AN INVESTIGATION INTO THE PREVALENCE AND GROWTH OF PHISHING ATTACKS AGAINST SOUTH AFRICAN FINANCIAL TARGETS

A thesis submitted in partial fulfilment of the requirements for the degree of

MASTERS OF SCIENCE

of

RHODES UNIVERSITY

by

DARSHAN MAGAN LALA

November 2015

Abstract

Phishing in the electronic communications medium is the act of sending unsolicited email messages with the intention of masquerading as a reputed business. The objective is to deceive the recipient into divulging personal and sensitive information such as bank account details, credit card numbers and passwords. Attacks against financial services are the most common types of targets for scammers. Phishing attacks in South Africa have cost businesses and consumers substantial amounts of financial loss.

This research investigated existing literature to understand the basic concepts of email, phishing, spam and how these fit together. The research also looks into the increasing growth of phishing worldwide and in particular against South African targets. A quantitative study is performed and reported on; this involves the study and analysis of phishing statistics in a data set provided by the South African Anti-Phishing Working Group. The data set contains phishing URL information, country code where the site has been hosted, targeted company name, IP address information and timestamp of the phishing site. The data set contains 161 different phishing targets. The research primarily focuses on the trend in phishing attacks against six South African based financial institutions, but also correlates this with the overall global trend using statistical analysis.

The results from the study of the data set are compared to existing statistics and literature regarding the prevalence and growth of phishing in South Africa. The question that this research answers is whether or not the prevalence and growth of phishing in South Africa correlates with the global trend in phishing attacks. The findings indicate that certain correlations exist between some of the South African phishing targets and global phishing trends.

Acknowledgements

The completion of this research thesis would not have been possible without the support and assistance of a number of people.

A special thanks to Prof. Barry Irwin for his supervision, invaluable assistance, patience and guidance of the course of this entire degree. Prof Irwin also obtained a valuable data set from the South African Anti-Phishing Working Group used within this research.

I want to thank the South African Anti-Phishing Working Group for providing the data set which allowed for the quantitative analysis to be performed.

I would like to thank my parents, Magan and Bhadra Lala, for instilling virtues of hard work and discipline in me. These tools featured immensely during my academic journey. Mom and Dad, thanks for letting me pursue my dreams and for laying the path.

Finally, I would like to thank my extended family and friends for their unconditional support and continuous motivation throughout this journey.

Table of Contents

List of Figures

List of Tables

Glossary

APWG	Anti-Phishing Working Group
DNS	Domain Name System
EMEA	Europe, Middle East and Africa
GNU	GNU's Not Unix
GPL	General Public License
IP	Internet Protocol
MDA	Mail Delivery Agent
MRA	Mail Retrieval Agent
MSA	Mail Submission Agent
MTA	Mail Transfer Agent
MUA	Mail User Agent
MX	Mail Exchanger (record)
PPMCC	Pearson Product-Moment Correlation Coefficient
RFC	Request For Comments
SA	South Africa
SMS	Short Message Service
SMTP	Simple Mail Transfer Protocol
SQL	Structured Query Language
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
URL	Uniform Resource Locator
VOIP	Voice Over Internet Protocol
XSS	Cross-Site Scripting

Chapter 1

Introduction

Phishing is the act of sending unsolicited email messages with the intention of masquerading as a reputed business (Chhabra, 2005; Pfleeger & Pfleeger, 2006). The objective is to deceive the recipient into divulging personal and sensitive information such as bank account details, credit card numbers and passwords (Grobler, 2010; Frauenstein & Von Solms, 2013). Attacks against financial services are the most common types of targets for scammers (Perez, 2012; Chaudhary, 2014). Phishing attacks in South Africa have cost businesses and consumers substantial amounts of financial loss (Goldstuck & Dagada, 2009; Grobler & Dlamini, 2012).

This research investigated existing literature to understand the basic concepts of email, phishing, spam and how these fit together. The research also looks into the increasing growth of phishing worldwide and in particular against South African targets. A quantitative study is performed and reported on; this involves the study and analysis of phishing statistics in a data set provided by the South African Anti-Phishing Working Group. The data set contains phishing URL information, country code where the site has been hosted, targeted company name, IP address information and timestamp of the phishing site. The data set contains 161 different phishing targets. The research primarily focuses on the trend in phishing attacks against six South African based financial institutions, but also correlates this with the overall global trend using statistical analysis.

This chapter introduces the research problem statement, research objectives and thesis structure.

1.1 Problem Statement and Research Goals

Phishing is a growing concern around the world (Symantec Corporation, 2015; RSA, 2015; Trend Micro Incorporated, 2012; Shcherbakova, Vergelis, & Demidova, 2015b). The aim of this research will be to understand whether or not the prevalence and growth of phishing against South Africa targets correlates with the global trend in phishing attacks. This research also aims to understand the possible reasons for the increase in phishing as well as where in the world phishing sites are hosted.

1.1.1 Problem Statement

At present, there is a minimal amount of research that attempts to correlate the global phishing trend to that of the South African phishing trend. Therefore, the purpose of this research is to investigate the prevalence and growth of phishing attacks against South African financial targets.

1.1.2 Research Objectives

The primary objective of this research is to:

- Investigate the current growth rates of phishing worldwide compared to the current rate of phishing against South African financial targets.

- Analyse and evaluate a data set of phishing web sites to ascertain correlation between South African financial phishing targets and global phishing trends.

- To add new knowledge to phishing trends of South African financial targets as compared to global phishing targets. The knowledge in this gap would allow for the relevant parties involved in the fight against phishing to better prepare and target stopping or minimising the effect phishing has on society and organisations.

The secondary objective of this research is to:

- To develop an understanding of email delivery, spam and various phishing techniques used.

1.2 Thesis Structure

This research has been divided into five chapters. The following describes the structure and order of the information in this research.

Chapter 2 discusses the literature review and considers existing literature to understand the basic concepts of email, phishing, spam and how these fit together. The chapter then briefly discusses email agents, email format and the SMTP protocol. Different phishing techniques are described together with examples to depict these different phishing scenarios. The chapter concludes by looking at existing phishing trend data, both globally and within a South African context.

Chapter 3 discusses the data set being used for the quantitative study. The chapter then discusses the research design and analysis process, followed by a brief description of the tools used for the analysis. The chapter concludes by briefly discussing the various study areas that will be focused on.

Chapter 4 discusses the analysis of the data set from the South African APWG as well as the statistics extracted regarding phishing trends seen in terms of phishing URLs targeting South African financial organisations. This chapter includes investigating the various themes and associated correlation which includes statistical analysis of the significance regarding the correlations observed.

Chapter 5 summarises the research thesis and presents the final conclusions from the research. The chapter concludes by providing suggestions for possible future research.

Chapter 2

Literature Review

2.1 Introduction

This chapter forms the research behind the study of phishing and existing trends observed. The chapter starts off by understanding the basics of email, spam and phishing followed by an example of email delivery. The chapter then delves deeper into understanding phishing together with the components thereof, and then discusses information flow in a phishing attack followed by a brief description of the consequences of phishing. The various types of phishing techniques that can be used by phishers are discussed in this chapter as well. The chapter concludes by discussing related work and statistics that are relevant to the research.

2.2 Understanding Email, Spam and Phishing

This section explains what email is and how it works, as well as provides a basic understanding of what spam and phishing are.

2.2.1 Email

Electronic mail, or email, is a method used to distribute and receive messages over electronic communication network systems. The modern day protocol for sending email messages is the Simple Mail Transfer Protocol (SMTP). A user can share various types on content within an email, including text, files, images or links (Chhabra, 2005).

RFC 5321 Simple Mail Transfer Protocol (Klensin, 2008) and RFC 5322 Internet Message Format (Resnick, 2008) are documents which describe the internet standards for relating to the syntax for text messages that are sent between computer users in the context of electronic mail or email. An example of an email address is:

user19@orangemail.com

The email address contains three portions which can be broken down as:

1. The local part 'user19' is a locally interpreted string, which identifies the local domain mailbox. This can be the alias, user, group or department of a company. The local part may be up to 64 characters long

2. The '@' (at-sign) is the divider, or address delimiter, in the full email address to distinguish between the local part and domain part. There can only be one '@' in an email address and it is required for all SMTP email addresses.

3. The domain part 'orangemail.com' portion is the domain hostname of where the mailbox exists and where email to this domain is delivered. The domain name cannot be longer than 254 characters.

2.2.2 Email Spam

Electronic spamming is the use of using electronic messaging to send out unsolicited messages. Electronic spamming includes the use of email, instant messaging, newsgroups, mobile phone messages, social networking, blogs as well as any electronic form where messages can be posted.

Email spam, also known as unsolicited bulk email, is the sending of unwanted email messages. According to Internet Security Threat reports by Symantec (2014, 2015) spam volume across the world dropped to 60% of all email traffic in 2014, as opposed to 66% in 2013 and 69% in 2012, as illustrated in Figure 1. This still equates to 28 billion email spam messages per day, as illustrated in Figure 2.

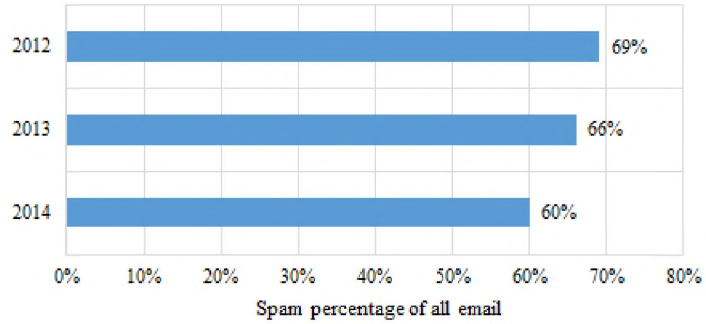


Figure 1: Percentage rate of email spam (Symantec, 2015)

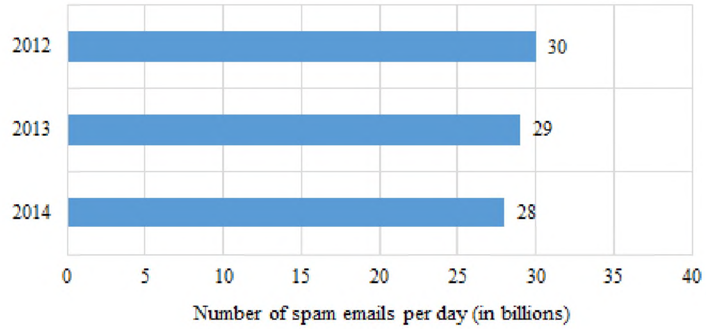


Figure 2: Estimated global spam volume per day (Symantec, 2015)

2.2.3 Phishing

The term phishing, shortened from 'password harvesting (ph) fishing', is a form of spam attack (Pfleeger & Pfleeger, 2006) whereby information such as usernames, passwords, and credit card details are solicited by a scam, usually an attacker(or phisher) falsely claiming to be a legitimate organisation (Grobler, 2010). The phishing email would contain a hyperlink, which directs the user to a site which is a fraudulent copy of the original intended site (Goldstuck & Dagada, 2009; Dagada, 2011). This site that opens from the malicious hyperlink is almost identical to the legitimate web site, and the user will believe the site to be genuine. The user, assuming he/she is unaware of phishing techniques, will then presume to login unaware that the site is recording his/her credentials and personal details which will then be used by the phisher (Frauenstein & Von Solms, 2013).

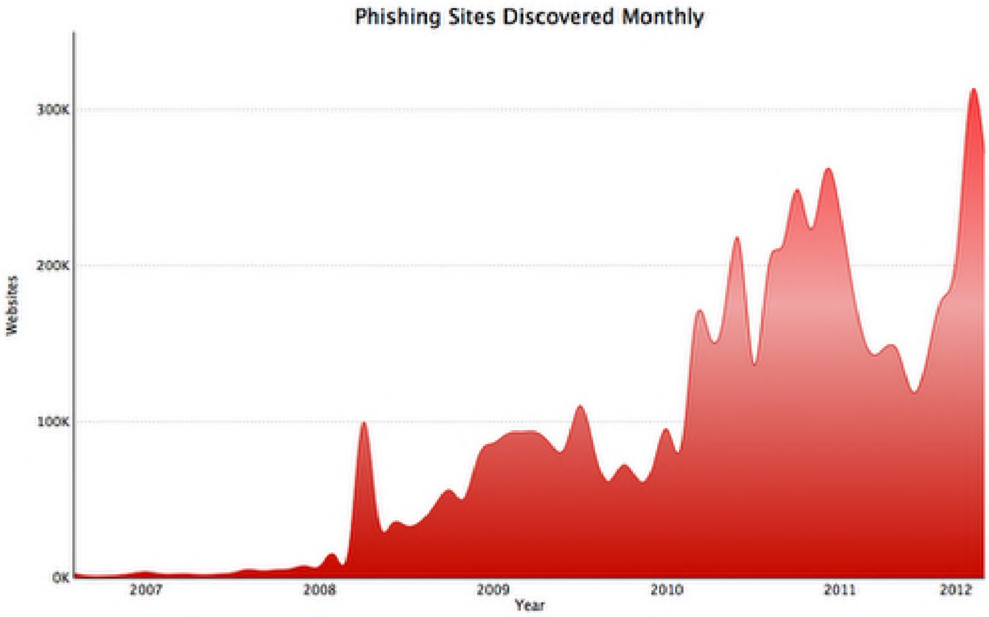


Figure 3: Increase in phishing sites from 2007 to 2012 (Perez, 2012)

Phishing was first reported in the mid 1990's when users of America Online were lured by phishers (Chaudhary, 2014) to divulge personal details through legitimate looking application notices from companies like Microsoft or PayPal (van der Merwe, Seker & Gerber, 2005). Phishing by email took off in 2003 with reported attacks growing; by 2004 the phishers started using copies of legitimate bank emails and websites (Anderson, 2007). By 2013, 37.3 million users around the world were subjected to phishing attacks. Figures 3 (Perez, 2012) and 4 (RSA, 2014a) illustrate the increase of phishing sites discovered over time.

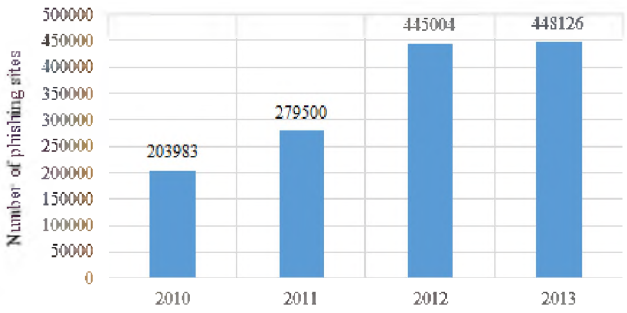


Figure 4: Increase in phishing sites discovered monthly, from 2010 to 2013 (RSA, 2014a)

2.3 Email delivery

This section briefly discusses the components that make up email flow and illustrates how basic email flow works over the internet. Understanding how email works over the internet will allow for a better grasp of how phishing works across the internet.

2.3.1 Email format

An email message contains two segments: a header and a body (Resnick, 2008). An email header consists of a structured set of fields such as from, to subject, etc. In an email message the header lines always precedes the body, as the header contains routing information including sender, recipient, date, time and subject (Resnick, 2008). The email body is the section of the email general users are all familiar with seeing. This contains the text and content of the message in the email.

RFC 5322 Internet Message Format contains the syntax required for understanding the breakdown and complexity behind an email message header and body. Table 1 shows the header fields and values as specified in RFC 5322.

Table 1: Header syntax as specified in RFC 5322

Header	**Value**
From	This field contains the email address of the sender
To	This field contains the email address of the recipient
Subject	This field contains a brief summary about the contents of the message
Date	This field contains the date and time of when the message was sent
CC	This field shows any email addresses who were carbon copied
BCC	This field shows any email addresses who were blind carbon copied
Received	Each mail server that handles an email adds its details to this field
Content-type	This field contains message format information

2.3.2 SMTP

Email has very little authenticity protection. There is nothing in the SMTP protocol which verifies that the listed sender, specified in the 'From:' address, is accurate or legitimate (Klensin, 2008). Therefore, spoofing the source address of an email address is not challenging for the phisher (Pfleeger & Pfleeger, 2006).

2.3.3 Email Agents

There are different types of mail agents which form part of the email infrastructure for sending and delivering email messages:

Mail user agent (MUA)

An email client, or previously known as a Mail User Agent, is a program that the end user makes use of to send, receive and read email (Chhabra, 2005; Almomani, Gupta, Atawneh, Meulenberg, & Almomani, 2013). Popular email clients include Microsoft Outlook, IBM Lotus Notes, Mozilla Thunderbird and Apple Inc.'s Mail. There are also web based email clients; these include Gmail, Yahoo Mail, and Hotmail.

Mail Submission Agent (MSA)

A Mail Submission Agent is a program that receives email messages from the Mail User Agent, and cooperates with a Mail Transfer Agent to initiate delivery of the email (Chhabra, 2005; Almomani et al., 2013). This generally makes use of the Simple Mail Transfer Protocol (SMTP).

Mail Transfer Agent (MTA)

A Mail Transfer Agent is software that transmits emails from one computer to another using SMTP (Almomani et al., 2013). The terms mail server, mail exchange or MX host may also refer to a computer system performing the role of a Mail Transfer Agent.

A Mail Delivery agent (MDA)

A Mail Delivery agent is a program which delivers the email message to the recipients' individual mailbox (Chhabra, 2005), and for arranging for it to be received by the local system (Almomani et al., 2013).

A Mail Retrieval Agent (MRA)

A Mail Retrieval Agent is a program that fetches email messages from a remote mail server (Chhabra, 2005; Almomani et al., 2013). The Mail Retrieval Agent works with a Mail Delivery Agent to deliver mail to a local mailbox. A Mail Retrieval Agent can be a stand-alone application, or can be incorporated into a larger Mail User Agent application. Examples of Mail Retrieval Agents include fetchmail, retchmail and getmail.

2.3.4 Email flow across the Internet

This section will now describe the flow of email communication across the internet using the components described in section 2.3.1. For illustrative purposes, Figure 5 is used to demonstrate what occurs when Alex sends an email to Bill.

Step 1: Sender composes email message and MUA transfers the message to local MTA

Alex composes an email message to be delivered to Bill, shown in step 1 in Figure 5. She makes use of the MUA on her local machine. The MUA transfers the message to the local MTA in internet message format. The local MTA is smtp.alpha.org.

Step 2: Sending MTA checks DNS for MX record of recipient

The sending MTA smtp.alpha.org queries DNS for the mx record from the recipient record bill@beta.org. This is shown as step 2 in Figure 5.

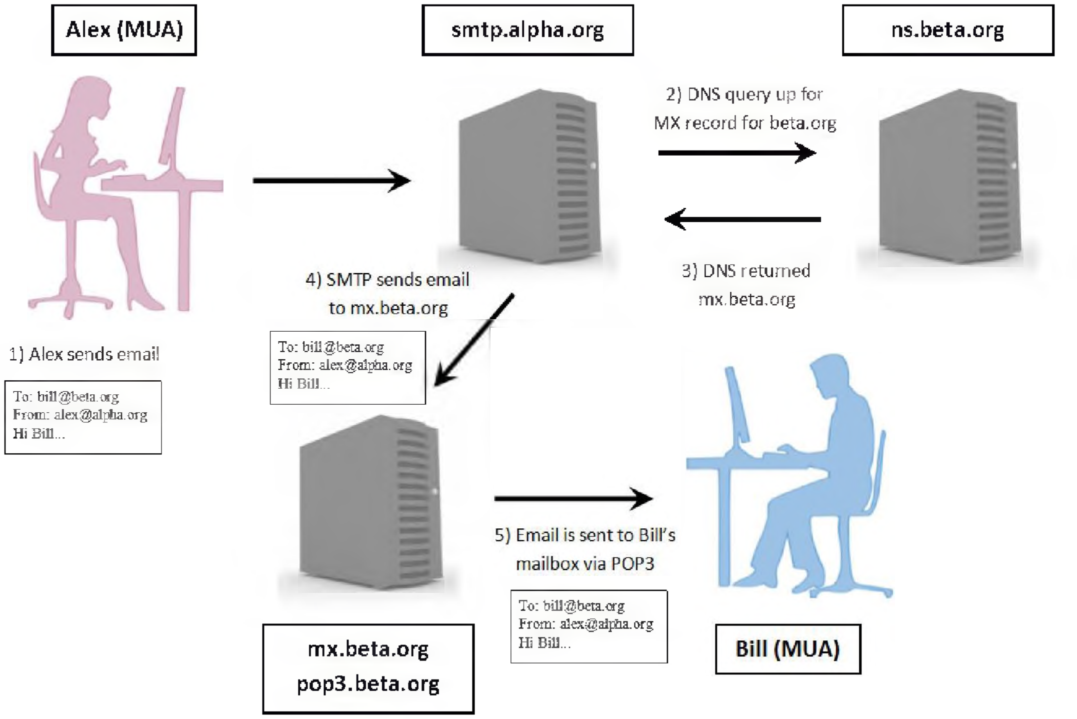


Figure 5: Basic flow of email from one user to another over the internet

Step 3: Sending MTA receives MX record of recipient domain

The DNS query made to the DNS server returns an MX record containing mx.beta.org as the receiving host for accepting emails. This is shown as step 3 in Figure 5.

Step 4: SMTP connection between sender MTA and recipient MTA

The sender MTA smtp.alpha.org uses SMTP to send the email message to the recipient's advertised MTA mx.beta.org as shown in step 4 of Figure 5.

Step 5: Recipient retrieves email message using MUA

As shown in Figure 5, Bill makes use of his MUA to fetch the message. Alternatively Bill can make use of his webmail service account to access his email.

2.4 Understanding Phishing

The section describes what a phishing attack is by understanding the components involved. This is illustrated by means of a sequence diagram the flow of information in a phishing attack. The section also looks into the motivation for phishing and concludes by discussing the consequences of phishing.

