximise the operation of the team. In a translucent team, some information may purposefully remain hidden in order to enhance the attractiveness of the team. Such translucent teams, although not part of this paper, may provide an interesting topic of study once the team formation problem in the DF sphere is well defined.

The following section formulates the TF task in DF.

III. THE TEAM FORMATION PROBLEM IN DIGITAL FORENSICS

Generally speaking, a (cyber-)criminal contemplating a crime has the same problem as a project manager: find a team that will successfully complete a project. The project requires a specific set of skills and/or knowledge related to the task. A project manager aims to find the best group of experts that the budget will afford. All the team members will have full knowledge of their role in the team. On the other hand, the criminal has a more complex notion of ‘afford’, in that the criminal should be able to convince or influence potential members to commit parts of the crime. This means that the team may well not consist of the ‘best’ experts. They are also likely to be team ‘members’ who are not aware of their role in the crime, or even be aware that a crime is being committed, through the simple execution of their jobs, or sharing of their knowledge. We define ‘aid’ as either the execution of a specific task, such as a job function, or the sharing of specific knowledge to assist in the execution of specific tasks.

The team formation problem is therefore formulated for DF investigations, as follows:

Definition I The Team Formation Problem in a Digital Forensics Context

Given a set of individuals Ψ , a set of topics they have knowledge about Θ , a graph depicting their communication habits $G = \langle V, E \rangle$, (where V is a set of vertices representing the individuals and E is a set representing the edge between the vertices from V) and a topical definition of a committed act, find $\Gamma \subset \Psi$ which depicts a likely team needed to either commit the act, or who will be able to provide aid in order for the act to be committed. ■

A formal definition of the notation in formulating the team formation problem in the DF context is provided in definition III-A.

It is important to understand the notion of a ‘likely’ team. The suspect may not have looked for the most influential people, or all the experts in order to commit a crime, any person who has the knowledge or can lead to knowledge may be sufficient. In particular the criminal may have had individuals in mind who had knowledge, and whom he would be able to influence.

This leads to a paradox in the existing definitions of team formation: teams may not consist of the best choices, and may more than likely resemble *translucent* teams [18] in which the criminal and co-conspirators hold a residual claim on the team. This paradox is defined as follows.

Definition II The Team Formation Problem Paradox

In order to accomplish the task at hand, the cyber-criminal’s choice in team may not consist of the experts, or seats of power in the organisation. Normal team formation analysis techniques rely on building a team from influential people or experts, meaning traditional team formation analysis techniques may be of limited use in this case.

Additionally, the suspect may not be part of the team produced during a traditional TF analysis. ■

This does not mean that traditional team formation analysis techniques are useless. Since traditional team formation coupled with Social Network Analysis (SNA) provides valuable information on the potential team that could be formed, they can act as a good guide during the investigative process.

The team formation problem as defined above therefore requires de-obfuscated data from which the following can be derived: a social graph for the persons under investigation, topics extracted from the data, and a framing of the act in terms of the topics. This last concept is important, since the investigator must have enough knowledge of the domain being investigated in order to frame the act in terms of the topics, which leads to the following definition of the act or crime.

Definition III The Crime as a Task

In the team formation problem for cyber-crime an *act*, is a task that can be defined based on knowledge that is required to complete it. Knowledge can be encoded into language phrases, of which several can be used to define the act. ■

Based on the above requirements, the team formation problem is considered with respect to seized email data. The choice of using email data aids in:

- 1) Constructing a social graph from the email data can be easily automated.
- 2) Extracting topics from the data can be approximated by performing noun-phrase-, and named entity extraction. Moreover, general IR techniques allows the easy indexing of large email corpora.
- 3) The terms used to define the act will correspond to the extracted terms and can thus be used during the guided investigation.

The following section considers the the examination of email data.

A. Examining Email Corpora

Given the team formation problem as defined in Definition III, this section considers the identification of what is termed a *candidate team*. This is a team that consists of all the individuals that could potentially form part of an *ideal team*. An ideal team is a team that may have fit the requirements of the suspect.

The Aardvark social search engine [13] attempted to find individuals that may have been able to answer questions from other individuals. It did so by determining the likelihood that a particular individual would be able to answer a question on a certain topic. Aardvark uses NLP techniques, as well as crafted profiles to build its model of users and their ability to answer question on particular topics.

The paper builds on this idea, by showing that an easy approximation for topics, and the social network of the individuals can be used to build a likely team (Definition III) for committing the crime.

To accomplish this the following is to be done prior to the analysis phase:

- 1) Create an index on topics for the corpus,
- 2) Create a communications network for the users of the mail system,
- 3) Define the act using nomenclature from the enterprise context,
- 4) Generate a sub-graph depicting the individuals involved in communication about the topic,
- 5) Use the sub-graph as a basis for further analysis and investigation.

The set of topics each team member is knowledgeable on is derived through IR techniques from the seized email corpus S .

For any corpus S , the following is defined for the team formation problem in cyber-security:

Definition IV Team formation problem notation

- 1) Θ represents all topics embedded in S ,
- 2) $\theta \in \Theta$ is the set of all topics that forms part of a search on S .
- 3) Ψ represents all the individuals within the corpus,
- 4) δ_u represents all the documents directly related to individual $u \in \Psi$. Directly related means that this individual has a copy of this document in their possession.
- 5) $\psi \subseteq \Psi$ is the set of individuals who are under consideration. It may be that certain individuals are excluded from the investigation from the start, therefore, although S may be about Ψ , only the set ψ is under consideration. As the investigation progresses more individuals may be added to Ψ and removed from ψ (or vice versa).
- 6) δ_u^t is the set of all documents for user u on topic $t \in \Theta$
- 7) $util(u)$ is a utility rating for u .
- 8) $G = \langle V, E \rangle$ is the social graph depicting the interaction between all $u \in \Psi$, with $V \subseteq \Psi$ and $E = \{(u_k, u_j) | u_k, u_j \in V\}$

For every individual in S , it is clear that their share of the mail will be a representation of the set of topics they deal with on a daily basis. Having no other information, it is reasonable to assume that this is a reflection of their knowledge on different topics. Consider for example the employee that spends ninety percent of their time corresponding about new contracts. It is reasonable to assume that they have knowledge on contracts and at least some of the process around them. The utility of this individual to the team is thus a function of the probability distribution given for the user given that topic t is discussed.

$$util(u) = p(u_i|t) \quad (1)$$

The utility function is purposefully provided as a function that could be used as part of an objective function calculation. Since 1 can be changed to represent specific constraints. As it stands, equation 1, assumes a steady state – that is, no new information as it becomes available during the investigation is considered. Consider for example a deposition which reveals beyond doubt that a particular individual had knowledge pertinent to the investigation. Thus, the utility function could be modified to reflect this, and the selection of *candidate team* would change.

Searching for the topic $\alpha \in \Theta$, the result corpus $s \in S$ will contain emails exchanged by individuals within the enterprise. Depending on the nature of the topic, the likelihood of an individual u_i corresponding (either receiving or sending an email) on the particular topic is (using Bayes' theorem): $p(u_i|t) = \frac{p(t|u_i)p(u_i)}{p(t)}$.

Since S is available as the sample space, it is easy to calculate $p(t|u_i)p(u_i) = p(u_i \cap t)$. Which in turn is calculated as in equation 2.

$$p(u_i \cap t) = \frac{|\Delta_u^t|}{|S|} \quad (2)$$

Here δ_u^t is the set of all documents covering topic t from individual u (as defined in III-A), and $|S|$ is the size of the entire corpus.

Individuals can now be ranked based on the utility they could potentially add to the team (since $\sum_{u_i \in \Psi} p(u_i|t) = 1$).

Based on the utility rank and the search result, it is possible to construct $G' = \langle V', E' \rangle$ where $G' \subseteq G$, with the constraint that $V' \subseteq V$. G' is thus a sub-graph of G which depicts only the correspondence on topics t . From the investigator's view point, G' presents the *candidate team* for aiding in a crime that requires knowledge on the subjects that will come from the individuals in the graph.

The resulting *candidate team* graph G' can then be used in well known social network techniques such as *centrality*, *span-tree's* to determine teams, and dense sub-graphs. However, at this point, the investigator can simply use the G' to guide the analysis of particular emails that could be evidence.

Now that the concepts behind the team formation problem have been articulated, the following section provides some initial samples in using the generation of G' on the Enron email corpus.

IV. EXPERIMENTAL RESULTS

In 2001, the Enron energy company was embroiled in a scandal relating to unlawful and unethical financial practices. Enron basically used complex financial techniques in order to hide their losses, thereby artificially boosting the company's stock value. During the investigation, the email of several hundred of the key employees in Enron was seized and analysed.

Subsequently, the corpus was purchased and released by Andrew McCallum who prepared the content and released the emails in a folder-based hierarchy, all in mbox (RFC4155) format [19]. Petitions by several individuals resulted in their emails being removed from the corpus, and the result is a corpus of one-hundred and fifty individuals spanning around 517,000 emails.

There has been a lot of research done on the corpus, including data mining, social network analysis based on the communication links between individuals, and so on. The ideas presented here are (as far as the authors are aware) the first examination of a team formation problem on the Enron corpus – specifically with the team formation problem framed in the DF context.

The purpose of the experiment for this paper was to consider the team formation problem on a real-world set of data. It is shown that very simple techniques can go a long way in providing guidance to the investigator when sifting through volumes of data.

The experiment was conducted based on the steps outlined in section III-A:

- 1) The entire email corpus (that was made available) was indexed, and an inverted index was created. This resulted in around 780,000 unique search terms for the 517424 emails all stored in RFC822 *mbox* format.
- 2) For the communications network (or social network graph) of the persons involved.
- 3) Several key phrases representing 'topics' were used to search the corpus (thus describing the act in terms of knowledge needed to commit or to aid in committing),
- 4) A sub-graph of the individuals who communicated about the topics was created, and merged into a graph that represents a *candidate team* for the act.

Some more comments on the techniques used are in order. The term dictionary constructed from the corpus contains terms stemmed using the Porter stemmer, and queries run against the term database are stemmed before the search is done. The social network graph for the employees consists of the interaction between Enron employees based on their in-box and sent mail folders.

Although the graph consists of all persons interacting based on the information from the mentioned sources, the visual graphs presented are restricted in two ways: firstly, only individuals from within Enron are displayed on the visualisation, and secondly, based on the likelihood calculation presented in equation 1, only a limited number of individuals are included in the graph. Both of these reasons are purely for a ease

of viewing consideration: a visual graph depicting too many vertexes and their links quickly degrades in readability and thus meaning (in printed format). It was thus decided to limit the number of nodes to something that would be meaningful and would be easily digestible.

Figures 1 (page 5) and 2 (page 5) represent a constrained sub-graph for the topics 'regulation' and 'service provider' (both provide the *utility value* for each individual in parenthesis).

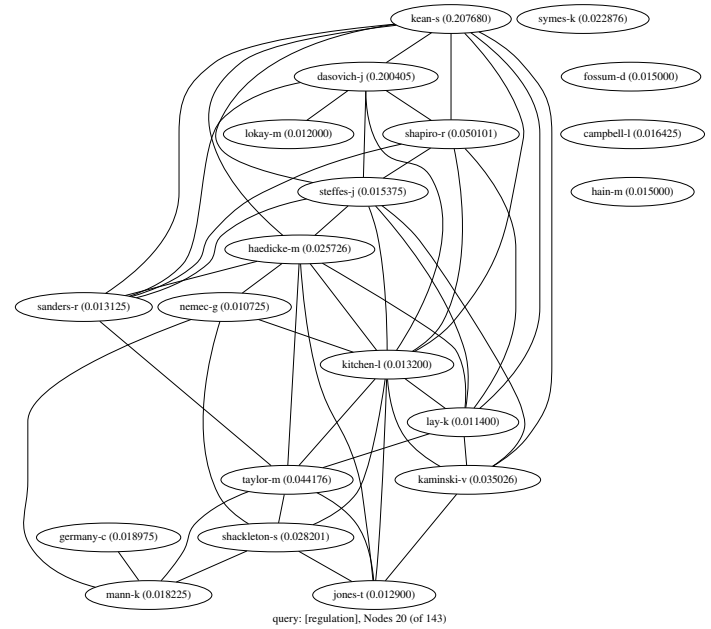


Fig. 1. Candidate Team for topic 'regulation'



Fig. 2. Candidate Team for topic 'service provider'

Figure 1 shows several vertexes that are disconnected – this revealed individuals who were corresponding about 'regulation' but likely not with parties in Enron.

Lack of space prevents the presentation of all the sub-graphs, however, the *candidate team* graph which includes the topics presented above is provided in 3. The following 'topics' were used for the generation: "Federal Energy Regulatory Commission", "Regulation", "Audit", "Contract", and "Service Provider".

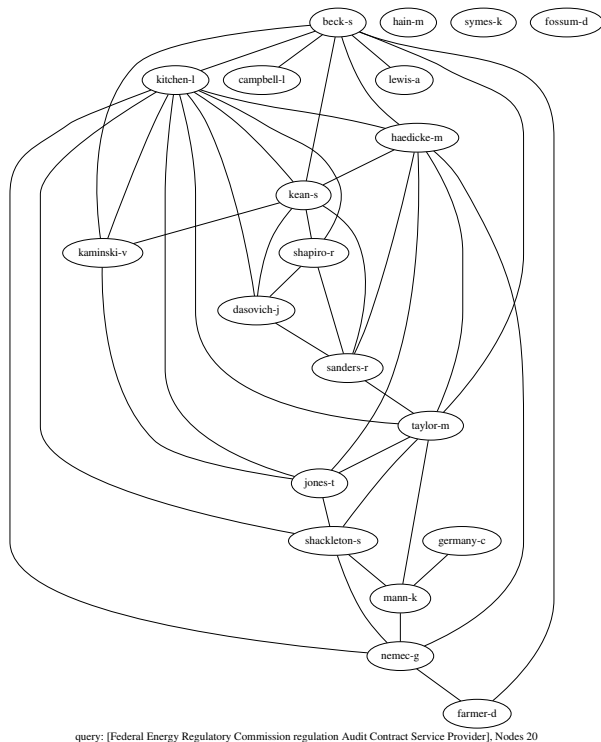


Fig. 3. Sub-graph for candidate team for query "Federal Energy Regulatory Commission", "Regulation", "Contract", and "Service Provider"

Just visual inspection of these graphs already provide good clues as to who the individuals with potential knowledge to help with the act are. Knowledge of the structure of the organisation would enable the investigator to follow potential leads – thus the sub-graph can provide guided investigation.

V. CONCLUSION

This paper reformulated the *team formation problem* within the DF paradigm. Since the team formulation problem is well defined outside of the DF paradigm, it is necessary to place it within the DF context in order to understand it properly. This allows the finer nuances and requirements dictated by the DF paradigm to be understood. In turn, this allows future work to aim specifically at solving particular problems in light of the reformulation. In addition, the team formation problem allows the investigator to be guided by the data within the system. It is important to understand that the proposed techniques should not be considered to be an automated system for solving a cyber-crime, these techniques should only act as a guide for the investigator.

The team formation problem is thus considered from the suspect's point of view: a crime is defined with respect to

topics that are covered by the individuals in the organisation. The team formation problem then identifies the *candidate team* which would likely be able to complete the task (i.e. commit the crime).

This *candidate team* provides the investigator with clues about the individuals within the organisation that may have formed part of the team, or those that may have been used by the suspect in order to complete his task. The important contribution is that the investigator is provided with a guided approach to investigate a large volume of data, thereby focussing the investigation. Additionally, there is an important benefit for privacy of third parties (persons whose emails form part of the seized corpus, but who have nothing to do with the act under investigation). There will be important implications for the investigator and investigation techniques, and further investigation here is also warranted.

The paper also defined formal notations and definitions as the starting point for reasoning and arguing about the team formation problem in the digital forensics perspective. This formal notation can be used as a foundation for future research in this paradigm.

Now that the team formation problem has been formulated for the DF paradigm, it becomes possible to define some future areas of research. These include: using NLP for better topic extraction, such as noun-phrases, or named entities. Once these have been extracted, the investigator can be presented with these 'topics' as a search filter. Such an approach would mean the investigator no longer needs to carefully craft the search terms, but can rely on the automated system.

Future work would also include comparing the results from the techniques proposed herein to regular social network analysis techniques.

Rahman introduced the concept of translucent team [18] in which a team has members that may withhold information from other team members. The effect of such a team within DF would be important to understand, since a cyber-criminal may employ such a team in order to commit a crime – thereby keeping knowledge of the crime away from those who may be able provide evidence.

REFERENCES

- [1] B. K. L. Fei, J. H. P. Eloff, M. S. Olivier, and H. S. Venter, "The use of self-organising maps for anomalous behaviour detection in a digital investigation." *Forensic Sci. Int.*, vol. 162, no. 1-3, pp. 33–7, 2006.
- [2] N. Beebe and J. Clark, "Dealing with Terabyte Data Sets in Digital Investigations," in *Advances in Digital Forensics*. Springer US, 2005, vol. 194, ch. IFIP — The International Federation for Information Processing, pp. 3–16.
- [3] G. Palmer, "A Road Map for Digital Forensic Research," DFRWS, Utica, NY, Tech. Rep., 2001.
- [4] M. Pollitt, "History, Histiography, and the Hermeneutics of the Hard Drive," in *Advances in Digital Forensics IX*, G. Peterson and S. Shenoi, Eds. Seneca, SC, USA: Springer, 2013, pp. 3–19.
- [5] N. Alherbawi, Z. Shukhur, and R. Sulaiman, "Systematic Literature Review on Data Carving in Digital Forensics," *Procedia Technology*, vol. 11, pp. 86–92, 2013.
- [6] A. Pal and N. Memon, "The Evolution of File Carving," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 59–71, 2009.
- [7] J. Kornblum, "Identifying almost identical files using context triggered piecewise hashing," *Digital Investigation*, vol. 3S, pp. 91–97, 2006.

- [8] V. Roussev, "An evaluation of forensic similarity hashes," *Digital Investigation*, vol. 8, pp. S43–S41, 2011.
- [9] T. Kohonen, "The Self Organising Map," in *IEEE*. IEEE, 1990, pp. 1464–1480.
- [10] N. L. Beebe and J. G. Clark, "Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results," *Digital investigation*, vol. 4, pp. 49–54, 2007.
- [11] K. Balog, "People Search in the Enterprise," Ph.D. dissertation, University of Amsterdam, 2008.
- [12] A. Bozzon, M. Brambilla, S. Ceri, M. Silvestri, and G. Vesci, "Choosing the Right Crowd: Expert Finding in Social Networks," in *Proceedings of EDBT/CDT '13*, ser. EDBT '13. New York, NY, USA: ACM, 2013, pp. 637–648. [Online]. Available: <http://doi.acm.org/10.1145/2452376.2452451>
- [13] D. Horowitz and S. D. Kamvar, "The Anatomy of a Large-scale Social Search Engine," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 431–440. [Online]. Available: <http://doi.acm.org/10.1145/1772690.1772735>
- [14] J. Zhang, J. Tang, and J. Li, "Expert finding in a social networks," in *Database Systems for Advanced Applications (DASFAA'2007)*, 2007.
- [15] J. J. Xu and H. Chen, "Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks." *Decision Support Systems*, vol. 38, pp. 473–487, 2004.
- [16] T. Lappas, K. Liu, and E. Terzi, "Finding a Team of Experts in Social Networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 467–476. [Online]. Available: <http://doi.acm.org/10.1145/1557019.1557074>
- [17] S. S. Rangapuram, T. Bühler, and M. Hein, "Towards realistic team formation in social networks based on densest subgraphs," in *WWW 2013*. ACM, 2013, pp. 1077–1088. [Online]. Available: <http://dblp.org/db/conf/www/www2013.html#RangapuramBH13>
- [18] D. M. Rahman, "Team Formation and Organization," Ph.D. dissertation, University of California Los Angeles, 2005.
- [19] E. A. Hall, "The application/mbox Media Type," Electronically, September 2005, <http://datatracker.ietf.org/doc/rfc4155/>.

Social Network Phishing: Becoming Habituated to Clicks and Ignorant to Threats?

Edwin D. Frauenstein¹ and Stephen V. Flowerday²

Department of Information Systems

University of Fort Hare

East London, South Africa

edwin.frauenstein@gmail.com¹, sflowerday@ufh.ac.za²

Abstract — With the rise in number of reported phishing cases in statistical reports and online news, it is apparent that the threat of phishing is not retreating. Phishers continuously seek new methods to deceive individuals into sharing their confidential information. As a result, today the traditional form of conducting phishing solely through email and spoofed websites has evolved. Social network phishing is a serious threat as it reaches a far wider audience, consequently affecting both business and private individuals. This paper argues that due to the constant updates of information users are engaged in on social networking sites, users may become habituated to clicking and sharing links, liking posts, copying and pasting messages, and uploading and downloading media content, thus resulting in information overload. This behavioral priming leads users to becoming more susceptible to social engineering attacks on social networks as they do not cognitively process messages with a security lens. This paper introduces social network phishing and briefly discusses activities users engage in on social networks sites, thus highlighting the formation of “bad” habits. Further, existing information processing models applicable to this context are discussed.

Keywords—social network phishing; social media phishing; phishing; social engineering; habits; information processing; heuristic processing; systematic processing

I. INTRODUCTION

With approximately two billion Internet users worldwide using social networking sites (SNSs) today [1], it is rare not to find individuals active on at least one social network (SN). The introduction of Web 2.0 technologies has given rise to widely popular SNSs such as Facebook, Twitter, LinkedIn, MySpace, Pinterest, Google Plus+, Tumblr, Snapchat, Instagram and Flickr. Facebook is the most popular SN and according to Facebook Corporation, it is used worldwide by approximately 1.55 billion monthly active users, increasing by 14% each year [1]. As of September 2015, an estimated 1.01 billion people log onto Facebook daily. Every minute on Facebook, 510 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded. Facebook Messenger, accessed mostly through smartphones, is used by 800 million users. Other SNSs have staggering figures too: WhatsApp is used by approximately 1 billion users; 400 million use LinkedIn; 307 million Twitter users, and

Instagram with 200 million [1]. These figures constantly rise and will soon be out-of-date.

Despite users having different levels of computer experience, backgrounds, cultures, race and gender, it is apparent that SNSs are not restricted to any particular type of user. Given the statistics, it should not be unexpected that SNSs present an opportunistic market for information security threat agents such as phishers.

One of the easiest ways of acquiring individuals’ information is through the popular SN Facebook. By having mutual friends, people can access a user’s profile. If the user concealed their information through privacy settings, an option would be to send them a friend request. Alternatively, for more specific information such as educational background and work history, other SNSs such as LinkedIn can be searched. If this fails, search for names using a search engine or lure them to open malicious links. This is the connected world we live in today. Information is not as private as one may perceive and this particular means of acquiring information and befriending strangers can be performed by any person, including phishers.

With the popularity of SNSs increasing and its extension to smartphone applications (apps), users may be subjecting themselves to a wider degree of security threat agents than anticipated. The traditional method of conducting phishing mainly through emails and spoofed websites has progressed to social platforms whereby it can infiltrate into organisation networks [2]. Since SNSs are widely popular, have an extensive number of users with diverse backgrounds, and encourage sharing of personal information, phishers use this as an ideal opportunity to gain confidential information, often made openly available by members of these sites. This information could then be used to conduct more targeted forms of phishing attacks (i.e. spear phishing, whaling and mishing) both on and off SNSs. Furthermore, phishers target users’ poor privacy habits or exploit their online behaviour by enticing them to click on links that is of interest to them.

The objective of the paper is to highlight social network phishing and related threats, SN habits, information processing models and its implications thereof to users.

II. SOCIAL ENGINEERING, PHISHING AND SOCIAL NETWORK PHISHING

A. Social Engineering Explained

Phishing is effective because it uses social engineering (SE) techniques to influence people into performing certain actions that will benefit the phisher. Reference [3] defines SE as using “[i]nfluence and persuasion to deceive people by convincing them that the social engineer is someone he is not, or by manipulation”. To persuade users, phishers make use of SE techniques that focus on prompting human emotions [4] such as greed, fear, heroism and desire to be liked. In general, people desire to obey authority such as a bank official or policeman. As such, scams use authoritative words or imitate organisations and authoritative persons in order to initiate a response. Another technique, typically used in traditional marketing, is making an opportunity seem scarce, or making the victim feel they have made a commitment by responding to the scam offer. Table 1 below by [5] presents a taxonomy of SE persuasion techniques with a comparison of persuasion principles by three key authors in this area.