2.4.1 Phishing attack

A phishing attack initiated by email generally consists of a few components (Butler, 2007; Chhabra, 2005):

The “alleged” sender

The sender may have a reputable name on the face of the email, but has questionable ’reply’ to address. They typically pose as figures of authority within financial industries, which happen to be in possession of sensitive information regarding their client.

The recipient

The phisher will try and target a maximum number of potential victims. Legitimate email address lists can be purchased on the internet. Phishers know that the majority of people will ignore the emails, or the user’s email security will move the email into the spam folder, but the financial gains of targeting the correct users make the continuation of such practices worthwhile(Butler, 2007).

The phishing email message

The phisher would use an email message to trick a potential victim into divulging sensitive information. In most cases, the phisher will include a sense of urgency to play on the user’s mind by creating a sense of urgency. E.g. informing the user that their bank account has a problem, and they need to login again to verify that everything is still working.

The spoofed phishing web site

In creating a new phishing web site, the phisher would have used the target company's images, logo, web page layout and similar font schemes to emulate the original site and mislead the user (Rajalingam, Alomari, & Sumari, 2012). They may even include working links to the original site for the most part of the document. The site will be hosted on a server that has not been blocked or blacklisted yet. Alternatively, the phisher may have hacked another website, and included the phishing site within the path of a legitimate domain.

Lure to the phishing web site

The phisher will also try to manipulate the links, as mentioned in section 2.4.1, to deceive the user into clicking the link and filling out a form.

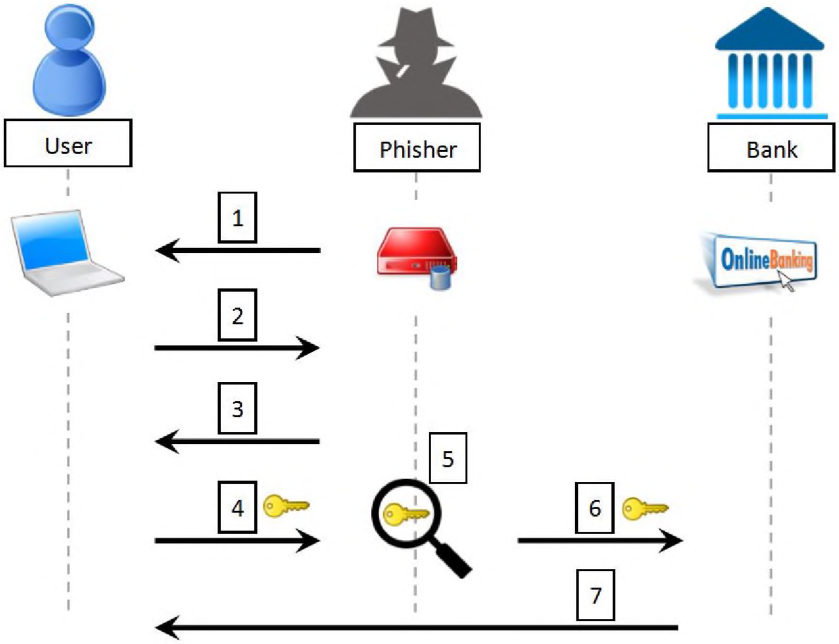


Figure 6: Sequence diagram to illustrate the flow of information in a phishing attack

To understand the basic flow of information in an email phishing attack, Figure 6 will be used, as per consensus reached between Rajalingham (2012), Butler (2007) and Chhabbra (2005), to illustrate the steps involved:

Step 1

Phisher sends out bulk phishing emails messages. For example, informing users that there is an issue with their account and they will have to log on.

Step 2

User receives the email, and is concerned about his online banking account. He unknowingly clicks on the link in the email.

Step 3

The web server where the phishing site is hosted receives the request to load the site, and responds back to the user with the phishing site.

Step 4

The user does not suspect anything wrong with the site, and logs on to his online banking profile.

Step 5

The phishing web server inspects the user's logon credentials. From here, the phisher can either return an error page, or can pass the request onto the actual online banking site and redirect to the user.

Step 6

If returning an error page, the user might become suspicious and call the bank or change his password. If the phisher is smart, he will pass the credentials onto the online banking site and redirect the user to the original web site.

Step 7

The user can log on successfully without knowing that the phisher has his online banking credentials.

This is a basic example of how an email phishing attack takes place. There are other variations which may include malware, mobile phishing and social media phishing. These are discussed in section 2.5.

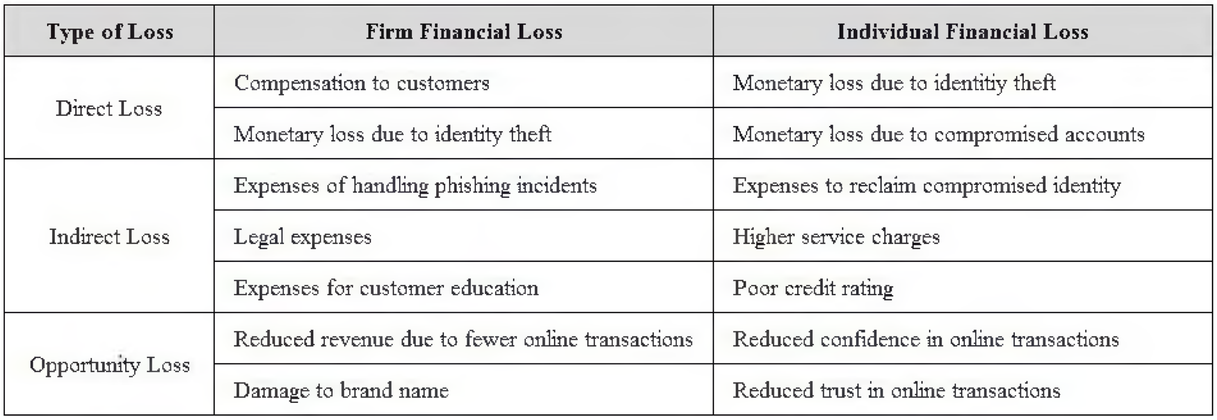
2.4.2 Motivation for Phishing

Megaw (2010) and Dlamini (2009) both agree that over the past decade, phishers have evolved from computer enthusiasts to professional hackers. Cyber criminals can now target people for either financial gains, access to email, access to network systems or even hack online games, as opposed to just for fun and showing off hacking skills (Anderson, 2007).

Strong competition in the technology marketplace have offered users more advanced services. This includes stealth-like technologies which allows for the users, and phishers, to earn additional income. Phishing attacks are low cost and simple to implement. RSA (2015) demonstrates this scenario where half a million email addresses can be purchased for as little as thirty dollars. This would imply that it would be relatively cheap to host a phishing site (RSA, 2015). These decreasing costs void any economic barriers to entering the phishing marketplace.

2.4.3 Consequences of Phishing

There are various types of associated financial losses with phishing attacks. Direct loss involves the firm compensating customers and the monetary loss affecting the customers. Indirect losses involve increased customer support for phishing victims as well as the efforts of customers to deal with credit-rating agencies to prevent themselves from being blacklisted due to the attacks. The types of loss are summarised in Figure 7 (Bose & Leung, 2014).



Type of Loss	Firm Financial Loss	Individual Financial Loss
Direct Loss	Compensation to customers	Monetary loss due to identitiy theft
	Monetary loss due to identity theft	Monetary loss due to compromised accounts
Indirect Loss	Expenses of handling phishing incidents	Expenses to reclaim compromised identity
	Legal expenses	Higher service charges
	Expenses for customer education	Poor credit rating
Opportunity Loss	Reduced revenue due to fewer online transactions	Reduced confidence in online transactions
	Damage to brand name	Reduced trust in online transactions

Figure 7: Impact of phishing attacks on firms and individuals

2.5 Variations of Phishing Techniques

The majority of phishing attacks make use of a certain type of technical manner in which to deceive the user (Singh, Somas, & Tambre, 2013). This section discusses a variety of phishing techniques being used:

2.5.1 Manipulating links

This includes making the link in an email appear to belong to the spoofed organisation, also called link obfuscation (Milletary, 2005; Hsu, Wang, & Pu, 2011; Singh et al., 2013). The use of extra subdomains and misspelled URLs are common techniques used by the phishers. Phishing URL's can be divided into three categories:

Host-domain Obfuscation The prefix of a URL contains the domain as the host of the spoofed organisation to mislead the user.

E.g. Original website: http://www.originalbank.co.za

E.g. Spoofed URL: http://www.originalbank.co.za.securesurfsites.co.za

Suffix Common Sequence This is where the URL contains the spoofed organisation's URL with an identical path.

E.g. Original website: http://secure.originalbank.co.za/banking/index.html

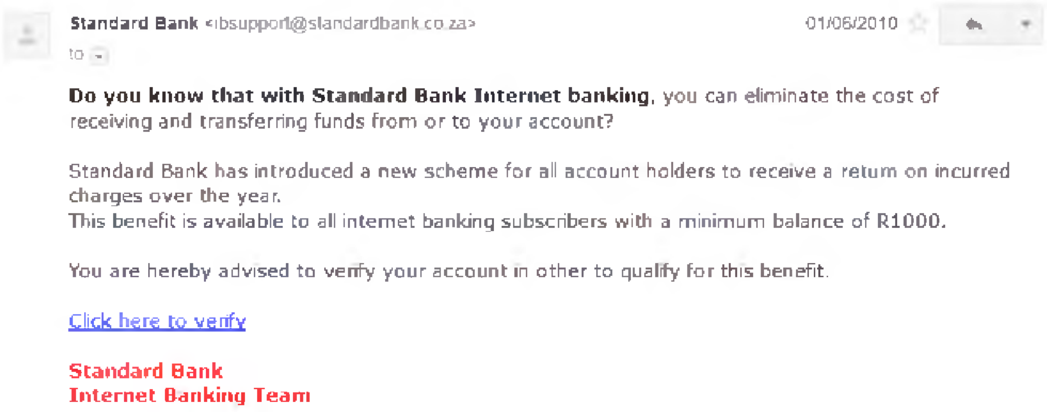
E.g. Spoofed URL: http://securesurfsites.co.za/secure.originalbank.co.za/banking/index.html

Variation This is where the URL is made up of combinations between the domain and path.

E.g. Original website: http://secure.originalbank.co.za/banking/index.html

E.g. Spoofed URL: http://secure.securesurfsites.co.za/originalbank/banking/index.html

Figure 8 shows an example of a phishing email. This email requests a user to log on to their banking profile site. The sender address has been spoofed to deceive the reader into believing that the email is legitimate.



Standard Bank <ibsupport@standardbank.co.za> 01/06/2010

to

Do you know that with Standard Bank Internet banking, you can eliminate the cost of receiving and transferring funds from or to your account?

Standard Bank has introduced a new scheme for all account holders to receive a return on incurred charges over the year.
This benefit is available to all internet banking subscribers with a minimum balance of R1000.

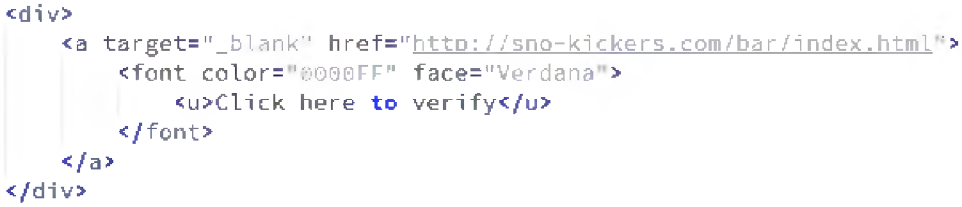
You are hereby advised to verify your account in other to qualify for this benefit.

Click here to verify

Standard Bank
Internet Banking Team

Figure 8: Example of phishing email

When looking at the source code of the phishing email, the link, shown in Figure 9 shows a URL address that is not related to the email content or company being represented in the email.



```
<div>
    <a target="_blank" href="http://sno-kickers.com/bar/index.html">
        <font color="#0000FF" face="Verdana">
            <u>Click here to verify</u>
        </font>
    </a>
</div>
```

Figure 9: The source code showing the fraudulent URL link in the phishing email

2.5.2 Filter Evasion

The use of an image in phishing emails is an evasion technique that is used by phishers to avoid text that is commonly used in phishing emails from detected by anti-phishing filters. Phishing evasion tactics include the use of optical character recognition (OCR) technologies to detect words used commonly in phishing (Singh et al., 2013). For example, an attacker would send a phishing email, similar to the example shown in Figure 8, as an embedded image in an email to avoid being detected by content scanning software. OCR technologies would scan the images for any phishing related text.

2.5.3 Spear Phishing

Spear phishing, also known as targeted phishing, are phishing attacks directed at a specific target or individual with the intent of getting access to sensitive information (Aaron & Rasmussen, 2014; Chaudhary, 2014). The phishers research information about the individual, this can be done via social networking or any place where you or someone else you know may have left personal information about you (Butler, 2005). The spear phishing email seems as if it were being originated from a personal contact, friend or trusted entity of the individual (Trend Micro Incorporated, 2012). Another variant of spear phishing attacks can be where the user will be asked to click on a link. This will in turn install malicious software which can steal information from the individual and his computer, and return this information back to the phisher. Figure 19 shows the decreasing trend of spear-phishing over the past three years (Symantec Corporation, 2015).

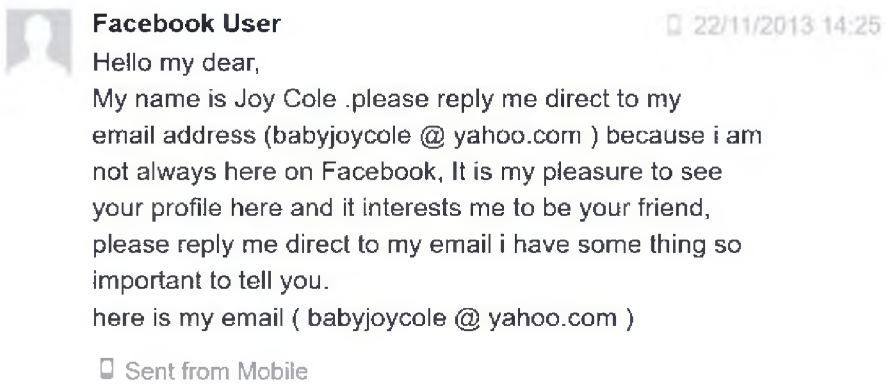
To indicate how spear phishing can be more successful, Halevi, Memon and Nov (2015) conducted a survey and a spear phishing exercise (Halevi, Memon, & Nov, 2015). This exercise involved sending a phishing mail to targeted users in an organisation, claiming to be from their IT manager. The email contained information about missing timesheets and required the user to click a link, after which a prompt to download a plugin would appear. To clear suspicious users who pay attention to phishing sites, the URL of the link was different to the company's web site. The results indicated that 62.5% of the participants clicked on the link believing that the email was intended for them from their IT manager, and 30% then went on to click the "download plugin" link without checking the URL.

2.5.4 Social Media Phishing

In a Symantec Internet Security Threat Report it is mentioned that phishing is turning to more advanced and sophisticated attack methods than ever before (Symantec Corporation, 2013). Phishers are now making use of social media sites to lure victims. If the phishers can capture a user's social media login details, they can then use the victim's account to send phishing emails and messages to all the user's friends. A message that seems to come from a friend or colleague might seem to be more trustworthy than an anonymous address. Figure 10 below is an example of a phisher attempting to make contact with a user using social media messaging.

2.5.5 Clone Phishing

Clone phishing is a type of phishing attack where an authentic email, which has been previously delivered to a recipient that contains a link, has been copied to create an almost identical email with the recipient address (Gupta, 2014). The link in the duplicate email will be a malicious link and the email is then sent off from an email address that has been spoofed as the original sender. The duplicate email may claim to be a resend of the original mail sent out or an updated version.



Facebook User 22/11/2013 14:25

Hello my dear,
My name is Joy Cole .please reply me direct to my email address (babyjoycole @ yahoo.com) because i am not always here on Facebook, It is my pleasure to see your profile here and it interests me to be your friend, please reply me direct to my email i have some thing so important to tell you.
here is my email (babyjoycole @ yahoo.com)

Sent from Mobile

Figure 10: Example of phisher contacting user through social media messaging.

2.5.6 Pharming

Pharming is a form of domain spoofing, typically achieved by compromising DNS servers or the hosts file on a user's computer which will in turn direct the user to a false website even though they have entered the correct URL of the intended website into their web browser (Grobler, 2010). Pharming is also known as DNS spoofing or DNS cache poisoning (Butler, 2005). There are malicious programs that exist which can change DNS server addresses. These types of Trojans allow the phisher to obtain the ability to force the user to any website regardless of the addresses that the user typed in the browser.

2.5.7 Google Phishing

Phishers can also make use of search engines to direct a user to a malicious shopping site. The phisher does not make contact with the user; the user instead makes use of a search engine to search for a product (Butler, 2005). The fake online shopping site will claim to be selling a product that the user searched for, usually at a promotional price to entice the victim. The phisher has no intention of making a legitimate sale, except for to collect credentials from the user. When purchasing from online stores, a user generally fills out a form and enters credit card information online. On submission of the form, an error notification is displayed, informing the user that there is a problem with the website and that the transaction has not occurred, meanwhile the phisher would have already received the information concerned, which can then be used for fraudulent purposes (Butler, 2005).

2.5.8 Mobile Phishing

Due to the rapid growth in mobile device use, the mobile domain has become a new target for phishing attacks. There are more companies offering services to user's mobile phones via applications as a M-Commerce communication medium (van der Merwe, Seker, & Gerber, 2005).

Wi-phishing involves the phishing of users who make use of wireless technology. The phisher will set up a Wi-Fi network in a public place, and a user who makes use of wireless broadband in the vicinity will connect assuming a legitimate connection. The phisher will then monitor the keystrokes and passwords entered by the user (Butler, 2005).

SMS phishing is another form of phishing via a mobile phone. The user will receive a message stating that they have signed up for a service, if not click on the link, or alternatively they will receive a message saying "your account has been suspended, please SMS your account number to activate." SMS phishing where a user received a link to a mobile site is similar to email phishing by asking the user to click on a link to a fraudulent site (van der Merwe & Seker, 2004). According to an Internet Security threat report by Symantec (2015), bank / account / loans phishing makes up 35% of all SMS spam, as illustrated in Figure 11.

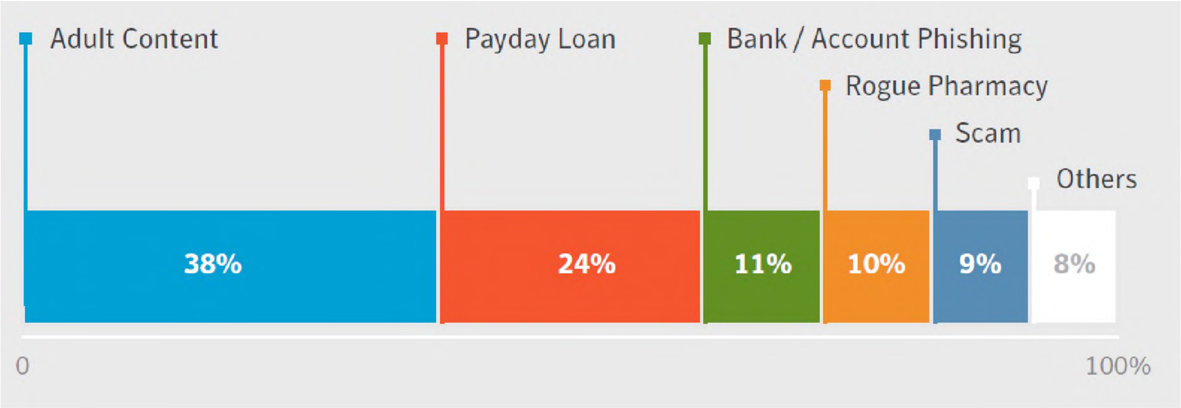
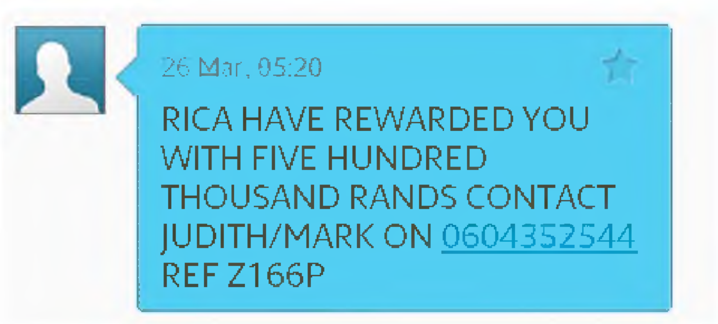


Figure 11: Categories of SMS spam (Symantec, 2015)

Alternatively, the user could receive a message quoting: "call us on 123 456 7890 to rectify your account issues." Once the fraudulent phone number is dialled, the voice prompts inform the user to enter their account numbers and PIN (Singh et al., 2013). Traditional phone equipment had dedicated fixed lines, but with Voice over IP (VOIP) being used increasingly, it offers a simpler digital platform for phishers to manipulate (Chaudhary, 2014). Figure 12 shows an example of an SMS phishing message asking the user to call the fraudulent phone number.



26 Mar, 05:20

RICA HAVE REWARDED YOU WITH FIVE HUNDRED THOUSAND RANDS CONTACT JUDITH/MARK ON 0604352544 REF Z166P

Figure 12: Example of SMS phishing

In 2015, Marforia (2015) performed a user study to evaluate the effectiveness of personalised indicators as a phishing detection mechanism. The study was focused on mobile banking applications. Results showed that users, who were fake phished with a fake banking logon screen accessed through the app, did not detect anything fraudulent. Marforia goes on to discuss how usage patterns of mobile applications are different to those of websites, as the mobile user interface is simplified (Marforio, Jayaram Masti, Soriente, Kostiainen, & Capkun, 2015). This makes it easier for cyber criminals to launch phishing attacks on mobile devices by the limitations of the mobile platform. The small screen size limits the user from fully inspecting the website for any anti-phishing security elements, perhaps even inspecting the URL itself. Users should be aware that smartphones are as much as capable as desktop browsers when visiting websites.

2.5.9 ’419’ scam

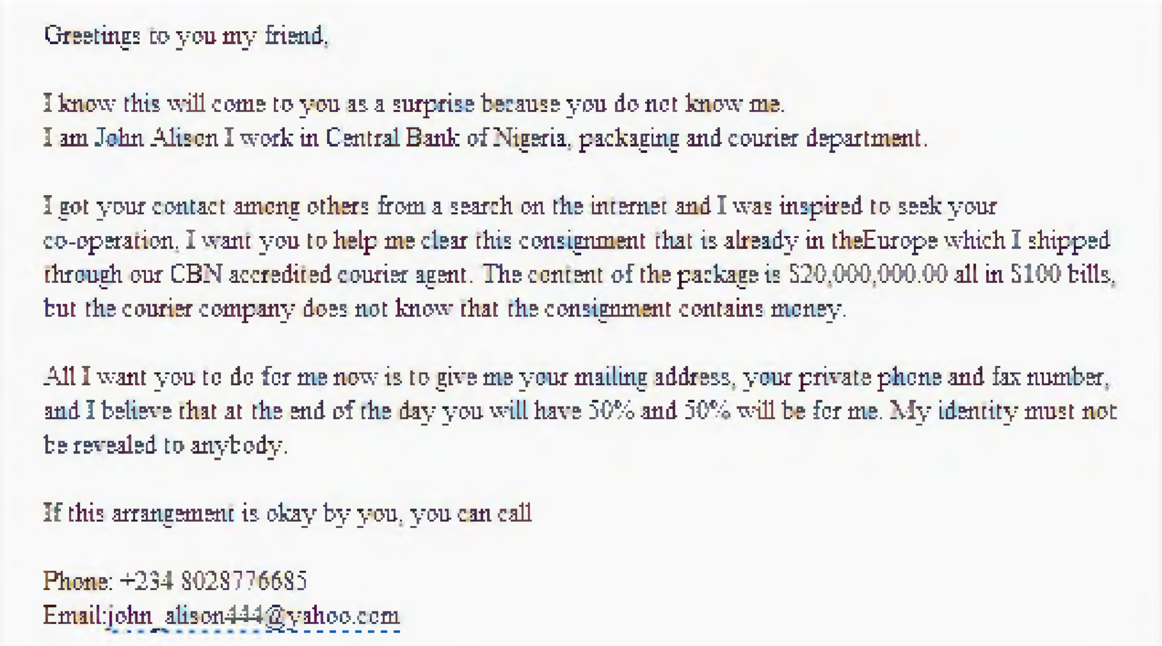
The notorious ’419’ scam is named after section 419 of the Nigerian Criminal Code (App 777 of 1990) that prohibits Advance Fee Fraud (Longe, Ngwa, Wada, Mbarika, & Kvasny, 2009). Types of Advanced Fee Fraud include “transfer of money from over-invoiced contracts”, “disbursement of money from wills/estates” and “charity relief organisation.” These involve the phisher asking the user to pay a fee into a foreign bank account to “make more money”, “claim from an inheritance” or donate to a fake “charity” cause. Figure 13 shows an example of a ’419’ scam whereby the phisher promises that the user will receive money (Gupta, 2014). Although this is out of scope, these emails can lead to phishing.

2.6 Technical: Session Hijacking

Most of the techniques mentioned above involve directing a user to a malicious web page. The user is vulnerable of being directed to a phishing site even if they have attempted to access a genuine site (Milletary, 2005). These include, but are not limited to:

2.6.1 Cross-site Scripting Attacks

Cross-site scripting (XSS) attacks can be run on web sites that requires input from the user. If the program does not correctly sanitise the input data, it is possible that malicious code



Greetings to you my friend,

I know this will come to you as a surprise because you do not know me.
I am John Alison I work in Central Bank of Nigeria, packaging and courier department.

I got your contact among others from a search on the internet and I was inspired to seek your co-operation, I want you to help me clear this consignment that is already in theEurope which I shipped through our CBN accredited courier agent. The content of the package is $20,000,000.00 all in $100 bills, but the courier company does not know that the consignment contains money.

All I want you to do for me now is to give me your mailing address, your private phone and fax number, and I believe that at the end of the day you will have 50% and 50% will be for me. My identity must not be revealed to anybody.

If this arrangement is okay by you, you can call

Phone: +234 8028776685
Email:john_alison444@yahoo.com

Figure 13: Example of a '419' scam

can be executed. For example, an attacked could include a JavaScript code onto a legitimate, but vulnerable, website to target user information and credentials.

2.6.2 Man-in-the-Middle Attacks

Man-in-the-middle attacks are when a 3rd party has the ability to access, read and edit internet traffic between two parties without either them having any knowledge.

2.6.3 Domain Name Resolution Attacks

Domain name systems (DNS) map domain names to IP addresses, and are therefore vital for users to surf the internet. One technique of targeting DNS queries is through the injection of malicious information into the authoritative DNS query responses. Another technique, also called 'Pharming' (Gupta, 2014), could be for the phisher to target the user's hosts file, which by some operating systems are checked by the local domain name resolver first, which could redirect a user to a malicious site. DNS cache poisoning attempts to feed the cache of a local DNS resolver with incorrect records (Chaudhary, 2014). This is not a complex process because DNS runs over UDP, and therefore easier to spoof UDP packets with different source addresses.

2.6.4 Domain Name Service Abuse

Dynamic DNS has enabled phishers to have a static domain name with a dynamic IP address allocated. This allows the phisher the ability to redirect traffic from one phishing site to another when the initial site gets shut down. The ease of use of dynamic DNS and the registration of multiple IP addresses for a single domain allows phishing sites to become more resilient (Milletary, 2005).

2.6.5 Domain Name Typos

Phishers can currently register domain names that are relatively close to a genuine domain. (Milletary, 2005). These sites attempt to install spyware or malware on the user who mistypes the intended domain name. For example, a user could mistakenly type www.hscbbank.com instead of www.hsbcbank.com, where www.hscbbank.com is a fraudulent website hosted by phishers.

2.6.6 Advanced Malware

With the recent advent of cyber security risks being identified, the number of possible malware attack methods have increased. Phishers can make use of electronic surveillance, password harvesters, self-contained scam pages, account siphoning tools and SQL injection scripts to steal data.

2.7 Related work

This section briefly discusses current trends of phishing in existing literature, both globally and within South Africa where information is available. Various technical reports were found useful in providing background and analysis for this research.

2.7.1 Sources of related work

This section briefly discusses the various sources of related work. Most of these sources are corporate work tied to large scale phishing and threat analysis, as minimal academic research was found around phishing trends and analysis.

Ramzan (2007), although 9 years ago, found observations in his phishing trends analysis of 2006 (Ramzan & Wuest, 2007). This is still relevant in terms of findings and possible trends associated with this research.

Symantec's Global intelligence Network has approximately 57 million attack sensors, and monitors 157 countries. On top of this they process over 8 billion email messages per month. This includes 5 million decoy accounts to detect spam phishing and malware. Every year, Symantec releases a report titled "Internet Security Threat Report", which reports back on vulnerabilities, targets, phishing. Spam, malware, data breaches, social networking as well as other new internet threats discovered. Symantec also collect phish information through an extensive anti-fraud community of enterprises, security vendors, and more than 50 million consumers (Symantec Corporation, 2013, 2014, 2015).

EMC RSA releases an annual review of phishing statistics. This includes information regarding takedowns by them, comparisons by geography and global trends (RSA, 2013, 2014a, 2014b, 2015).

APWG releases a report, titled "Global Phishing Survey: Trends and Domain Name Use" bi-annually. This reports investigates trends and their significance according to various factors which includes target distribution, phishing by uptime, top-level domains and domain registrars (Rasmussen & Aaron, 2012; Aaron & Rasmussen, 2013b, 2013a, 2014; Rasmussen & Aaron, 2014).

Kaspersky Lab have developed tools for over 16 years to protect against all types of cyber-attacks, including financial attacks. These technologies require detailed analysis of malicious samples. A summary of the analysis and findings are published in two types of reports, namely 'Kaspersky Lab Report: Financial Cyber Threats' (Kaspersky Lab, 2014, 2013) which is published annually, and a 'Spam and Phishing Statistics Report' (Gudkova & Demidova, 2014; Shcherbakova, Vergelis, & Demidova, 2014; Shcherbakova et al., 2015b; Shcherbakova, Vergelis, & Demidova, 2015a) which is published quarterly.

2.7.2 Key findings from related work

This section points out the key findings from related work. The findings are sorted by observation type relevant to the research.

Annual and monthly trend findings

The following findings are related to phishing trends observed monthly and annually.

i. According to an internet security threat report (Symantec Corporation, 2015), there was a dip in the email phishing rate between June 2014 and September 2014. As shown in Figure 14 and Figure 15, the overall phishing rate in 2014 was 1 in 965, compared with 1 in 392 in 2013, and 1 in 414 for 2012. Note: this figure represents an inverse graph, where the smaller the number the greater the risk.

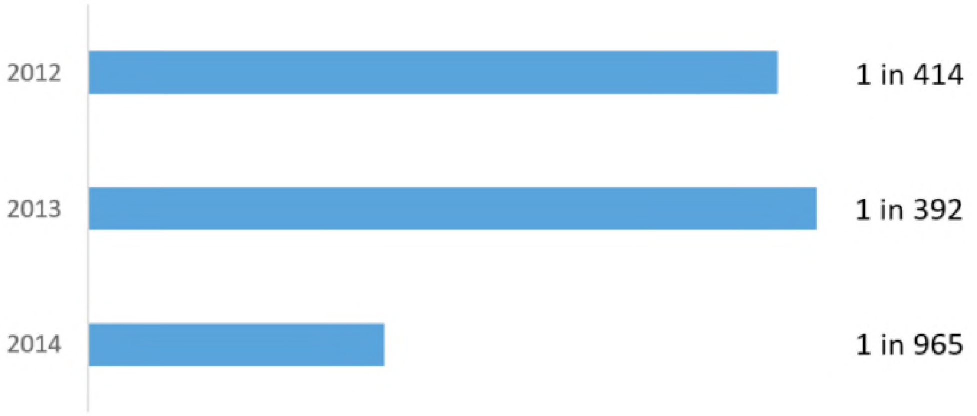


Figure 14: Email Phishing Rate for the past three years (Symantec, 2015)

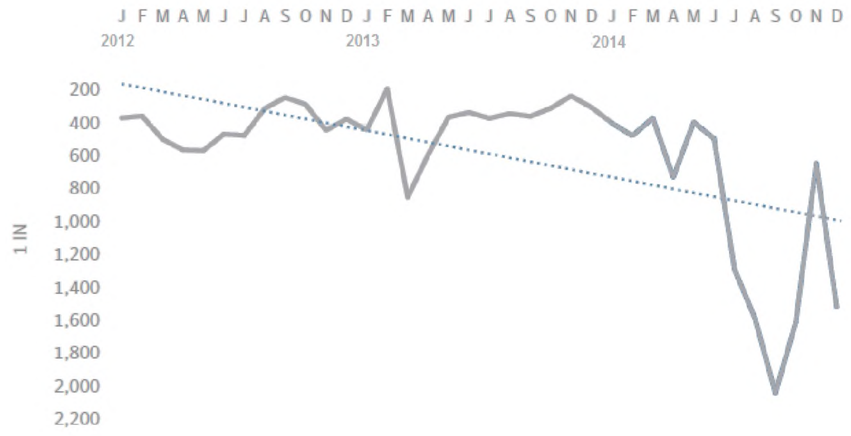


Figure 15: Phishing rate 2012 to 2014 (Symantec, 2015)

ii. Figure 16 illustrates that there was an increase in 2012 onwards of social media phishing URLs (Symantec Corporation, 2015).

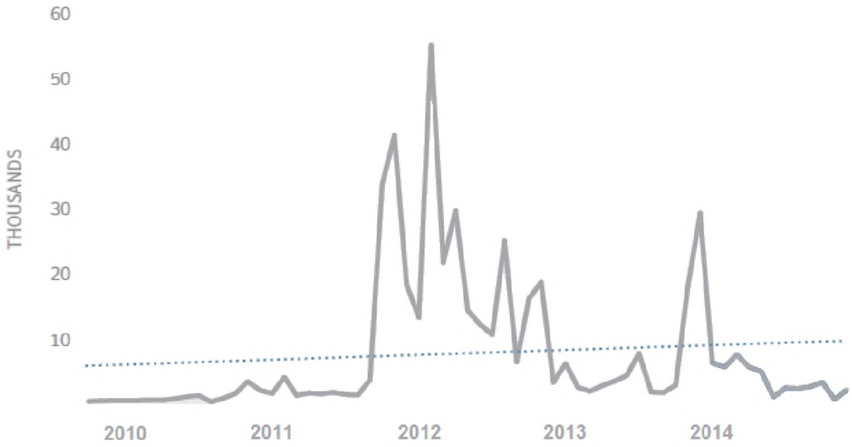


Figure 16: Phishing URLs used on Social Media for the past three years (Symantec, 2015)

iii. Figure 17 illustrates the APWG phishing trend per half-year between 2011 and 2014 for global sites. According to APWG (Aaron & Rasmussen, 2014) and RSA(RSA, 2014a) there has been an increase of 60% in phishing attacks in second half of 2013 seen both globally and in South Africa (Figure 18), as opposed to first half of 2013. In the first half of 2014, there is a noticeable difference between the global trend and South African trend. As per Figure 17 where the global trend increases, Figure 18 shows the decrease for the South African trend. The trend from APWG reports by Aaron and Rasmussen (2012, 2013, 2014) are depicted in Figure 17 and 18.

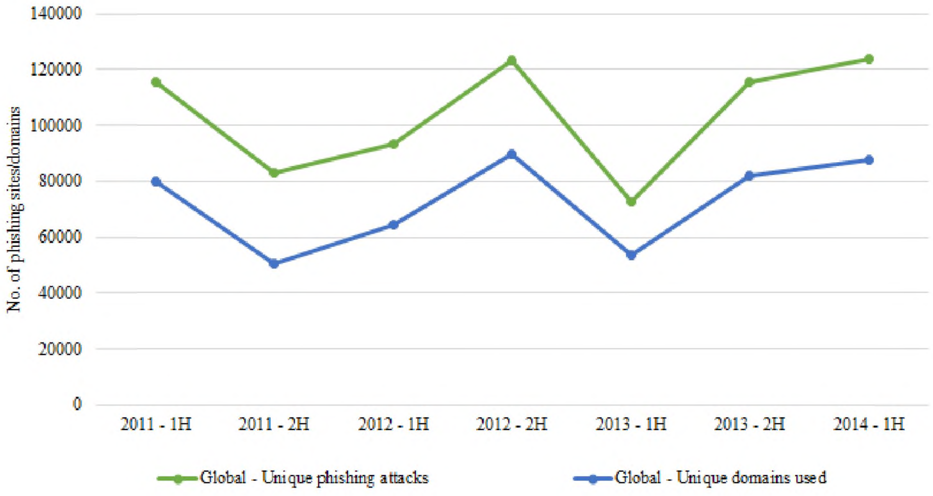


Figure 17: Global phishing sites and domains (Aaron, 2012, 2013, 2014)

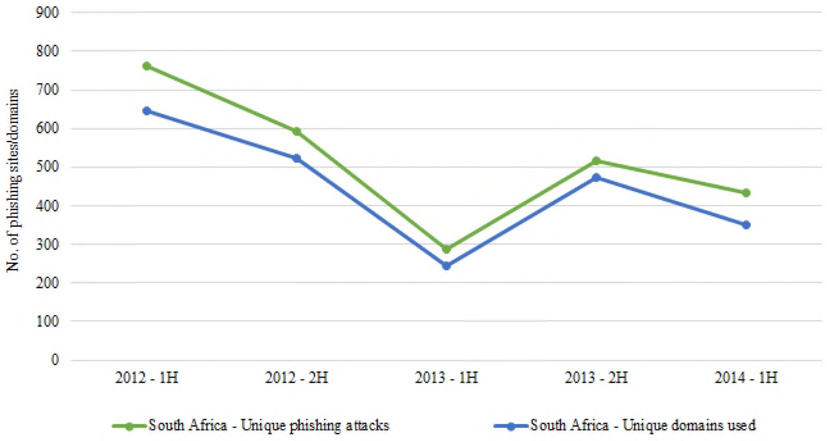


Figure 18: South Africa phishing sites and domains (Aaron, 2012, 2013, 2014)

iv. Kasperky Lab and Shcherbakova (2014, 2015) found that in Q2 of 2013, phishing fell to 0.0024% (1 in 417) of emails compared to Q1 2013 of 0.004% (1 in 250), however in Q3 2013 this grew to 0.0071% (1 in 141). By Q1 2014, the percentage of phishing emails was still at 0.0071% (1 in 141).

v. According to Symantec, the average number of spear-phishing attacks per day (between 2012 and 2014) has decreased, as shown in Figure 19.

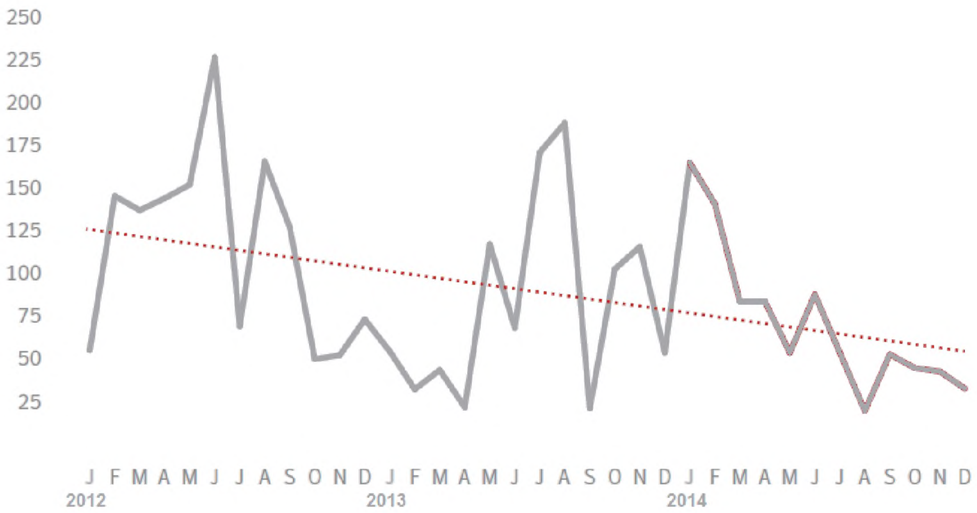


Figure 19: Spear-phishing attacks per day for the past three years (Symantec, 2015)

Daily trend findings

The following finding is related to phishing trends observed daily:

i. Ramzan (2007) noted that phishing declines considerably on weekends with more than a 20% dip in number of phishing emails (Ramzan & Wuest, 2007).

Findings related to phishing target sector

The following findings illustrate the trends related by phishing target sector:

i. 71 percent of phishing attacks in 2013, were related to spoofed financial organisation, compared to 67 percent in 2012 (RSA, 2014a).

ii. Kasperksy Lab (2014, 2015) noted that financial phishing attack, including phishing against banks, payments systems, and online shops, accounted for 28.73% of all phishing attacks in 2014, a decrease of 2.72% when compared to 2013(31.45%). In 2012, this was 22.95% of all phishing. Banking, accounted for 16.27% of all attacks in 2014, a

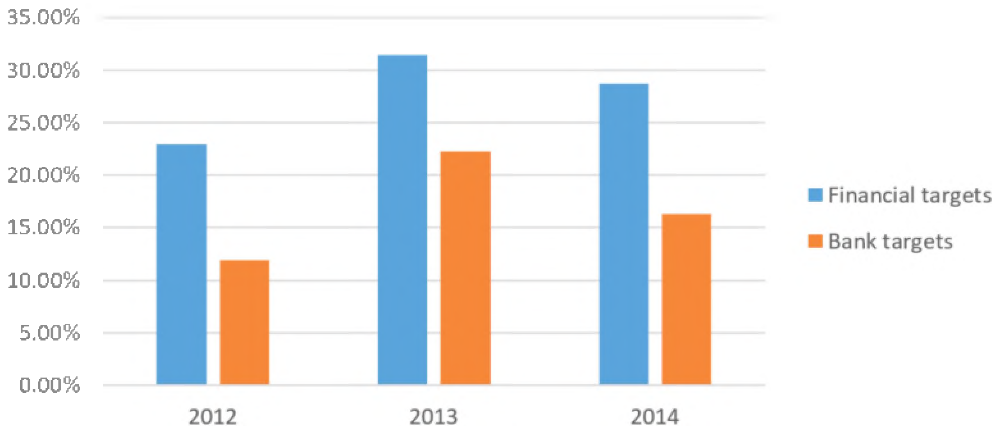


Figure 20: Phishing comparison of financial targets vs. banks

decrease of 5.93% compared to 2013(22.2%). In 2012, this was at 11.92 of all phishing. Both financial targets and banks saw an increase in 2013 over 2012, with a decrease in 2014 as illustrated in Figure 20.

iii. Gudkova and Shcherbakova (2014) notices in 2014, Visa (31%) Paypal (30%) and Mastercard (5%) make up 66% of phishing attacks on payments systems (Gudkova & Demidova, 2014; Shcherbakova et al., 2014).

Geographical findings

These findings are related to geographical trends observed:

i. In 2012, South Africa was ranked second in terms of phishing destination by geography, where 1 in 177 emails were phishing related (Symantec Corporation, 2013).

ii. In 2013, South Africa was ranked fifth, in terms of phishing targeted countries (RSA, 2013). In terms of comparing the EMEA (Europe, Middle East and Africa), South Africa ranked third, behind United Kingdom and Germany in terms of phishing attacks, making up 15% of phishing in the region (RSA, 2014b).

iii. In terms of phishing geography, Kaspersky (2014, 2015) noticed that Over half (56%) of all identified unique attack sources were found in just 10 countries, which means the attackers have a small set of preferred “home bases” to launch their attacks. Shcherbakova (2014, 2015), found that in Q1 2015, in terms of proportion: 72% of all phishing were hosted from 20 countries. In Q2 2015, this figure was 74% (Shcherbakova et al., 2015a).