TABLE I. PRINCIPLES OF PERSUASION IN SOCIAL ENGINEERING [5]

Principles of Influence [6]	Psychological Triggers [7]	Principles of Scams [8]
Authority	Authority	Social Compliance
Social Proof	Diffusion Responsibility	Herd
Linking and Similarity	Deceptive Relationship	Deception
Commitment and Consistency	Integrity and Consistency	Dishonesty
Scarcity	Overloading	Time
Reciprocation	Reciprocation	Need & Greed
	Strong Affect	Distraction

From the table above, it is evident that there are common techniques overlapping in each of the principles used by social engineers (e.g. authority, reciprocation). This is necessary to persuade users into performing actions instructed by the social engineer. Some of these techniques and principles are applied in other forms of SE attacks such as baiting, pretexting, ransomware and phishing.

B. Phishing Explained

Definitions of phishing constantly change, especially since phishers seek new practices to carry out their attacks. The Anti-Phishing Working Group define phishing as “a criminal mechanism employing both social engineering and technical subterfuge to steal consumers' personal identity data and financial account credentials” [9]. Reference [10] define it as “a form of social engineering in which an attacker, also known as a phisher, attempts to fraudulently retrieve legitimate users' confidential or sensitive credentials by mimicking electronic communications from a trustworthy or public organization in an automated fashion.” Using phishing definitions from 2458 publications, [11] defines phishing as “a scalable act of deception whereby impersonation is used to obtain information from a target.” As noticed from all the definitions, the specific channel(s) used by phishers to exploit attacks is

not mentioned. This is not unexpected given that today phishers continue to use a variety of methods to conduct a phishing attack and these methods constantly change too. As a result, this presents challenges to educate users effectively to identify new techniques that aim to scam them [4]. More especially since phishing is designed to focus on exploiting human weaknesses, in particular cognitive biases, instead of technology vulnerabilities [12].

For many years, phishers typically use email messages and spoofed websites designed to appear as if they originate from a recognised and trusted source/authority (e.g. financial institution). By imitating as a legitimate source, phishers gain the victim's trust who then carry out the actions instructed by the phisher. The phisher would gain confidential information which can be used by the phisher or sold to other illegal entities. Confidential information is usually login information such as usernames and passwords. However, more sensitive information is also sought after such as identification/social security numbers and credit card details.

Phishing is regarded as a socio-technical attack. The “social” aspect uses SE techniques, as seen in Table 1, to convince users into performing actions which in turn benefits the phisher. Timing is also an important factor as phishers will take advantage of events such as religious festivities, holidays and tax season. This establishes urgency on the part of the victim as they may believe it is an opportunity that is available for a limited time period only. As depicted in Table 1, this preys on SE techniques of scarcity, overloading, distraction, need and greed.

Typically, the phisher directs an email to the victim expressing a fabricated event in the message. The message can take the approach of alerting the victim of an imminent threat or danger. For example, the victim may be warned that his/her bank account may be “hacked” as the organisation being imitated has been experiencing fraudulent activity such as a security breach. Alternatively, the message may convince the victim that they have won a substantial reward or prize. In both cases this is regarded as the “bait”, and would require the user to open an attachment or click on a hyperlink for verification purposes. If the user clicks on the hyperlink, they are subsequently directed to the spoofed website that appears identical in design to the genuine website of the institution being imitated. This is seen as the “hook”. They then unsuspectingly log-in to the spoofed website with their personal account information. This is known as the “catch”. For this to be effective, it must convince the victim and establish trust. This can be accomplished by the strength of the message arguments and the authority of where the email purportedly originates from such as a recognised or reputable organisation. In this regard, institutional logos or branding are used within the email thereby convincing victims of its authenticity. To further convince the user to comply with the request, the phisher might add an element of fear in the message. For example, the email may state that should the victim decide to ignore the request, it may result in their account or membership being suspended or terminated in a certain time frame. This preys on SE techniques of authority, distraction and time. The addition of fear may increase the likelihood of the victim following through with the phisher's

request. From these instructions provided in the email, the user “thinks” they have a choice to decide on which action to take; however, ultimately there are no correct choices as the instructions supplied are false to begin with. From this discussion, much importance is placed on email design to carry out phishing attacks successfully.

C. Social Network Phishing Explained

Phishers make use of SNSs to carry out attacks on their victims. Currently, there is no common accepted definition for social network phishing as the terms “social media phishing”, “social phishing” and “social networking phishing” are used interchangeably in literature. The common element is that SE techniques are used to conduct attacks in SN environments. Security experts of the company Proofpoint, determined in the past year that the number of phishing attempts on popular SNSs have increased by 150% [13]. According to [14], 22% of phishing scams on the web target Facebook (see Fig.1), and phishing sites imitating SN websites consist of more than 35% of all cases whereby Kaspersky anti-virus software was triggered.

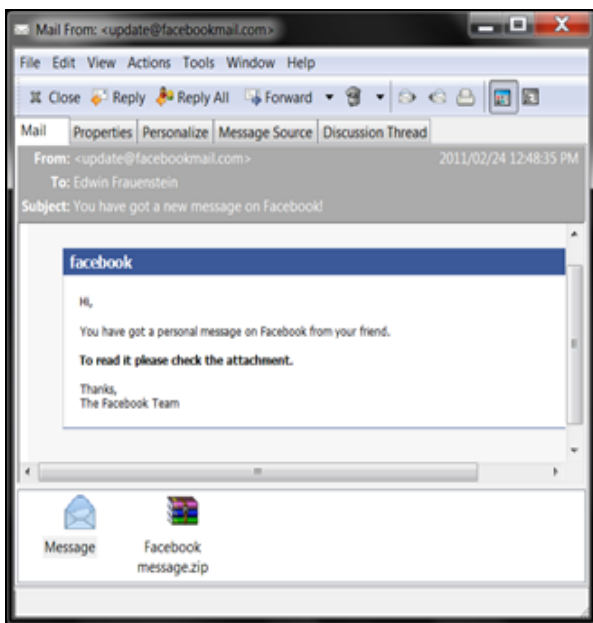


Fig. 1. Phishing email purportedly originating from Facebook

Users of SNSs like Facebook or Twitter have a greater risk of being targeted for SE attacks because of the vast amount personal information people openly share about themselves [15]. Furthermore, it is easy for phishers to impersonate a friend of the victim to gain their trust. This can lead to more targeted forms of attacks such as spear phishing and clickjacking, which are discussed in the next section.

III. SOCIAL NETWORK PHISHING TECHNIQUES

This section reveals that SNSs are a playground to conduct various forms of phishing attacks. The SE techniques used frequently in phishing emails is also employed in a SN environment. Since Facebook is the most used SN, one can expect it to be the prime target for phishing attacks. Scams on

Facebook include cross-site scripting, clickjacking, survey scams and identity theft [16]. Ways in which these scams can be carried out can be in the form of fake comments, fake media content, or fake promotion discounts (see Fig. 2).

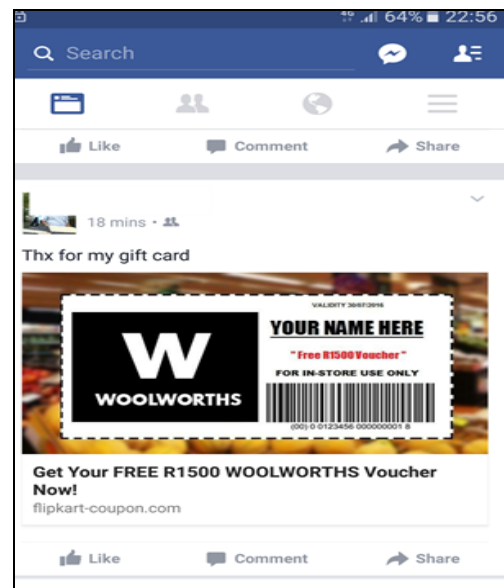


Fig. 2. Fake free shopping voucher found on Facebook

Scams such as the one depicted in Fig. 2 prey on SE techniques of authority, scarcity, need and greed. Trust can be further enhanced if these fake vouchers are shared by trusted friends.

A. Spoofing

Much like standard phishing, SN users are enticed to click on a link which subsequently directs them to a fake webpage to log-in. Victims may have been enticed through a message originating from a hijacked friend’s account, malware infected links or attachments, or a phishing email with a link to log-in to a spoofed SN webpage. Preying on the authority SE technique, imagine the strength of phishing if the phisher impersonates a celebrity. The fake profile would then include a network of bogus friends associated with that particular fake account. This is used to persuade the victim into believing that the account is genuine because a profile consisting of few friends may be suspicious to the victim.

B. Identity Theft (Cloning)

Cloning is when a phisher creates a SN account imitating the victim’s account and is not regarded as hacking because the victim’s account was not compromised. Cloning is common on Facebook and is made easier for the phisher if the victim has made their profile information and images publicly visible. Through the cloned account, the phisher submits friend requests attached with a convincing message to the victim’s friends. For example, “My account has been hacked, please delete my other account you have and communicate with me using this one only. If you don’t delete it, you may be hacked too!” Once accepted, the phisher begins sending messages to the friends connected to the victim instructing the

recipient to click on a link, consequently acquiring personal information from them. Similar to phishing emails, various SE techniques will be employed in the fabricated message to convince the user to click on the link.

Business orientated SNSs such as LinkedIn offer phishers the opportunity to collect data on companies and their employees. They can then use that information to launch spear phishing attacks, targeting employees specific to that organisation [16]. On LinkedIn, phishers could pose as prospective job recruiters, requesting documentation from the victim such as a curriculum vitae. This would contain a range of personal information the phisher could then use to conduct identity theft. Moreover, the phisher may request the victim to provide them with a copy of their identity document which can be used to conduct other crimes. The victim would willingly give out this information as it is not unusual to receive such requests from recruiters.

C. Malware-based

Malware-based phishing refers to a spread of phishing messages by using malware. For example, the victim installs a rogue Facebook app which automatically sends messages to all their Facebook friends. Such messages often contain links allowing the recipients also to install the rogue Facebook app on their computers or smartphone devices. Other deceptive techniques include promising Facebook users that by installing a particular app, which is malware, will allow them to see a list of people who visit their Facebook profile page. Another example: Enticing users with the option of installing the Facebook Color Changer app that will allow the user to change the colour of their Facebook account from the standard blue to a colour of their choice [17].

Other forms of malware-based phishing include content-injection which is malicious content. The malicious content can often be in the form of bogus posts (e.g. Facebook or LinkedIn posts, tweets) published by users whose accounts were affected with rogue apps. In many cases, victims are unable to see the bogus posts posted by the malware apps on their behalf. The bogus posts, for example, may contain a photo of the user and the statement: "I am injured and in the hospital. If you would like to help me, please sign up by clicking on the following link." When the victim clicks on the link, they will be requested to provide their personal data, which may be used by the phisher to commit identity theft and other scams. A post may contain malicious content and hoax text that requests the user to share the post. For example, a distributed hoax message stating that Facebook founder Mark Zuckerberg is giving away \$45 million to ordinary users and to be selected as one of the thousand lucky entrants, the message must be copied and pasted to one's wall along with five friends tagged in the post [18]. Again, this preys on SE techniques of scarcity, need and greed and so on.

Given the variations of how scams are conducted on SNSs, it is difficult to expect users to be updated with phishers' new techniques on exploiting technology and the methods in which they carry out their attacks on SNSs.

IV. BRIEF BACKGROUND ON PHISHING LITERATURE

A brief literature survey of "phishing" reveals that the most cited articles were published more than a decade ago. Even so, most researchers in the area of phishing continue to cite these published works. Although much research is available, the problem of addressing phishing still remains and as such continues to be an area of interest amongst scholars.

Early phishing research focused on technological controls with the testing and measurement of anti-phishing detection tools such as web browser toolbars [19], email detection filters [20], and URL detection [21,22,23,24]. Technological controls certainly perform a vital role in detecting the majority of phishing emails as they are automatically detected and filtered, preventing it from reaching the user. Other technological tools, such as web browser warnings, indicate to the user potentially malicious webpages.

Despite the technological tools available to assist users to identify phishing appropriately, they did not meet expectations. Research turned towards investigating user responses in how users interpret web browser warnings [25,26,27,28].

Information security literature highlights that humans are vulnerable to social engineering attacks and that security is a "people problem" [29]. As a result, research efforts were directed towards educating the human element to change their current behaviour [30, 31]. Educational approaches used were online games [32] and embedded email training systems [33]. Furthermore, research aimed at improving users' security awareness in phishing [34].

Recently, to understand this problem more, research focused on exploring differences in gender and personality traits with regard to phishing susceptibility. Openness, conscientiousness, extraversion, agreeableness, and neuroticism are considered the Big Five personality traits [35]. Other research applied a scenario-based design to study both the relationships between demographics and phishing susceptibility, and the effectiveness of several anti-phishing educational materials [36]. The relationship between the Big Five personality traits and email phishing response and how these traits affect users' privacy behaviour on Facebook was examined [4]. Additionally, [37] assessed the basic demographics of personality characteristics, dispositional trust, impulsivity, and web/computer based behaviour, beliefs, and previously experienced phishing consequences. Their study examined two behavioural/consequence factors: experiencing a monetary loss without reimbursement, and a belief that one may receive a legitimate request to confirm account information via email. A conceptual phishing susceptibility framework that utilises the Big Five personality traits and links the level of social engineering security-exploit susceptibility to an individual's personality traits was proposed by [38], and [39] investigated users' behaviour response when presented with phishing emails. They found personality traits of extraversion and openness were better at detecting phishing emails.

One of the earliest phishing experiments in the context of SNSs was by [40]. They discussed how phishing attacks can

be more effective by exploiting personal information found from SNSs. In a study by [41], they analysed data recorded from different parts of the world which compared the phishing emails used by phishers to lure victims in 2008 and 2014 respectively. They found that phishers have recently shifted their focus towards targeting online social media such as Facebook and YouTube to spread their phishing links. There is a lack of research dedicated to the susceptibility of social engineering victimization in SNSs, or to understanding which demographic factors correlate with falling for social engineering tricks in the context of SNSs [42]. As a result, a study by [42] attempts to predict a person's vulnerability to SE based on demographic factors (i.e. age, gender and educational level), relationship status, and personality type.

This section revealed that phishing research began addressing technological aspects, moved towards educating users, and finally examining the relationship of certain personality traits with phishing susceptibility on SNSs. The subsequent sections discuss emerging areas of interest in phishing research, namely habits and information processing.

V. SOCIAL NETWORK ACTIVITIES THAT MAY LEAD TOWARDS HABIT FORMATION

This section begins by discussing some popular activities SN users are engaged in and how its usage of such may develop into habits. As a result, it may affect them not to pay particular attention to suspicious information such as phishing scams. Furthermore, these habits may influence their behaviour to such an extent, that it may influence them on other online applications.

In public, it is not abnormal to see majority of people glued to their smartphones, most of whom are most likely using social apps to update their status or post pictures. This behaviour may be reinforced as it is repeated frequently. How one behaves in the physical world may not be much different compared to online SN environments. For example, industry professionals exchanging business cards with others is not uncommon practice – even if one has had no prior history with that person. Thus, receiving “invites” through a professional SN such as LinkedIn, members may behave similarly in this environment too by accepting invitations from strangers as the user has been conditioned to operate in this manner. Phishers may use this as an opportunity to gather personal information from the user.

Interconnectivity between smartphone apps gives users the freedom to broadcast their activities or messages across to other SNSs. For example, the Strava™ app, a social running and cycling app, allows users to publish their run to Facebook by simply clicking on the embedded Facebook icon. Other runners, who can be strangers, can follow one's run and view the map of the route. In another example, LinkedIn updates can also broadcast as tweets on Twitter. For the latter, a Twitter audience could be anyone provided they follow the user. If phishers have access to this specific information, they can perform spear phishing attacks on their targets.

Members of SNSs can also be notified via email of any activities linked to their preferences, e.g. tagged in a friend's post. Thus, receiving an email appearing to originate from the

SNS (as depicted in Fig. 1) may not appear suspicious to the user. As most users have a SN account, many websites, including e-commerce websites, allow users the option to log-in with their SN credentials (i.e. username and password). Thus, if a SN account has been hijacked, it may provide a means for phishers to conduct other forms of cybercrime using those credentials.

SNSs have common “social” functions that users have grown accustomed to. For example, most SNS, including social apps, have features such as inserting profile pictures, a status, mood, commenting on and liking posts. Most of these features exist in Facebook, LinkedIn and further instant messenger applications such as Whatsapp. “Following” users and pages is a standard function across most SNSs such as Facebook, LinkedIn, Twitter and others. It may not be concerning to members to be followed back or to receive a friend request from a stranger. As a result, receiving an email, purportedly originating from a SNS to accept a friend request may increase the chances for users to not be suspicious. As pointed out by [14], users are more likely to click on links in suspicious emails if it originates from a Facebook friend rather than from a bank.

YouTube is a popular media platform to watch online video content ranging from amateur footage that users have uploaded to various channels dedicated to particular areas of interest. These videos can be shared to other SNSs such as Facebook. To view the video, users have to click on the play icon, something which most users would be accustomed to. The latter poses a problem if phishers are sharing spoofed media content to other SNSs to lure users into viewing the video – especially if it is of interest to the user.

SN users also appear to post insensitive messages without thinking of its consequences. Users may think because they are not dealing directly with people, they are in a protective ‘bubble’. Recently, the South African public has been outraged by racial comments made by Penny Sparrow referring to Black people as “monkeys” [43]. Shortly thereafter, other prominent figures such as former Standard Bank investment strategist Chris Hart and radio personality Gareth Cliff were also accused of arguably “tweeting” racial utterances. In all cases, the organisations for which these individuals worked were pressured into taking action against them. This emphasises that users' behaviour on SNSs can put organisations' reputation at risk. As a result, organisations have seen the need to introduce social media policies [2].

From the SN activities described in this section, users may develop SN habits which can influence their ability to detect potential phishing attacks.

VI. DISCUSSION

Anti-phishing educational interventions typically focus on educating users on email messages and spoofed websites. Users are made aware to examine the message content for poor grammatical and spelling errors, not to click on hyperlinks within emails, not to open attachments from unknown sources and so on. However, phishers can take advantage of each of these education aspects on SNSs. For example, with the diverse language cultures of users in SNSs,

some users may not be able to identify grammatical errors in phishing emails. On the other hand, research uncovered that in some cases the grammatical errors, known as scammer grammar, may be crafted intentionally by the scammers [44] on the assumption that less educated users may be more susceptible to fall for scam offers.

SN users are inundated with links on their Facebook, Twitter and LinkedIn profiles and have grown accustomed that these shortcuts will lead them directly to content within the webpage or externally to other sources. Phishers are using URL shorteners not only for reducing space but also to hide their identity [45]. It is difficult for Twitter users to know whether the URLs they have received are legitimate [46]. Since Twitter limits any messages (i.e. tweets) posted to 140 characters, link shortening services, such as bit.ly, are used to shorten longer Internet addresses. Despite Twitter recently announcing that usernames, quoted tweets, photos and other media attachments will not count against the 140 character limit, these link shortening services are still currently being used. This presents educational concerns because consequently users will be unable to identify the website name of where these shortened links lead to, thus making it even more difficult to establish whether they are potentially dangerous. Phishers also use this opportunity to create shortened URLs to redirect users to malicious sites [16]. Furthermore, smartphone browsers display limited security information due to its small screen size. As a result, users who have been educated in phishing to look for the secure https:// protocol in the URL bar may not be able to see this directly on their smartphone. Furthermore, users may be engaged with other information seeking activities using other software applications thus distracting them. From these distractions, users may not be in the right frame of mind when presented with security attacks, thus leaving them vulnerable [47]. Users may also be overloaded with emails. Reference [48] found an increased likelihood of falling victim to phishing by the volume users receive.

VII. HABITS

What if user SN behaviour has become automated, to a certain degree, by going through the motions of scrolling through posts by friends, liking and sharing posts, clicking on various links and pages and not processing this information with more consideration to detail? It was found that users who habitually engage on Facebook are significantly more likely to fall prey to a social media phishing attack [49].

It was reported that one of the main reasons for social media usage is for self-distraction and boredom relief [50]. Receiving continual support in the form of comments and “likes” reinforces users’ behaviours and as such will be repeated by them [50]. Habitual clicking may lead to the user building a schema which leads them to instant gratification. As a result, it may become difficult for users to break this habit. It may be possible that these habits affect users to process information found on SNSs in a more systematic manner. This is elaborated more in Section VIII.

According to [51], almost no Information Systems research has investigated the potential importance of

subconscious (automatic) behaviours known as habits. Users’ “habitual pattern of email use is an issue that has yet to be examined within the phishing-based deception context” [49]. Overtime, when enacted repeatedly, behaviours become action-scripts that are applied without conscious reflection about its antecedents, consequences, or even its enactment [52]. In the context of Information Systems (IS) usage, [51] define “habit” as “the extent to which people tend to perform behaviours automatically because of learning.” These authors suggest that continued usage of Information Systems is not only a consequence of intention, but also of habit. As pointed out earlier, SNSs are exceedingly popular and as such, users are engaged for many hours on these sites. However, habit is not the same as behaviour [51]. It should be understood as a type of mindset that enhances the perceptual readiness for habit related cues, and prevents an individual from being distracted and from adopting other, less efficient courses of action. A stable context promotes habit formation in that it only requires a minimum of the individual’s attention in reacting adequately to certain situations. In the context of phishing, this stable context could be engaging in SNSs or checking email. Once a habit is established, behaviour is performed automatically to such a degree that it requires little or no conscious attention and minimal mental effort. Thus, if users continuously open links on Facebook and Twitter without any fear of consequence, it may cross other environment too, for example email or banking websites.

VIII. INFORMATION PROCESSING MODELS

According to [53], social-psychological research on phishing has implicated ineffective cognitive processing as the key reason for individual victimization. As such, it is important to consider models related to this problem. Therefore, this section focuses on persuasion theories applicable to the phishing context.

The heuristic-systematic model (HSM) is a model of information processing that originated from persuasion research in social psychology [54]. The model attempts to explain individual information processing and attitude formation in persuasive contexts. The HSM and elaboration likelihood model (ELM) are closely related models and are recognised as dual process models because they both propose two major approaches to persuasion, namely: the central route and the peripheral route. The key difference between the two models is that HSM explicitly recognises dual processing (i.e. parallel or jointly), while ELM suggests information processing occurs on a continuum. Researcher [55] found that the ELM offers an encouraging framework for understanding the ways in which social engineers gather sensitive information or get unwitting victims to comply with their request.

Since its introduction, dual-process models remain today’s most influential persuasion paradigms [56]. Compared to systematic processing, [54] define heuristic processing as “a limited mode of information processing that requires less cognitive effort and fewer cognitive resources.” Heuristic processing draws upon simple decision cues, often termed “rules of thumb”, and occurs when individuals lack motivation or cognitive resources. This processing occurs at a

superficial level, allowing the receiver to form judgments based on cues such as credibility, attractiveness, and message length [57] – all of which are key SE techniques. Additionally, heuristic processing takes advantage of the factors embedded within or surrounding a message (heuristic cues) such as its source, format, length and subject in order for the user to perform a validity assessment quickly [12]. Phishing emails exploit these factors the most. If receivers are able and properly motivated, they will elaborate, or systematically analyse, persuasive messages. If the message is well reasoned and logical, it will persuade them [56]. Further, systematic processing takes place when users carefully analyse the message's content and may also conduct further research to validate the message [12].

According to [12], persuasion research studies how received messages can change users' attitudes. The model suggests that people either use heuristics and short-cuts in decision-making, or they systematically process the merits and demerits of a given argument. The HSM and the theory of planned behaviour was linked by [58] through a model of risk information seeking and processing model (RISP). They proposed that the method of information processing users apply to risk information from media and other sources affects their beliefs, evaluations and attitudes.

According to [59], systematic processing is more likely when careful thought is likely to generate judgment confidence. Further, if the message is particularly relevant to the person on a personal level such as their goals or interests.