Reasons for phishing

These findings are general observations from the literature review that possibly describe the reasons behind phishing trends:

i. One of the findings from the analysis of phishing activity in 2013 (Symantec Corporation, 2014) was the increase of campaigns which would target information not usually associated with standard phishing activities. These could include attempts to steal frequent flyer and loyalty card accounts, online credentials for utility accounts, and account details.

ii. According to Kaspersky (2014), in 2013 there was an in increase in banking Trojans that could see a victim’s balance, which ensures maximum profit for the phisher, especially within the mobile space. This could have added to the phishing increase in 2013. In Q3 2013 (Shcherbakova et al., 2014) there was a major increase in notifications from social networking sites were most frequently imitated in phishing emails.

iii. Shcherbakova (2014), also notice that in Sep and Nov 2013, there was an increase in phishing, particular related to the launch of Apple iPhones and iPads, as phishers used keywords of hot topics to lure users to malicious sites.

iv. Gudkova (2014) noticed a trend, as part of Kaspersky, that the football world cup caused a seasonal theme in phishing. This started from the beginning of 2014 enticing users to win tickets or prizes by entering their details, including credit card information.

v. Ramzan also noted that phishing activity increases when people are pre-occupied with other events, with almost a 30% increase noticed around the FIFA football world cup of 2006 (June to July 2006), Superbowl (January to February 2006) as well as the Christmas and New Year's holidays(Ramzan & Wuest, 2007).

vi. Kaspersky Lab noted that the number of attacks and affected users decreased by 20% in 2014, which could be due to law enforcement agencies around the world working together to prosecute criminals who were spreading financial malware and phishing (Kaspersky Lab, 2013).

2.8 Summary

This chapter discussed what phishing is, and indicated the various ways in which phishing attacks can be conducted. The chapter concluded off by looking at existing literature regarding phishing trends. In light of there being minimal information regarding phishing attacks and trends on South African targets, this research is relevant as it fits in with the existing global phishing trends. The significance of this research is that new knowledge will be created regarding phishing trends against South Africa targets and the correlation to global phishing trends.

The next chapter focusses on data collection and the data analysed for the research objectives.

Chapter 3

Data collection

3.1 Introduction

As introduced in Chapter 1, the purpose of this study is to understand phishing growth trends in South Africa as compared to global phishing growth rates. This chapter describes the data set used for the data analysis, followed by a brief description of the analysis process and research design for the data analysed in this thesis. A brief description of the tools used in the analysis of the data set are then described. The chapter concludes by discussing the study areas that will be investigated as part of the data set analysis followed by the factors of comparison that will be investigated.

3.2 Data set

The starting point of the analysis is the South African Anti-Phishing Working Group (APWG) data set, hereafter 'the data set.' This was obtained by Prof. Barry Irwin from the South African Anti-Phishing Working Group (APWG) team. This data set is an extract of the ongoing exercise performed and owned by the South African APWG. The data set contains information around phishing URLs, date-time, host country and targeted company in a MySQL database format. This database contains a mixture of global and specifically South African phishing targets augmented with data from PhishTank[1]. Table 2 shows a description of the columns in the database used by the researcher in processing this data.

Table 2: Description of database columns

Column	Data
id	represents the unique id associated to the entry
phish_id	represents the phishing id associated to the entry
url	the phishing URL reported
submission_time	date-time as to when the phishing site was reported
verified	represents whether or not the phishing site has been verified
verification_time	date-time as to when the phishing site was verified
online	represents whether or not the phishing site was online
target	represents which firm was targeted
cidr_block	indicates the IP range the site was found to be hosted within
announcing_network	represents the announcing network
rir	represents the associated internet registry
detail_time	no information for this entry
ip_address	IP address of phishing site detected
done	represents whether the site has been taken down or not
cc	indicates the hosting country of the phishing site
screened	represents whether the site has been removed

3.2.1 Data set metadata

To understand the data set, this section discusses the metadata regarding the data set being analysed. Table 3, indicates the meta data information regarding the data set.

Table 3: Data set meta data

Size	3.14GB
Number of records	579 955 phishing records
Number of unique targets	161 targeted global brands
Number of unique hosting countries	174 hosting countries
Data set time range	25 October 2006 to 17 March 2014
Working sample date range*	December 2011 to March 2014
Number of records in working sample	567 209 phishing records
Number of unique targets(brands) in working sample	159 targeted global brands
Number of unique hosting countries in working sample	174 hosting countries

[1] https://www.phishtank.com/

Due to the first 34 months of the data set, October 2006 to November 2011, only containing an average of 34 phishing emails per month, it was decided that this should be excluded from the data analysis. This would greatly skew the trends and stats. From December 2011 onwards, the average per month was 20 257 phishing sites per month, a far more realistic figure. Therefore for the purpose of this research, the* ***working sample* *of phishing statistics will be for the date range December 2011 to March 2014.*

3.2.2 Data set schema

The pre-processing of the data set included mapping the input data to tables, as well as extracting sample information, shown in table 4, to ascertain whether or not the data will be useful. From the information available on the data sets, the following columns of data were used from the schema: url, verification_time, target, cc, as this proved to be the most fruitful in terms of understanding the trends to meet the research objectives.

Table 4: Extract of data from the South African APWG data set

# id	phish_id	url	verification_time	target	cidr_block	ip_address	cc
15001	1328889	https://onsn.nazwa.pl...	2011-12-19T16:38:53+00:00	PayPal	77.55.0.0/16	77.55.0.35	PL
15002	1328888	https://mnst.nazwa.pl...	2011-12-19T16:39:54+00:00	PayPal	77.55.0.0/16	77.55.12.208	PL
15003	1328886	http://www.sign-in-us.info...	2011-12-19T16:39:23+00:00	AOL	82.165.0.0/16	82.165.202.90	DE
15004	1328885	http://paypal.com.cgi-bin.webscr...	2011-12-19T16:39:23+00:00	PayPal	50.22.64.0/18	50.22.81.93	US
15005	1328884	http://111.118.214.102...	2011-12-19T16:39:24+00:00	PayPal	111.118.212.0/22	111.118.214.102	IN
15006	1328883	http://www.golfnomad.co.uk...	2011-12-19T16:35:50+00:00	PayPal	178.250.48.0/21	178.250.54.2	GB
15007	1328882	http://nsbr-111.net...	2011-12-19T16:35:50+00:00	PayPal	66.96.128.0/18	66.96.146.30	US
15008	1328881	http://209.236.121.88...	2011-12-19T16:35:50+00:00	Other	209.236.112.0/20	209.236.121.83	US
15009	1328880	https://www.bizzkonnect.com...	2011-12-19T16:34:49+00:00	AOL	174.120.0.0/14	174.121.93.139	US
15010	1328879	http://www.cynergycomm.net...	2011-12-19T16:54:08+00:00	Cartasi	74.204.6.0/23	74.204.7.130	US

3.3 Research Design and Analysis Process

This study entitled “an investigation into the prevalence and growth of phishing attacks against South African targets” was conducted as a controlled quantitative research study that analysed the South African APWG data set for the period December 2011 to March

2014. The primary reason for analysing this data set was to compare the phishing trends of South African phishing trends against global phishing trends as per the research objectives (Section 1.1). The analysis investigated various factors including trend per month, per week and per day. These are discussed in section 3.3.5. SQL queries were used to explore data from the data set. The actual SQL statements are excluded from this research.

3.3.1 Hypothesis based analysis

Since this research is a quantitative study in an attempt to find correlation, a hypothesis based analysis of the data was used in this research. According to Nenty (2009) and Cresswell (2013), the idea behind quantitative research does not encourage the idea of accepting a null hypothesis, but more of retaining it (Nenty, 2009; Creswell, 2013). This is because failure to reject the null hypothesis still has some probability that it might be incorrect. Several hypothesis are deduced and then tested inductively based on the results of correlation and significance.

To test the significance of the correlation, a null hypothesis and alternate hypothesis were prepared:

Null Hypothesis

The null hypothesis H_{o}: ρ states that population correlation coefficient is not significantly different from 0. This will indicate that there is not enough evidence to suggest a significant linear relationship (correlation) between x and y in the data set analysed.

H_{o}: $\rho = 0$

Alternate Hypothesis

The alternate hypothesis H_{a}: ρ states that the population correlation coefficient is significantly different from 0. This will indicate that there is sufficient evidence to conclude that a significant linear relationship (correlation) exists between x and y in the data set analysed, because the correlation is significantly different from 0.

H_{a}: ρ 0

To test the hypothesis, correlation values are calculated. Correlation is a form of analysis that measures the strengths of association between two variables. There are three major types of correlation used in statistics: Pearson correlation, Kendall rank correlation and Spearman correlation. For the purpose of this research, the Pearson correlation coefficient was chosen based on examples and previous research (Salar, 1989; Hauke & Kossowski, 2011; Shumway & Stoffer, 2013). The Pearson correlation coefficient measures the degree of the relationship between linear related variables (Creswell, 2013), which describes our research objective of finding a relationship between South African phishing trends and global trends.

3.3.2 Pearson Product-Moment Correlation Coefficient

To understand the correlation statistically, the Pearson Product-Moment Correlation Coefficient (PPMCC) will be used as a correlation measurement. The primary reason for choosing the PPMCC was based on other research papers which used the PPMCC for statistical correlation to great effect (Fister, 2009; Kopsovich, 2001; Saad et al., 2009). The PPMCC indicates the strength and direction of a linear relationship between variables (Salar, 1989; Shumway & Stoffer, 2013). This coefficient is represented by the letter r, Greek symbol ρ for rho, for which the formula is:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

where, cov is the covariance, σ_X is the mean of X, and E is the expectation. In terms of measuring and understanding the scale, the correlation coefficient ranges between -1 and 1:

Positive correlation If X and Y have a strong positive linear correlation, r will be close to +1. An r value of exactly +1 indicates a perfect positive fit. Positive values indicate a relationship between x and y variables such that as values for x increases, values for y also increase. If r = +1, the slope of this line is positive.

Negative correlation If X and Y have a strong negative linear correlation, r will be close to -1. An r value of exactly -1 indicates a perfect negative fit. Negative values indicate a relationship that all data points lie on a line for which Y decreases as X increases. If r = -1, the slope of this line is negative.

No correlation If there is no linear correlation or a weak linear correlation, r is close to 0. A value near zero means that there is a random, nonlinear relationship between the two variables.

As shown in table 5, a correlation coefficient greater than 0.8 is generally described as strong, whereas a correlation less than 0.5 can be described as weak (Shumway & Stoffer, 2013). A perfect correlation of ± 1 occurs when the data points all lie exactly on a straight line. These values vary based upon the type of data being examined, and can vary with outliers. ρ is a dimensionless quantity where it does not depend on the units used. In the next chapter, the research will show the correlation coefficient calculated for the various factors of comparison.

Table 5: Strength of correlation based on coefficient

strong negative correlation	weak negative correlation	no correlation	weak positive correlation	strong positive correlation
ρ < -0.8	-0.8 < ρ < -0.5	-0.5 < ρ +0.5	+0.5 < ρ +0.8	+0.8 < ρ

3.3.3 Significance of the correlation

To test the significance of the correlation coefficient to decide whether the linear relationship is strong enough to use to model the relationship, numerous hypothesis tests have been performed (Illowsky, 2015). This will measure the correlation coefficient against a table of critical values to determine if the correlation coefficient is significant.

Table 6: Critical values used to test for correlation significance

df = n - 2 (n = sample size)	**critical value for α = 0.05**
5	0.754
10	0.576
26	0.374
29	0.355
50	0.273
100	0.195
363	0.103

For the purposes of this research, an α (alpha) level of 0.05 will be used, this is also known as the significance level. The α value 0.05 is a commonly used significance level for testing ρ for correlation (Shumway & Stoffer, 2013; Illowsky, 2015). The significance value found on the critical values correlation table using α and degrees of freedom (sample size - 2) would be the deciding point to determine whether or not ρ would indicate that 95 out of a 100 times the relationship found exists in the data set. The critical values are based on the Neyman Pearson Lemma (Von Mises, 2014; Illowsky, 2015). Table 6 shows the critical values used to test the correlation value to determine significance. The degrees of freedom (df) equals the number of values in the sample size minus two. Only the df values that were used in this research are shown in table 6, along with the associated critical values.

3.4 Analysis tools

To assist with the extraction of data and running of SQL queries on the data set, graphing, plotting and statistical calculations, the following tools were used in this study.

3.4.1 MySQL Workbench

MySQL Workbench is a visual database design tool that integrates SQL development, administration, database design, creation and maintenance into a single integrated development environment for the MySQL database system[2]. This is freely available tool. This was used to interact with the data set housed in a MySQL database.

3.4.2 Microsoft Excel

Microsoft Excel is a spreadsheet and graphing tool developed by Microsoft[3]. This is a paid for tool, and requires a license. This was used for all the creation of graphs.

3.4.3 R

R is also known as the R project for statistical computing[4]. R is a software environment for statistical computing, data analysis and graphics, and is freely available under the GNU General Public License (R Core Team, 2013). This was used for the quick calculations of the correlation coefficients and significance values.

3.5 Study Areas

To investigate if the data had an observable relationship, the research pursued the analysis of the data from different perspectives, meaning different processing steps at various levels of granularity. First, a definition of the components of the data had to be done, and then the different comparative methods that the data will be compared by had to be described. This section concludes off by showing the extracted data from the data set.

3.5.1 What will be compared

The objectives were to compare phishing trends of the South African financial phishing targets to that of global trends. A group of the combined six South African financial phishing

[2] https://www.mysql.com/products/workbench/

[3] https://products.office.com/en/excel

[4] https://www.r-project.org/

targets has been created. For the global portion, two groups were defined: global for all firms in the data set, and global top fifteen financial phishing targets as defined below.

South African financial phishing targets The comparison will be between the six combined South African financial phishing targets as shown in table 7, and as named under target in the data set. For the purposes of this research, and anonymity as requested by South African APWG, these will be called SA #1, SA #2, SA #3, SA #4 SA #5 and SA #6, as determined by a random draw. This group is referred to as the **South African combined 6** (SA combined 6) firms. This SA combined 6 group totals 4 223 phishing URLs which makes up 0.74% of the 567 209 phishing URLs for all targets in the sample size. The individual phishing targets, as well as the combined six targets will be investigated. This will allow for the creation of additional knowledge with regards to the South African financial phishing targets, which is in line with the objectives mentioned in Chapter 1.

Table 7: The six South African financial phishing targets

ABSA Bank
Capitec Bank
First National Bank
Nedbank Ltd.
South African Revenue Service
Standard Bank Ltd.

All Global phishing targets This will be all targets in the sample size, which includes 161 targeted organisations around the world, as found in the data set. There are 567 209 phishing URLs for all 161 targets for the sample size.

Top 15 Global financial phishing targets The objective of the research is to compare South African financial phishing targets with global phishing statistics. It would be interesting to see the South African combined six financial phishing targets compared to the global financial phishing targets in case there were any trends observed between financial organisations. Thus, a group of the top fifteen global financial targets have been created. This will be the fifteen highest financial targets, in terms of number of phishing sites, according to the data set. This group will be referred to as the **global top fifteen** firms. These fifteen phishing targets from the data set are shown in table 8. There are 287 692 phishing URLs for these fifteen phishing targets, which makes up 50.7% of the 567 209 phishing URLs for all targets in the sample size. Blinding is not done as individuals are not reported from this grouping.

Table 8: Global top fifteen financial phishing targets

ASB Bank Limited
Banco De Brasil
Bank of America Corporation
Barclays Bank PLC
Bradesco
Halifax
Internal Revenue Service (US)
Itau
JPMorgan Chase and Co.
Mastercard
Natwest bank
Paypal
Santander UK
Visa
Wells Fargo

3.6 Factors of comparison

The phishing statistics from the data have been compared by various factors over different time periods to find correlation. These themes of comparison are depicted in table 9.

Table 9: Themes of comparison for the phishing statistics

Theme Number	Comparison by	Comparison time period	Statistics reference
1	month	sample size time period	Table 10
2	month	month of the year	Table 11
3	week	sample size time period	Tables 24 and 25
4	week	week of the year	Table 26
5	day	day of the year	Figure 38 and 39
6	day	day of the month	Table 12
7	day	day of the week	Table 12
8	hosting country	sample size time period	Table 14 and 15

3.7 Time series analysis

The data set is a time series set of data. This research looked for various ways that correlation exists which includes comparison by day, week and month as seen in previous phishing trends, analysis and statistics (Ramzan & Wuest, 2007; Aaron & Rasmussen, 2013a; Rasmussen & Aaron, 2014; Symantec Corporation, 2015; RSA, 2015). This was in comparison between each of the six South African financial phishing targets, the combined South African six phishing targets, all global phishing targets and global top fifteen financial phishing targets as defined above.

3.8 Phishing statistics from the data set

This section presents the phishing statistics extracted from the data set. Table 10 represents the extracted data depicting the number of phishing attacks per month over the sample size time period.

Table 10: Phishing statistics per month over time

Month-Year	**Global**	**Global Fin Top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
Dec-11	7023	3032	61	46		2		13	
Jan-12	10112	5057	28	5	1	3		19	
Feb-12	10836	6241	50	18	3	22	1	6	
Mar-12	9107	4830	126	22		19	2	29	54
Apr-12	11101	6747	182	17		16		63	86
May-12	13435	7630	360	11		76	1	118	154
Jun-12	22599	16238	249	15		33	2	77	122
Jul-12	22502	15449	98	19		34		14	31
Aug-12	22911	14700	71	8	4	12	2	36	9
Sep-12	20813	12366	73	16	7	4	1	31	14
Oct-12	20299	10326	93	21	1	15	5	26	25
Nov-12	17205	5626	98	9		33	6	32	18
Dec-12	19497	8512	79	4		11	12	27	25
Jan-13	24541	14341	114	5	9	20	15	65	
Feb-13	21165	12049	154	5	5	14	16	65	49
Mar-13	18630	9908	302	7	75	41	39	58	82
Apr-13	15471	7263	317	18	101	43	37	68	50
May-13	25985	10324	200	17	91	18	18	37	19
Jun-13	18755	7773	170	8	93	11	10	30	18
Jul-13	22156	9561	156	10	77	8	14	27	20
Aug-13	27852	13869	93	16	46	2	10	11	8
Sep-13	26408	14471	170	12	59	20	22	42	15
Oct-13	27016	14001	155	16	56	12	17	30	24
Nov-13	27151	10184	224	15	51	4	25	31	98
Dec-13	36986	18924	147	17	30	7	51	11	31
Jan-14	32961	15003	176	19	28	12	52	6	59
Feb-14	24383	8559	228	12	22	15	81	9	89
Mar-14	10309	4708	49	4	6	1	21	3	14

Table 11 represents the extracted data depicting the number of phishing attacks per month of the year.

Table 11: Phishing statistics per month

Month	**Global**	**Global Fin Top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
January	34653	19398	142	10	10	23	15	84	0
February	32001	18290	204	23	8	36	17	71	49
March	27737	14738	428	29	75	60	41	87	136
April	26572	14010	499	35	101	59	37	131	136
May	39420	17954	560	28	91	94	19	155	173
June	41354	24011	419	23	93	44	12	107	140
July	44658	25010	254	29	77	42	14	41	51
August	50763	28569	164	24	50	14	12	47	17
September	47221	26837	243	28	66	24	23	73	29
October	47315	24327	248	37	57	27	22	56	49
November	44356	15810	322	24	51	37	31	63	116
December	56483	27436	226	21	30	18	63	38	56

Table 12 represents the extracted data depicting the number of phishing attacks per day of the week.

Table 12: Phishing statistics per day of the week

Day in Week	**Global**	**Global Fin Top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
Sunday	65921	33907	216	27	27	32	36	66	28
Monday	89536	47898	694	63	117	89	73	160	192
Tuesday	84408	44421	731	82	135	72	80	163	199
Wednesday	90469	46076	784	70	131	103	74	179	227
Thursday	90581	43731	790	58	175	99	72	148	238
Friday	80441	39071	697	61	121	84	58	177	196
Saturday	65853	32588	311	31	59	29	67	91	34

Table 13 represents the extracted data depicting the number of phishing attacks per day of the month.

Table 13: Phishing statistics per day of the month

Day in Month	**Global**	**Global Fin Top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
1	24431	11173	93	17	13	12	9	20	22
2	24612	14894	70	10	13	6	8	16	17
3	20501	12536	120	11	13	16	10	44	26
4	21633	12714	154	15	25	12	16	56	30
5	17525	8429	124	14	26	12	6	31	35
6	21919	10153	123	15	31	8	13	37	19
7	16605	8345	121	10	34	9	15	22	31
8	17310	8923	116	8	34	13	13	29	19
9	17622	9331	131	16	27	29	17	29	13
10	15858	7608	112	9	22	23	10	28	20
11	17892	9031	138	12	23	17	9	39	38
12	17679	9914	137	23	32	6	8	27	41
13	14837	7712	121	22	12	19	15	20	33
14	17486	8130	174	10	23	28	22	31	60
15	16749	8413	198	16	34	20	26	39	63
16	17784	8762	153	10	30	25	10	36	42
17	19593	10144	132	16	23	9	12	40	32
18	16646	7865	171	9	34	22	30	33	43
19	16585	8207	176	22	35	24	17	33	45
20	19600	11059	173	15	23	23	17	42	53
21	18698	9392	146	4	20	18	20	38	46
22	17694	9212	170	13	30	18	16	44	49
23	20437	9336	185	17	29	16	36	34	53
24	17124	8425	155	19	24	18	17	35	42
25	19551	10320	146	3	36	13	14	19	61
26	20262	11438	130	12	32	15	15	24	32
27	19746	9961	132	13	24	11	14	32	38
28	18435	8496	178	14	27	28	28	32	49
29	16987	7104	102	3	17	18	10	36	18
30	14867	6551	86	10	12	12	5	26	21
31	10541	4114	56	4	7	8	2	12	23

Table 14 represents the extracted data depicting the number of phishing attacks by hosting country for the top 20 hosting countries for global phishing targets as well as the global top fifteen phishing targets.

Table 14: Phishing statistics by hosting country for global targets

Country Code	**Global**	**Country Code**	**Global Fin Top 15**
US	231891	US	120710
GB	26097	GB	17985
FR	25746	CA	16909
DE	24963	DE	13254
CA	24798	FR	12874
BR	20547	BR	8135
CN	17503	AU	7461
AU	11836	RU	5147
RU	11218	TR	4723
NL	10073	NL	4538
TR	9438	PL	4227
PL	8095	CL	3794
IT	7569	MY	3570
EU	7415	EU	3414
CL	6993	IT	3198
UA	6376	SG	2655
MY	6202	UA	2629
ES	5738	CZ	2626
CZ	4819	ES	2615
SG	4623	RO	2388
Other (153)	95269	Other (137)	44840
Sum of Top 20	**471940**	**Sum of Top 20**	**242852**
Total	**567209**	**Total**	**287692**

Table 15 represents the extracted data depicting the number of phishing attacks by hosting country for the top 20 hosting countries for the six individual South African phishing targets.

Table 15: Phishing statistics by hosting country for South African targets

Country Code	**SA #1**	**Country Code**	**SA #2**	**Country Code**	**SA #3**	**Country Code**	**SA #4**	**Country Code**	**SA #5**	**Country Code**	**SA #6**
US	159	US	316	US	221	US	140	US	359	US	582
CA	30	CA	158	CA	37	CA	45	CA	121	CA	190
DE	18	DE	31	DE	35	FR	39	RO	52	DE	34
FR	18	FR	24	RO	31	KR	32	KR	52	PL	24
BR	18	RO	16	FR	30	RU	21	BR	42	MY	21
NL	16	GB	12	EU	20	RO	20	RU	41	TR	19
RU	12	RU	11	NL	15	BR	19	DE	39	AU	17
GB	9	MY	11	RU	11	DE	16	NL	30	GB	17
IE	9	SI	10	GB	8	DK	11	PL	22	FR	17
HK	8	NL	9	IT	7	ES	11	GB	20	NL	15
CZ	7	ES	9	DK	7	NL	11	FR	17	ES	14
IT	7	TR	9	ES	6	GB	11	ES	13	EU	14
RO	6	PL	9	TR	6	AU	8	IT	12	BR	13
EU	5	ID	7	VN	5	BY	8	TR	10	IT	11
PL	5	DK	7	AU	5	JP	5	DK	10	ZA	11
AU	5	BR	7	CL	4	IT	4	CL	9	NZ	10
JP	4	IT	7	BR	4	IN	4	ID	9	HU	8
UA	4	CL	7	JP	4	PA	4	AU	9	UA	7
CN	3	BE	5	IN	4	PL	3	UA	9	EE	6
SE	3	NZ	4	KZ	3	CL	3	PT	9	SI	5
Other (24)	46	Other (23)	96	Other (14)	45	Other (15)	45	Other (21)	99	Other (22)	79
Sum of Top 20	**346**	**Sum of Top 20**	**669**	**Sum of Top 20**	**463**	**Sum of Top 20**	**415**	**Sum of Top 20**	**885**	**Sum of Top 20**	**1035**
Total	**392**	**Total**	**765**	**Total**	**508**	**Total**	**460**	**Total**	**984**	**Total**	**1114**

3.9 Summary

In this chapter, the research discussed the South African APWG data set used in the analysis. The research design and analysis process was then discussed, which included brief descriptions of the hypothesis based analysis that was used together with the Pearson Product-Moment Correlation Coefficient and significance of correlation calculations. The chapter then briefly highlighted the various tools used before moving on to the focus study areas. The section on study areas defined which phishing targets were compared, as well as what factors of comparison were used to find correlation. The chapter concluded off by presenting the phishing statistics, extracted from the South African APWG data set, which was used in the analysis phase.

In the next chapter, the analysis of the various themes of comparison are discussed.

Chapter 4

Data Analysis

4.1 Introduction

In this chapter the results of the data analysis are presented. The data extracted from the data set is processed in order to address the objectives posed in Chapter 1. Two fundamental goals drove the subsequent data analysis. Those goals were to develop a base of knowledge about the phishing trends in South Africa, and to determine if the current perception about these trends are consistent with global trends. These objectives were accomplished. The findings presented in this chapter demonstrate the potential for merging theory and practice.

The analysis has been divided according to eight themes. These themes of comparison are as follows:

1. **Theme 1** - Phishing statistics by month over sample size time period. The purpose of Theme 1 is to understand the monthly trend over the sample size time period.

2. **Theme 2** - Phishing statistics by month within a calendar year. The purpose of Theme 2 is to understand the monthly trend within a calendar year, which would compare phishing statistics between January and December.

3. **Theme 3** - Phishing statistics by week over the time period. The purpose of Theme 3 is to understand the weekly phishing statistics over the sample size time period.

4. **Theme 4** - Phishing statistics by week within a year. The purpose of Theme 4 is to understand the weekly phishing statistics within a calendar year. This would compare

phishing statistics by week number within a year.

5. **Theme 5** - Phishing statistics by day within a year. The purpose of Theme 5 is to understand the daily phishing statistics for each day within a year. This would compare phishing statistics by day number within a year.

6. **Theme 6** - Phishing statistics by day within the month. The purpose of Theme 6 is to understand the daily phishing statistics within a month. This would compare phishing statistics by the 31 days within a month.

7. **Theme 7** - Phishing statistics by day of the week. The purpose of Theme 7 is to understand the daily phishing statistics within a week. This would compare phishing statistics by day of the week.

8. **Theme 8** - Phishing statistics by hosting country. The purpose of Theme 8 is to understand phishing statistics by hosting country.

The research now focuses on each of the themes. In each theme the research introduces the theme with a brief description, followed by plotting graphs in the discussion. The hypothesis of the theme is then defined, followed by the calculation of the Pearson correlation coefficient together with the significance values. The hypothesis is then decided upon with a brief conclusion. A summary of the theme is then briefly presented. These steps are repeated for each theme. All data and statistics indicated in this analysis chapter are relevant to the data set.

4.2 Theme 1 - Phishing statistics by month over sample size time period

The purposes of Theme 1 is to understand the monthly trend over the sample size time period among the six South African financial phishing targets, and compare these trends to the global phishing targets and the global financial top fifteen phishing targets. Table 10, depicts the monthly comparison of phishing sites over the sample size time period.

4.2.1 Theme 1 - Statistics discussion

Figure 21 compares the six South African phishing targets over the sample size time period. Visibly one can see, the increase in May 2012 for SA firm 3, SA firm 5 and SA firm 6 as shown at point A. In May 2012, point A, two of the firms launched two new products: a new payments method and an internet access product. Across all six South African phishing targets between February and April 2013, point B, there is an increase in the number of phishing sites. At the same time as point B, two of the South African phishing targets launched mobile banking applications. During August 2013, point C, there is an overall decrease. In November 2013, one of the firms launched a rewards programme which caused a rise in phishing sites. Point D, December 2013, shows a decrease for only two of the South African phishing targets.

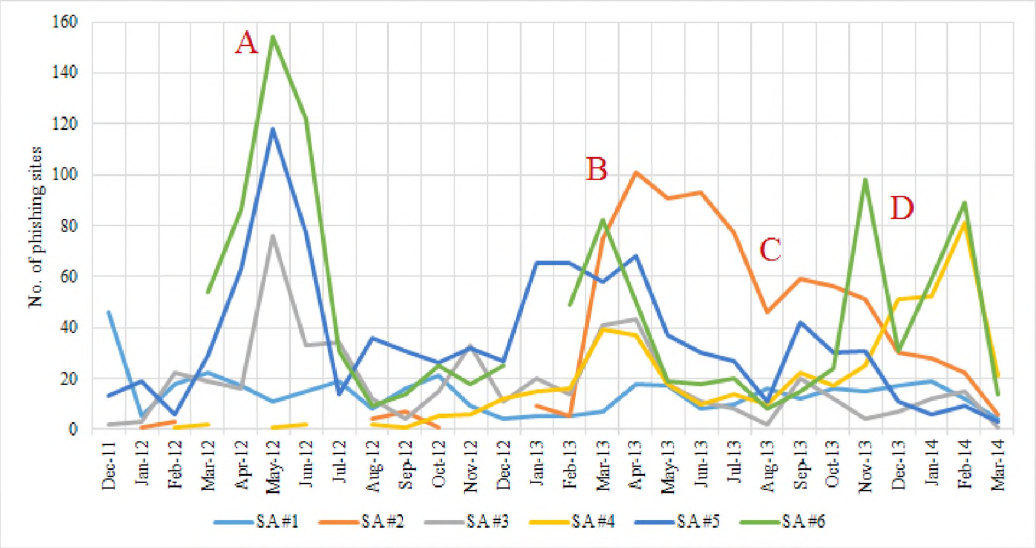


Figure 21: Phishing comparison monthly: South African combined 6

Figure 22 shows the comparison between the six combined South African phishing targets, compared to the global phishing targets and the global financial top fifteen phishing targets.

This graph has been normalised to allow for comparison on the same scale. There is an increase in phishing sites for the combined South African phishing targets from January 2012 to May 2012, at point A. The global trend increases till June 2012. Both the combined South African trend and the global trends have a spike in January 2013, shown at point B. At point C, the combined South African phishing targets have a decreasing trend from April 2013 till August 2013, whereas the decrease for the global firms only start from May 2013 onwards. While the combined South African firms have a decreasing trends towards December for both 2012 and 2013, the global firms have an increasing trend for the same month, as shown in point D.

As mentioned in Chapter 2, the trend noticed by Ramzan in 2006 indicated a general increase in phishing sites during the month of December (Ramzan & Wuest, 2007). This increase in phishing towards December 2012 and December 2013 trend still exists for global sites, but not for the South African phishing targets as illustrated in Figure 22.

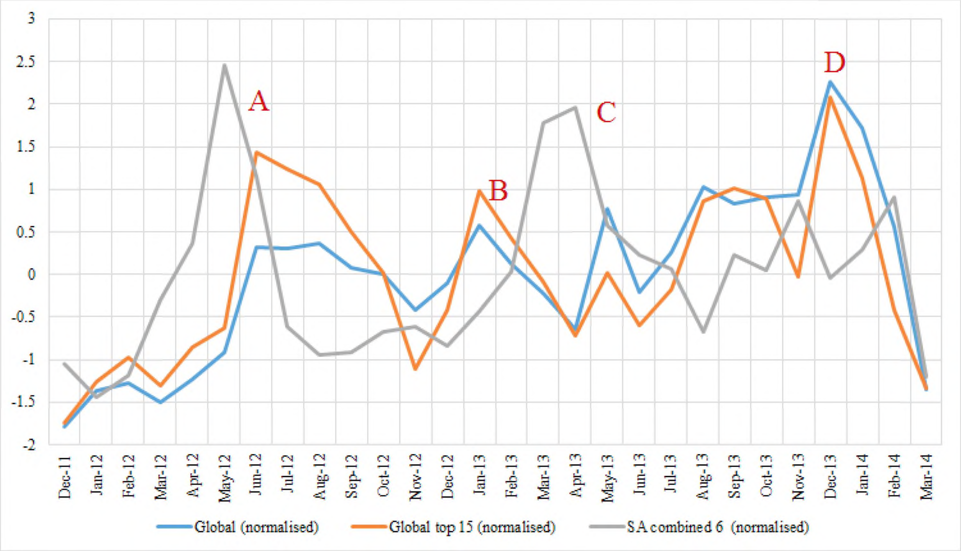


Figure 22: Phishing comparison monthly: South African combined 6 vs global (normalised)

Figure 23 shows the trend comparison of the individual South African phishing targets as compared to the six combined South African financial phishing targets. When analysing this graph, one can clearly observe that two of the South African targets have an increasing trend, whereas the other four have a decreasing/negative trend.

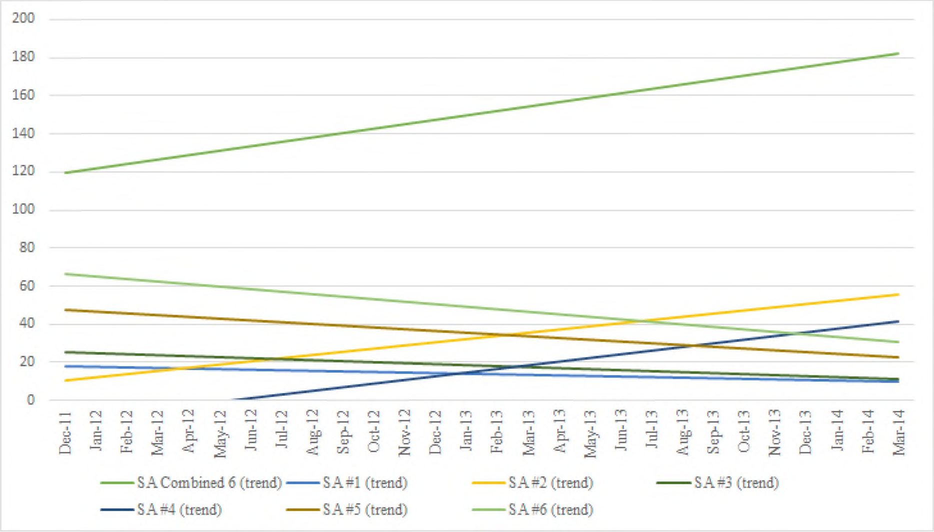


Figure 23: Trend of South African phishing sites per month

The sum of the South African combined 6 phishing targets trend indicates a positive or growing trend, which can be seen in Figure 24 when compared to the global trend and global top 15 financial phishing targets trend. However, the incline is not as steep as the global trends. This graph has been normalised for comparison on the same scale.

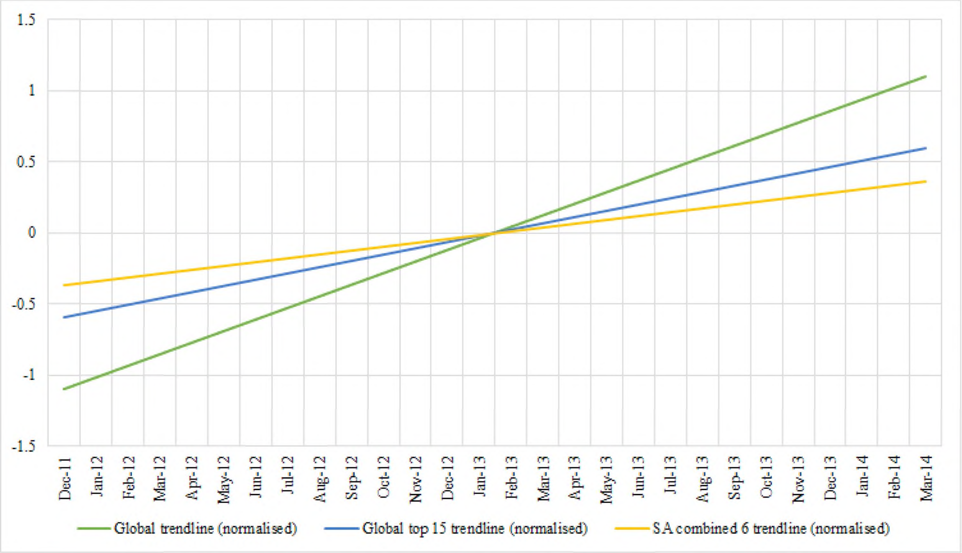


Figure 24: Trend of global and South African phishing sites per month

The research now focuses on the hypothesis to determine if any statistical correlation exists between the South African trend and the global trend.

4.2.2 Theme 1 - Hypothesis

Determine if any correlation exists between the monthly trend over the sample size time period for global phishing targets, global top 15 financial phishing targets and the South African financial phishing targets.

In table 16, the correlation values have been calculated and are shown in comparing between the data sets. With a df value of 26, and $\alpha = 0.05$, the critical value range is determined to be: -0.374 < critical value < +0.374. The values **highlighted** in table 16 indicate relationships that fall outside of the critical range.

Table 16: Correlation coefficient values for comparison by month over time

	Global	**Global top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
Global	-	**0.865**	0.170	-0.151	0.162	-0.174	**0.463**	-0.143	-0.190
Global top 15	**0.865**	-	0.103	-0.134	-0.063	-0.037	0.168	0.043	-0.117
SA combined 6	0.170	0.103	-	-0.070	**0.735**	**0.687**	0.351	**0.685**	**0.774**
SA #1	-0.151	-0.134	-0.070	-	0.167	-0.100	0.039	-0.227	0.083
SA #2	0.162	-0.063	**0.735**	0.167	-	**0.8433**	0.143	0.320	0.072
SA #3	-0.174	-0.037	**0.687**	-0.100	**0.8433**	-	-0.105	**0.736**	**0.610**
SA #4	**0.463**	0.168	0.351	0.039	0.143	-0.105	-	-0.286	0.174
SA #5	-0.143	0.043	**0.685**	-0.227	0.320	**0.736**	-0.286	-	**0.656**
SA #6	-0.190	-0.117	**0.774**	0.083	0.072	**0.610**	0.174	**0.656**	-

The following statements can be made:

Since the correlation values of the above relationships show that $\rho > 0.374$ or $\rho <$ -0.374, ρ is significant.

Decision: Reject H_o: $\rho = 0$. For Theme 1, there are two South African phishing targets that have a correlation with the global phishing trend. Figure 36 shows the negative correlation between the global trend and SA #5 in a normalised graph. Visually, the global trend increases as the week number of the year increases, as opposed to the SA #5 which decreases throughout the year. This is depicted in a normalised graph for comparative purposes, as well as the values highlighted in table 16.

Conclusion: There is sufficient evidence to conclude that there is a linear relationship between the pairs of data highlighted in table 16 when compared monthly over the sample size time period, because the correlation is significantly different from 0. For the remaining relationships, the null hypothesis cannot be rejected, as ρ is not significant.

4.2.3 Theme 1 - Summary

From Theme 1, there is only one South African phishing target that has a significant correlation with the global phishing trend. Figure 25 shows the relationship between SA #4 and the global trend, for which a significant correlation was found. The graph illustrated in Figure 25 has been normalised to allow for comparison on the same scale.

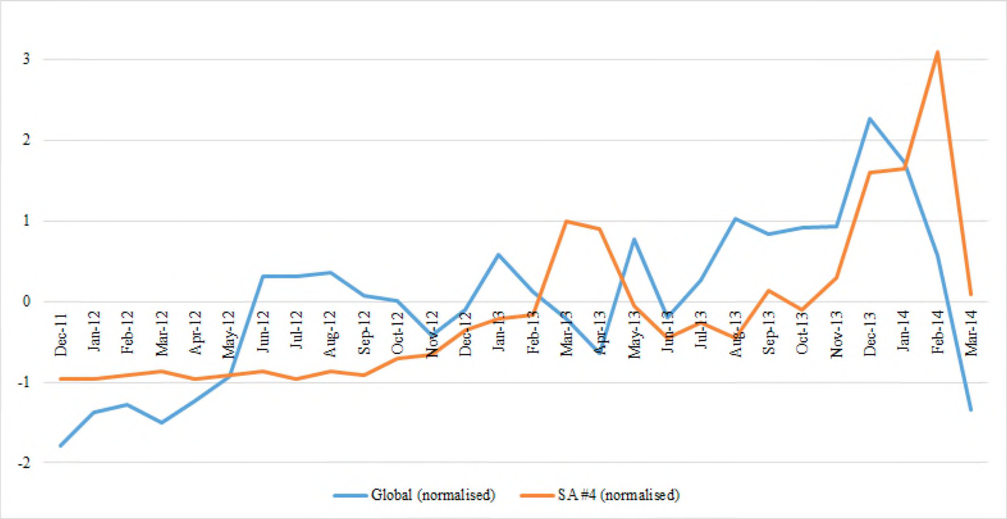


Figure 25: Phishing comparison monthly: global and SA #4 (normalised)

The next theme also focuses on monthly statistics, but by month within a calendar year.

4.3 Theme 2 - Phishing statistics by month within a calendar year

The statistics and graphs in Theme 1 compared the phishing targets monthly over the sample size time period. Theme 2 presents the comparison of phishing sites compared monthly within a year, to understand if any correlation existed during the year.

4.3.1 Theme 2 - Statistics discussion

Figure 26 illustrates the phishing trends between the South African combined six financial phishing targets and global firms by month. Visibly, one can see that all six South African phishing targets had an increase in phishing sites from February to March, as shown in point A. For the month of May, point B, three of the South African phishing targets had their peak amount of phishing sites. Between July and August, the majority of the South African targets had a decrease in phishing sites, shown at point C. At point D, five out of the six South African phishing targets had a decreased amount of phishing sites. This is in contrast with global trends, which have an increase in December as shown in Figure 27. This is also mentioned by Ramzan, that phishing increases in December (Ramzan & Wuest, 2007).