Ideally, systematic processing would be the preferred method of choice when users are presented with phishing. However, this type of processing requires more effort, time and cognitive resources. As such, users may limit systematic processing unless they are motivated to do so by following motivational factors by [12]: perceived risks, perceived importance of decision outcome, skills level, time and other pressures and the presence/absence of heuristic cues.

Users may process information concerning risks intensively, superficially, or not at all [58]. Unfortunately, users typically trust phishing messages on superficial cues like design and author. If users consider determining the validity of a phishing message or messages received via a SNS as being too time consuming, difficult or unimportant, this may influence users to resort to heuristic processing. As such, this will put them at risk to phishing attacks. Ideally, if users were motivated to systematically process information they receive, checking it for validity, there would presumably be less victims of phishing.

IX. SUMMARY

This paper introduced social network phishing and discussed its similarities with traditional email phishing. It is evident that phishers continue to make use of SE effectively to persuade their victims into performing certain actions, including on SNSs. The paper further discussed a brief literature background of phishing research and how it has progressed from technological controls to psychological theories. It also discussed how SNS users are becoming habituated to behaving in a certain manner which may

influence them not to pay closer attention to certain deceptive methods employed on SNSs. This behaviour may influence usage on other platforms such as email. The paper also highlighted habits and information processing models as areas in phishing research that require more attention by researchers. It is only recently that models have been developed which take into consideration habits and information processing. Future research aims to develop a user susceptibility model which considers investigating the linkages between SE techniques, habits and information processing.

REFERENCES

- [1] Statista, "Statistics and facts about Social Networks," 2015. <http://www.statista.com/topics/1164/social-networks/>
- [2] H. Wilcox and M. Bhattacharya, "Countering Social Engineering through Social Media: An Enterprise Security Perspective," 7th International Conference on Computational Collective Intelligence Technologies and Applications (ICCCI 2015), LNAI, Springer, vol. 9330, 2008, pp. 54-64.
- [3] K.D. Mitnick and W.L. Simon, *The art of deception: Controlling the human element of security*. New York, NY: Wiley, 2002.
- [4] T. Halevi, J. Lewis, and N. Memon, "A Pilot Study of Cyber Security and Privacy Related Behavior and Personality Traits," WWW 2013 Companion, Rio de Janeiro, Brazil. ACM, 2013.
- [5] A. Ferreira, L. Coventry, and G. Lenzini, "Principles of persuasion in social engineering and their use in phishing." In: *Human Aspects of Information Security, Privacy, and Trust. Lecture Notes in Computer Science*, 9190. Springer, Cham, 2015. pp. 36-47.
- [6] R. B. Cialdini, *Influence: The Psychology of Persuasion*. Harper Business, 2007.
- [7] D. Gragg, "A Multi-Level Defense Against Social Engineering", SANS Institute - InfoSec Reading Room, Tech. Rep, 2003.
- [8] F. Stajano and P. Wilson, "Understanding Scam Victims: Seven Principles for Systems Security", *Commun. ACM*, vol. 54, no. 3, pp. 70-75, Mar. 2011.
- [9] APWG, *Phishing Activity Trends Report, 4th Quarter 2014*, https://docs.apwg.org/reports/apwg_trends_report_q4_2014.pdf
- [10] M. Jakobsson and S. Myers, "Phishing and countermeasures: understanding the increasing problem of electronic identity theft." 2006
- [11] E.E.H. Lastdrager, "Achieving a Consensual Definition of Phishing Based on a Systematic Review of the Literature," *Crime Science*, 3, 2014.
- [12] X. Luo, W. Zhang, S. Burd, and A. Seazzu, "Investigating phishing victimization with the Heuristic-Systematic Model: A theoretical framework and an exploration," *Computers and Security*, Vol.38, October, 2013, pp. 28-38.
- [13] H. King, "Top 5 social media scams to avoid," *CNN Money*, <http://money.cnn.com/2016/04/22/technology/facebook-twitter-phishing-scams/>
- [14] A. Stern, "Social Networkers Beware: Facebook is a Major Phishing Portal," *Kaspersky Lab*, 23 June 2014, <https://blog.kaspersky.com/1-in-5-phishing-attacks-targets-facebook/5180/>
- [15] J. Allen, L. Gomez, M. Green, P. Ricciardi, C. Sanabria, and S. Kim, "Social Network Security Issues: Social Engineering and Phishing Attacks," *Proceedings of Student-Faculty Research Day*, CSIS, Pace University, 2012.
- [16] Sophos, "Social Networking Security Threats," 2011. <https://www.sophos.com/en-us/security-news-trends/security-trends/social-networking-security-threats/facebook.aspx>.
- [17] S. Khandelwal, "Warning – Facebook Color Changer App Is Just A Scam, Infects 10000 Users," 2014. *The Hacker News*, http://thehackernews.com/2014/08/warning-facebook-color-changer-app-is_9.html
- [18] K. Wagstaff, "Hoax Alert! No, Zuckerberg Isn't Giving Millions to Facebook Users," *NBC News*, 28 December 2015,

<http://www.nbcnews.com/tech/social-media/hoax-alert-no-zuckerberg-isnt-giving-millions-facebook-users-n476551>

- [19] Y. Zhang, J.I. Hong, and L.F. Cranor, "Cantina: a content-based approach to detecting phishing web sites," In: 16th international conference on World Wide Web, ACM. 2007, pp. 639-648.
- [20] I. Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," Proceedings of the 16th international conference on World Wide Web. ACM. 2007.
- [21] S. Garera, N. Provos, M. Chew, and A.D. Rubin, "A framework for detection and measurement of phishing attacks," Proceedings of the 2007 ACM workshop on Recurring malware. Alexandria, VA: ACM. 2007, pp.1-8.
- [22] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "A Comparison of Machine Learning Techniques for Phishing Detection," APWG eCrime Researchers Summit, Pittsburgh, PA, USA. 2007, pp. 60-69.
- [23] L. Wenyin, G. Huang, L. Xiaoyue, Z. Min, and X. Deng, "Detection of Phishing Webpages based on Visual Similarity," WWW 2005, ACM
- [24] Y. Zhang, S. Egelman, L.F. Cranor, and J. Hong, "Phishing phish: Evaluating anti-phishing tools", 2006.
- [25] R. Dhamija, J.D. Tygar, and M. Hearst, "Why phishing works," In: SIGCHI conference on Human Factors in computing systems, Montreal, Canada: ACM. 2006, pp. 581-590.
- [26] M. Wu, R.C. Miller, and S.L. Garfinkel, "Do security toolbars actually prevent phishing attacks?," Proceedings of the SIGCHI conference on Human Factors in computing systems: ACM. 2006.
- [27] J.S. Downs, M.B. Holbrook, and L.F. Cranor, "Decision strategies and susceptibility to phishing," Proceedings of the second symposium on Usable privacy and security. Pittsburgh, Pennsylvania: ACM. 2006, pp. 79-90.
- [28] S. Egelman, L.F. Cranor, and J. Hong, "You've been warned: An empirical study of the effectiveness of web browser phishing warnings," Twenty-sixth annual SIGCHI conference on Human factors in computing systems. Florence, Italy: ACM. 2008, pp. 1065-1074.
- [29] R. West, C.B. Mayhorn, J. Hardee, and J. Mendel, "The Weakest Link: A Psychological Perspective on Why," Social and Human Elements of Information Security: Emerging Trends, 2009, pp. 43-60
- [30] M.B. Burns, A. Durcikova, and J.L. Jenkins, "What kind of interventions can help users from falling for phishing attempts: a research proposal for examining stage-appropriate interventions," 46th Annual Hawaii International Conference on System Sciences, 2013, pp. 4023-4032.
- [31] I. Kirlappos and M.A. Sasse, "Security education against phishing: A modest proposal for a major re-think," 2009.
- [32] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L.F. Cranor, J. Hong, and E. Nunge, "Anti-Phishing Phil: The design and evaluation of a game that teaches people not to fall for phish," Proceedings of the 3rd symposium on Usable privacy and security. Pittsburgh, Pennsylvania: ACM. 2007.
- [33] P. Kumaraguru, Y. Rhee, A. Acquisti, L.F. Cranor, J. Hong, and E. Nunge, "Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System," CHI 2007. San Jose, California: ACM. 2007, pp. 905-914.
- [34] C.R. Dodge, C. Carver, and A.J. Ferguson, "Phishing for user security awareness," Computers & Security, 26, 2007, pp. 73-80.
- [35] S. Gosling, P. Rentfrow, and W. Swann Jr. "A very brief measure of the Big-Five personality domains," Journal of Research in Personality, 2003, pp. 504-528.
- [36] S. Sheng, M. Holbrook, P. Kumaraguru, L.F. Cranor, and J. Downs, "Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions," Proceedings of the 28th international conference on Human factors in computing systems. Atlanta, Georgia, USA: ACM. 2010.
- [37] C.B. Mayhorn, A.K. Welk, O.A. Zielinska, and E. Murphy-Hill, "Assessing Individual Differences in a Phishing Detection Task," Proceedings 19th Triennial Congress of the IEA, Melbourne. 2015.
- [38] J. Parish, J. Bailey, and J.F. Courtney, "A Personality Based Model for Determining Susceptibility to Phishing Attacks," Southwest Decision Sciences Institute, 2009.
- [39] M. Pattinson, C. Jerram, K. Parsons, A. McCormac, and M. Butavicius, "Why do some people manage phishing e-mails better than others?," Information Management & Computer Security, Vol. 20 Iss 1 2012. pp. 18-28.
- [40] T.N. Jagatic, N.A. Johnson, M. Jakobsson, and F. Menczer, "Social Phishing," Communications of the ACM. 50 (10). 2007, pp. 94-100.
- [41] S. Gupta and P. Kumaraguru, "Emerging Phishing Trends and Effectiveness of the Anti-Phishing Landing Page," 2014.
- [42] A. Algarni, Y. Xu, Yue, T. Chan, and T. Yu-Chu, "Social engineering in social networking sites: how good becomes evil," In Proceedings of the 18th Pacific Asia Conference on Information Systems. 2014.
- [43] J. Flanagan, "Estate agent forced to go into hiding after Facebook post deriding black people as 'monkeys' for dropping litter on beaches causes a storm in South Africa," 2015, <http://www.dailymail.co.uk/news/article-3383844/Estate-agent-forced-hiding-Facebook-post-deriding-black-people-monkeys-dropping-litter-beaches-causes-storm-South-Africa.html>.
- [44] B. Nikiforova and D.W. Gregory, "Globalization of trust and internet confidence emails," Journal of Financial Crime, vol. 20 Iss: 4, 2013, pp. 393-405.
- [45] S. Chhabra, A. Aggarwal, F. Benevenuto, and P. Kumaraguru, "Phish/Social: The Phishing Landscape through Short URLs," CEAS '11, Perth, Australia: ACM. 2011.
- [46] C. Everett, "Social media: opportunity or risk?," Computer Fraud & Security, vol 2010, Issue 6, June 2010, pp. 8-10.
- [47] K. Ivaturi, L. Janczewski, and C. Chua, "Effect of Frame of Mind on Users' Deception Detection Attitudes and Behaviours," 2014, CONF-IRM.
- [48] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H.R. Rao, "Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model," Decision Support Systems, 51, 2011, pp. 576-586.
- [49] A. Vishwanath, "Habitual Facebook Use and its Impact on Getting Deceived on Social Media," Journal of Computer-Mediated Communication, 20. 2015, pp. 83-98.
- [50] T. Priestley, "Is Social Media Just Another Bad Habit To Break?" Forbes, August 2015, <http://www.forbes.com/sites/theopriestley/2015/08/13/is-social-media-just-another-bad-habit-to-break/#2c533f9e6f99>
- [51] M. Limayem, S.G. Hirt and C.M.K. Cheung, "How Habits Limit The Predictive Power of Intention: The Case Of Information Systems Continuance," MIS Quarterly, vol. 31 (4), 2007, pp. 705-737.
- [52] R. LaRose and M.S. Eastin, "Social cognitive theory of Internet uses and gratifications: Toward a new model of media attendance," Journal of Broadcasting & Electronic Media, 48(3), 2004, pp. 358-377.
- [53] A. Vishwanath, B. Harrison, and Y.J. Ng, (in-press). "Suspicion, Cognition, Automaticity Model (SCAM) of Phishing Susceptibility," Communication Research.
- [54] A.H. Eagly and S. Chaiken. The psychology of attitudes. FortWorth, TX: Harcourt Brace and Jovanovich. 1993.
- [55] M. Workman, "Wisecrackers: A Theory-Grounded Investigation of Phishing and Pretext Social Engineering Threats to Information Security," Journal of the American Society For Information Science And Technology, 59 (4), 2008, pp. 662-674.
- [56] W.D. Crano and R. Prislin, "Attitudes and Persuasion. Annu. Rev. Psychol, 57, 2006, pp. 345-74.
- [57] K.A. Cameron, "A practitioner's guide to persuasion: An overview of 15 selected persuasion theories, models and frameworks," Patient Education and Counseling, 74, 2009, pp. 309-317.
- [58] R. J. Griffin, K. Neuwirth, J. Giese, and S. Dunwoody, "Linking the Heuristic-Systematic Model and Depth of Processing," Communication Research, vol. 29 No. 6, 2002, pp. 705-732.
- [59] ChangingMinds, "Heuristic-Systematic Persuasion Model," http://changingminds.org/explanations/theories/heuristic-systematic_persuasion.htm

Identity Management for e-Government

Libya as a case study

Othoman ELASWAD, Christian Damsgaard JENSEN
Department of Applied Mathematics and Computer Science,
Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark
E-mail: {otel,cdje}@dtu.dk

Abstract-Governments are strengthening their identity (ID) management strategies to deliver new and improved online services to their citizens. Such online services typically include applications for different types of permissions, requests for different types of official documents and management of different types of entitlements. The ID management scheme must therefore be able to correctly authenticate citizens and link online presence to real world identities.

Many countries, in particular in the developing world, are currently introducing national ID management schemes for the first time. While most of these countries have paper based records, many of these are regionally based and few of these have been consolidated, so these records may contain incorrect, incomplete, inconsistent or redundant information.

In this paper, we explore the design space for national ID management and online authentication schemes, in this context. In particular, we propose a simple model for issuing national ID numbers that satisfy these goals and use this model to examine two different ID management schemes implemented in Libya, which allows us to compare different approaches to national identity management. The two schemes were implemented within a fairly short time, so we may assume that the cultural, social, educational and technological factors remain unchanged. This allows a direct comparison of objectives and means. Based on this examination, we evaluate the current Libyan ID number system with respect to the identified objectives. Our evaluation of the two Libyan NID schemes show that if National Identity Management does not fully meet the requirements identified in our simple model, then it may be vulnerable to various forms of online risks such as impersonation and identity theft attacks. Considering online crime, during the design of an Identity Management system, is especially important in developing countries, where such crimes have not previously existed in the society.

Keywords: Identifiers, Authentication & Identity Management

I. INTRODUCTION

Nationality is defined as “persons having a common language and culture form a nation and, as such, ought to be entitled to self-government as a state” [1]. Identity means “the condition of being a specified person in which one’s attitudes and actions can define” [2]. Therefore, national Identity means identifying a specific person in a society which has a common language and culture. Various countries’ governments around the world have been working on providing a unique identity to their citizens. Many of them have already implemented national identity schemes, e.g. the Scandinavian countries

Sweden, Denmark and Norway introduced Personal Identity Number (PIN) schemes before computerization in 1947, 1968 and 1970 respectively [3]. There are, however, other countries, which have only introduced such schemes in the last decades. One example is India, which in 2009 decided to introduce a Unique Identification (UID) for all its citizens and hence launched the UID program called “Aadhaar” [4]. Another example is the Libyan national number initiative which was implemented in 2013, but neither of these schemes is fully implemented yet. This means that citizens in those countries have no standard means to prove who they are. This becomes problematic in the transformation to e-Government, where government services are provided to citizens online. The national ID is increasingly becoming the cornerstone of a secure and trusted ecosystem. In fact, many countries have extended their traditional National ID Number (NIDN) schemes by introducing national ID card schemes to support new functions such as identification, authentication and digital signatures that integrate with existing technologies. Thus, digital identity is now the primary means by which a natural person can access government e-services [5].

The main drivers behind implementing national Identity Management (IDM) are to improve the identification and authentication mechanisms in order to reduce crime, combat terrorism, eliminate identity theft, control immigration, stop benefit fraud, and provide better service to both citizens and legal immigrants [1][2][6]. Therefore, the introduction of a single national identifier is generally considered an essential step towards the introduction of a technology to integrate data about the individual citizens quicker and far more easily.

There are various characteristic of good identifiers including universality of coverage, each person should have an identifier, uniqueness each person should have one identifier and no two persons have the same identifier, and , permanence through the lifetime of the individual [1][7]. In order for an identifier to be operational, it is also required that each person can be linked to their identifier in a verifiable way. While this is not a characteristic of the identifier per se, we include “verifiability” as a separate goal in our list of requirements. Digitalization and electronic records are fairly new in the context of government records, so we also need to consider how PINs are assigned and used in e-Government. The typical PIN lifecycle consists of three phases: creation, use and retirement. A PIN is normally created and assigned at birth or when a person immigrates and/or becomes naturalized in a

country, but people born before the start of electronic record keeping, may have to apply for a PIN later in life. A PIN may be used to link all official records regarding a single person; so many different authorities will have access to the PIN. This means that it is extremely important not to use the PIN as a knowledge based authentication factor, as it is done in some countries, because knowledge of a person's PIN will be distributed among all government agencies that the person has interacted with, i.e. *the more useful a PIN is as an identifier the less useful it is as an authenticator*. Finally, a PIN will be retired, but not necessarily deleted, when a person dies or loses his citizenship, e.g. as a result of emigration and/or naturalization in a different country. The quality of a digital identity created as a result of an application from an existing person depends on the accuracy of data information in the originally filed paper records that are to be transferred to electronic form.

One of the main issues of digital identity is the possibility of fraud, i.e. that one or more of the four requirements listed above are not met. Unique number projects, however, introduce a number of complex risks, such as duplication, impersonation and ID crimes [2][5][8]. Although the benefits of most national ID schemes are fairly similar, the culture, social norms, citizens' skills and historical context in term of digital infrastructure, may be quite different, so those factors impact on the design of national ID systems. We therefore need to understand how these factors influence the design space for IDM for e/Government services. In the past decades, the Libyan government has undertaken two separate efforts to develop a national ID with similar infrastructure and culture. This allows us to examine different design choices in a context, where culture and the legacy infrastructure remain largely unchanged. This paper is going to introduce a simple model of issuing a national ID number, which considers the features of good identifiers in terms of uniqueness, universality, permanence and verifiability and then evaluate the first and second Libyan national ID by analyzing both systems to identify strengths and weaknesses of both systems. Part of this paper is based on an interview with 5 members of the Libyan national ID team project to get information about why the first national ID project was cancelled and what are the new features of the new project.

The structure of the rest of this paper is as follows. Related work is presented in the next section, and the new model is introduced in the following section. The forth section presents the Libyan national number systems. Finally, analysis and discussions are presented in the last section.

II. RELATED WORK

Many developing countries do not have national identity systems in place, and many of the ones that do, suffer from high rates of under-registration [4]. The implications of this are societal exclusion that limits access to education, health, banking, and opportunities for personal economic growth. Zelanyr evaluates the Indian Universal Identity (UID) project and identifies the main driver as linking people to various applications including passport, driver's license, tax, bank

accounts and elections, i.e. the Indian UID is primarily used as an infrastructure for identification and verification. All the records from the old system were converted into electronic records. Once the new system was operational, it took a period of two months for the new system to de-duplicate the 86M demographic database as well as the 56M iris database that had been established until that time [4]. This study indicates that one of the main issues in developing countries is that citizen's information in paper based registers may be incomplete, duplicated or inconsistent, which results in a poor digital infrastructure.

The online environment allows for the collection and interconnection of larger amounts of information than ever before. This may have tremendous benefits to both governments and citizens, but it also creates several risks that did not exist in the more traditional paper-based systems. For instance, one Norwegian study reported that, in 2004, members of Norwegian Public Service Pension Fund (NPSPF) could apply online for loans by simply entering their social security number (SSN). If the SSN was valid and belonged to an NPSPF's member, then the sender received a message containing the person's name, address and zip code. The author of the Norwegian study showed that it was possible to determine valid SSN using NPSPF's loan web page by implementing a script to build a database containing the previously mentioned information and furthermore, it is possible to classify SSN to a set of people based on specific area using zip code[9]. Another study summarizes the implication of loss or compromise of digital identity in terms of financial loss, emotional distress and reputational damage. The author refers the implication of digital identity to the presumption that the digital identity recorded and used under the scheme is authentic, accurate and exclusive [5]. In practice, issues exist where systems do not correctly recognize the identity of a citizen or where it permits the identity to be misused by another person. Countries that recently adopted such projects should benefit from other countries' experience and take this into account to avoid such mistakes when implementing a digital identity scheme.

One study [9]reported that Government identity management can be implemented in different ways so it is useful to assess these differences against historical, cultural and social backgrounds and these elements can often be as important as technology in determining an approach to identity management. Accordingly, the public and private sector are now producing a wide range of reference frameworks aimed at achieving consistency in designing privacy and security into identity management systems and as result, they are gaining greater community acceptance. Authors took New Zealand's identity management as an example case. New Zealand's identity management system is based on a Government Login system to provide both single and multifactor authentication to support services with different transaction values and associate risks, while identification is performed via the Identity Verification Service (IVS). Authors concluded that New Zealand example demonstrates the value of starting from a sound understanding

of the policy environment and a clear vision of what is to be achieved. Therefore, identifying the function of digital identity and mechanisms to achieve these functions are important as well as legislation that regulates the privacy of individual's information and data protection. All these components and others make the vision clear..

Al-Khouri studied the implementation of the national ID programme in the United Arab Emirates (UAE) and provides some suggestions to increase public acceptance and consequently increase project success chances. The author reported that the implementation of the project must take place in three stages. In the first stage, the project must attempt to enroll the population for the new ID card with a minimal set of data; as only primary identification data will be required for first time enrollment, it was proposed to eliminate the application form and rather make use of the existing electronic data in the Ministry of Interior's database to obtain and verify the citizen's personal information. In stage two, efforts must be directed towards promoting and enforcing the presentation of the new ID card for identity verification and as a pre-requisite to access the most frequently visited government services. The organizations that provide these services then need to maintain the new ID numbers in their databases, which should be used when moving to stage three of the strategy which requires the national ID database to interface and integrate with these databases. In the lessons learned section, Al-Khouri identifies some points that should be considered when implementing IDM, including that proper planning is essential to the success of such projects. The project team should take enough time to focus on the procedure of enrolment of the whole population and the issuance of the new ID card. Such programmes also need to put much effort into promoting e-identity and e-verification services using the new ID card. It would be interesting to measure the impact of national ID programmes on the overall government economy[10].

The above paragraphs show the absence of national identity schemes in developing countries and shortcomings of such scheme when they exist. Lack of a clear vision to introduce National identity leads to threats to privacy and security of national identity schemes. As a result various online risks might exist, which could cause financial loss, emotional distress and reputational damage. The integrity and accuracy of national ID management schemes that transfer the individual's existing records to electronic form depend on the quality of the existing paper based records. Therefore, it is necessary to validate the individual's paper based records in terms of completeness, correctness, non-duplication and verifiability, before they are transferred to electronic form. The following section will introduce a simple model that identifies national identity requirements and properties of good individual's record, which will be used to evaluate Libya's national number scheme.

III. SIMPLE MODEL FOR ISSUING NATIONAL NUMBER

A. *Concepts of Personal Identity Number (PIN):*

A personal identity number is a number that acts as a unique identifier, which represents an individual during remote interactions with public and private sector agents. According to Clarke [2], human identity has become central to our modern conception of mankind since the renaissance. The original needs for identification were social rather than economic. Relatives and friends recognize a person on a contextual basis, in which physical appearance, voice characteristics, knowledge of private information all play a part. As the complexity of economic transactions increased, the need arose for parties to know with whom they were dealing. It became normal for parties to provide one another with information about themselves, appropriate to the nature of the transaction[2].

Organizations often assume that there exist a one-to-one relationship between persons and identities, no matter how many different roles he or she may play, or choose to adopt[2]. Multiple identity management is a significant issue for individuals and organizations. Beynon-Davies [7] studied the issue of IDM in terms of a semiotic framework consisting of three interrelated processes: authentication, identification and enrolment. The study examined different forms of personal identifiers and characteristics of good identifiers. Identifiers are grouped into natural identifiers and surrogate identifiers. Natural identifiers are often fail to provide the uniqueness demanded by organizations and their information systems. Surrogate identifiers include additional features such as codes and tokens used to uniquely identify individuals. Our model is based on the characteristic of good identifier identified by the studied mentioned above [2][7]. The difference between the above studies and our study is that the above studies investigate the issues of identifier facing multiple identity management while we focus on national identity scheme issue.

Blume [3] studies the legal system in which the personal identity number (PIN) is used. The PIN in Scandinavian countries is described as a piece of paper with the name of the individual and the unique number. There is no other information on the paper, e.g. a photograph, to help authenticate the person. The study identifies the important elements and principles that should be considered when introducing personal identity number schemes. These elements include purpose, expiration, issuer, data protection, regulation and legislation. Therefore, based on these studies we developed a simple model that will be used to evaluate both of Libya's national ID schemes.

B. *Simple Model*

To achieve the goals identified in the introduction section, our simple ID management model must satisfy the following requirements.

1) *Characteristics of good identifier*

- Universality: every citizen should have an identifier.
- Uniqueness: every citizen should have one identifier and no two citizens have the same identifier.

- Permanence: the identifier should not change, nor be changeable without authority.
- Verifiability: there should be reliable means to verify the mapping between citizen and identifier.

2) Properties of individual's records

- Accuracy: Individual's records should have correct information.
- Complete: Individual's record should have all required information, such as name, surname, date of birth and contact information.
 - Consistency: Individual's records within various government bodies should have the same information.
 - Singularity: Every individual should have one record to avoid duplication.

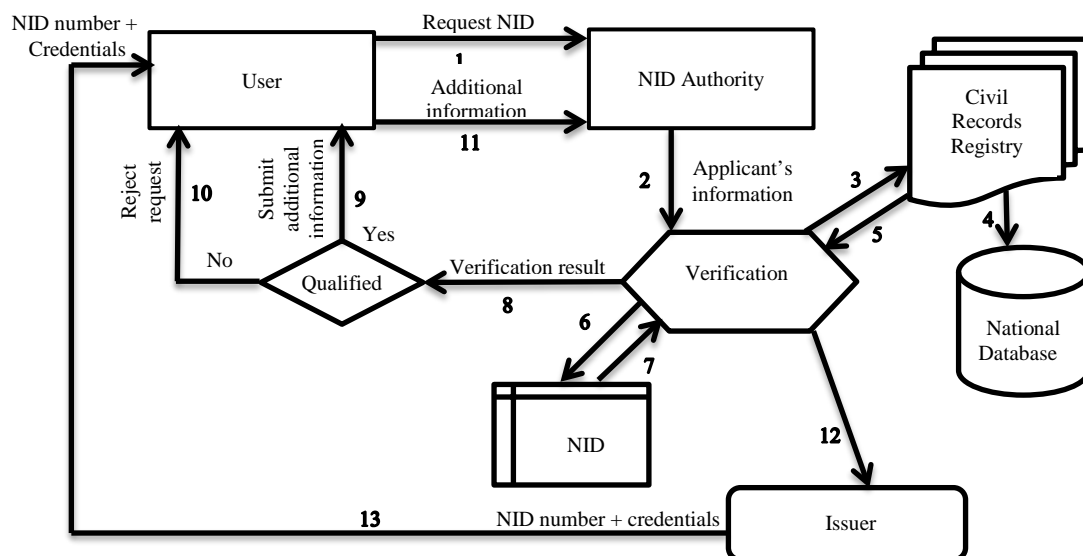
3) Requirements for National ID Management

- Build a digital national database.
 - Capture all individual's information to ensure universality.
 - Ensure that only correct information is recorded and keep in registers. This requirement to ensure accuracy.
 - False information should be amended to increase accuracy.
 - Missing required information must be completed, e.g. mobile number, email and postal address, to ensure completeness of the required information.
 - Consistency of individual's information within various government bodies should be verified to avoid inconsistency of the individual's records.
 - Remove individual's duplicate records to ensure uniqueness of records.
- Legislations and regulations
 - Issue rules to distinguish private information from public information.
 - Issues rules concerning illegal use of national identity.
 - Issues rule to regulate distribution of information between government bodies and organizations.
 - Issue rules that concerns illegal use of national identity.

- Enrollment requirement
 - An applicant needs to prove her/his entitlement to a NID.
 - An applicant needs to submit an application requesting NID.
 - An Authority needs to validate applicant's entitlement.
 - An Authority needs to ensure that the applicant has not been issued a NID before.

4) Model description

We assume that a national database as well as legislation already exists. Therefore, we focus on the process of issuing a NID in the following. First, an applicant fills in a form with his or her basic personal information including name, surname, date of birth, contact information and official documents proving his personal information. The National Identity Authority verifies whether the applicant qualifies for a NID number or not. An applicant qualifies for a NID number if he has an official document that has already been registered in a civil registration authority or any other national registration that contains (part of) a population register. The other requirement is that the applicant should not have been issued a NID number before - this can be checked from NID database. If an applicant does not qualify, then his request will be rejected, otherwise, he will be asked to submit additional information to check that he has not been issued a NID with different official documents, i.e. exploiting inconsistencies in the paper based records. This additional information could be biometric data, such as finger print, photograph or DNA. The NID authority then compares the biometric data submitted by the applicant with the biometric data previously stored in the NID database and rejects the application if a match is found. Finally, the NID authority provides qualified applicants with a national id-number and credentials that can subsequently be used to verify the applicant during online authentication. Figure 1 shows the process of issuing a national number that will be unique and reliable. Figure1 shows the individual steps when a user applies for a national ID number from the NID Authority: 1) a citizen shows up in person to submit the application form and required documents, 2) applicant's information is sent to the verification process, 3) verify authenticity of submitted documents from the civil records



978-1-5090-2473-8/16/\$31.00 ©2016 IEEE Figure 1 A simple model for issuing a national number

registry, 4) civil records registry checks records against the national database, 5) civil records registry responds to the NID authority, 6) if the documents are genuine, the verification authority check NID's database to see if it has issued a NID number to the person before based on biometric information, 7) return result of check to NID Authority, 8) verification result will determine if user qualified to get NID number or no, 9) if yes, more information will be requested, 10) if no, citizen's request will be rejected, 11) a citizen will submit additional information, 12) NID will issues NID number + credentials, 13) credential will send to citizen.

IV. LIBYAN NATIONAL ID SCHEMES

As mentioned earlier, the Libyan government has implemented two different IDM schemes; the first scheme was based on an ID smartcard which contains biographical and biometric information about the citizen. The second scheme is just a number that links a citizen with his recorded information. The duration between the discontinuation of the first project and the initiation of the second project was about a year, so technological and non-technological factors can be assumed to remain constant. The question is whether implementing two different IDM approaches with the same digital infrastructure, culture, citizen skills will eliminate the risks of online transaction. This section gives an overview of existing Libyan official documents and a summary of an interview we made with Libyan officials, before we describe the first and second Libyan national ID projects.

A. Libyan official documents

A family book, shown in Figure 2, is an official document, which is issued by the civil registry authority to each Libyan family. It contains information about all members of the family (husband, wife and children) including name, surname, date and place of birth for each member of the family. Each family book contains a unique six digits number used to differentiate each family book; this number is called the auditing number.

Identity plastic card is another compulsory official document issued to all Libyan citizens older than 18. The Immigration Office is the responsible government body that



Figure 2 Libyan Family book

issues identity plastic cards. The card contains citizen's information including name, middle name, surname, date of birth, photograph, hand written signature and a fingerprint. The card is used to prove the identity of the holder when travelling from one city to another. It is also used as an identity proof for all government transactions. Traveling outside the country needs another document which is a passport. Libyan passport contains all the previous information. All the information included in both ID card and passport is hand written.

B. Summary of interview with NID number project team

The aim of the interview is to collect some information about the process of issuing new Libyan national number. A general discussion was done with some members of the project team about the aim of the project and how it is implemented. We asked various questions to understand the main objectives of the new Libyan NID, possible weaknesses of the previous NID scheme and the motivation for developing a new scheme. Also, our question focused on the mechanisms and process they made to achieve their goals. For example, we asked how they verified the correctness and completeness of an individual's records. Similar questions were on how they verified the uniqueness of the issued NID numbers. Their answer regarding uniqueness was that they assumed every Libyan family only has one family book. We summarized the result of the interview as follows:

- The goals of the second scheme appear quite similar to the objectives of first Libyan national ID project.
- There were no reasons mentioned in terms of technical weaknesses, such as security issues, usability issues or any other technical weaknesses.
- They assumed that every family has one family book.
- They have not verified social benefit fraud by people who may have more than one family book.
- They have not validated that family members who married or died have been removed from the family book.
- They have not validated correctness, uniqueness, consistency and completeness of individual's record at civil registry.
- There are few errors during input data and most of those errors were date of birth.
- Citizens can correct input error data by contacting their civil registry.

C. First Libyan National Id-number Project

This sub section will summarize the first Libyan national ID scheme in terms of motivation and drivers, requirements to issue national ID.

1. Motivation and drivers

- Eliminate social benefit fraud.
- Improve quality of public services for citizens.
- Build a digital infrastructure for Libyan e-government services.
- Build a national database to store a citizen’s information.

Items	First Libyan NID number	Second Libyan NID number
Aim	To prevent duplication, illegal immigration, improve digital infrastructure, reduce id theft	Prevent duplicate salary, start point of Libyan e-G, improve performance and efficiency of public sectors.
Registration procedure	Applicants come in person	Based on the civil register records
Requirements	Libyan family book and fill a form	has a record in civil register and has a family book
Getting NID number	Issues smart card contains NID number	Check by SMS message or through NID web site
Number of digits	13 digits	13digits
Biometric data	Finger print and DNA	In future
Uses	Cancelled before use	Registration of election

Table1 summary of the information about two Libyan NID number schemes

2. Requirements to issue first Libyan national ID

To get an ID-card in the first Libyan ID scheme, an applicant has to meet the following requirements:

- An applicant has to fill in a specific form for the national id.
- An applicant needs to provide an official government document which contains all the names of the applicant, i.e. first name, middle name, third name and surname.
- An applicant needs to submit the applicant’s biometric information including the fingerprints of all ten fingers, scanning the applicant’s signature, a photograph of the applicant’s face and the DNA of the applicant through a saliva swab.
- The applicant verifies the correctness of the information on the printed paper and pays about 3US\$ as a fee for the ID-card.

D. SECOND LIBYAN NATIONAL ID SCHEME

In 2012, the Central Bank of Libya expressed concerns about identity fraud. At the time, there was no definitive official database of Libyans and according to the prime minister, “300,000 Libyans were in receipt of more than one state salary and some had more than 60 salaries” [11].

1. Motivation and drivers

- Central Bank of Libya has concerns about identity fraud.
- Increasing number of citizens receive more than one state salary.
- Provide equality of opportunity access to information and prevent duplication and corruption.
- Improve e-services.

2. Requirements

The only requirement to get the new Libyan NID number is the family book. Therefore, all Libyan citizens who have a family book have been issued a NID number by the

government. There are two methods to get your new national number including SMS messages and the national identity number web site. In the first method, an SMS message, containing a family book registration number and the date of birth, is sent to the NID system, which replies with a message containing the NID number. In the second method, a family book registration number, the date of birth and a random number generated by the system is entered through the NID web site, which then displays NID number. After a citizen gets a NID number all the family members included in the family book will be issued NID numbers and all new born babies will be issued a NID number during the registration of birth.

V. ANALYSIS:

Before we start to analyze the two Libyan NID number schemes, we are going to summarize some points in the table above:

1. Duplication of Libyan new NID

The new Libyan NID is based on the existing civil registry records. The civil registry was working manually and not centralized, so it is difficult for the civil registry’s officers to discover if some families have more than one family book. The manual registration of information in the civil registry generally means that it is open to human error and there are cases where some family members are not removed from the family book when they marry or die. These issues negatively reflect on the new NID and we realized them by contacting some citizens who have two NID. When we asked the NID project team members, during our interview, about such cases, they acknowledged that there are a few such cases, but that it is not a system error, but rather a civil registry database error. Simply dismissing such errors as civil registry database errors, however, fail to acknowledge the team project plan error that such problems have not been taken into account, e.g. by sanitizing and normalizing the civil registry database, before the data was imported into the new NID database. Furthermore, the enrollment procedure contains an error because the project team has not required citizens to apply for new NID, rather they produced the NID automatically to all citizens who have a family book. In our model (fig 1, a citizen needs to explicitly apply for a NID by filing a special form and submitting additional biometric information, e.g. a fingerprint, so the system can verify the submitted information against government records to check the correctness of the

document and subsequently check the NID database to prevent duplicated information. One study reported that in many countries, the national identity systems are based on a civil registry, but it is quite difficult to come up with a high integrity civil registry [4]. However, this problem is exacerbated in third world countries, such as Libya, where corruption in public administration is possible and where records are maintained manually. When such countries decide to introduce technology they must identify main challenges of their country in term of culture, regulations, policies and infrastructure. A few third world countries have planned well for the introduction of IDM technology, such as the United Arab Emirates (UAE). In 2003, when UAE government decided to develop a modern identity management system to improve the performance of public administration, they clearly identified all processes for the enrolment, processing, production and delivery of ID cards. The UAE national ID system has avoided duplication of individual's identity by implementing biometrics based on fingerprints [12]. The Indian Universal Identity program (UID) eliminates duplication of identity by implementing a multi-modal system¹ of biometric data to ensure the highest accuracy levels and the smallest room for error [4].

2. *Authenticity of new NID*

The structure of the Libyan NID consists of thirteen digits, the first digit represent the sex (one for men and two for women), the next four digits represent year of birth and the last seven digits should make the id-number unique to differentiate them from others. In the above sections we mentioned how citizens obtain their NID by sending an SMS message which contains the family book registration number and date of birth or through NID website by entering the same information and a random number given by NID system. It is clear that NID system cannot verify citizens based on those information for example, a citizen can send any six digits as a family book registration number and date of birth randomly by SMS or through the NID website, then the NID system will send back NID number if this information corresponds to a record in the database without checking if this information belongs to that particular citizen or not. This issue is quite similar to the issues of SSN at Norwegian Public Service Pension Fund that mentioned in related work section. Another limitation is that the NID website has no limit on the number of attempts to get a NID number, so it is possible to make a brute force attempt at guessing one piece of information if the other piece is known. It is generally considered good practice to have a limited number of attempts for entering authentication information, such as online banks, which often give users only three times of retrying to enhance security of the system and prevent an exhaustive search to get all national number/family book numbers. Furthermore, the social society, Libyan culture, and the rate of illiteracy, means that friends and family members often carry out administrative duties for others. Moreover, social engineering or corruption in some

¹ In this context, Multi Modal Biometric including finger prints, iris and facial recognition to ensure high accuracy.

public administration makes it easy to get family book registration number and date of birth for friends, cousins and neighbors. As a result, the citizen's personal information, as well as family member's information, are not protected those information include name, surname, date and place of birth and national ID number. Often, the identity number is not considered as secrecy but it should be considered as private information. For example, Danish data protection provides legislation that protect identity number against misuse [3]. The random number, provided by the system, cannot prevent citizens from retrying to get other's NID, but is only resist automated impersonation attacks from malicious software from impersonation. One of the main reasons that Libyan NID system could not verify citizens because NID team project has not provided any credentials that can be used to link individual with his personal information. Implementing one of authentication methods including something you know (such as passwords), something you have (such as smart cards), and something you are (such as fingerprints, iris) is important in such projects.

3. *Application*

In 2014, the Committee for the UN high Commissioner for election to the committee of the founding of the Constitution announced that citizens can register for election by sending SMS containing NID and constituency number to the registration center. First, as some citizens have two NID that means that they can vote twice. The second issue is that citizens can send NID of another citizen along with the number of a remote constituency, and thereby prevent them from voting because the NID system has not provided sufficient verification processes as well as election registration system. As a result, a number of citizens, when they tried to register by sending SMS message, received a message that they had already registered and that they should use the previous mobile number to update information, such as changing constituency number. However, this is impossible without knowing who registered them or which mobile number sent their information. When such cases increased, the Committee suggested that any citizen who could not register through sending SMS needed to come to the registration center with proof of identity and then they will cancel the previous registration and give a new registration number for the vote, so they solved the issue manually.

Based on our interview with project team members of the new Libyan national number, and information we collected from different sources regarding first Libyan national number, we realized that, the objectives of the new Libyan national ID and the previous Libyan national ID are quite similar, but the implementation were very different, e.g. the registration process and the method of getting a national ID as illustrated in Table 1. Based on our model, the weakness of new Libyan NID was identified as relying totally on the civil registry without sanitizing or normalizing information records. The registration process in our model requires a citizen to come in person for enrollment and to submit biometric data. This requirement will prevent duplication of data and partially

based on civil registry to check the validity of documents. The second requirement was to provide a credential that will be verified by the system to ensure the authenticity of a citizen. This requirement is also missing from the new Libyan NID number, and we have seen the negative effect of that during e-voting registration example. Therefore, there is no clear technical reasons make Libyan government to change national ID especial the process of registration and verification of the previous project looks much more secure than the new project. Furthermore, personal information of citizen is not protected. In new Libyan NID system

VI. CONCLUSION

The introduction of a national ID is considered an initial step to the introduction of a technology to enable private and public sector organizations to integrate data about the individuals far more easily. Identity systems have a wide range of uses, such as reducing incidents of identity theft, combating terrorism and providing better services to citizens and residents. On the other hand, implementing such projects has a number of risks that did not exist in more traditional paper-based systems. This paper introduced a simple model for issuing a national id-number that satisfies characteristics of good identifiers, such as uniqueness and verifiability. To study whether the implementation of different IDM schemes with the same digital infrastructure, culture and social background will reduce risks of online transaction, we evaluated two implementations of national ID numbers in Libya. We found some weaknesses including vulnerability of personal information, duplication of NID, i.e. that a person could have more than one NID. For example, we have seen how the election registration in Libya is affected by national ID issues and at the end they were forced to resolve problems through manual processes. It is clear from the study that new Libyan NID system has the following limitation:

- Failed to link online presence to real world identity and as result exposed to various forms of fraud. For example with few tries through NID website, it is possible to get other's NID.
- New Libyan NID database based on non-normalized records of civil register. It means issues as incomplete, inconsistency and duplication of individual's record will be inherited in the new digital database. For example, some citizens have more than one NID.
- Privacy of NID has not been considered and as a result it is easy to get other citizen's NID.
- Incompleteness of contact information at individual's record such as mobile phone, home address and emails makes it difficult to contact citizens.

Developing countries that adopt such projects need to plan them well, such as defining criteria for choosing a project team member, identifying the main objectives and the requirements of the project, and considering the digital infrastructure, social culture, and availability of experts. All those factors, and others, need to be considered closely, to benefit from advanced technology. Finally, the project team of

the new Libyan national ID needs to make a review and makes some updates to the processes, such as enabling verification, to make the Libyan NID number a digital infrastructure that meets the planned requirements and avoids negative effects of NID numbers.

VII. REFERENCES

- [1] [1] S. Arora, "National e-ID card schemes: A European overview," *Information Security Technical Report*, vol. 13, no. 2, pp. 46–53, 2008.
- [2] [2] R. Clarke, "Human identification in information systems: Management challenges and public policy issues," *Information Technology & People*, vol. 7, no. 4, pp. 6–37, 1994.
- [3] [3] P. Blume, "The personal identity number in Danish law," *Computer Law & Security Review*, vol. 5, no. 3, pp. 10–13, 1989.
- [4] [4] F. Zelazny, "The Evolution of India's UID Program," *Center for Global Development*, 2012.
- [5] [5] C. L. Sullivan, "Digital Citizenship and the Right to Digital Identity Under International Law," *Sullivan, Clare (2014)'Digital Citizenship and the Right to Digital Identity under International Law', in'Information Ethics and Security' ed Kierkegaard S, ISBN10:87-994854-3/ ISBN 13:978-87-994854-4-4*, 2014.
- [6] [6] S. Chander and A. Kush, "Unique Identification Number and E-Governance Security," *International Journal of Computing and Business Research*, vol. 1, no. 1, 2010.
- [7] [7] P. Beynon-Davies, "Personal identity management in the information polity: The case of the UK national identity card," *Information polity*, vol. 11, no. 1, pp. 3–19, 2006.
- [8] [8] G. Aichholzer and S. Straub, "The citizen's role in national electronic identity management: A case-study on Austria," *2nd International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services - CENTRIC 2009*, pp. 45–50, 2009.
- [9] [9] R. McKenzie, M. Crompton, and C. Wallis, "Use cases for identity management in e-government," *IEEE Security & Privacy*, no. 2, pp. 51–57, 2008.
- [10] [10] A. M. Al-Khouri, "PKI in Government Digital Identity Management Systems," *European Journal of ePractice*, vol. 4, pp. 4–21, 2012.
- [11] [11] "Libyan national identity." [Online]. Available: <http://www.irb.gc.ca/Eng/ResRec/RirRdi/Pages/index.aspx?doc=454961&pls=1>.
- [12] [12] A. M. Al-Khouri, "UAE National ID Programme Case Study," *International Journal Of Social Sciences*, vol. 1, no. 2, pp. 62–69, 2007.

Recognizing Surgically Altered Faces using Local Edge Gradient Gabor Magnitude Pattern

Chollette C. Olisah

Department of Computer Science and IT
Baze University
Abuja, Nigeria
chollette.olisah@bazeuniversity.edu.ng

Peter Ogedebe

Department of Computer Science and IT
Baze University
Abuja, Nigeria
peter.ogedebe@bazeuniversity.edu.ng

Abstract— For humans, every face is unique and can be recognized amongst similar faces. This is yet to be so for machines. Our assumption is that beneath the uncertain primitive visual features of face images are intrinsic structural patterns that uniquely distinguish a sample face from those of other faces. In order to unlock the intrinsic structural patterns, this paper presents in a typical face recognition framework a new descriptor, namely the local edge gradient Gabor magnitude (LEGM) descriptor. LEGM first of all uncovers the primitive inherent structural pattern (PISP) locked in every pixel through determining the pixel gradient in relation to its neighbors. Then, the resulting output is embedded in the pixel original (grey-level) pattern using additive function. This forms a pixel's complete structural pattern, which is further encoded using Gabor wavelets to encode the frequency characteristics of the resulting pattern. From these steps emerges an efficient descriptor for describing every pixel point in a face image. The proposed descriptor-based face recognition method shows impressive results over contemporary descriptors on the Plastic surgery database despite using a base classifier and without employing subspace learning. The ability of the descriptor to be adapted to real-world face recognition scenario is demonstrated by running experiments with a heterogeneous database.

Index Terms—plastic surgery face descriptor face recognition.

I. INTRODUCTION

For humans, every face is unique and can be recognized amongst similar faces. In having machines execute human operations with high precision is the general direction of research for all fields. Most importantly is the area of information security where prevention of unauthorized access electronically or physically cannot be compromised. Therefore, careful consideration into scenarios such as plastic surgery and its effect to face recognition should be of optimum concern to the research community. Why? Because modified faces due to plastic surgery appear distinct or begin to resemble the face of another individual. In such a case, existing feature extraction approaches might fail. So there should be a way to identify those features that remain unchanged after a face undergoes plastic surgery and still does not intercept with features of the face of another individual. However, it might be difficult to identify such features.

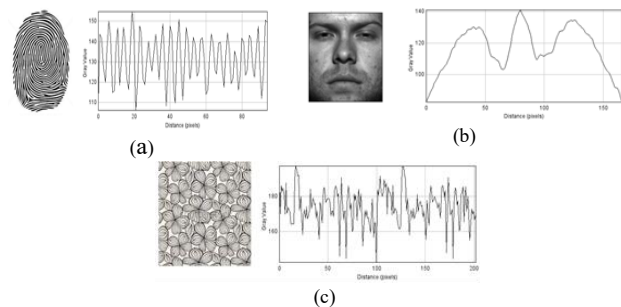


Fig. 1. Profile plot of various patterns.