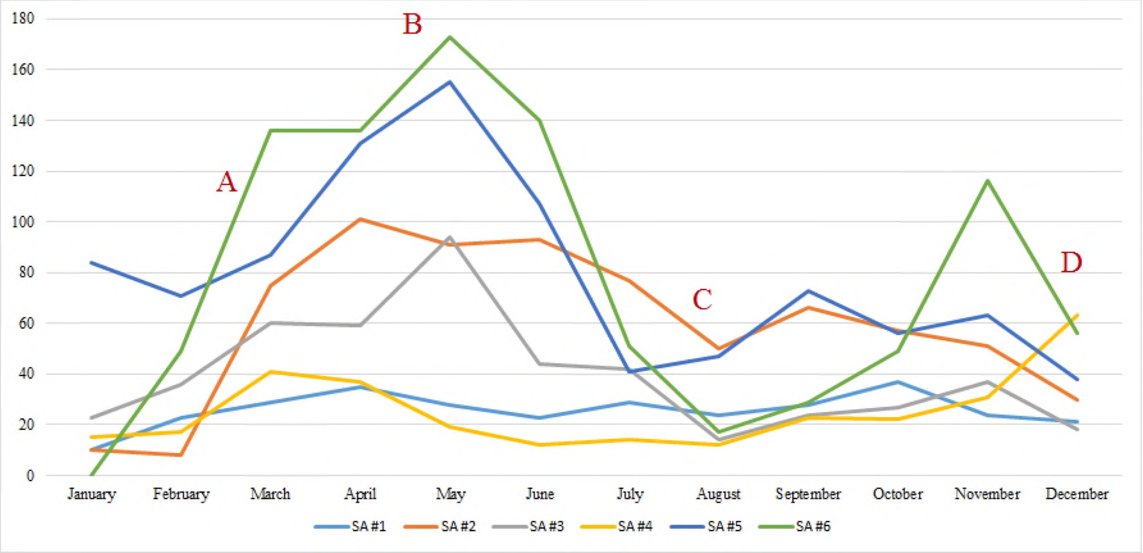


Figure 26: Phishing comparison by month for South African phishing targets

Figure 27 shows the same monthly comparison, but for the South African combined six phishing targets compared to the global statistics. At point A in Figure 27, The South African trend increases for the first 5 months of the year, whereas the global trends show decreasing trend for the first 4 months of the year. With May being the peak month for South African phishing targets, there is also an increase noticed in the global trends, as shown in point B in Figure 27. In August (point C), the global number of phishing sites increases, as opposed to South African phishing sites which show a decrease, as shown in point C in Figure 27. The trend for the last three months of the year is different for global trends as compared to South African combined trend. At point D, the global trends show a decrease from October to November, and an increase from November to December. For the South African combined targets, the case is the inverse, where there is an increase from October to November, and a decrease from November to December.

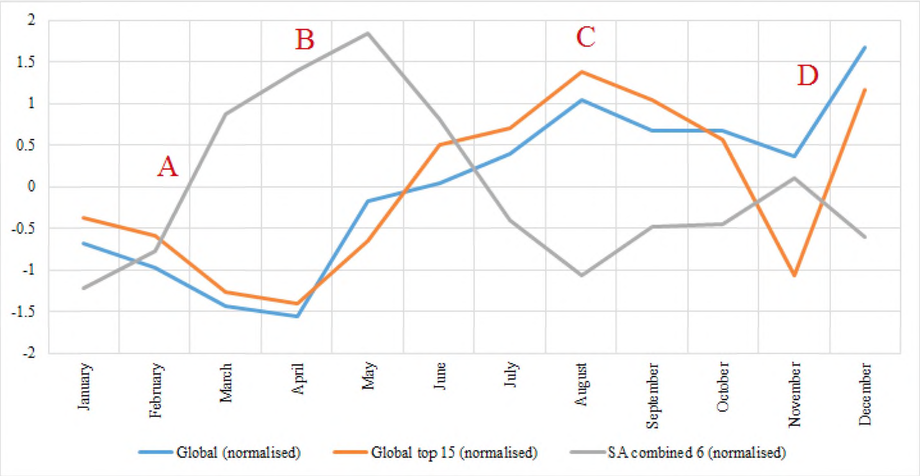


Figure 27: Phishing comparison by month: SA combined 6 and global (normalised)

The research now focuses on the hypothesis to determine if any statistical correlation exists between the South African monthly trend and the global monthly trend.

4.3.2 Theme 2 - Hypothesis

Determine if any correlation exists between the monthly trend within a year for global phishing targets, global top 15 financial phishing targets and the South African financial phishing targets.

In table 17, the correlation values have been calculated and are shown in comparison of the data sets. With a df value of 10, and $\alpha = 0.05$, the critical value range is determined to be: -0.576 < critical value < +0.576. The values **highlighted** in table 17 indicate relationships that fall outside of the critical range.

Table 17: Correlation coefficient values for comparison by month

	Global	**Global top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
Global	-	**0.847**	-0.469	-0.079	-0.175	-0.547	0.106	**-0.610**	-0.406
Global top 15	**0.847**	-	-0.525	-0.098	-0.157	**-0.634**	-0.137	**-0.592**	**-0.595**
SA combined 6	-0.469	-0.525	-	0.512	**0.837**	**0.898**	0.113	**0.859**	**0.888**
SA #1	-0.079	-0.098	0.512	-	**0.652**	0.352	0.104	0.154	0.385
SA #2	-0.175	-0.157	**0.837**	**0.652**	-	**0.633**	-0.042	0.550	**0.713**
SA #3	-0.547	**-0.634**	**0.898**	0.352	**0.633**	-	-0.024	**0.826**	**0.854**
SA #4	0.106	-0.137	0.113	0.104	-0.042	-0.024	-	-0.131	0.200
SA #5	**-0.610**	**-0.592**	**0.859**	0.154	0.550	**0.826**	-0.131	-	**0.724**
SA #6	-0.406	**-0.595**	**0.888**	0.385	**0.713**	**0.854**	0.200	**0.724**	-

The following statements can be made:

Since the correlation values of the above relationships show that $\rho > 0.576$ or $\rho < -0.576$, ρ is significant.

Decision: Reject H_o: $\rho = 0$, for the values **highlighted** in table 17.

Conclusion: There is sufficient evidence to conclude that there is a linear relationship between the pairs of data **highlighted** in table 17 when compared monthly over the sample size time period, because the correlation is significantly different from 0. For the remaining relationships, the null hypothesis cannot be rejected, as ρ is not significant.

4.3.3 Theme 2 - Summary

From Theme 2, comparison per month in a year, there were three South African phishing targets that showed a significant relationship with the global trends. Figure 28 shows the global phishing trend and SA #5 compared in a normalised graph. The correlation was a negative correlation, which means an inverse relationship existed. At point A, the global trend for April was at its lowest point of the year, whereas with SA #5, May was the highest point of the year in terms of number of phishing sites. At point B, between May and July, the trend of SA #5 decreases where the global trend shows an increase. At point C, as mentioned before in Figure 27, the global trend increases from November to December to the global peak by month, as compared to SA #5 where the trend decreases from November to December to the lowest point in the year.

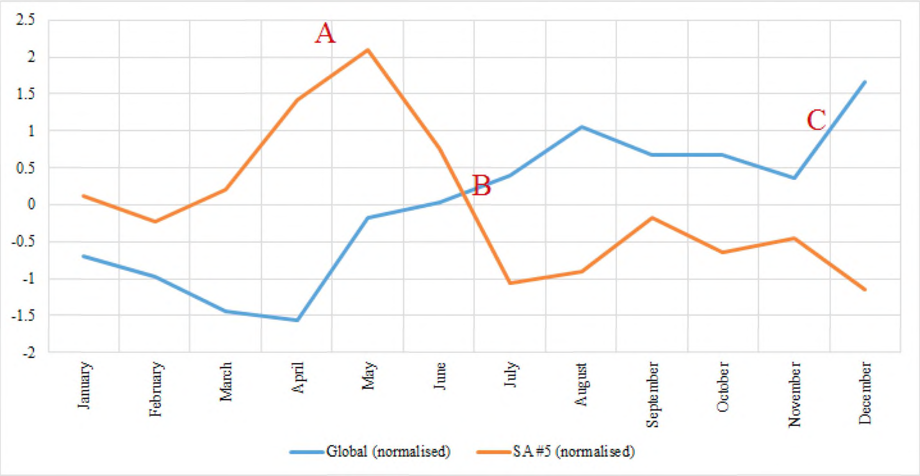


Figure 28: Monthly comparison between global and SA #5 within a year (normalised)

In Theme 2, a negative correlation was also observed between the combined global top 15, SA #3, SA #5 and SA #6. The normalised comparison of this correlation is shown in Figure 29. At point A, there was an increase from January to May for the three South African phishing targets to there respective peaks in the year, as opposed to the decreasing trend of the combined global top 15 to the lowest point of the year. From May to August, the inverse relationship continued as shown in point B. At point B, the three South African phishing targets decreased from May to August, whereas the combined global top 15 increased from

April to August. As noticed in Figure 27 and Figure 28, there was an inverse relationship observed over the last three months of the year, as indicated at point C of Figure 29.

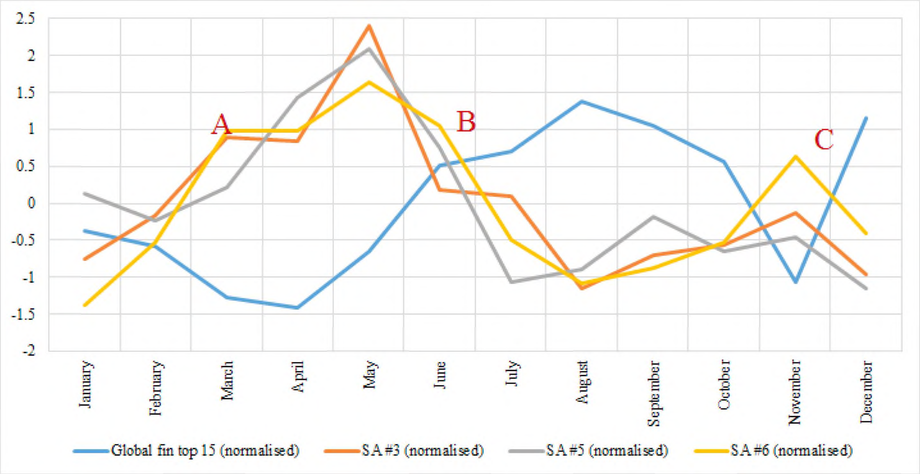


Figure 29: Monthly comparison between global, SA #3, SA #5 and SA #6 (normalised)

In Themes 1 and 2, the research looked for correlation of phishing statistics by month. The next two themes will focus on phishing statistics by week.

4.4 Theme 3 - Phishing statistics by week over the sample size time period

The purpose of Theme 3 is to understand the weekly trend over the sample size time period among the six South African financial phishing targets and compare these trends to the global phishing targets. The sample size time range is 121 weeks.

4.4.1 Theme 3 - Statistics discussion

Table 24 and table 25 in Appendix A, depicts the weekly comparison of phishing sites. Figure 30 shows the weekly trend of the global phishing targets in a normalised graph over the 121 week period of the sample size. At week 36 of 2013, point A, and week 49 of 2013, point B, there was an increase in the phishing sites for both global trends and the global top fifteen financial phishing targets as shown in both Figure 30 and Figure 31.

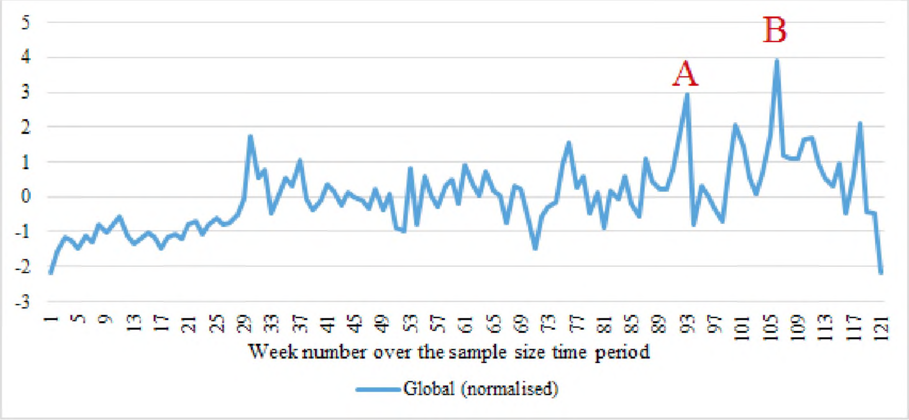


Figure 30: Trend of global phishing targets over 121 weeks (normalised)

Figure 31 shows the weekly trend of the global top fifteen financial phishing targets in a normalised graph over the 121 week period of the sample size.

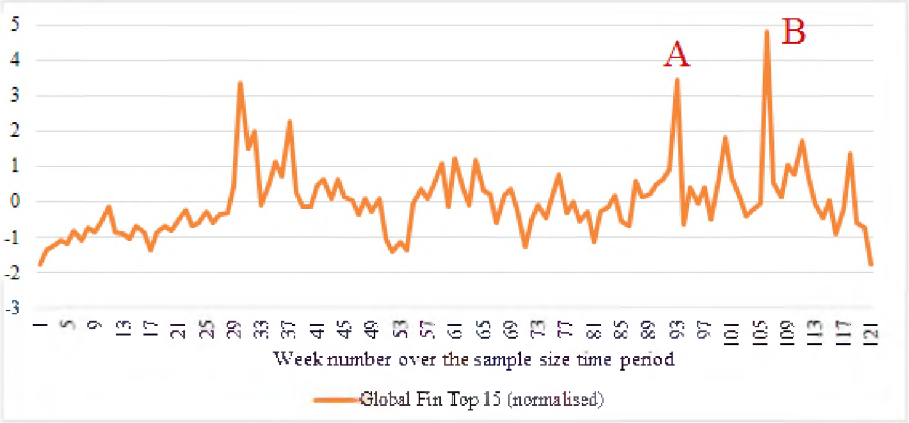


Figure 31: Trend of global financial top 15 phishing targets over 121 weeks (normalised)

Figure 32 shows the weekly trend of the six South African combined phishing targets in a normalised graph over the 121 week period of the sample size. Unlike Figure 30 and Figure 31 which show global trends, Figure 32 shows the South African combined six phishing targets with different peaks. At point A in Figure 32, the peaks are at week 11 and week 17 of 2013. At point B, the graph does not follow the same trend as the global trend for December statistics.

From Theme 4, there are two South African phishing targets that have a correlation with the global phishing trend. Figure 36 shows the negative correlation between the global trend and SA #5 in a normalised graph. Visually, the global trend increases as the week number of the year increases, as opposed to the SA #5 which decreases throughout the year. This is depicted in a normalised graph for comparative purposes.

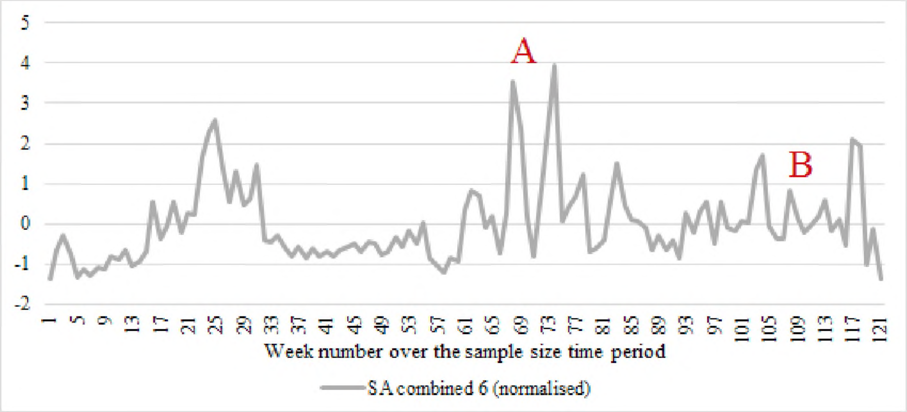


Figure 32: Trend of South African combined 6 phishing targets over 121 weeks (normalised)

The research now focuses on the hypothesis to determine if any statistical correlation exists between the South African trend and the global trend.

4.4.2 Theme 3 - Hypothesis

Determine if any correlation exists between the weekly trend over the sample size time period for global phishing targets, global top 15 financial phishing targets and the South African financial phishing targets.

In table 18, the correlation values have been calculated and are shown in comparing between the data sets. With a df value of 26, and $\alpha = 0.05$, the critical value range is determined to be: -0.195 < critical value < +0.195. The values highlighted in table 18 indicate relationships that fall outside of the critical range.

Table 18: Correlation coefficient values for comparison by week over time

	Global	Global top 15	SA combined 6	SA #1	SA #2	SA #3	SA #4	SA #5	SA #6
Global	-	**0.858**	0.187	-0.169	-0.076	-0.103	**0.261**	-0.095	-0.070
Global top 15	**0.858**	-	0.099	-0.202	-0.131	-0.058	0.119	-0.027	-0.019
SA combined 6	0.187	0.099	-	0.035	**0.685**	**0.567**	**0.542**	**0.638**	**0.62**
SA #1	-0.169	-0.202	0.035	-	0.043	-0.072	0.268	-0.081	0.094
SA #2	-0.076	-0.131	**0.685**	0.043	-	0.496	-0.033	**0.357**	0.126
SA #3	-0.103	-0.058	**0.567**	-0.072	0.496	-	0.106	**0.443**	**0.315**
SA #4	**0.261**	0.119	**0.542**	0.268	-0.033	0.106	-	-0.028	**0.439**
SA #5	-0.095	-0.027	**0.638**	-0.081	**0.357**	**0.443**	-0.028	-	**0.389**
SA #6	-0.070	-0.019	**0.692**	0.094	0.126	**0.315**	**0.439**	**0.389**	-

The following statements can be made:

Since the correlation values of the above relationships show that $\rho > 0.195$ or $\rho <$ -0.195, ρ is significant.

Decision: Reject H_o: $\rho = 0$, for the values **highlighted** in table 18.

Conclusion: There is sufficient evidence to conclude that there is a linear relationship between the pairs of data **highlighted** in table 18 when compared weekly over the sample size time period, because the correlation is significantly different from 0. For the remaining relationships, the null hypothesis cannot be rejected, as ρ is not significant.

4.4.3 Theme 3 - Summary

From Theme 3, there is only one South African phishing target that has a significant correlation with the global phishing trend. Figure 33 shows the relationship between SA #4 and the global trend, for which a correlation was found. The graph has been normalised to allow for comparison on the same scale.

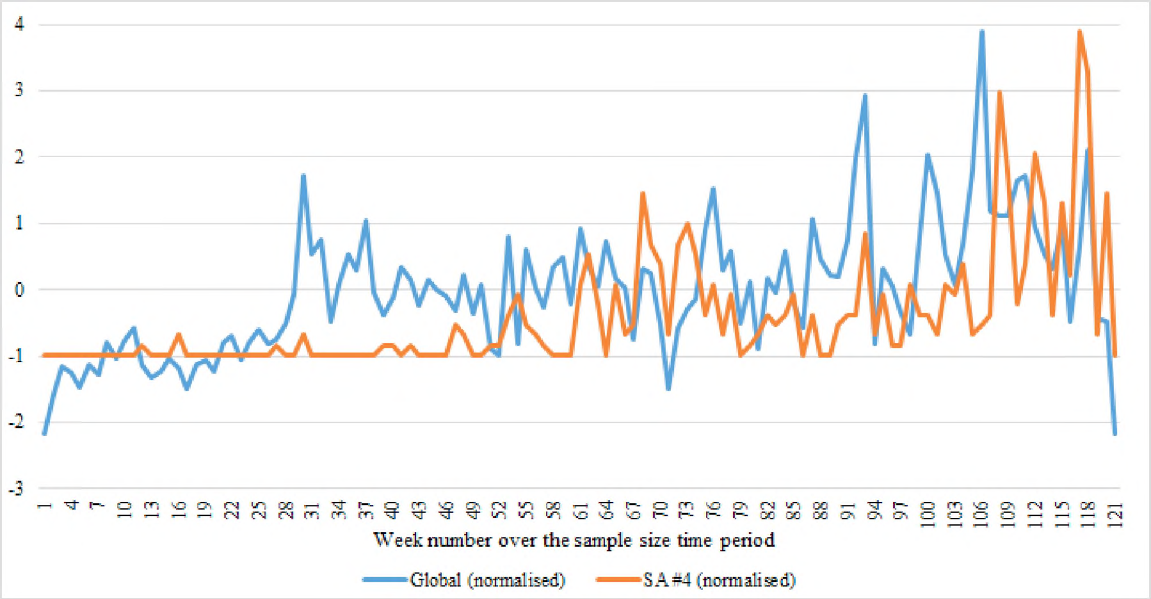


Figure 33: Global and SA #4 phishing trends over 121 weeks (normalised)

Theme 3 investigated the weekly phishing statistics over the sample size time period. In the next theme, the research discusses phishing statistics per week within a calendar year.

4.5 Theme 4 - Phishing statistics by week within a calendar year

Theme 3 investigated the weekly phishing statistics over the sample size time period of 121 weeks. The purpose of Theme 4 is to understand the weekly trend within a calendar year among the six South African financial phishing targets, and compare these trends to the global phishing targets and the global top fifteen financial phishing targets.

4.5.1 Theme 4 - Statistics discussion

Table 26, in Appendix A, depicts the weekly comparison of phishing sites over the period of a year. The normalised graph comparison of the global phishing targets and the South African combined 6 phishing targets are shown in Figure 34.

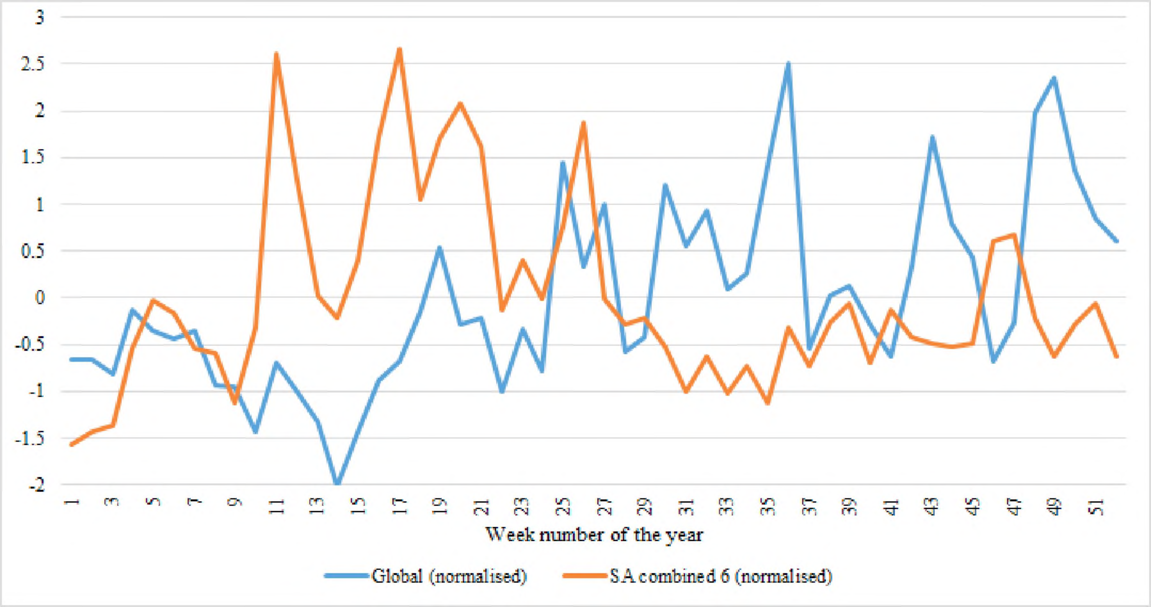


Figure 34: Phishing comparison by week: Global and South African combined 6

Figure 35 shows the weekly statistics for the individual South African phishing targets.