The face image pattern (two dimensional: 2D) unlike other patterns such as fingerprint image, or a natural scene image, has more uncertain primitive visual features, that is, there isn't clear distinctiveness of facial features on intensity description (grey-level). To demonstrate this claim, given in Fig. 1 is the profile plot of fingerprint, a natural scene image and face image drawn in order to interpret the distinctiveness of their visual features with respect to pixel information.

As can be observed in Fig. 1 the fingerprint, natural scene images show to possess some form of distinctive features such as lines, contours, points, edge, texture or shape patterns and it translates to the distinctiveness shown at pixel level information through the profile plot. For the very reason of uncertainty of face image primitive visual features at intensity grey (level), pattern representation still remains an important problem in face recognition and related areas of image understanding.

With the common goal to tackle the problems in face recognition, a number of research disciplines have emerged with numerous face recognition methods. The holistic based representation methods such as the principal component analysis (PCA) [1] and its classification counterpart, the linear discriminant analysis (LDA) [2]. However, the holistic based representation methods are generally known to perform poorly as feature extraction methods, but are mostly applied as dimensionality reduction methods.

Other methods include: methods representing local appearance information. The local binary patterns and its variants such as the local binary pattern histogram Fourier features (LBP-HF) [3], completed local binary pattern (CLBP), which comprises of CLBP-M-S (magnitude and phase), CLBP-

S (phase), CLBP-M (magnitude) [4]. The Gabor representation and its variants such as histograms of Gabor ordinal measures (HOGOM) [5], local Gabor binary pattern histogram sequence (LGBP) [6], local Gabor XOR patterns (LGXP) [7]. These methods retain different levels of information that are not usually apparent in grey-level (intensity description) face images. However, the type of the local details retained plays a vital role in face recognition tasks, especially in complex instances where many appearance variation factors are entangled. In such cases, the representation method that best disentangles the variation factors in order to represent only significant features will suffice.

Our emphasis is that since LBP, Gabor and some of their variants are texture based descriptors (only varying in magnitude from each other) they might not be able to explore the face image information that suggest useful discriminative cues against plastic surgery effects on the face image with possible expression modality. For instance, let's take the case of a face that must have been subjected to plastic surgery and at image capture may suffer from either expression or variations in lighting conditions. So it isn't only the problem of the uncertainty in primitive visual features of a face image, but also of the ability to exploit facial features that are useful to recognition.

Therefore, this paper proposes a new facial shape and appearance descriptor namely, local edge gradient Gabor magnitude (LEGGM) pattern that exploits a sample face primitive inherent structural pattern (PISP).

The rest of the paper is organized as follows. In section II introduces the art of describing a person's face using LEGGM. In Section III, the experimental application scenario is presented in order to reflect a typical real-world experimental setting. In section IV is the experimentation and analysis, while in Section V is the conclusion.

II. LEGGM DESCRIPTOR

In a given face recognition framework, a plug-in of the local edge gradient Gabor magnitude (LEGGM) descriptor for extracting essential features for face recognition in the event of plastic surgery separated faces is proposed. The LEGGM algorithmic process for extracting essential features is illustrated in Fig. 2 and discussed subsequently. Given an illumination normalized face image, the actual processing for LEGGM descriptor comprises of five major steps: a) PISP computation, b) Complete face structural pattern computation, c) Information encoding using Gabor wavelets, d) Down-sampling, and e) Normalization. They are described briefly as: To detect the PISP of the face image, the edge gradient of each pixel based on its surrounding neighbor is first determined. Successively, an embedding process that uses an additive function to calculate at each pixel point the complete face structural pattern follows. This is explained as the integral of the structural pattern information to the global appearance information (which is of a normalized image). The resulting information from the preceding step, known as the complete face structural pattern is further process in the frequency domain using Gabor wavelets. This is to express at various

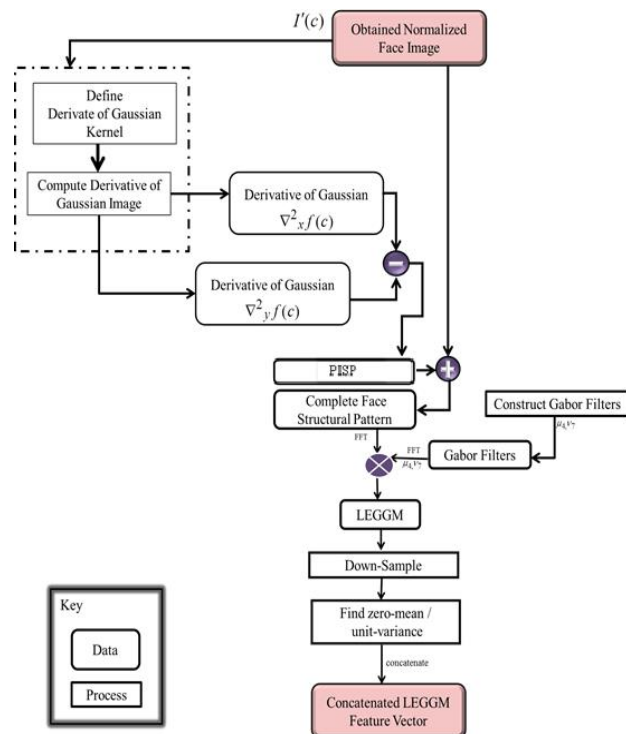


Fig. 2 The descriptor algorithmic process

frequencies of the discriminative properties of the complete face structural information. The resulting information forms LEGGM for describing a face sample. On applying Gabor wavelets at 5-scales and 8-orientations, a face image is described by forty (40) LEGGM features, which are further down-sampled using an interpolation dependent down-sampling approach. This is to mitigate the problem of redundancy resulting from Gabor wavelet. Given that there are forty (40) independent down-sampled LEGGM features for describing a sample face, their respective data will have to be standardized. Therefore, a zero-mean/unit-variance standardization method is employed on the forty (40) down-sampled LEGGM features. Finally, this standardized LEGGM features are further concatenated along the scale to obtain the augmented LEGGM feature vectors used for describing a single sample face image.

The use of the illumination normalized image as opposed to the original grey-level image is due to the fact that edge gradient distribution of an image is a function of illumination and surface reflectance [8]. This means that the image surface properties can limit the distribution of the image gradient. In other words, to be able to capture the actual edge gradients, which reflect an objects surface properties such as shape, curves, regions, boundaries and/or outlines, an illumination insensitive image is preferred. However, it should be noted that the use of the illumination normalized image in the designed descriptor architecture is only on the basis that the image that is an input to the face recognition system might be illumination deficient. Otherwise, the image in its original grey-level is sufficient.

The local edge gradient Gabor magnitude (LEGGM) pattern

at pixel position c for an i th image sample is formally defined as follows:

$$LE \quad M_{u,v}^i(c) = [LE \quad M_{0,0}^{(T)}(c), LE \quad M_{0,1}^{(T)}(c), \dots, LE \quad M_{4,7}^{(T)}(c)]^T \quad (1)$$

and simplified as,

$$= \quad {}^i \mu_{u,v} \quad (2)$$

where ${}^i \mu_{u,v}$ is the augmented features of the forty (40) down-sampled and normalized LEGGM features, which can be used to describe a face image. T is the transpose operator.

III. EXPERIMENTAL APPLICATION SCENARIO

Face recognition task for application purposes can be defined as a function of face identification and verification. While some of the application areas can be strictly categorized under identification task or verification task, some of them cut across the two tasks. The category within which each application area can be described is illustrated in Fig 3.

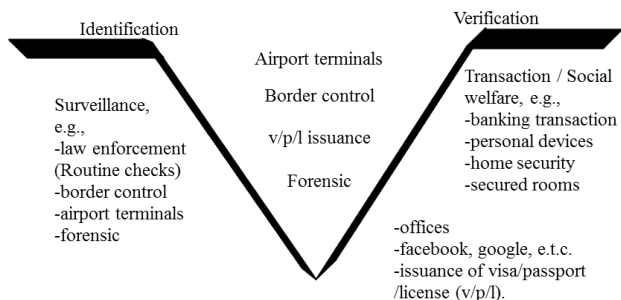


Fig. 3. Face recognition application domains and their respective task at a glance

The application scenario such as the airport scenario as mentioned in Fig. 4 cuts across everyone's living affairs and embodies the two tasks of identification and verification. Imagine after long years of hard work in the busy work-force and someday out of the blues you decided to reward yourself by taking a shot at plastic surgery. Having undergone a facial aesthetic plastic surgery procedure, you decide to go on a casual trip to a tourist destination. Or perhaps you were called in to attend to some work demands at branch miles away that requires you to fly. Now, considering the outcomes of the surgery, two incidents are possible: 1) your facial appearance becomes different after undergoing plastic surgery procedures, and 2) your facial appearance could also tend towards the appearance of a different individual. Suppose then that on your causal trip, incident 2) occurs on your stop at the airport terminal resulting in you being identified as a wanted criminal from a list of suspects, or on your official trip incident 1) causes a denial of your travel rights, that is, your identity could not be verified. What then could possibly be your fate? Let us leave the answer to you, but from a technology point-of-view, the

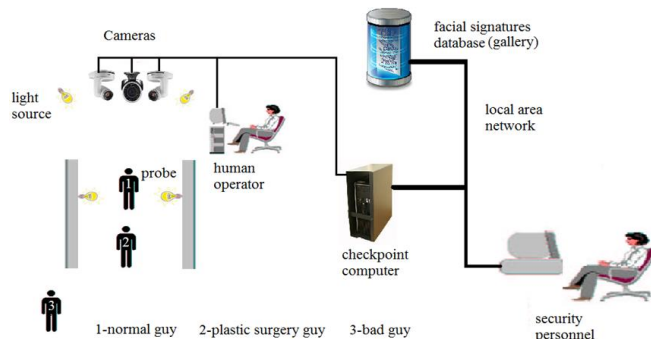


Fig. 4. Overview of typical airport application of face recognition

occurrence of such incidents should be 99.00% avoided. Therefore, recognition of a person even after undergoing plastic surgery should be moving towards such percentile.

In view of the presented arguments above, two scenarios are evaluated. The case of recognizing plastic surgery separated faces and a heterogeneous scenario where different sets of real faces that have undergone plastic surgery and usually experimented-on faces are combined. The heterogeneous case tries to model a practical airport scenario as demonstrated in Fig. 4. To effectively represent such a scenario, four data sets are used, which are the plastic surgery data set [9], the Georgia Tech face (GT) data set [10], the labelled faces in the wild (LFW) data set [11], and a heterogeneous data set. The heterogeneous data set is created by combining subsets of the plastic surgery data set, GT data set, LFW data set with a subset of the Essex data set [12].

A. Plastic Surgery Data Set

The plastic surgery data set [9] contains near frontal faces of real people who have undergone plastic surgery. In all, there are a total of 1800 face images of 900 subjects (excluding cheek and chin surgery procedures with 21 subjects, i.e., 42 samples). A mirror samples of the 921 subjects face images is created making it a total of 3684 face samples. The experimental scenario (ES): Four images per subject, three images are used to make up the train set and also make-up the gallery set. The remaining image is used to make up the test set. It should be noted that there is no subspace learning employed for this experiment.

B. Heterogeneous Data Set

In this data set, images of different subjects from the plastic surgery data set are selected arbitrarily, a total of 321 subjects with plastic surgery. Then full frontal faces are selected from various data sets. From the Essex data set [12] are 231 subjects with illumination problem. An additional 50 subjects are added from the GT data set [10], and 38 subjects from the LFW data set [11]. This brings the total number of subjects to 640, with every subject having 3 images. Experimental Scenario (ES) with subspace learning is given as: 2 images are used to make up the train/gallery set, while the remaining image makes up the test set (probe). For all the subjects the image selected for the test set is unseen during the training phase. Some sample faces from the heterogeneous data set that make-up the heterogeneous database are given in Fig. 5.



Fig. 5. Sample faces from the heterogeneous database

IV. EXPERIMENTAL RESULTS

This section reports the experimental results of applying the LEGGM to face recognition. In all the experiments the identification results and verification results are reported using the cumulative match characteristics (CMC) curve, receiver operating characteristics (ROC) curve or points from the ROC curve, and the equal error rate (EER) evaluation metrics.

A. Evaluation and Benchmarking of LE M with Contemporary Face Descriptors

Using ES of the plastic surgery database, the identification results of different descriptor-based face recognition methods are presented without employing any subspace learning/training. The descriptors are used in their original feature-dimension. The facial descriptors under comparison are the LBP variants, which are the CLBP-M-S, CLBP-M and CLBP-S, while the Gabor variants used are LGBP and LEGGM. The identification rates are reported on Rank basis, where the Ranks 1-10 are considered. The results of employing different facial descriptors in the recognition of faces that have undergone plastic surgery are reported for various plastic surgery procedures and their results shown in Fig. 6. From the figure the following observations are made.

The Gabor based descriptors are observed to be more robust against non-reversible facial appearance changes due to plastic surgery procedures. Their robustness is shown by their above 65% Rank-1 recognition rate that they achieved in a number of the experiments, which is more than what the LBP based descriptors achieved. The identification accuracy of LBP based descriptors is rather disappointing. They failed to reach a satisfactory recognition rate despite existing in a much lower-dimensional space. Overall, LEGGM, a facial shape and appearance descriptor, shows to have achieved the best Rank-1 identification rates. Its highest Rank-1 identification rate is above 87%, which is achieved for the case of recognizing faces that have undergone Dermabrasion surgery.

While surgery procedures to some facial features such as the eye, nose, forehead and the entire-face (which have been found

in psychophysics and computer vision, to contribute largely to face recognition accuracy [13]) minimally affects outlines of the facial features. More of the effects are to the skin regions surrounding the features where the stretching of skin is done to achieve aesthetics. For surgeries that involved such procedures only a minimum-maximum of 8% and 76% correct identification rates were observed for all the descriptors compared. Though, the best performing descriptor is LEGGM facial shape and appearance descriptor, its Rank-1 identification capability did not go beyond 76% for the cases of Blepharoplasty (eye), Rhytidectomy (entire-face), brow-lift (forehead and eye) and Rhinoplasty.

Observed also in Fig. 6 is that LEGGM is mostly unaffected by skin texture changing plastic surgery procedures. The identification rates for texture changing procedures reached 87.50%. The closeness in performance of LGBP to LEGGM shows that they share something in common in comparison with the CLBP-M-S [4], CLBP-M or CLBP-S [4]. The CLBP-S performed surprisingly well from Rank 5 to 10 in the recognition of faces that have undergone Blepharoplasty surgery, while LGBP [6] performed the best from Rank-2 to Rank-10 in the recognition of faces that underwent cheek and chin surgery. Both identification performances of LEGGM and LGBP for the cheek and chin surgery altered faces may not be unconnected with their performances achieved for the texture changing procedures because the region that is modified after chin surgery is not included in the cropped face image.

From Table I, LGBP, CLBP-M-S, CLBP-M and CLBP-S show that they are most appropriate for face verification task than recognition task. Their performances in verification task differ greatly from their performances in the identification task. For instance, take the case of Rhytidectomy where the CLBP-M achieved as low as 8.44% identification rate. In the verification task it achieved as high as 84.09%, 52.60%, 69.81% and 76.62%, verification rates at points on the ROC curve where FAR is 0.1591 (EER), 0.01, 0.05 and 0.1, respectively. Similar performances are observed for the other descriptors such as LGBP, CLBP-S, and CLBP-M-S.

B. Experiments on Heterogeneous Database

Here, it is of expectation that the designed descriptor-based face recognition method will be robust against a number of image formation factors that are present in the system because of its invariant property. The results of the designed descriptor-based face recognition methods in this subsection are on the basis of the subspace learning using principal component analysis plus linear discriminant analysis (PCA plus LDA) [14], locality sensitive discriminant Analysis (LSDA) [15] and supervised locality preserving projection (sLPP) [16]. The results are reported in terms of identification rate, verification rate and EER. The plots of the results are shown in Fig. 7 and Table II.

From Fig. 7 and Table II, it can be seen that the use of PCA plus LDA performed best in all the experiments by a large margin, which can be observed from the Rank-1 up to Rank-10. The use of sLPP performed second best followed by LSDA. In

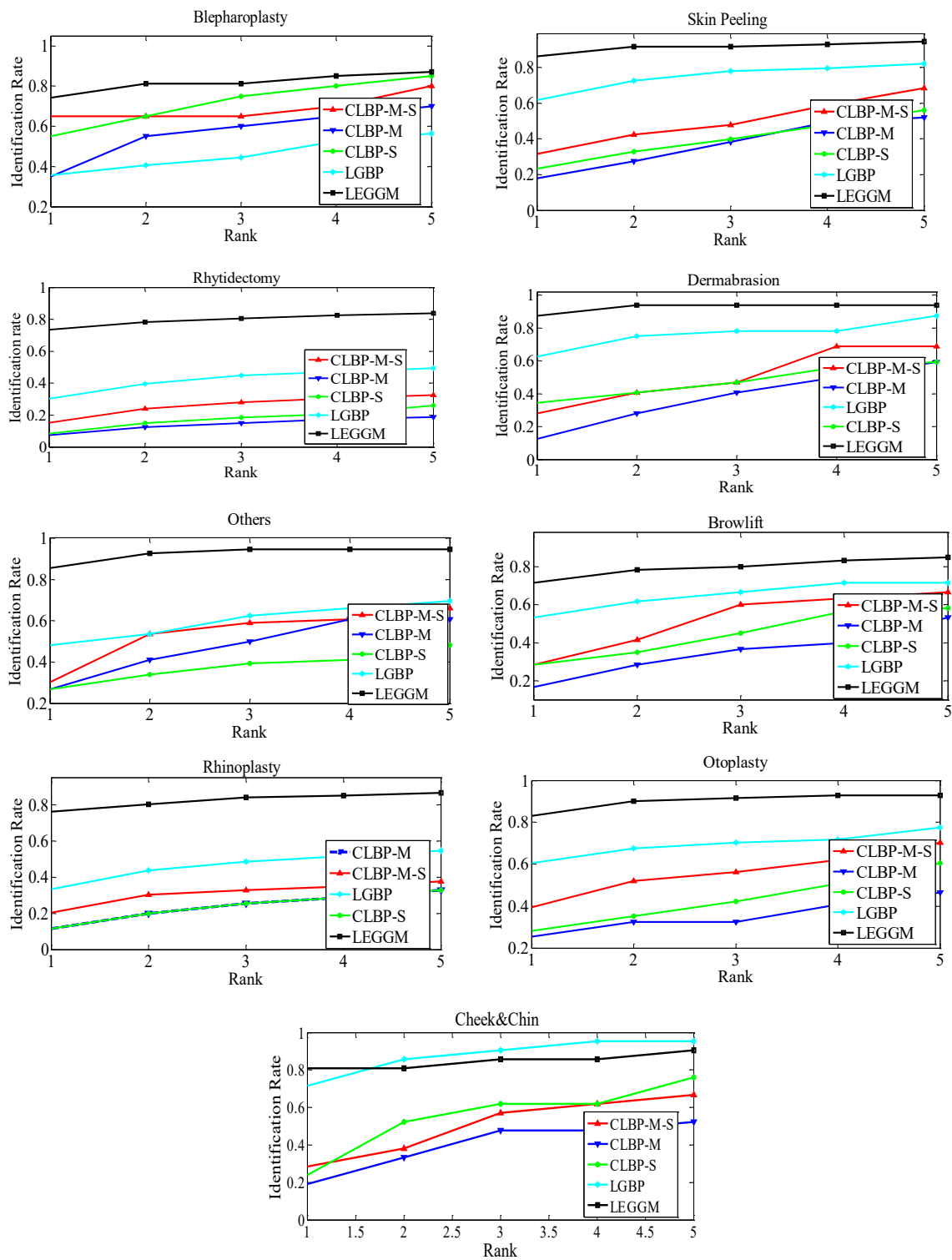


Fig. 6. Identification performances of LEGGM descriptor and existing descriptors without employing subspace learning for different plastic surgery procedures

Table I
Recognition performances of LEGGM descriptor and existing descriptors for different types of plastic surgery procedures

PSP	Method	@ FAR 0.01	@FAR 0.05	@FAR 0.1	EER	Rank-1
BL	CLPB-M-S	75.00	80.00	100	0.0921	65.00
	CLBP-M	60.00	70.00	80.00	0.1500	35.00
	CLBP-S	85.00	85.00	90.00	0.1000	55.00
	LGBP	73.29	90.10	95.05	0.0693	35.64
SP	LEGGM	71.29	86.14	89.11	0.1089	74.26
	CLPB-M-S	80.82	87.67	95.89	0.0818	31.51
	CLBP-M	60.27	80.82	82.19	0.1384	17.81
	CLBP-S	71.23	87.67	91.78	0.0947	23.23
RY	LGBP	94.52	97.26	100	0.0274	61.64
	LEGGM	83.33	88.89	93.06	0.0695	86.11
	CLPB-M-S	74.03	86.04	90.26	0.0973	15.26
	CLBP-M	52.60	69.81	76.62	0.1591	8.44
DE	CLBP-S	69.16	81.49	87.34	0.1135	18.48
	LGBP	84.42	93.83	95.78	0.0559	30.19
	LEGGM	76.62	87.99	92.53	0.0844	73.38
	CLPB-M-S	68.75	87.50	96.88	0.0670	28.13
OT	CLBP-M	53.13	75.00	84.38	0.1563	12.50
	CLBP-S	59.38	81.25	84.38	0.1250	34.38
	LGBP	87.50	90.63	90.63	0.0938	62.50
	LEGGM	75.00	84.38	87.50	0.1250	87.50
BR	CLPB-M-S	78.57	89.29	92.86	0.0893	30.36
	CLBP-M	67.86	80.36	82.14	0.1250	26.76
	CLBP-S	71.43	89.29	91.07	0.0899	26.76
	LGBP	78.57	87.50	89.29	0.1071	48.21
RH	LEGGM	76.36	89.09	90.91	0.0928	85.45
	CLPB-M-S	73.33	83.33	90.00	0.1000	28.33
	CLBP-M	60.00	75.00	75.00	0.1700	16.67
	CLBP-S	66.67	81.67	86.67	0.1169	28.33
OTO	LGBP	85.00	91.67	93.33	0.0814	53.33
	LEGGM	58.33	78.33	83.33	0.1333	71.64
	CLPB-M-S	73.44	87.50	91.15	0.0889	20.31
	CLBP-M	68.23	82.29	88.02	0.1095	11.46
CC	CLBP-S	68.23	82.29	88.02	0.1095	11.46
	LGBP	83.33	94.27	95.31	0.0573	33.33
	LEGGM	78.65	88.54	92.71	0.0876	76.04
	CLPB-M-S	81.69	91.55	95.77	0.0704	39.44
TOTAL	CLBP-M	56.34	78.87	85.92	0.0985	25.35
	CLBP-S	69.01	85.92	90.14	0.1126	28.17
	LGBP	84.51	92.96	92.96	0.0704	60.56
	LEGGM	76.06	88.73	92.96	0.0739	83.10
TOTAL	CLPB-M-S	76.19	95.21	100	0.0470	28.57
	CLBP-M	52.38	57.14	66.67	0.1905	19.05
	CLBP-S	71.43	85.71	90.48	0.0952	23.81
	LGBP	95.24	95.24	100	0.0476	71.43
TOTAL	LEGGM	80.95	95.24	95.24	0.0476	80.95
	CLPB-M-S	75.76	87.57	94.76	0.0815	31.88
	CLBP-M	58.98	73.37	80.10	0.1441	19.23
	CLBP-S	70.17	84.48	88.88	0.1064	26.62
TOTAL	LGBP	85.15	92.61	94.90	0.0678	50.76
	LEGGM	75.18	87.48	90.82	0.1653	79.83

PSP-plastic surgery procedure, BL-Blepharoplasty, SP-skin peeling, RY-Rhytidectomy, DE-Dermabrasion, OT-Otoplasty, BR-brow lift, RH-Rhinoplasty, OTO-others, CC-cheek&chin, EER=equal error rate, FAR=false acceptance rate

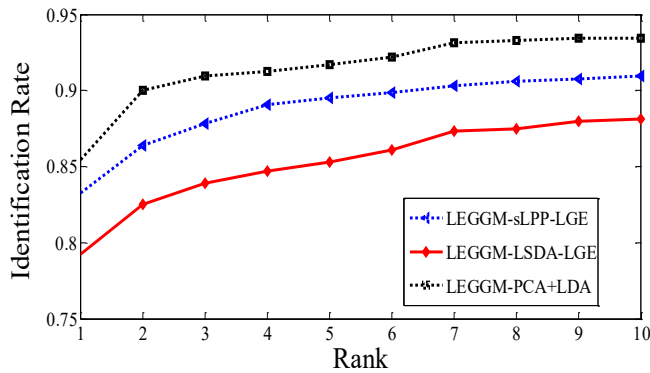


Fig. 7. Identification performance of the descriptor-based face recognition method for the heterogeneous database

Table II

Performance of the descriptor-based face recognition method in a heterogeneous case

Method	@FAR 0.01 (%)	@FAR 0.05 (%)	@FAR 0.1 (%)	EER (%)	VR (%)	Rank-1 (%)
LEGGM-sLPP-LGE	88.12	92.34	94.69	6.86	93.14	83.28
LEGGM-LSDA-LGE	83.13	89.06	91.56	8.89	91.11	79.22
LEGGM-PCA+LDA	93.44	95.78	96.88	4.21	96.79	85.47

FAR=false acceptance rate, EER=equal error rate, VR=verification rate

comparison with the previously reported experiments, LSDA can be seen to have significant increase in recognition accuracy. The obvious reason one could point at is the fact that there are more percentages of frontal-view images in the heterogeneous database than is included in the other databases (GT or LFW). That notwithstanding, far better recognition accuracies are envisaged to be achieved for the entire system if the image sets in the database are restricted to only the frontal-view images as it is commonly practiced in literatures, but that will make the system less practical.

Overall, the experiment on the heterogeneous data sets validates that the intrinsic facial characteristics of the descriptor-based face recognition method captured and retains for recognition can, to a good extent, be robust against a wide range of facial variation that is possible in a real-world face recognition scenario.

V. CONCLUSION

Through experimental analysis it was shown that the essential cues at local points of the face image LEGGM encodes are more effective for describing faces that have undergone plastic surgery than existing descriptors. It was further shown that the contemporary descriptors, which are either dependent on pixel intensity (greyscale) or texture dependent, do not sufficiently address face recognition problem in the event of plastic surgery. It is also observed that the proposed descriptor-based face recognition method showed that it can be adapted to real-world face recognition scenarios.

REFERENCES

- [1] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA, Springer-Verlag, 1986.
- [2] D. Swets J. Weng, Using Discriminant Eigenfeatures for Image Retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1996), 831-836. [19] L. Zhang, R. Chu, S. Xiang, S. Liao, S. Li, "Face Detection Based on Multiblock lbp Representation, Proceedings of International Conference on Advances in Biometrics, 2007, pp. 11-18.
- [3] T. Ahonen, J. Matas, C. He, M. Pietikäinen, Rotation Invariant Image Description with Local Binary Pattern Histogram Fourier Features, *Proceedings of the 16th Scandinavian Conference on Image Analysis*, 2009, pp. 61-70.
- [4] Z. Guo, D. Zhang, A Completed Modeling of Local Binary Pattern Operator for Texture Classification, *IEEE Transactions on Image Processing*, 19(2010), 1657-1663.
- [5] Z. Chai, R. He, Z. Sun, T. Tan, H. Mendez-Vazquez, Histograms of Gabor Ordinal Measures for Face Representation and Recognition, *Proceedings of the 5th International Conference on Biometrics*, 2012, pp. 52-58.
- [6] W. Zhang, S. Shan, W. Gao, X. Chen H. Zhang, Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-statistical Model for Face Representation and Recognition, *Proceedings of the Tenth IEEE International Conference on Computer Vision*, 2005, pp. 786-791.
- [7] S. Xie, S. Shan, X. Chen, J. Chen, Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition, *IEEE Transactions on Image Processing*, 19(2010), 1349-1361.
- [8] W. Chen, M. Er, S. Wu, Illumination Compensation and Normalization for Robust Face Recognition using Discrete Cosine Transform in Logarithm Domain, *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 36(2006), 458-466.
- [9] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A. Noore, S. Nooreydzan, Plastic Surgery: A New Dimension to Face Recognition, *IEEE Transactions on Information Forensics and Security*, 5(2010), 441-448
- [10] A. V. Nefian, Georgia Tech Face Database, http://www.anefian.com/research/face_reco.htm. (accessed 02.02. 2013)
- [11] G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, *Proceedings of the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008, pp. 1-14.
- [12] A. V. Savchenko, Directed Enumeration Method in Image Recognition, *Journal of Pattern Recognition*, 45(2012), 2952-2961.
- [13] S. J. Lederman, R. L. Klatzky, R. Kitada, Haptic Face Processing and Its Relation to Vision, in: *Multisensory Object Perception in the Primate Brain*, New York, USA, Springer, pp. 273-300, 2010.
- [14] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1997), 711-720.
- [15] D. Cai, X. He, K. Zhou, J. Han, H. Bao, Locality Sensitive Discriminant Analysis, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007, pp. 708-713.
- [16] Z. Zheng, F. Yang, W. Tan, J. Jia, J. Yang, Gabor Feature-based Face Recognition using Supervised Locality Preserving Projection, *Journal of Signal Processing*, 87(2007), 2473-2483.

Adaptable Exploit Detection through Scalable NetFlow Analysis

Alan Herbert
Rhodes University
Grahamstown, RSA
Email: g09h1151@campus.ru.ac.za

Barry Irwin
Rhodes University
Grahamstown, RSA
Email: b.irwin@ru.ac.za

Abstract—Full packet analysis on firewalls and intrusion detection, although effective, has been found in recent times to be detrimental to the overall performance of networks that receive large volumes of throughput. For this reason partial packet analysis technologies such as the NetFlow protocol have emerged to better mitigate these bottlenecks through log generation. This paper researches the use of log files generated by NetFlow version 9 and IPFIX to identify successful and unsuccessful exploit attacks commonly used by automated systems.

These malicious communications include but are not limited to exploits that attack Microsoft RPC, Samba, NTP (Network Time Protocol) and IRC (Internet Relay Chat). These attacks are recreated through existing exploit implementations on Metasploit and through hand-crafted reconstructions of exploits via known documentation of vulnerabilities. These attacks are then monitored through a preconfigured virtual testbed containing gateways and network connections commonly found on the Internet. This common attack identification system is intended for insertion as a parallel module for Bolvedere in order to further increase the Bolvedere system's attack detection capability.

Index Terms—Digital forensics, Network security, Intrusion detection

I. INTRODUCTION

This research builds on the module repertoire that is compatible with the Bolvedere platform currently in development by the authors of this paper. The Bolvedere platform is a highly adaptable and scalable NetFlow processor intended for distributed identification of malicious network activity. All modules developed for this platform run concurrently, be it remote to the Bolvedere host system or within it. The implementation brought forward in this research is in respect to development of a new module to run on the Bolvedere system [1].

A. Problem Statement

Issues that need to be addressed within this area of research are that of latent security in large scale networks. Included in this is network traffic monitoring to protect users from vulnerabilities that can cause their systems to perform malicious tasks within the network. Exploitation of these vulnerabilities can lead to inclusion within botnets, proxying of malicious data, and other malicious activities that lead to disruption of natural flow of data within the Internet.

For this reason automated systems are required to monitor interactions within networks that source traffic onto the Internet. These systems should then provide a feedback mechanism to better detect and mitigate malicious activities introduced into the Internet.

B. Research Goals

This research in its entirety aims to develop a modular platform for which modules that process network flows can interface in order to discern events on a network. Adaptability is taken into account in the form of system resources required to run a Bolvedere system. Bolvedere is able to execute itself on a single host machine that is Linux OS (Operating System) compatible as well as further scale out over multiple threads, multiple processes, multiple processors and multiple separate physical hosts. This adaptability and scalability also includes support for multiple languages as well as hardware technologies such as GPUs (Graphical Processor Unit) and FPGAs (Field-Programmable Gate Array). Furthermore, this scalability ensures the ability for Bolvedere to take on the task of Internet level network flow discernment as to whether a network flow is malicious or not.

The first question proposed by this research was if one could use NetFlow logs to detect a malicious exploit. With this question in mind this research's first goal was to collect NetFlow logs generated from network flows created by malicious network exploits. Sections IV-B1 and IV-B2 discuss how these resultant NetFlow logs were observed and what distillation process occurred when attempting to produce rule sets from different exploits.

The application of these rule sets within a Bolvedere module is then brought to light in Section IV-C which discusses the accuracy of such an automated mechanism, as well as where this module falls short and how these failings can be mitigated.

This paper begins with a literature review in Section II which deals with background knowledge required to better understand this research and why it is necessary. Following this, Section III discusses how this Bolvedere module was implemented through an understanding of the tools and libraries used within it. Finally, results are discussed in Section IV and a final conclusion is presented in Section V.

II. LITERATURE REVIEW

As the main data descriptor of this research's implementation is based on NetFlow [2], the first topic dealt with in this paper is the aforementioned technology. Once this technology is outlined this paper then moves on to discuss and explain current malicious activities that occur on the Internet and how they are accessed and acted upon.

A. NetFlow

The NetFlow protocol is best described as a means of logging network flows that pass through a flow monitoring

device in a communication pair's route. A network flow is defined as a unidirectional connection and communication between a host and any other host, multicast group, or broadcast domain in the form of a sequence of packets [3]. A flow monitor implementing the NetFlow protocol can collect fields out of these communications, write them into predefined fields (restricted either by NetFlow protocol version or by use of a known template) and then transmit them to a logging host for analysis or storage [4]. There have been multiple versions of NetFlow with wide-spread support over multiple firewall and routing devices on the Internet.

The need to update the NetFlow protocol over the years arose from multiple factors. First, the addition of IPv6 (Internet Protocol version 6) [5] that was brought about by the IP address exhaustion [6] of the IPv4 (Internet Protocol version 4) [7] space required amendments to be added to the NetFlow protocol. Furthermore, the need for better use of network resources grew as the amount of traffic passing through flow monitor points increased. Finally, the requirement to adapt these records to one's needs gave way to updating the NetFlow protocol to give users the ability to break out of the predefined logging fields determined by older versions of NetFlow into a dynamic space that allows for field collection through a user created predefined and distributed template [4]. This method of log collection also has additional unused record space for later allocation of new network protocols released at future dates.

Major updates to the original NetFlow standard have included the addition of new fields and further standardisation and this resulted in version 5 of the protocol. This version allowed for logging subnet masks and AS (Autonomous System) numbers [8]. Version 8 saw inclusion of aggregation of records that were first defined in version 5 [9]. More recently, version 9 continued to build on the freedom brought forth by version 8 through the addition of the template packet to the NetFlow protocol. This template packet allowed one to define the fields to be logged in a record and the order in which they are logged from a flow [4].

These templates are coupled with template identification numbers that allow for use of multiple templates and is defined by the 2-byte-long template identification field within the NetFlow protocol. Although IDs 0 through to 255 are reserved for use by specific predefined flow templates, template identification numbers 256 through to 65535 are available for public use; a fairly large template space. This template count further extends the memory requirements of these devices and most devices supporting NetFlow version 9 limit the number of templates that can be stored to a count far less than the available 65535 due to memory and performance limitations [2].

B. Malicious Attacks

A malicious event acted upon something or someone is simply defined as an action with intent to do harm to that entity. Within the Internet these entities are commonly end-point hosts. These hosts are typically targeted in order to collect information or prevent other entities on the Internet from accessing the information provided by the host; be it private or public. This section intends to focus on two main reasons for gaining access to a system through malicious means and these are denial of service and information theft.

1) *Denial of Service*: The goal of a DDoS (Distributed Denial of Service) attack is the same as that of a (DoS Denial of Service) attack in that they both aim to bring down a service that exists on a network. The key difference between a DDoS attack and a DoS attack is that a DDoS attack uses multiple physical source hosts rather than a singular host. These hosts usually exist within a botnet [10]. Note that a DoS attack can appear to be a DDoS attack through the spoofing of multiple IPs; this makes detecting a DDoS attack difficult.

2) *Identity Theft*: Identity theft refers to any form of theft that enables the thief to perform actions of that of the victim while holding all the credentials needed to be identified as that victim. This on the Internet includes usernames and passwords, private identifying information and man-in-the-middle attacks to gain access to tokens passed between hosts in order to gain access to the victims session [11].

3) *Information Theft*: This is simply stealing information that is private, be it from a single person, group or company. This information is usually targeted and sensitive to public viewing or viewing by a competitor.

4) *Information Destruction*: On the other end of the spectrum when compared to information theft, information destruction can be performed through ransomware or simply destruction of a targets information store, be it deletion of a database or project. Ransomware aims to encrypt data with a key that the attacker holds and typically requires payment in order to retrieve the key to decrypt one's data with. If one fails to acquire this key, be it because of the event timing out or one's unwillingness to pay for the data, then the data that is encrypted is effectively as good as destroyed [12]. Complete outright destruction of data is usually a method used to bring down companies as if a company were to lose its database of operations, it would be as if the company never existed from that point on. It is note worthy that the cost of data recovery in the event of data destruction can be enough to put a company under in itself [13].

C. Common Attack Vectors

There are multiple ways to perform analysis on potential targets and multiple methods to go about executing an attack on these targets. These all fall into specific categories of which the ones this research is aimed to analyse and identify are listed below.

1) *Brute-Force Attacks*: This form of an attack is applicable occurs in multiple forms and is defined by an attack method having little to no intelligence [14]. The method of this attack is to simply to try every combination from start to finish of the attack space until a solution is found. As this attack intends to attempt every combination as input to a system in order to gain access to information held by the system, this approach is time consuming and thus attempts to leverage high throughput hardware in order to achieve this task in a acceptable time [15] (this hardware includes GPU through the use of CUDA and OpenCL).

This form of attack can however be broken down into two subsets, these are offline and online brute-force attacks. A offline brute-force attack is typically performed against a data that exists on the attackers storage devices and is locally accessible; these include databases of hashed passwords. For this example an attacker would typically attempt to recover the password used to generate a hash through generating every

possible input for the respective hashing algorithm until the output matches that of the hash that one is trying to recover a password for. At this point the input used is the original password that the attacker is looking for.

An online brute-force attack refers to an attack on a remote service or system. Typically this is done through guessing login credentials until access is gained. This kind of approach can be seen used on systems that require a username and password such as SSH or a website. One would try to generate combinations of username and password until a successful login occurs.

2) *Vulnerability Exploitation:* Humans are not perfect and as hardware and software are developed by people, there are imperfections introduced into the systems. These imperfections lead way to unexpected behaviour that when acted upon lead to results that fall outside of the systems intended operation. These result can range from simple errors in output to code being remotely uploaded and executed on the system or the system being shut down completely.

Malware such as Blaster Worm [16], Conficker [17] and SQL Slammer [18] use these vulnerabilities to upload and execute themselves upon remote hosts. These vulnerabilities exploited by these malware include MS03-026 [19], MS03-039 [20], MS08-067 [21] and the Microsoft SQL Server Resolution Service [18]. Even though these are well known vulnerabilities that have long since been fixed there are still many existing systems on the Internet that are still vulnerable to these attacks due to improper maintenance of said systems. These iconic attack methods were chosen to show their age (dating back to January 2003) and further exclaim the neglect shown by some system administrators.

3) *Social Engineering:* There are many methods both physical and digital of social engineering but this text will focus on two methods used on the Internet to explain what it is. These methods are phishing and baiting. The idea behind social engineering is to attack the human psyche through misleading someone to act in a way they would usually not, or exploiting one's natural characteristics into acting upon something that should not be acted upon. The former is prevalent in phishing where the latter is used in baiting.

Phishing gets someone to give up private information by pretending to be something or someone it is not. A simple example of a phishing attack is a website pretending to be an existing bank that it is not. If it were to successfully trick someone into entering their banking details the third-party that set up the website would then gain access to that private information [22].

Baiting on the other hand relies on a human trait known as greed. If someone really wants something that they can't get and a malicious source offers that something, it opens up a vector of attack. This is commonplace in pirated software available on the Internet. Someone wants to use a piece of software that they have to pay for, why not offer it for free and attach malware to the executable. Why even hide behind the faade of a executable when one could just rename the malicious executable and change the display icon to mimic that of the original software; once the user clicks run it doesn't really matter what the user sees next as the attack is already successful [23].

D. Availability of Attacks

The large number of malicious attacks occurring on the Internet on daily base is due to two reasons. The first is the ease of which one can perform these attacks. For the most part someone with little to no knowledge of how a vulnerability is exploited on a network can simply get hold of tools and scripts to run at the press of a button that one points at a target. The second is that many malicious attacks are automated, be it via botnet [10] or by a malicious preconfigured system.

There is a market for these systems, botnets and zero day attacks (a vulnerability that is yet to be exploited and is unknown to the vendor) and these can fetch a high price depending on the capabilities of the exploit, however there are freely available tools for configuring and performing malicious exploits of systems on a network [24]. These exploits are for the most part well known and fixed and for a target to fall victim to these exploits is due to their own negligence in terms of keeping their system up to date. This section will deal with the two common ways exploits are performed in a legal penetration testing environment.

1) *Metasploit:* This software suit that is directed at penetration testers to test the security of networks and users on it. It houses a wide variety of tools that allow for assessment of software and systems running on a network, as well as the awareness of the users on a network through features like the generation of phishing campaigns to test users on a network. Coupled with these features is a database of fully functional exploits that are known to the penetration testing community. This means that one can install Metasploit, which is free, and launch these attacks on a target host on a network at one's leisure [25]. This can of course be for Metasploits intended purpose, that being security conciousness, or for malicious reasons.

For intended reasons Kali Linux exists as a penetration testing operating system that comes pre-installed with Linux based software that one would use for penetration testing of a system or network [26]. Furthermore, the Metasploit community supports this Linux distribution and thus regular updating of the distribution and tools on it is freely available.

2) *Reuse of Code in the Wild:* There is a need to understand existing malware on the Internet and for this collection and reuse of such malware in a safe environment, while monitoring the characteristics of the captured malware, can be invaluable in combating it [27]. There are multiple methods of capturing malware and the class of software that typically performs this action is referred to as a honeypot; although this is not the only way to capture a piece of malware. A honeypot acts as a vulnerable system in order to attract malicious attacks [28]. Any attacks that are targeted at the honeypot are then logged and any data upload is too. From this point one can set up an environment such as a virtual network, or if one has the resources a live, environment in which to rerun these malware and analyse their characteristics.

Other methods of malware collection range from malware sharing communities to collecting the remnants of uploaded scripts and programs to a server that may have resulted in a failed or successful attack. Either way, deletion of these malware are a loss to the community trying to combat these forms of malicious attack and one should attempt to pass on the malware to a third-party that has a use for it (hopefully not malicious in nature).

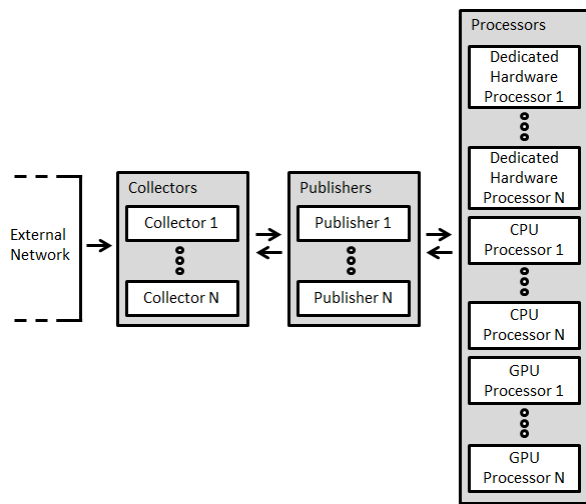


Fig. 1: Bolvedere System Overview

III. BOLVEDERE MODULE IMPLEMENTATION

This research collects and analyses NetFlow logs in order to discern whether a network flow could have attempted a malicious attack or not. It is understood that a non-malicious network flow can give the same NetFlow log results as NetFlow log generated from a malicious flow. It was decided that dealing with false positives was better than dealing with false negatives in this system as this would mean that a successful attack would go unrecognised. The results of these true and false positives would both be presented to the user with all related information for the user to discern, be it through further processing or if the traffic is low enough by one's self, as this module is intended to identify malicious attacks and not to take mitigating action.

In order to achieve this and ensure adaptability and scalability, this module intended for use with Bolvedere required some careful planning. A short discussion on tools used and method of configuration will be pointed out in this section as to help the reader better understand how this intended adaptability and scalability was achieved.

A. Tools Used

The two major potential bottlenecks identified in this system was on the network interface and the rule set processor. Referring to Figure 1 with the understanding that this module is a processor in the Bolvedere system means that the network component of the system has to act in a distributed manner between a publisher and the connected processors. For this the use of a broadcast groups was used to allow for a single network packet sent by a publisher to arrive at multiple destinations.

The library used for handling these broadcast groups was ZMQ (Zero Message Queue). The ZMQ library gives access to a distributed networking model that makes use of sockets and broadcasting to allow for increased concurrency within a system. Furthermore, these transmitted messages ensure atomicity over the entire broadcast group. The ZMQ transport layer can be set to in-process, inter-process, TCP and multicast modes depending on whether the communications are happening within a process or between processes on a single host, or

between processes on separate hosts [29]. Furthermore, ZMQ is a library that is supported by over 30 languages and the entirety of the protocol is documented. This means that no matter what processor module one intends to implement for Bolvedere, one can choose the best language for the job.

In order to quickly access the rule sets to discern whether a network flow is malicious or not, a SQL (Structured Query Language) database was used. There are many variants of SQL databases which range from a file on disk, as SQLite [30] implements, to entire databases loaded into RAM to achieve maximum throughput, as implemented MemSQL [31]. Given these implementations all share a common language, one can swap out the back-end of this module according to the needs and/or limitations of the host system.

The rest of the software which applies the rule set to incoming NetFlow logs of this module was written in C and a Python based prototype also exists.

B. Configuration

Three steps are required for configuration which are listed below:

- Select publishers that the module should listen to in order to receive processed NetFlow logs.
- Load in which rules the module should implement in order to discern these received NetFlow logs so that detection potential malicious activity can occur.
- Select a method of notification for when detection occurs.

The first point expanded upon will be that of publisher selection. A module in the Bolvedere system is allowed to subscribe to more than one publisher to receive its processed NetFlow logs from. The format in which the information arrives is predetermined at configuration time of a publisher and is intended to be in a format best suited for use by listening modules.

The second point is that of loading the rule sets. One of these malware detection modules may attempt to detect one or many forms of malware signatures. If the rule set gets too large and one wishes to break up the rule set into smaller chunks of work in order to better scale the systems ability to process this module, all one simply needs to do is start more of these modules that only deals with a subset of the entire rule set. Furthermore, these new modules can run concurrently with the rest of the modules on a separate physical host completely, or on a separate processor thread within the same host.

Lastly the method in which these findings are presented to the user is optional. Currently the options of receiving daily reports via an email or presenting findings to a terminal at runtime are available for selection. The use of a terminal allows piping of outputs from this module into other applications for further processing. Any option in between can trivially be implemented at a later stage, however the question of did this implementation achieve its goal did not rely on this functionality.

IV. RESULTS

This paper sought to collect NetFlow logs generated by malicious network flows on a network and then analyse them for generating rule sets for use in a Bolvedere system module to detect further attempts of these attacks. For this reason this results section will be broken down into two major parts. The first part will deal with the collection and analysis of NetFlow

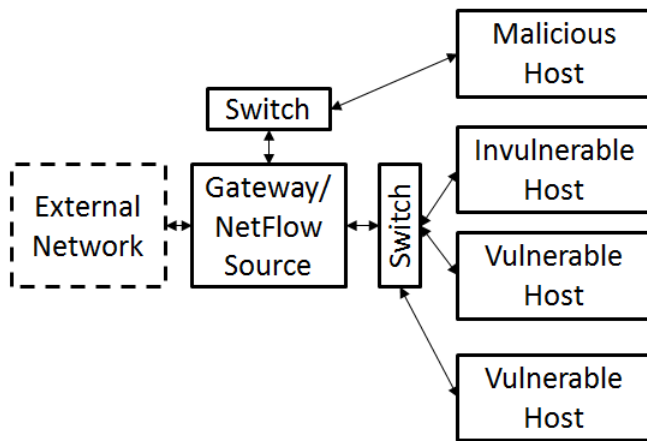


Fig. 2: Virtual Network Overview

logs generated by controlled malicious attacks in order to build rule sets to detect further attempts of these attacks. The second part of these result will be targeted at the running of the automated Bolvedere system module that implements these rule sets to detect repetitions of these recorded malicious attacks.