The research now focuses on the hypothesis to determine if any statistical correlation exists between the South African trend and the global trend.

4.5.2 Theme 4 - Hypothesis

Determine if any correlation exists between the weekly trend within a calendar year for global phishing targets, global top 15 financial phishing targets and the South African financial phishing targets.

In table 19, the correlation values have been calculated and are shown in comparing between the data sets. With a df value of 50, and $\alpha = 0.05$, the critical value range is determined to be: -0.273 < critical value < +0.273. The values highlighted in table 19 indicate relationships that fall outside of the critical range.

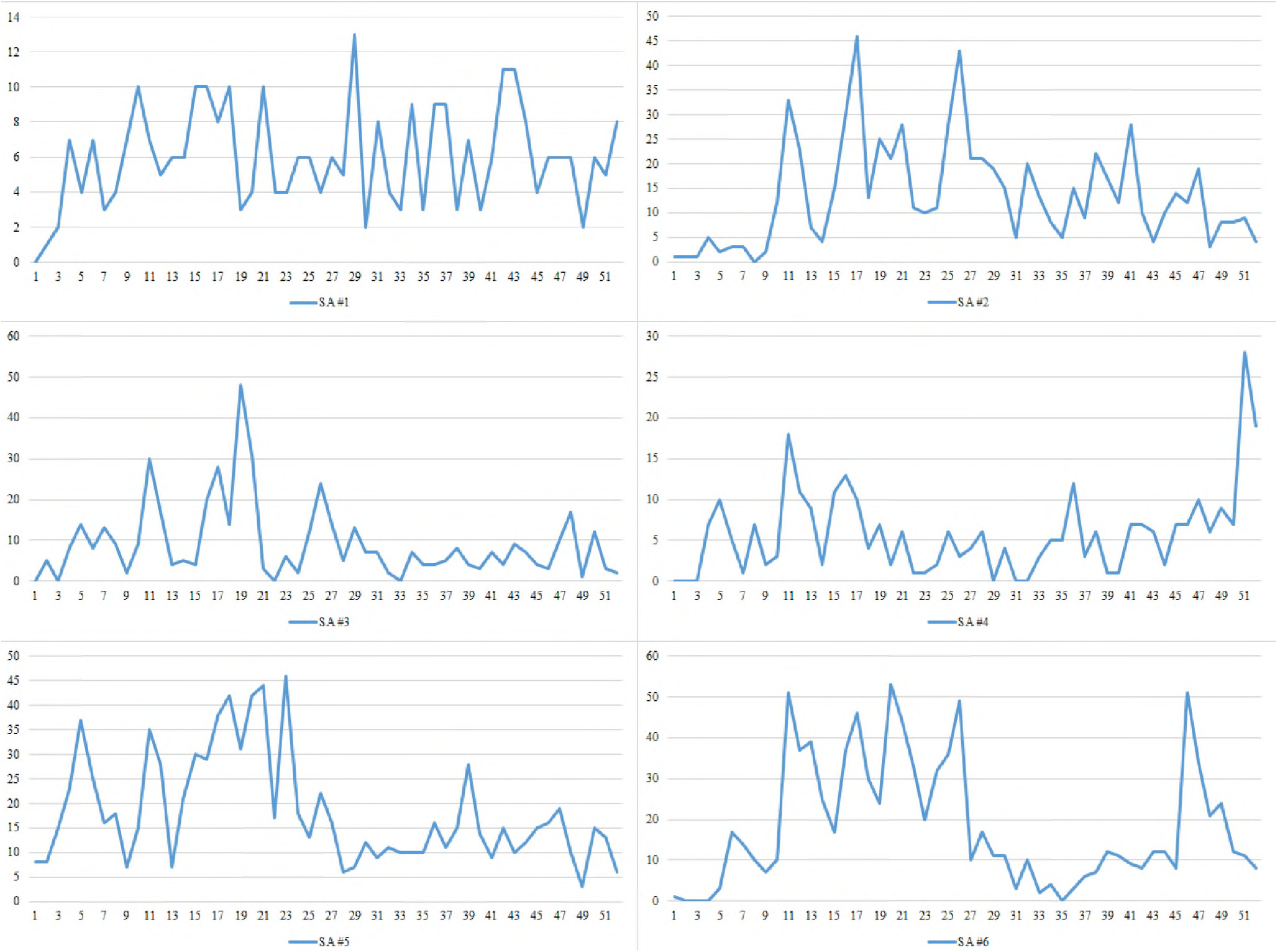


Figure 35: Phishing comparison by week of the year: South African phishing targets

Table 19: Correlation coefficient values for comparison by week within a year

	Global	**Global top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
Global	-	**0.809**	-0.157	-0.046	-0.025	0.012	0.150	**-0.284**	-0.226
Global top 15	**0.809**	-	-0.211	-0.091	0.066	-0.097	0.042	**-0.319**	**-0.291**
SA combined 6	-0.157	-0.211	-	0.237	**0.793**	**0.417**	**0.345**	**0.736**	**0.854**
SA #1	-0.046	-0.091	0.237	-	0.161	0.046	0.184	0.142	0.111
SA #2	-0.025	0.066	**0.793**	0.161	-	**0.551**	0.188	**0.398**	**0.612**
SA #3	0.012	-0.097	**0.417**	0.046	**0.551**	-	0.161	**0.504**	**0.482**
SA #4	0.150	0.042	**0.345**	0.184	0.188	0.161	-	0.105	0.191
SA #5	**-0.284**	**-0.319**	**0.736**	0.142	**0.398**	**0.504**	0.105	-	**0.511**
SA #6	-0.226	**-0.291**	**0.854**	0.111	**0.612**	**0.482**	0.191	**0.511**	-

The following statements can be made:

Since the correlation values of the above relationships show that $\rho > 0.273$ or $\rho <$ -0.273, ρ is significant.

Decision: Reject H_o: $\rho = 0$, for the values **highlighted** in table 19.

Conclusion: There is sufficient evidence to conclude that there is a linear relationship between the pairs of data **highlighted** in table 19 when compared weekly within a calendar year, because the correlation is significantly different from 0. For the remaining relationships, the null hypothesis cannot be rejected, as ρ is not significant.

4.5.3 Theme 4 - Summary

From Theme 4, there are two South African phishing targets that have a correlation with the global phishing trend. Figure 36 shows the negative correlation between the global trend and SA #5 in a normalised graph. Visually, the global trend increases as the week number of the year increases, as opposed to the SA #5 which decreases throughout the year. This is depicted in a normalised graph for comparative purposes.

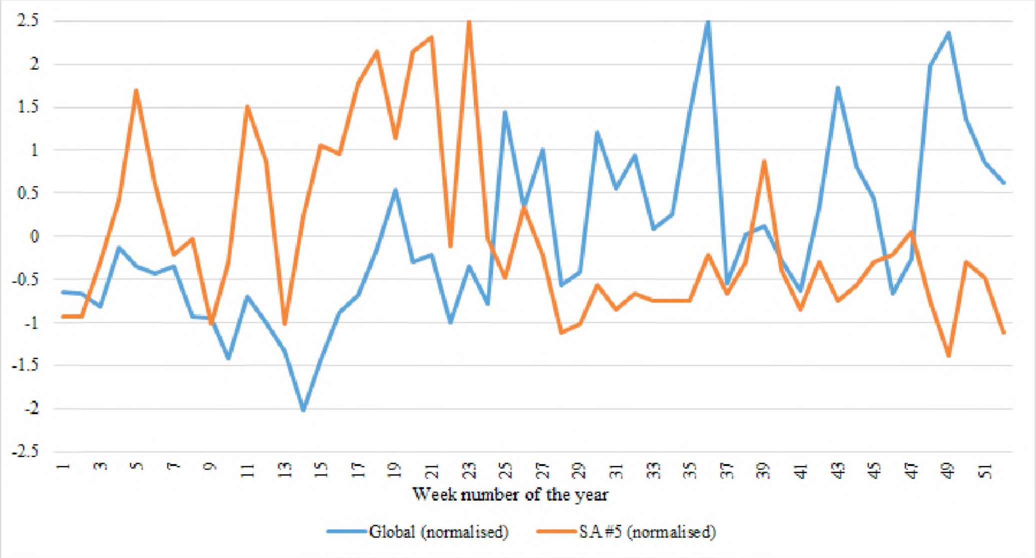


Figure 36: Phishing comparison per week: global and SA #5 (normalised)

Table 26, in Appendix A, depicts the weekly comparison of phishing sites per week of 2012 and 2013. The normalised graph comparison of the global phishing targets and the South African combined 6 phishing targets are shown in Figure 36. Figure 37 shows the global top 15 financial phishing targets trend, as compared to the negative correlation of the SA #5 and SA #6 trends. This is depicted in a normalised graph for comparative purposes.

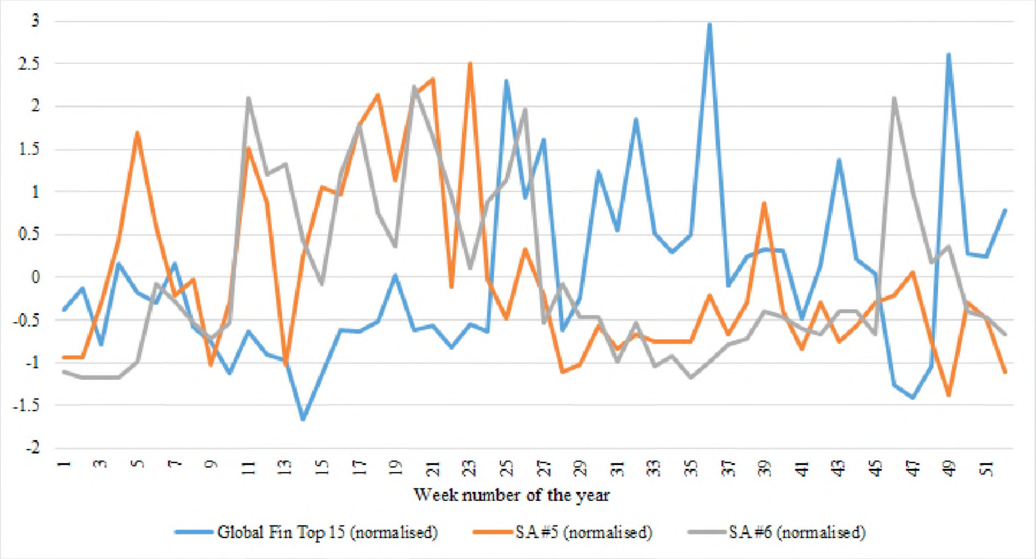


Figure 37: Phishing comparison per week: global top 15, SA #5 and SA #6 (normalised)

Themes 3 and 4 investigated phishing statistics by week. Themes 5, 6 and 7 investigate phishing statistics by day.

4.6 Theme 5 - Phishing statistics by day within a calendar year

The purpose of Theme 5, 6 and 7 is to understand the daily trend among the South African phishing targets as compared to the global trend. The daily trend would be investigated by 'day of the year' in Theme 5. Theme 6 and 7 will illustrate the 'day of the month' and 'day of the week' statistics respectively.

4.6.1 Theme 5 - Statistics discussion

Figure 38 illustrates the daily trend of phishing statistics for the global and global top fifteen phishing targets within a calendar year.

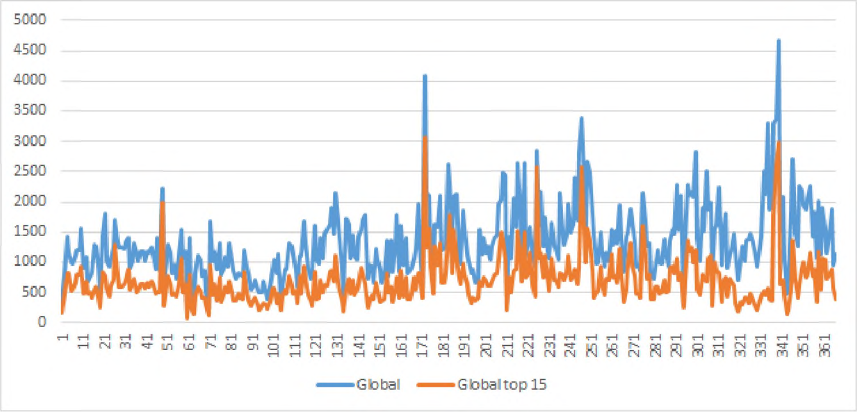


Figure 38: Phishing comparison per day of year: global

Figure 39 shows the daily trend statistics for the individual South African phishing targets within a calendar year.

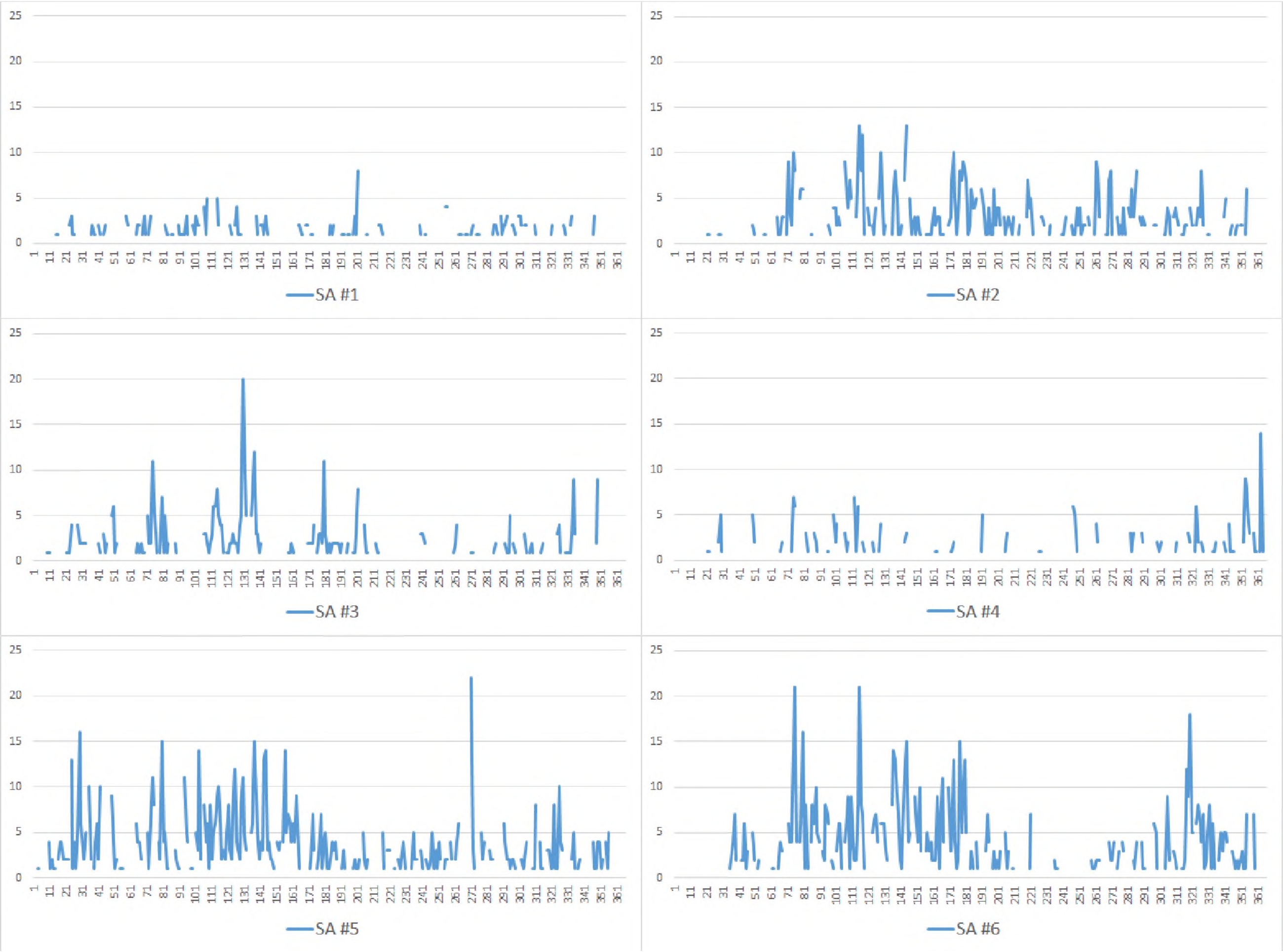


Figure 39: Phishing comparison per day of year: SA individual 6

The research now focuses on the hypothesis to determine if any statistical correlation exists between the South African trend and the global trend.

4.6.2 Theme 5 - Hypothesis

Determine if any correlation exists between the daily trend within a calendar year for global phishing targets, global top 15 financial phishing targets and the South African financial phishing targets.

In table 20, the correlation values have been calculated and are shown in comparing between the data sets. With a df value of 50, and $\alpha = 0.05$, the critical value range is determined to be: -0.103 < critical value < +0.103. The values **highlighted** in table 20 indicate relationships that fall outside of the critical range.

Table 20: Correlation coefficient values for comparison by day within a calendar year

	Global	**Global top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
Global	-	**0.850**	**0.113**	-0.018	**0.144**	0.067	-0.050	-0.032	-0.041
Global top 15	0.850	-	0.041	-0.061	**0.146**	-0.018	-0.011	-0.084	-0.035
SA combined 6	**0.113**	0.041	-	**0.262**	**0.664**	**0.623**	**0.400**	**0.687**	**0.761**
SA #1	-0.018	-0.061	**0.262**	-	0.139	**0.328**	0.102	0.090	0.054
SA #2	**0.144**	**0.146**	**0.664**	0.139	-	**0.249**	**0.198**	**0.175**	**0.454**
SA #3	0.067	-0.018	**0.623**	**0.328**	**0.249**	-	**0.246**	**0.326**	**0.336**
SA #4	-0.050	-0.011	**0.400**	0.102	**0.198**	**0.246**	-	**0.179**	**0.216**
SA #5	-0.032	-0.084	**0.687**	0.090	**0.175**	**0.326**	**0.179**	-	**0.382**
SA #6	-0.041	-0.035	**0.761**	0.054	**0.454**	**0.336**	**0.216**	**0.382**	-

The following statements can be made:

Since the correlation values of the above relationships show that $\rho > 0.103$ or $\rho <$ -0.103, ρ is significant.

Decision: Reject H_o: $\rho = 0$, for the values **highlighted** in table 20.

Conclusion: There is sufficient evidence to conclude that there is a linear relationship between the pairs of data **highlighted** in table 20 when compared daily within a calendar year, because the correlation is significantly different from 0. For the remaining relationships, the null hypothesis cannot be rejected, as ρ is not significant.

4.6.3 Theme 5 - Summary

From Theme 6, the South African combined six phishing targets and SA #2 showed a correlation with the global trend, as per table 20. Figure 40, 41 and 42 illustrate the daily statistics for the global, South African combined six and SA #2 trends in normalised graphs.

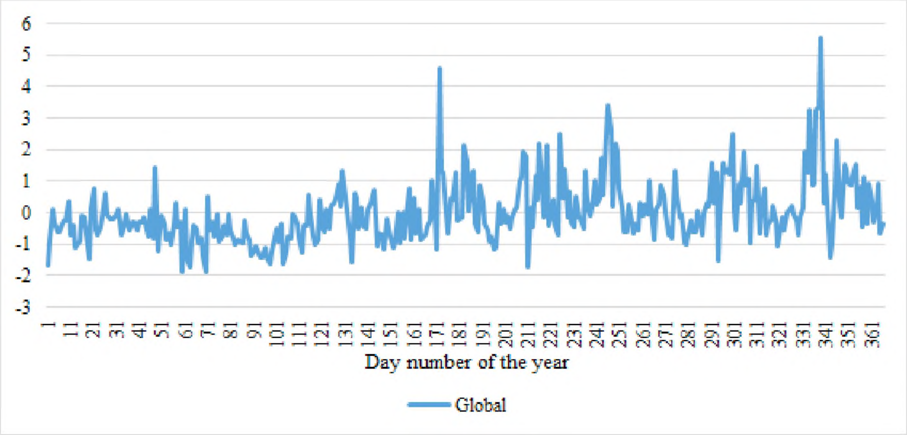


Figure 40: Daily trend of global phishing targets (normalised)

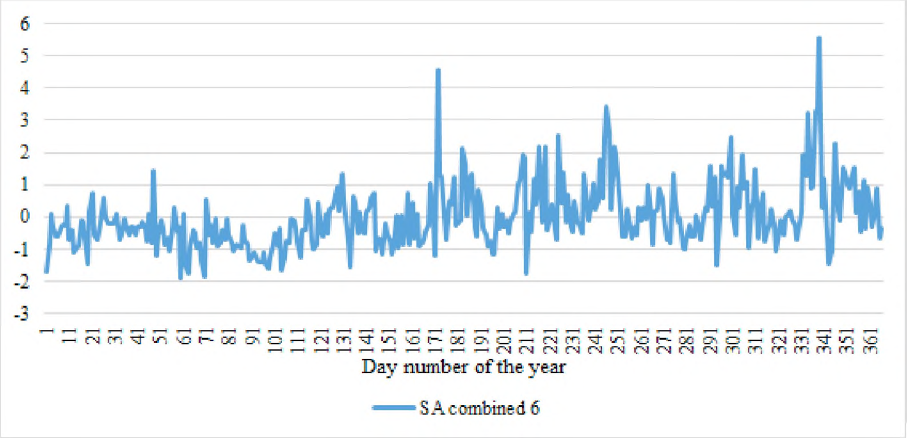


Figure 41: Daily trend of the South African combined 6 phishing targets (normalised)

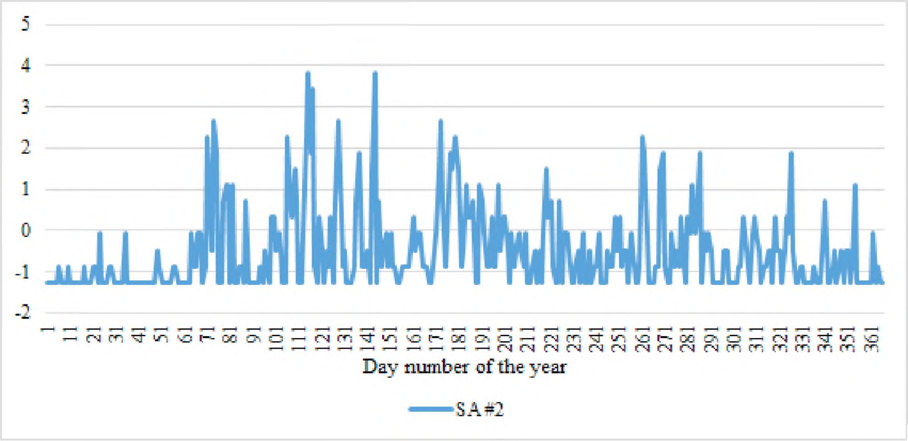


Figure 42: Daily trend of SA #2 (normalised)

The correlation values in table 20 also indicated that SA #2 had a correlation with the global top fifteen financial phishing targets. Figure 43 illustrates the normalised graph of the global top fifteen financial phishing targets.