A. Environment

These tests are all performed within a virtual environment. A basic configuration of the virtual network this virtual environment uses can be seen in Figure 2. One can see that hosts are broken up into 4 groups and these are listed below with a short description:

- Malicious Hosts: These hosts launch malicious attacks on vulnerable and invulnerable hosts typically through the use of Metasploit or custom written code as to the documented exploitation attack vector.
- Vulnerable Hosts: These hosts are built to be vulnerable to a monitored attack.
- Invulnerable Hosts: These hosts are built to be made resistant to a monitored attack.
- Gateway/NetFlow Source: This host acts as a gateway to an external network (advertises it is connected to the Internet) and also runs the NetFlow log generator to store the network flows that pass through it. This host uses softflowd to generate all NetFlow logs [32].

One can observe that in order for the malicious host to communicate with a vulnerable or invulnerable system it has to first pass through the gateway system running softflowd. This means that the network flows generated between these hosts can be fully logged and analysed into rule sets at a later point.

B. NetFlow Logs and Rules Generated

The purpose of this section is to collect the NetFlow logs generated by softflowd when observing the network flows caused by malicious attacks. The defining features of these generated logs are then extracted by an automated rule generator and used to form rule sets that can discern further attempts of the attacks that generated the network flow. The rule generator simply observes a repeated attack and collects relevant metadata about the attack before generating a rule

for that form of attack. The NetFlow logs generated that the automated rule generator observed have been tabulated and can be referred to in Tables I to VIII. These results were gathered over 6 iterations of which 3 iterations were designed to be successful and 3 were designed to fail. The failures did not show a large variance in results and so due to the space limitations of this paper have been omitted but will still be discussed later in this section. Also, due to the nature of networking technologies there is some variance in the collected results, this is handled through displaying the result as a range rather than a set value where necessary.

Terminology used in these results is explained below:

- 1) Attacker: The host that is implementing an exploitation.
- 2) Target: The host which the attacker is attempting to exploit.
- 3) Victim: A third-party that is affected due to an Attacker's exploit.
- 4) A, B and C: These refer to randomly assigned ports by the operating system when a connection is created without being told to use a specific port.
- 5) N: This refers to all numbers after the last until process is terminated.
- 6) X and Y: These refers to counters of varying size relating to packet and byte counts.
- 7) Exploit: Used to define the stage of the exploit in which the exploitation is being attempted.
- 8) Payload: Used to define the stage of the exploit in which the payload is being transferred and executed on the system.
- 9) Runtime: Used to define the stage of the exploit in which the payload is running.

1) *Results:* Unsurprisingly the first point to note is that it is the attacker that always starts the communications in these exploits. The method is usually performed through fingerprinting a target to identify which services are running on a system (these logs are not interesting and so have been omitted). Once a vulnerable service has been identified the exploit then is executed and if successful the payload is then uploaded to the vulnerable host and an attacker gains access to the target in their chosen method. As these vulnerabilities are found in services running on a host and these services run on specific ports, it is noteworthy that these specific ports is what an exploit targets.

A significant point that arose when an exploit was repeated was that the initial NetFlow log's packet count, byte count and service port were consistent (the service port consistency is important as some services utilize multiple ports). This means that one can say that for a new NetFlow log between two hosts, if a set port is connected to that receives a set packet count with set total byte count, one should check that targeted host for an occurrence of an attack that is represented by this signature. Although one should also note that as a NetFlow log only contains the metadata of a network flow, a perfectly legitimate network flow could also cause this NetFlow log to be generated.

Some finer details to notice is that MS08-067 (exploitation of Microsoft RPC (Remote Procedure Call) service) exploitations tend to have their packet count and byte count vary more than exploits utilizing other vulnerabilities in these results. Another point is that the NTP (Network Time Protocol) monitor list attacks were generated using 3 separate monitor

TABLE I: NetFlow Logs Generated by a Successful ms08_067_shell Exploit

Flow Index	Exploit Stage	Direction	Source Port	Destination Port	Packet Count	Byte Count
1	Exploit	attacker → target	A	445	43 - 47	9900 - 10100
2	Exploit	target → attacker	445	A	42 - 44	7600 - 7700
3	Payload	attacker → target	B	Set in Exploit	8	695
N	Runtime	Bi-Directional	B/C	Set in Exploit	X	Y

TABLE II: NetFlow Logs Generated by a Successful ms08_067_vnc Exploit

Flow Index	Exploit Stage	Direction	Source Port	Destination Port	Packet Count	Byte Count
1	Exploit	attacker → target	A	445	43 - 47	9900 - 10100
2	Exploit	target → attacker	445	A	42 - 44	7600 - 7700
3	Payload	attacker → target	B	Set in Exploit	278	416549
N	Runtime	Bi-Directional	B/C	Set in Exploit	X	Y

TABLE III: NetFlow Logs Generated by a Successful java_rmi_server Exploit

Flow Index	Exploit Stage	Direction	Source Port	Destination Port	Packet Count	Byte Count
1	Exploit	attacker → target	A	1099	6 - 7	358
2	Exploit	target → attacker	1099	A	7	567
3	Payload	attacker → target	A	Set in Exploit	7	7400 - 7500
N	Runtime	Bi-Directional	A/B	Set in Exploit	X	Y

TABLE IV: NetFlow Logs Generated by a Successful distcc_exec Exploit

Flow Index	Exploit Stage	Direction	Source Port	Destination Port	Packet Count	Byte Count
1	Exploit	attacker → target	A	3632	7	656
2	Exploit	target → attacker	3632	A	4	276
3	Payload	attacker → target	B	Set in Exploit	4	216
N	Runtime	Bi-Directional	B	Set in Exploit	X	Y

TABLE V: NetFlow Logs Generated by a Successful samba_symlink_traversal Exploit

Flow Index	Exploit Stage	Direction	Source Port	Destination Port	Packet Count	Byte Count
1	Exploit	attacker → target	A	445	10	975
2	Exploit	target → attacker	445	A	8	790 - 800
N	Runtime	Bi-Directional	B	Set in Exploit	X	Y

TABLE VI: NetFlow Logs Generated by a Successful samba_usermap_script Exploit

Flow Index	Exploit Stage	Direction	Source Port	Destination Port	Packet Count	Byte Count
1	Exploit	attacker → target	A	139	7	733
2	Exploit	target → attacker	139	A	4	356
3	Payload	attacker → target	B	Set in Exploit	3	164
4	Payload	target → attacker	Set in Exploit	B	2	135
N	Runtime	Bi-Directional	B	Set in Exploit	X	Y

TABLE VII: NetFlow Logs Generated by a Successful unreal_ircd_3281_backdoor Exploit

Flow Index	Exploit Stage	Direction	Source Port	Destination Port	Packet Count	Byte Count
1	Exploit	attacker → target	A	Set in Exploit	3	164
2	Exploit	target → victim	Set in Exploit	A	2	135
N	Runtime	Bi-Directional	A	Set in Exploit	X	Y

TABLE VIII: NetFlow Logs Generated by a Successful ntp_mon_list Exploit

Flow Index	Exploit Stage	Direction	Source Port	Destination Port	Packet Count	Byte Count
1	Exploit	attacker → target	A	123	1	60, 90 or 234
2	Exploit	target → victim	123	A	up to 10	up to 4460

list request packets, these were of size 60, 90 and 234 as found out in the wild [33]. As this attack is UDP based reflection attack, the attacker did not receive any feedback as to success or unsuccess of their attack and instead only a response was generated by an NTP server to the victim which the attacker intended to DDoS. For this reason the attacker also requires the uses of a third-party discovery tool, such as ping, to see whether the victim was still reachable or not (these ping logs were not shown as they are not part of the exploit tested however did exist in the communications).

The failures of these exploits for the most part resulted in a TCP reset at some point in the exploit attempt. The resulting NetFlow logs depict this with an initial flow from the attacker with a response flow of 1 packet that is 46 bytes in length (this represents a TCP reset). The only two exceptions to this were the MS08-067 based attacks, which showed a response flow from the target before a follow up flow was generated in order to access the payload of which was responded to with a flow of the aforementioned TCP reset. The second was the NTP based attacks which because they were UDP based, showed no response to the exploit in any form.

2) *Rule Sets Generated*: Tables I to VIII in the Results section, Section IV-B1, are in the format of the rule sets that will be given to the Bolvedere module that will attempt to discern these exploits¹. It is notable that these rules outlined by these tables require far fewer checks in an attempt to discern a network flow than deep packet analysis does; this is due to the fact that every packet in a network flow doesn't get analysed but rather the existence of a network flow. Coupling this with the sheer reduction in the amount of throughput the overall system has to handle as a NetFlow log only contains the metadata of a network flow. This allows for multiple NetFlow source nodes to sink their generated logs into a fewer hosts running Bolvedere than the equivalent amount of hosts required for a deep packet analysis solution.

C. Automated Module in Action

In order to check proper functionality and usability of this Bolvedere module, one has to provide control data for the results to be compared against. For this reason legitimate network traffic is required to the services running on the vulnerable host. In this testing, the legitimate connections and use of the services on the vulnerable host was performed by bots. These bots were programmed to perform simple tasks that required use of these services at random times ranging between 500 milliseconds and 10 seconds. One must note that the vulnerable target host was running every exploitable service in which the rule sets were generated for allowing for ease of testing². Furthermore, Microsoft Windows services were made available on this system through use of a Windows

¹The tables were developed this way to save space.

²This system is provided by RAPID7 and is available for download at <https://information.rapid7.com/metasploitable-download.html>

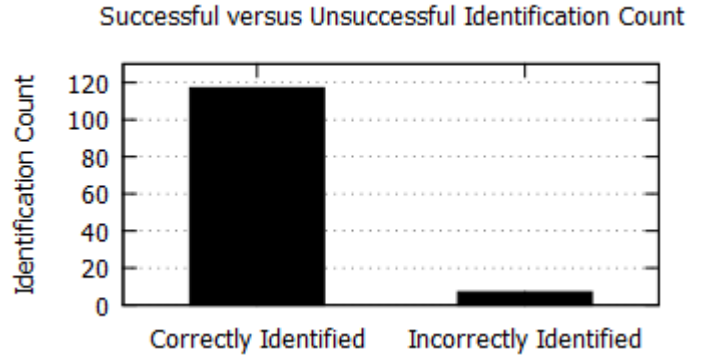


Fig. 3: Comparison of Success versus Failure of Network Flow Identifications

XP virtual machine running on the vulnerable host that was configured to bridge its network interface with that of the vulnerable host system. At runtime, the bots were first enabled to start communicating with the vulnerable host and then the attacks were manually launched and results observed through the terminal display of the Bolvedere module.

Listing 1: Terminal Output of Bolvedere Module

```
[10.42.0.45:45677 -> 10.42.0.33:445,
Size:9991,
Count:44]: Potential ms08_067_shell

[10.42.0.13:34782 -> 10.42.0.33:445,
Size:15232,
Count:113]: No malicious activity found

[10.42.0.68:40029 -> 10.42.0.33:123,
Size:60,
Count:1]: Potential ntp_mon_list

[10.42.0.13:37087 -> 10.42.0.33:139,
Size:23011,
Count:146]: No malicious activity found

[10.42.0.103:28928 -> 10.42.0.33:445,
Size:18002,
Count:174]: No malicious activity found
```

Testing occurred over 124 separate connections consisting of multiple network flows depending on the task at hand. The success or failure of a result was considered on a per connection bases and were discerned as to whether a connection was malicious by the NetFlow logs generated by the entire connection. Terminal output of this module can be referred to in Listing 1 which includes a false positive regarding the

detection of a ntp_mon_list attack. This was in fact a legitimate request for the monitor list from the NTP server. Referring to Figure 3 one can see the results produced by these 124 separate connections.

Of these 124 connections 117 were successfully identified as either a malicious or legitimate connection where the only 7 failures were false negatives produced when trying to determine whether a monitor list request from the NTP service was legitimate or part of a DDoS attack. This means that the rule set produced for this Bolvedere module is 94.355% accurate when attempting to discern the exploits recorded in Tables I to VIII under controlled test conditions.

These results suggest that detection of malicious activities when legitimate network flows closely resemble that of malicious flows becomes difficult. In the case of ntp_mon_list, this is because a legitimate request is used to exploit an amplification attack on a victim which is near impossible to detect against other legitimate requests. For this reason it is suggested that further revisions take into account previous connections made by an IP address, however memory usage should be considered before this step is taken.

Considering the high level of accuracy produced by this module when considering the given rule set and non-ntp_mon_list exploit and legitimate connections, these results hold promise into extension into detection of other malicious connections as well as extension into real world implementation. In all, the fact that no malicious connections were missed even though there were false positives means that this Bolvedere module has successfully achieved the goal set out by this research.

V. CONCLUSION

This research aimed to answer two questions, this being whether NetFlow logs can be used to discern malicious exploits and whether rule sets can be generated from these results and automated as a Bolvedere module. After execution of known exploits through a NetFlow node in a controlled environment, NetFlow logs were recorded and distilled into a rule set. Once this rule set was implemented within a Bolvedere module the accuracy of this module was shown to be 94.355% when discerning whether a network flow was legitimate or a specific malicious exploit. Although there were false positives, no executed exploits were recorded as false negatives. Given that every network flow containing a malicious exploit was detected by this Bolvedere module, this research was deemed successful and further development into its rule sets and fine tuning of the module itself shows promise for future iterations and use in live environments.

ACKNOWLEDGMENT

The authors wish to acknowledge the joint support of the Council for Scientific and Industrial Research (CSIR) and Rhodes University for the financial support and access to facilities for this research.

REFERENCES

- [1] A. Herbert and B. Irwin, "FPGA Based Implementation of a High Performance Scalable NetFlow Filter," in *Southern Africa Telecommunication Networks and Applications Conference*, D. F. Otten and M. R. Balmahoon, Eds., 2015, pp. 177 – 182.
- [2] Cisco. (2003) NetFlow V9 Export Format. Cisco Systems, Inc. Accessed 13th February 2015. [Online]. Available: <http://tinyurl.com/ond3pe8>

- [3] D. R. Kerr and B. L. Bruins, "Network flow switching and flow data export," Jun. 5 2001, uS Patent 6,243,667.
- [4] B. Claise, "Cisco systems NetFlow services export version 9," *IEEE Networking Group RFC*, 2004.
- [5] S. E. Deering, "RFC 2460: Internet Protocol, version 6 (IPv6) specificat," 1998.
- [6] G. Huston. (2014, February) IPv4 Address Report. Accessed 6th February 2014. [Online]. Available: <http://www.potaroo.net/tools/ipv4/index.html>
- [7] J. Postel, "RFC 791: Internet Protocol," IETF, Tech. Rep., 1981.
- [8] G. Huston. (2006) NetFlow Packet Version 5 (V5). Cisco Systems, Inc. Accessed 13th February 2015. [Online]. Available: http://netflow.caligare.com/netflow_v5.htm
- [9] ——. (2006) NetFlow Packet Version 8 (V8). Cisco Systems, Inc. Accessed 13th February 2015. [Online]. Available: http://netflow.caligare.com/netflow_v8.htm
- [10] G. Gu, R. Perdisci, J. Zhang, W. Lee *et al.*, "Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection." in *USENIX Security Symposium*, vol. 5, no. 2, 2008, pp. 139–154.
- [11] Y. Desmedt, "Man-in-the-middle attack," in *Encyclopedia of Cryptography and Security*. Springer, 2011, pp. 759–759.
- [12] A. Gazet, "Comparative analysis of various ransomware virii," *Journal in computer virology*, vol. 6, no. 1, pp. 77–90, 2010.
- [13] D. M. Smith, "The cost of lost data," *Journal of Contemporary Business Practice*, vol. 6, no. 3, pp. 1–9, 2003.
- [14] D. A. Leslie, *Legal Principles for Combatting Cyberlaundering*. Springer, 2014, page 7.
- [15] C. Paar and J. Pelzl, *Understanding cryptography: a textbook for students and practitioners*. Springer Science & Business Media, 2009.
- [16] S. H. Chad Dougherty, Jeffrey Havrilla and M. Lindner. (2003, August) W32/Blaster worm. CERT. Accessed 9th March 2015. [Online]. Available: <http://www.cert.org/historical/advisories/CA-2003-20.cfm>
- [17] Microsoft. (2009, April) Worm:Win32/Conficker.E. Accessed 31st October 2015. [Online]. Available: <http://tinyurl.com/hdlorbc>
- [18] C. Shannon and D. Moore, "The spread of the witty worm," *Security & Privacy, IEEE*, vol. 2, no. 4, pp. 46–50, 2004.
- [19] Microsoft. (2003, July) Microsoft Security Bulletin MS03-026 - Critical. Microsoft. Accessed 27th January 2016. [Online]. Available: <https://technet.microsoft.com/library/security/ms03-026>
- [20] ——. (2003, September) Microsoft Security Bulletin MS03-039 - Critical. Microsoft. Accessed 27th January 2016. [Online]. Available: <https://technet.microsoft.com/en-us/library/security/ms03-039.aspx>
- [21] ——. (2008, October) Microsoft Security Bulletin MS08-067 - Critical. Accessed 31st October 2015. [Online]. Available: <https://technet.microsoft.com/en-us/library/security/ms08-067.aspx>
- [22] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 581–590.
- [23] S. Staniford, V. Paxson, N. Weaver *et al.*, "How to own the internet in your spare time." in *USENIX Security Symposium*, 2002, pp. 149–167.
- [24] L. Bilge and T. Dumitras, "Before we knew it: an empirical study of zero-day attacks in the real world," in *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012, pp. 833–844.
- [25] D. Maynor, *Metasploit toolkit for penetration testing, exploit development, and vulnerability research*. Elsevier, 2011.
- [26] J. Muniz, *Web Penetration Testing with Kali Linux*. Packt Publishing Ltd, 2013.
- [27] R. Perdisci, A. Lanzi, and W. Lee, "Mcboost: Boosting scalability in malware collection and analysis using statistical classification of executables," in *Computer Security Applications Conference, 2008. ACSAC 2008. Annual*. IEEE, 2008, pp. 301–310.
- [28] N. Provos *et al.*, "A virtual honeypot framework." in *USENIX Security Symposium*, vol. 173, 2004, pp. 1–14.
- [29] P. Hintjens, *ZeroMQ: Messaging for Many Applications*. " O'Reilly Media, Inc.", 2013.
- [30] SQLite Consortium. (2016) Sqlite: Small. fast. reliable. choose any three. Accessed 30th April 2016. [Online]. Available: <https://www.sqlite.org/>
- [31] MemSQL Inc. (2016) Make every moment work for you. Accessed 30th April 2016. [Online]. Available: <http://www.memsql.com/>
- [32] D. Miller. (2016) Softflowd. Mindrot. Accessed 30th April 2016. [Online]. Available: <http://www.mindrot.org/projects/softflowd/>
- [33] L. Rudman and B. Irwin, "Characterization and analysis of ntp amplification based ddos attacks," in *Information Security for South Africa (ISSA), 2015*. IEEE, 2015, pp. 1–5.

Unsupervised Learning for Robust Bitcoin Fraud Detection

Patrick Monamo*, Vukosi Marivate†, Bheki Twala‡,

*Council for Scientific and Industrial Research
Email: pmonamo@csir.co.za

†Council for Scientific and Industrial Research
Email: vmarivate@csir.co.za

‡University of Johannesburg
Email: btwala@uj.ac.za

Abstract—The rampant absorption of Bitcoin as a cryptographic currency, along with rising cybercrime activities, warrants utilization of anomaly detection to identify potential fraud. Anomaly detection plays a pivotal role in data mining since most outlying points contain crucial information for further investigation. In the financial world which the Bitcoin network is part of by default, anomaly detection amounts to fraud detection. This paper investigates the use of trimmed k -means, that is capable of simultaneous clustering of objects and fraud detection in a multivariate setup, to detect fraudulent activity in Bitcoin transactions. The proposed approach detects more fraudulent transactions than similar studies or reports on the same dataset.

Keywords—cybercrime, anomaly, outlier, trimmed k -means, data mining.

I. INTRODUCTION

The latest technological advancement in the global financial system is the establishment of an internet-based payment system capable generation and minting of currency without the use financial institutions as trusted third parties responsible for transaction processing known as the Bitcoin peer-to-peer network [16]. Based on the fact that Bitcoin is recent and involves money or money-equivalents, scepticism relating to usage expansion and its proneness to fraud have been a matter of concern. This research study is commissioned to detect anomalous transactions based on pattern recognition on the Bitcoin network

Bitcoin offers several advantages compared to conventional currency. It eliminates the third party and thus helps lower the transaction fees. This has paved a way for governments across the world to explore the use of Bitcoins for remittance purposes, given the amount of remittance flows that exceed foreign direct investments in most developing countries [14]. Although other currently used technologies for remittance flows managed to lower costs, less than those of conventional banking, a report by the World Bank [22] and G20 [11] shows that the costs are still exorbitant for both the sender and the receiver. The global average cost of sending money is about 8% with the figure escalating to 12% in Sub-Saharan Africa [14], which is still below the global target of 5% thus; the call to explore alternative technological development such as Bitcoin.

Bitcoin is also open to anyone across the globe with
978-1-5090-2473-8/16/\$31.00 ©2016 IEEE

no arbitrary legal fees. One more property of Bitcoin, that distinguishes it from other digital currencies and payment systems, is the fact that parties can make payments and transfers anywhere in the world without divulging their true identity [6]. The network makes use of pseudonyms which are addresses derived from public keys. The provision of pseudonyms by the Bitcoin network can be considered to pave a way for cybercriminals to conceal the nature of location, source, ownership, or control of these financial proceeds [15] [5]. This defeats recommendations made by Financial Action Task Force (FATF), which is responsible for standards and promotion of implementation of regulatory and operational measures to combat money laundering and related threads to the global financial system. One of the key requirements deemed to be important towards combating money laundering and terrorism financing relate to identification, verification and reporting. The anonymity of the Bitcoin network bypasses the requirements of FATF according to [22], [5], [14], and hence Silk Road scheme was able to launder approximately \$1.2 billion.

Some of the major businesses like Mt Gox and Bitcoinica suffered a loss due to weaknesses attributable to a compromise of one key. To mitigate such weaknesses, latest technological developments that include Hierarchical Deterministic Multisignature (HDM) established from Bitcoin Improvement Proposal such as BIP32 help enhance financial security to an extent that compromising a single party cannot be equivalent to a compromise of funds of other users involved in the system [17] [3].

In the advent of a steady absorption of Bitcoin by developing economies for remittances flows, there exist a need for research to develop techniques that will assist regulatory authorities and related law enforcement entities in the fight against cybercrime. The main objective of this paper is to find and classify anomalies on the Bitcoin network based on transaction patterns. This will serve as an aide to detect financial fraud and associated activities such as money laundering. Secondary to that, the paper also seeks to assess performance of the anomaly detection algorithms using publicly available Bitcoin transaction data from blockchain. Furthermore, the study will make an assessment results in relation to the impact of HDM.

This paper is organized as follows: Section II provides an

overview of studies related to unsupervised anomaly detection in general as well as specific to the Bitcoin network. The proposed methodological orientation is detailed in Section III. Section IV is dedicated to analysis of experimental results while in Section we provide conclusion and related discussion.

II. RELATED WORK

Most studies involving anomaly detection adopt data with instances labelled as either fraudulent or legitimate [7], [21], [18], [4]. The labelling assists researchers to train anomaly detection algorithms, and assess algorithmic performance with test data. Due to novelty of the Bitcoin network, only a limited number of transactions have been reported as fraud¹ and as such makes the use of supervised technique infeasible.

On the premise of detecting both rogue users and their associated transactions, [19] used k -means clustering, Mahalanobis distance and unsupervised Support Vector Machines (SVM) on 100,000 data points of the Bitcoin dataset due to lack of computational power. [19] employed two types of graphs to model behavioural patterns in the network. Their study considered users as nodes with transactions between them serving as edges and *vice versa*. The algorithms used were able to detect 3 out of 30 known cases. In their subsequent study, using the full dataset that was extracted from the network on the 7th of April 2013, [20] attained similar results through the adoption of k -means clustering, Local Outlier Factor(LOF) as well as laws of power degree and densification to attain similar objectives. According to [24], based on Bitcoin network transaction from the genesis block until 13th July 2013, k -means clustering detected two interesting clusters. On the one hand a large cluster contained all good users as well as the known victims while on the other hand the three rogue users were clustered together in the smaller group. Furthermore, [24] generated synthetic node data that resembled the patterns of the three heists under investigation. The performance of the model based on synthetic data was found to attain an accuracy level of 76.5 percent in terms of detection rate. The improvement on findings can be attributable to the not yet uncovered properties of rogue users of the Bitcoin network given the disparities with the real-world data.

K -means clustering has the ability to group instances together, but lacks the prowess of detecting outliers. While LOF is popular for outlier detection, it does not scale well in large datasets with computational time. This paper proposes an approach that will compensate for the above-mentioned weaknesses.

III. METHODOLOGY

In this section we provide a brief outline of the dataset used, followed by a description of all features that were extracted from the dataset. The section is concluded by describing the machine learning algorithm proposed for the study. For the anomaly detection techniques, it is assumed that the majority of the transactions on the network are legitimate with at most only 1% being fraudulent.

A. Data Description

This study will use the Bitcoin dataset housed by the Laboratory for Computational Biology at the University of Illinois². All transactions from the genesis block to blockchain 230686 dated 7 April 2013. The blockchain under study contains 6 336 769 users which in our case we refer to as nodes. In between the users are 37 450 461 edges (transactions) that link interactions among users.

B. Feature Extraction

Based on the measured variable provided by the Bitcoin network, we attempt to build more meaningful features that will assist our learning algorithm in terms attaining the desired objectives. A total of 14 features were derived from the transaction data of the Bitcoin network. The following 14 features were derived from the dataset:

- Currency features: total amount sent, total amount received, average amount sent, average amount received, standard deviation received, standard deviation sent
- Network Features: in degree, out degree, clustering coefficient, number of triangles,
- Average neighbourhood (source target) whereby with reference to each query node: source refers to origin on incoming transaction and target is the destination. The four features identified: in-in, in-out, out-out, out-in.

C. Pre-Processing

Given that the dataset lists all transactions that took place during the period under study, cognisance of cases whereby some nodes were involved only in sending or receiving is noted. This led to the existence of missing values in our final dataset and hence imputation was in this regard exercised. We replaced missing values with zeroes based on the premise of equivalence to sending or receiving 0 BTC.

To have appropriate metrics between instances in our multivariate environment, we opted to transform our data. Our transformation resulted in each instance centred around mean zero and unit variance.

D. Proposed Method

The objects contained by the Bitcoin network dataset as described in Section 3.1 are unlabelled, hence this study opt for algorithms that are capable of outlier detection by considering the underlying structure of groupings existing within the network. The compared algorithms are standard k -means clustering and its robust version by [8]. The methods are discussed in the following subsections.

1) *K-Means Clustering*: On the basis of limited known number of transactions reported as fraud, we opt for k -means clustering. The clustering algorithm comprises of three key steps [23]:

- initialization of the centroids, followed by

¹<https://bitcointalk.org/index.php?topic=576337>

²<http://compbio.cs.uic.edu/data/bitcoin/>

- segmenting the dataset into k groups, and
- update the centroids until convergence is attained after several number of iterations

Although [19] cautioned that k -means clustering is not a technique for outlier detection, it lays the basis to evaluate methods given that outliers will be found furthest from the centroids of clusters they are associated with. In k -means, the average behaviour of objects in each cluster is represented by the calculated centroids while the Euclidean distance of each object in a cluster provide us with a measure of location relative to the centre. In this manner the method paves a way towards optimum outlier detection in any given cluster. Based on proposed features extracted, the algorithm is adopted to further confirm evidence provided by previous literature with regard to the potential number of clusters existing within the network.

2) *Trimmed K-Means Clustering*: The second algorithm is based on partial trimming that is more robust than classical k -means clustering in [8]. Given the trimming level α with the lowest possible variation penalized by Φ , the procedure is formulated as follows:

Let $\alpha \in (0, 1)$, the number of clusters k , and the penalty function Φ be given. For any set A such that $P(A) \geq 1 - \alpha$ and any k -set $M = m_1, m_2, \dots, m_k$ in \mathbb{R}^d , the method considers the variation of M given A to be :

$$V_{\Phi}^A(M) = \frac{1}{P(A)} \int_A \Phi(\inf_{i=1, \dots, k} \|X - m_i\|) dP$$

- Obtain k -variations given A , $V_{k, \Phi}^A$ by minimising in M :

$$V_{k, \Phi}^A = \inf_{M \subset \mathbb{R}^d, |M|=k} \Phi^A(M)$$

- Obtain the trimmed k -variation, $V_{k, \Phi, \alpha}$ by minimizing in A :

$$V_{k, \Phi, \alpha} = V_{k, \Phi, \alpha}(X) = V_{k, \Phi, \alpha}(P_X)$$

$$= (\inf_{A \in \beta^d, P(A) \geq 1 - \alpha} V_{k, \Phi}^A)$$

The primary objective of the algorithm is to obtain a trimmed set A_0 , if it exists, and a k -set $M_0 = m_1^0, m_2^0, \dots, m_k^0$, if it exists, through the condition:

$$V_{\Phi}^A(M_0) = V_{k, \Phi, \alpha}.$$

While standard k -means clustering provides only the properties of the resultant clusters, this robust version assumes the existence of specific proportion of outlier within the network. The technique trims out those objects that are furthest from the centroids to be anomalous in nature, hence the name. According to [12], the general concept of k -means combined with impartial trimming provides robust result in terms of influence function, breakdown point and qualitative robustness as the key performance measures.

IV. RESULTS

This paper used R implementations that have been developed for the above algorithms³. The full dataset was aggregated to a node level and the algorithms applied to the first 1 000 000 nodes when instances are listed according to the increasing node number as per the original transaction data. It should be noted that for model building, all the extracted features were used for both classical k -means and trimmed k -means clustering algorithms. The rest of this sections discusses the results from our outlier detection approaches.

A. Classical k -means clustering

In the absence of knowledge regarding nature of clustering structure of the Bitcoin network, the first step was to estimate the number of clusters k , by using the within sum of squares as the key clustering performance metric. The choice was motivated by its popularity in clustering problems as well as the ease of interpretation. As a result, the clustering algorithm was iterated over a range of values of k to determine the best number of clusters. Given the nature of initial randomization attributable to k -means, the algorithms were ran five times to gauge the stability of the value of k in this regard as well as the centres. Whilst k can be infinitely large, it was only restricted to a maximum of 15 in this particular study to reduce time complexity.

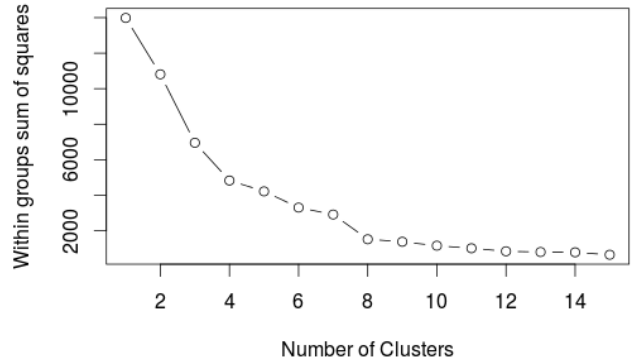


Fig. 1. The elbow chart showing optimal clustering attained at $k=8$

Figure 1 shows that the optimal number of clusters is realized at $k=8$. Although $k=4$ appears to be a good choice, there is a large gain in variability up to $k=8$ which flattens thereafter, hence, the choice of 8 clusters, which is in agreement with previous reports [19].

In Figure 2 cluster distribution to visualize relative frequencies is provided, while Figure 3 reflects a 2-dimensional plot all resultant clusters as represented by various colour codes for k -means using the two largest components. The distribution shows that we have cluster 8 being the largest with almost 60% of all instances. Due to the algorithm's sensitivity to outliers, this results shows two outright extreme points belonging to a single cluster. Table IV-A below provides a

³The main packages used include *mclust*, *fpc* and *tclust* [10], [9]

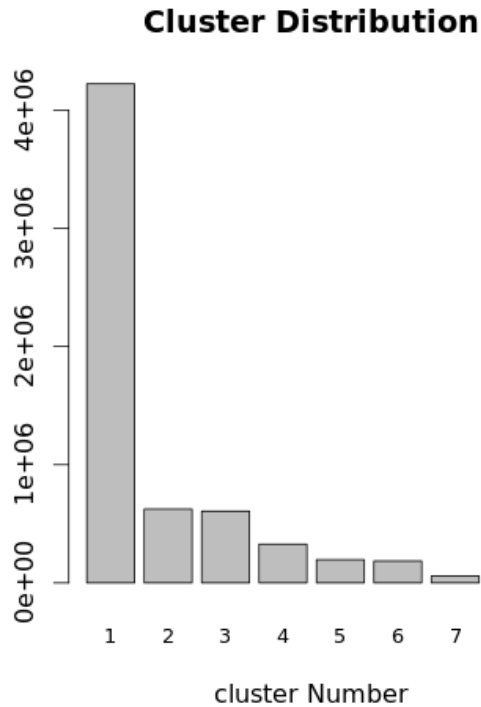


Fig. 2. Distribution of of attained clusters through k -means clustering algorithm.

snapshot of centroids associated with the each of the 8 clusters. The table shows average value of Bitcoin received and sent by each user together with associated user out-degree and clustering coefficient. The number *out-degree* describes the total quantity of outgoing transactions for each user. This figure appeared fairly moderate among the first 6 clusters followed by an abrupt difference when considering cluster both cluster 7 and 8. The proportion of users associated with each query node that transact together (*clustering-coefficient*) tend to be strong in the top 3 clusters and poor in the bottom ones. The summary shows that clusters with higher clustering coefficients tend to transact with relatively small amounts of Bitcoins. While this clusters show good connectivity among users as vindicated by higher clustering coefficients, they also exhibit lower values regarding user degrees. In contrast, clusters with poor cluster coefficients were found to have abnormally large node degree and transacting on higher amounts.

TABLE I. CLUSTER CENTROIDS k -MEANS USING SELECTED ATTRIBUTES

ClusterLabel	AverageSent	AverageReceived	ClusterCoeff	OutDegree
1	2.99	2.99	0.50	9.24
2	1.92	1.97	0.61	4.77
3	0.26	0.24	0.70	2.21
4	99.63	107.44	0.23	5.56
5	87.27	64.98	0.31	7.38
6	41.00	36.44	0.12	16.00
7	98.51	67.48	0.00	532 534.00
8	9.50	17.90	0.00	477 035.60

kmeans clustering results

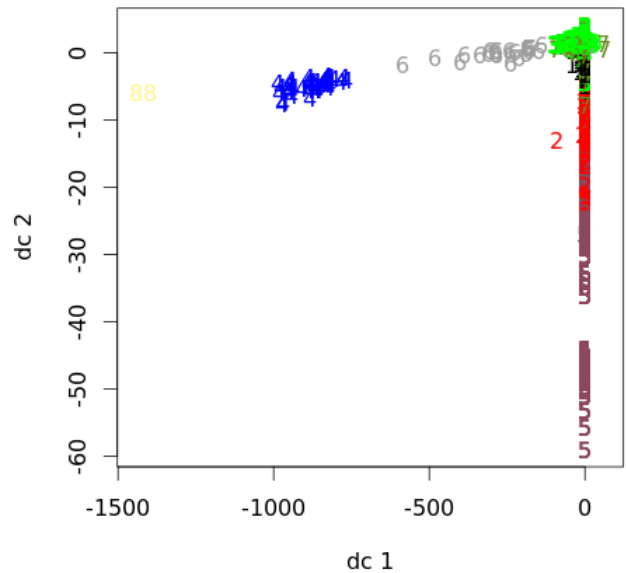


Fig. 3. A graphical representation of k -means clustering results

B. Trimmed k -means for robust clustering and outlier detection

Similar to classical k -means, the first steps in this algorithm will be to estimate the number of groups existing within the network as well as the trimming level, α . Although the actual contamination level is unknown, in this paper it is fixed at $\alpha = 0.01$ as guided by similar studies involving financial fraud and intrusion detection. For comparison purposes with classical k -means, the value of k is restricted to a maximum of 15. The value associated with optimal number of clusters in this method is realized when the difference between the log-likelihood of $k + 1$ and k approximates to 0 given a specified trimming level using BIC [13]. The algorithm achieved optimal clustering at $k=8$ and the distribution with an additional cluster containing outliers is shown on Figure 4

The proportion of objects trimmed as outliers are marked with an "o" on Figure 5.

TABLE II. CLUSTER CENTROIDS TRIMMED k -MEANS USING SELECTED ATTRIBUTES

ClusterLabel	AverageSent	AverageReceived	ClusterCoeff	OutDegree
1	30.01	29.14	0.26	6.10
2	0.04	0.03	0.70	2.00
3	1.70	2.29	0.55	8.19
4	39.84	35.24	0.14	8.63
5	30.51	42.54	0.60	4.92
6	77.04	51.59	0.98	3.81
7	1.02	1.05	0.61	4.51
8	25.86	24.73	0.01	10.45
9	2217.36	1886.70	0.49	1508.89

1) *Linking Results to HDM*: Although HDM was not yet implemented at the time of cybercriminal incidents that took place, this research take a further step to assess results in relation to such developments. It is noted that only 5 of the 30

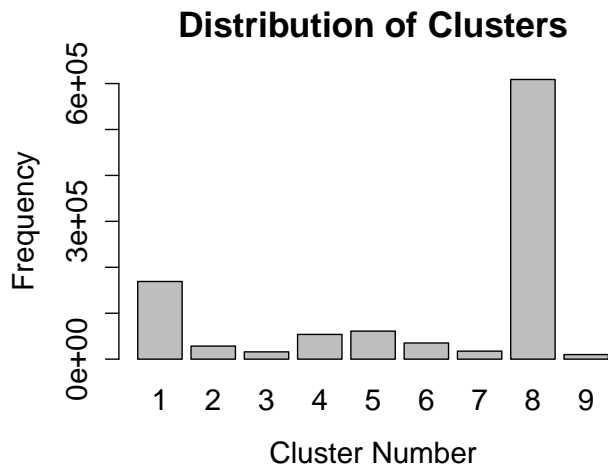


Fig. 4. Distribution of trimmed k -means clustering results

trimmed kmeans cluster results

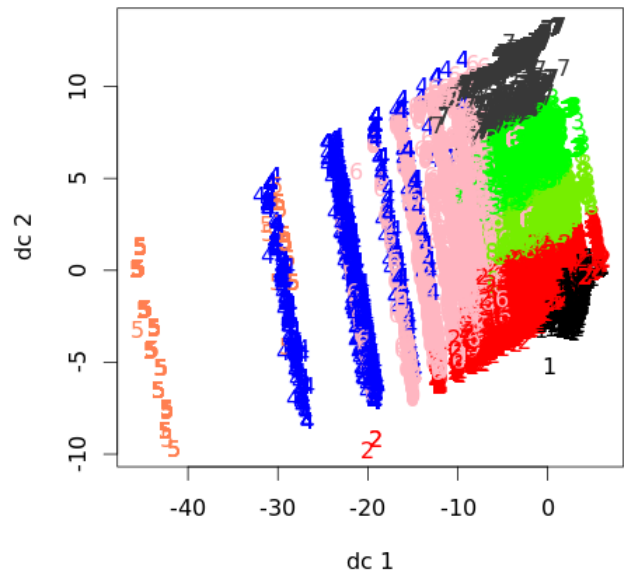


Fig. 6. Clustering results from trimmed k -means with trimmed outlier left out

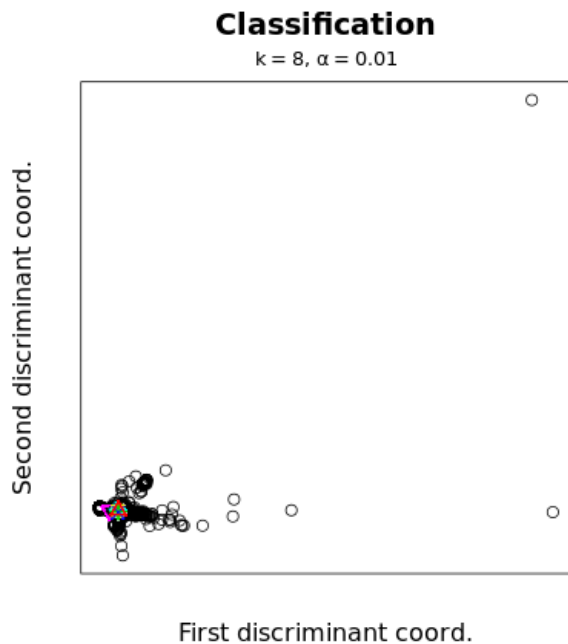


Fig. 5. Clustering results from trimmed k -means with outliers marked 'o'

known anomalies were successfully detected by the algorithm. The detected nodes involved the following entities: 1. Mt Gox, 2. Linode Hack, 3. Stone Man Loss, 4. Allinvain and 5. 50 BTC Theft. Anomalies associated with the first four entities are reported to have originated from poor backups. This fraudulent activities occurred in the absence of HDM which has been developed to mitigates vulnerabilities attributable to wallet backups. The adoption of HDM could have significantly reduced the number criminal incidents experienced by the network users.

V. DISCUSSION

The results shows some disparities between the two clustering algorithms in the presence of outliers. K -means due to its sensitivity to anomalies, formed a spurious cluster containing only two objects as depicted in Figure 3. When applying trimmed k -means to the dataset, spurious cluster attained from k -means is filtered out and as a result improvements in group structures is realized. From Table IV-B, it is evident that the detected outliers were spread across multiple clusters as elucidated by moderated attribute values in non-anomalous clusters. The presence of anomalies obstruct visuals of the underlying structure in Figure 5. In Figure 6, all trimmed outliers are left out and an approximately clear structure can be seen.

The detected outliers represented by trimmed proportion was compared with the 30 known fraud cases to assess the performance of the proposed approach. To realize this objective, we use all transactions of the Bitcoin network for the period under study. The raw data of the Bitcoin network contain similar attributes to the list of the 30 known fraudsters. The said attributes in this regard include sender, receiver, date, time (up to seconds level) and value in BTC. Although the list of known criminal elements is made of 30 users, it should be noted that the raw list is composed of 76 transactions which were ultimately matched against the 37 million edges of the network.

Finally, to detect anomalies the two datasets were matched by date, time and amount. Furthermore, the resultant outliers were matched against the outliers detected by the trimmed k -means algorithm. Of the 30 known bale users, the algorithm successfully detected 5 of them.

The findings in this paper proves to be an improvement to results on similar reports in terms of the number known anomalies that were detected successfully. Furthermore, this paper vindicate that the adoption of recent technological developments (e.g. BIP and HDM) when coupled with good performing fraud detection algorithms will enhance financial security of users on the peer-to-peer network. This combination serves as mitigating factors on sceptical attitude towards Bitcoin and thus provide a platform for acceptance by the global village into mainstream economy.

VI. FUTURE STUDIES

The main challenge in this study that instances are unlabelled and hence becomes difficult to validate results. On the basis of reliance on known criminal elements, future studies will consider comparing results with neighbourhood-based algorithm. In the absences of validation methods in this type of situation, a look at algorithms undersampling majority instances while oversampling majority groups will be explored. Based on the fact that the Bitcoin dataset in in the form of a network, graph-based algorithms for anomaly detection is also an option in this regard.

One of the important challenge in data mining is the ever increasing amount of data which is the case of the transactions in blockchain. From an analytics perspective, it is proposed that techniques on streaming data be considered as well as segmenting the data according to major developments such prior/post HDM.

VII. CONCLUSION

In this paper, we evaluated the use of trimmed k -means clustering for unsupervised cybercrime detection in the Bitcoin network. Although unavailability of labels which makes it difficult to evaluate algorithmic performance with regard to flagged suspicious activities, the algorithms successfully detected some of the known fraudulent activities. In comparison to previous studies on fraud detection on Bitcoin network, trimmed k -means provided promising results with improvements of detection rate with regard to known fraudulent elements. There is still more work to be done. There is a larger need for advanced feature extraction [1], [2]. With more informative feature extraction, we might be able to train supervised learning methods on the known fraudulent cases and explore if it will reveal other similar behaviour in the network.

REFERENCES

- [1] Leman Akoglu, Mary McGlohon, and Christos Faloutsos. Oddball: Spotting anomalies in weighted graphs. In *Advances in Knowledge Discovery and Data Mining*, pages 410–421. Springer, 2010.
- [2] Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery*, 29(3):626–688, 2015.
- [3] Anon. Reclaiming financial privacy with hd wallets, 2013.
- [4] Siddhartha Bhattacharyya, Sanjeev Jha, Kurian Tharakunnel, and J Christopher Westland. Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3):602–613, 2011.
- [5] Adrian Blundell-Wignall. The bitcoin question: Currency versus trustless transfer technology. *OECD Working Papers on Finance, Insurance and Private Pensions*, 2014.

- [6] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A Kroll, and Edward W Felten. Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *Security and Privacy (SP), 2015 IEEE Symposium on*, pages 104–121. IEEE, 2015.
- [7] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.
- [8] JA Cuesta-Albertos, Alfonso Gordaliza, Carlos Matrán, et al. Trimmed k -means: An attempt to robustify quantizers. *The Annals of Statistics*, 25(2):553–576, 1997.
- [9] Chris Fraley and Adrian E Raftery. Mclust version 3: an r package for normal mixture modeling and model-based clustering. Technical report, DTIC Document, 2006.
- [10] Heinrich Fritz, Luis A García-Escudero, and Agustín Mayo-Iscar. tclust: An r package for a trimming approach to cluster analysis. *Journal of Statistical Software*, 47(12):1–26, 2012.
- [11] G20. G20 plan to facilitate remittance flows. Technical report, G20, 2014.
- [12] Luis Ángel García-Escudero and Alfonso Gordaliza. Robustness properties of k means and trimmed k means. *Journal of the American Statistical Association*, 94(447):956–969, 1999.
- [13] Luis Angel García-Escudero, Alfonso Gordaliza, Carlos Matrán, and Agustín Mayo-Iscar. Exploring the number of groups in robust model-based clustering. *Statistics and Computing*, 21(4):585–599, 2011.
- [14] Ralph C Maloumy-Baka, Christian Kingombe, et al. The quest to lower high remittance costs to africa: A brief review of the use of mobile banking and bitcoins. *Centre for Finance and Development Working Paper. Graduate Institute. Geneva*, 2015.
- [15] G Krishnapriya MCA and M Prabhakaran. An multi-variant relational model for money laundering identification using time series data set. unpublished paper.
- [16] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Consulted*, 1(2012):28, 2008.
- [17] Coinbeyond News. Anatomy of hdm structure, July 2014.
- [18] EWT Ngai, Yong Hu, YH Wong, Yijun Chen, and Xin Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3):559–569, 2011.
- [19] Phillip Thai Pham and Steven Lee. Anomaly detection in bitcoin network using unsupervised learning methods. Unpublished Report.
- [20] Phillip Thai Pham and Steven Lee. Anomaly detection in the bitcoin system-a network perspective. Unpublished Report.
- [21] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014.
- [22] Dilip Ratha, Supriyo De, Ervin Dervisevic, Christian Eigen-Zucchi, Sonia Plaza, and Kirsten Schietler. Migration and remittances: Recent developments and outlook-special topic: Financing for development. *Migration and Development Brief*, 24, 2015.
- [23] Archana Singh, Avantika Yadav, and Ajay Rana. K-means with three different distance metrics. *International Journal of Computer Applications*, 67(10):13–17, 2013.
- [24] Deepak Zambre and Ajey Shah. Analysis of bitcoin network dataset for fraud. Unpublished Report, 2013.