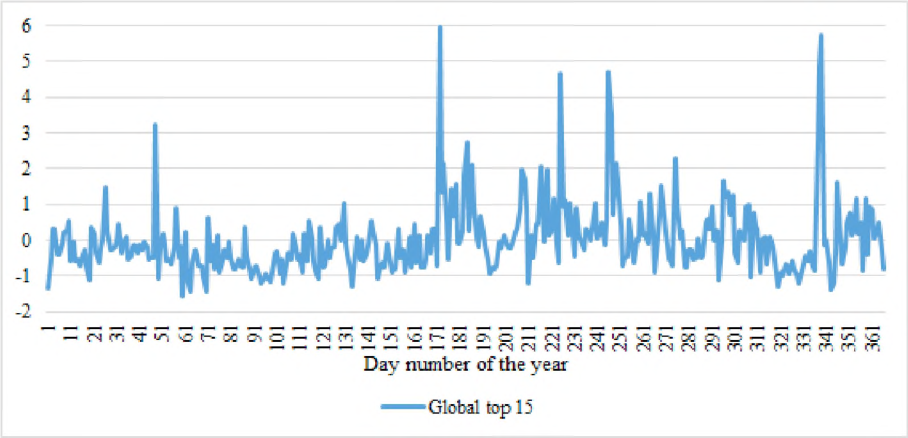


Figure 43: Daily trend for the global top 15 phishing targets (normalised)

Theme 5 investigated phishing statistics by day of the year. Theme 6 investigated phishing statistics by day within the month.

4.7 Theme 6 - Phishing statistics by day within the month

The purpose of Theme 6 is to understand the daily trend within the month among the six

South African financial phishing targets, as compared to the global phishing targets and the global top 15 financial phishing targets.

4.7.1 Theme 6 - Statistics discussion

Day of the month phishing statistics are the summed up number of phishing sites of each day of the month. Table 13, depicts the day of the month comparison of phishing sites. Figure 44 shows the day of the month comparison between the global phishing targets and the global top 15 financial phishing targets.

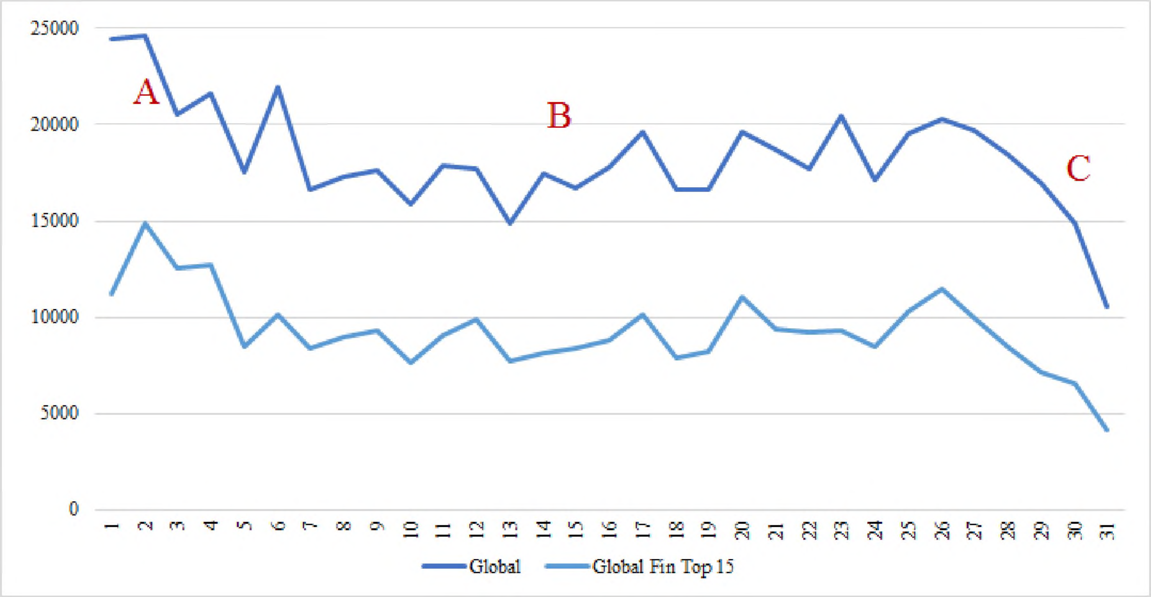


Figure 44: Phishing comparison by day of the month: global

Figure 45 shows the day of the month comparison for the six combined South African phishing targets. Point A shows that the beginning of the month has low phishing numbers for South African targets, whereas point B on the 15th, has the peak amount of phishing sites. Point C, shows a decrease towards the end of the month, but this could also be due to there being only 7 months in the year which have 31 days.

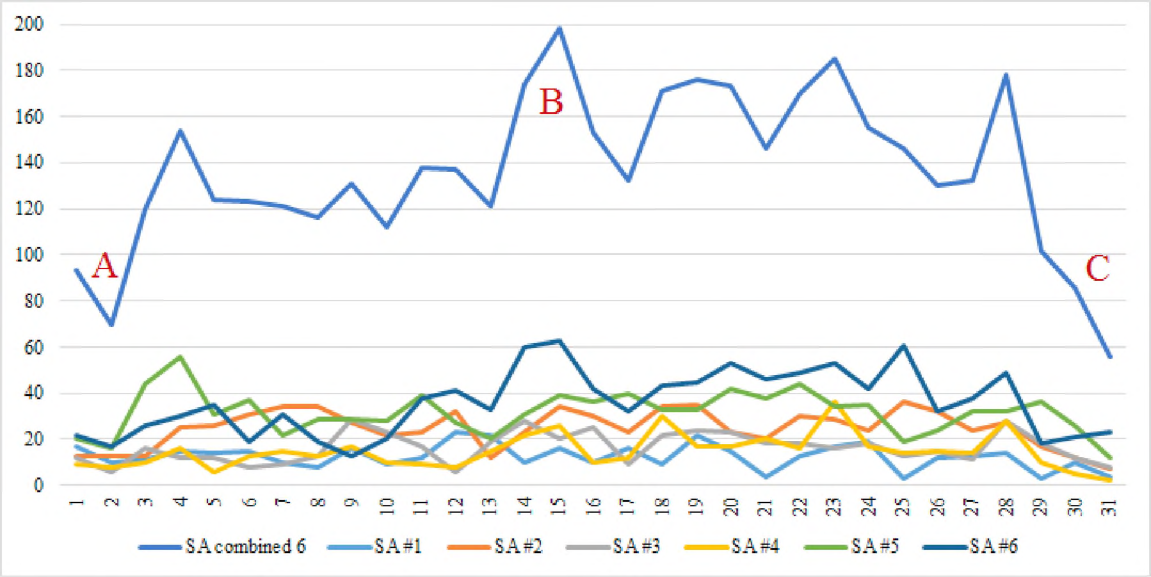


Figure 45: Phishing comparison by day of the month: South African phishing targets

4.7.2 Theme 6 - Hypothesis

Determine if any correlation exists between the daily trend (day of the month) for global firms vs global top 15 financial firms vs South Africa combined 6 financial firms vs each of the individual 6 South African financial firms.

In table 21, the correlation values have been calculated and are shown in comparing between the data sets. With a df value of 29, and $\alpha = 0.05$, the critical value range is determined to be: -0.355 < critical value < +0.355. The values highlighted in table 21 indicate relationships that fall outside of the critical range.

The following statements can be made:

Since the correlation values of the above relationships show that $\rho > 0.355$ or $\rho <$ -0.355, ρ is significant.

Decision: Reject H_o: $\rho = 0$, for the values highlighted in table 21.

Conclusion: There is sufficient evidence to conclude that there is a linear relationship between the pairs of data highlighted in table 21 when compared monthly over the sample size time period, because the correlation is significantly different from 0. For the remaining relationships, the null hypothesis cannot be rejected, as ρ is not significant.

Table 21: Correlation coefficient values for comparison by day of the month

	Global	**Global top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
Global	-	**0.901**	0.053	0.157	0.063	-0.233	0.106	0.215	-0.057
Global top 15	**0.901**	-	0.037	0.181	0.056	-0.251	0.010	0.245	-0.060
SA combined 6	0.053	0.037	-	0.343	**0.657**	**0.576**	**0.794**	**0.577**	**0.817**
SA #1	0.157	0.181	0.343	-	0.160	0.052	0.183	0.167	0.124
SA #2	0.063	0.056	**0.657**	0.160	-	0.176	**0.471**	0.256	**0.454**
SA #3	-0.233	-0.251	**0.576**	0.052	0.176	-	**0.500**	0.256	**0.357**
SA #4	0.106	0.010	**0.794**	0.183	**0.471**	**0.500**	-	0.298	**0.600**
SA #5	0.215	0.245	**0.577**	0.167	0.256	0.256	0.298	-	0.247
SA #6	-0.057	-0.060	**0.817**	0.124	**0.454**	**0.357**	**0.600**	0.247	-

4.7.3 Theme 6 - Summary

For Theme 6, there were no correlation coefficients significant enough for linear correlation to exist.

Theme 6 investigated the phishing statistics by day of the month. Theme 7 will investigate phishing statistics by day within the week.

4.8 Theme 7 - Phishing statistics by day of the week

The purposes of Theme 7 is to understand the day of week trend among the six South African financial phishing targets, and compare these trends to the global phishing targets and the global top fifteen financial phishing targets.

4.8.1 Theme 7 - Statistics discussion

Table 12, depicts the day of week of phishing sites over the sample size time period. Figure 46 shows the day of the week comparison among the global phishing targets and the global top fifteen financial phishing targets. When observing Figure 46, the global statistics indicate that Saturday and Sunday are the two lowest days of the week, in terms of phishing sites.

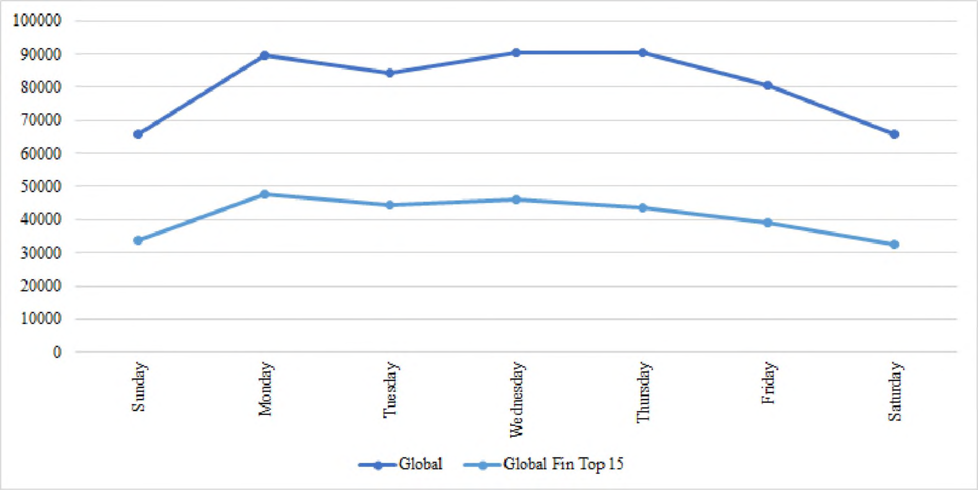


Figure 46: Phishing comparison by week day: South Africa vs global

Figure 47 shows the day of the week comparison for the six South African financial phishing targets. Five out of the six South African phishing targets have decreased numbers for Saturday and Sunday. The exception being SA #4, which has an increase on Saturday as compared to Friday phishing numbers.

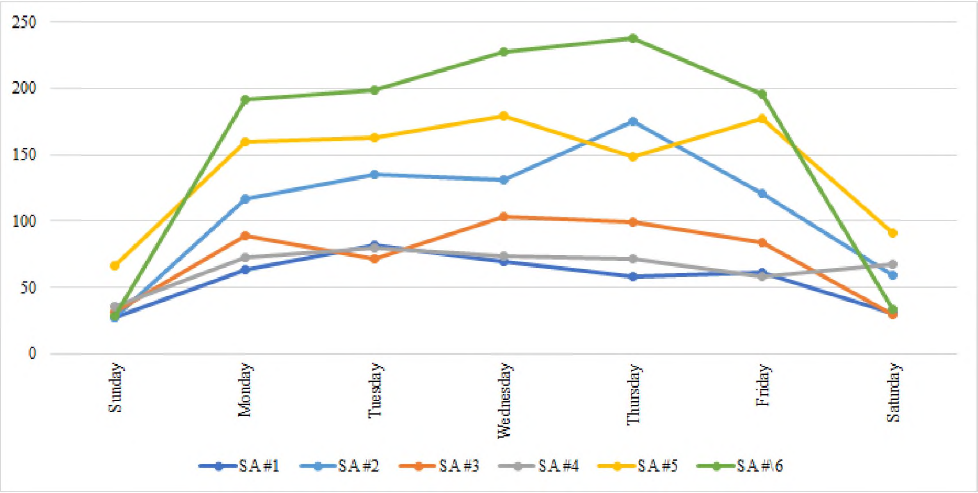


Figure 47: Phishing comparison by week day: South African firms

4.8.2 Theme 7 - Hypothesis

Determine if any correlation exists between the day of week for global phishing targets, global top 15 financial phishing targets and the South African financial phishing targets.

In table 22, the correlation values have been calculated and are shown in comparing between the data sets. With a df value of 26, and $\alpha = 0.05$, the critical value range is determined to be: -0.755 < critical value < +0.755. The values **highlighted** in table 22 indicate relationships that fall outside of the critical range.

Table 22: Correlation coefficient values for day of the week comparison

	Global	Global top 15	SA combined 6	SA #1	SA #2	SA #3	SA #4	SA #5	SA #6
Global	-	**0.959**	**0.957**	**0.842**	**0.908**	**0.969**	0.701	**0.868**	**0.965**
Global top 15	**0.959**	-	**0.879**	**0.856**	**0.785**	**0.882**	0.700	**0.815**	**0.880**
SA combined 6	**0.957**	**0.879**	-	**0.907**	**0.957**	**0.956**	0.742	**0.950**	**0.994**
SA #1	**0.842**	**0.856**	**0.907**	-	**0.809**	**0.791**	0.754	**0.911**	**0.884**
SA #2	**0.908**	**0.785**	**0.957**	**0.809**	-	**0.896**	0.747	**0.838**	**0.953**
SA #3	**0.969**	**0.882**	**0.956**	**0.791**	**0.896**	-	0.589	**0.899**	**0.973**
SA #4	0.701	0.700	0.742	0.755	0.747	0.589	-	0.688	0.673
SA #5	**0.868**	**0.815**	**0.950**	**0.911**	**0.838**	**0.899**	0.688	-	**0.931**
SA #6	**0.965**	**0.880**	**0.994**	**0.884**	**0.953**	**0.973**	0.673	**0.931**	-

The following statements can be made:
Since the correlation values of the above relationships show that $\rho > 0.755$ or $\rho <$ -0.755, ρ is significant.

Decision: Reject H_o: $\rho = 0$, for the values **highlighted** in table 22.

Conclusion: There is sufficient evidence to conclude that there is a linear relationship between the pairs of data **highlighted** in table 22 when compared by day of week over the sample size time period, because the correlation is significantly different from 0. For the remaining relationships, the null hypothesis cannot be rejected, as ρ is not significant.

4.8.3 Theme 7 - Summary

From Theme 7, five out of the six South African phishing targets have a linear correlation with the global trend, as indicated in table 22. With the exception SA #4, the South African phishing targets have decreasing phishing sites on the weekend. This same trend of decreased phishing numbers of the weekend exists for the global and global top 15 financial phishing targets as well

The next theme briefly investigated the geography of hosting country.

4.9 Theme 8 - Phishing statistics by hosting country

The purpose of Theme 8 is to understand the hosting countries for phishing sites. Table 14, shows the top 20 phishing hosting countries for the global firms and the global top 15 financial phishing targets. Table 15, shows the top 20 phishing hosting countries for the six individual South African financial phishing targets. The United States and Canada are 1st and 2nd respectively for all the South African financial firms. This is followed closely by Germany, France, Romania and Korea. When compared to the global sites and global top 15 financial firms, United States remains as top host, followed by Great Britain and then Canada, Germany and France.

Observing the top 20 hosting countries for the South African phishing targets, at least 85% of all phishing mails for that organisation come from the top 20 host countries for that organisation, as shown Figure 48. This same applies to the global phishing targets and global top 15 financial phishing targets.

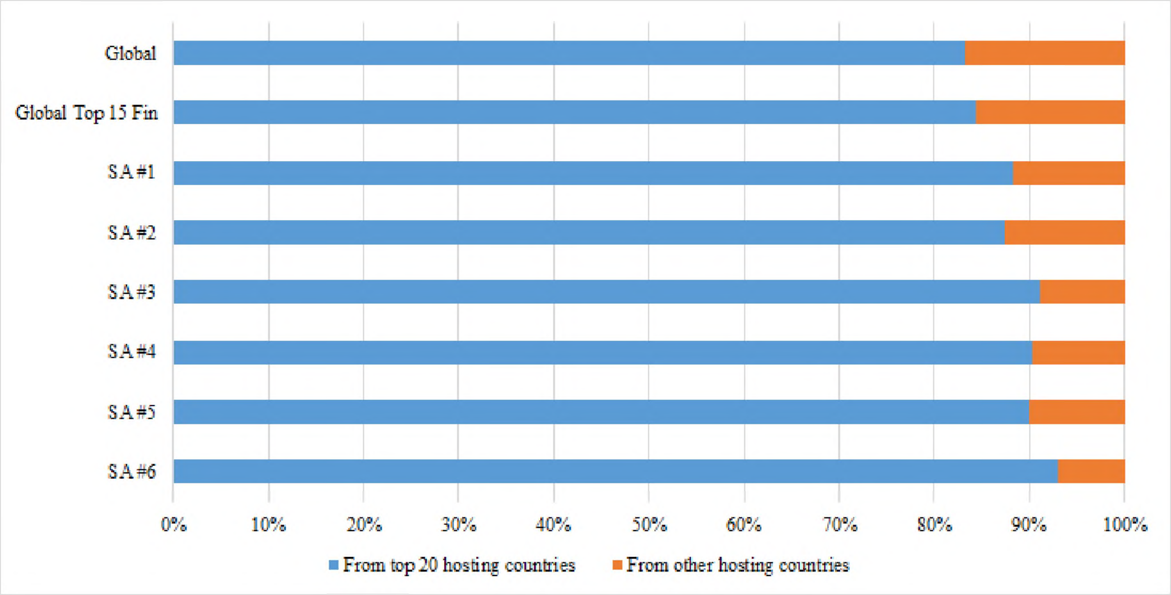


Figure 48: Phishing comparison: Distribution of hosting countries

4.10 Summary

In this chapter, the research focussed on the analysis of the date extracted from the data set. Various themes were investigated as possibilities for finding correlation. These themes included understanding phishing by month, week, day and geographical hosting country. The analysis was broken down into three sections, namely statistical analysis, hypothesis calculation and a summary of the findings from the correlation calculations.

In the next chapter, the research concludes with a summary of work, findings and suggestions for future research.

Chapter 5

Conclusion

This chapter provides the discussion of the results in terms of observations noticed and assesses the attainment of the research goals. Finally, possible future research areas are discussed.

5.1 Summary of Work

The purpose of this research thesis was to investigate a data set of phishing URLs to determine whether or not the phishing trend of South African financial targets correlates with the global trend in phishing attacks. This was achieved by analysing existing literature regarding trends and then performing an evaluation of the APWG phishing data set, where both South African phishing targets and global targets were investigated and statistically analysed to test for correlation.

The APWG phishing data set was in a SQL database format, for which MySQL was used to host the database. Microsoft Excel and R were used for graphing and statistical analysis of the data set. The data set analysed, involved eight themes of comparisons, namely: phishing trend by month, phishing trend by week, phishing trend by day of the month and week, and finally phishing by hosting country. It was noticed that there were a few outliers in the data set for the South African phishing targets investigated, which affected the correlation coefficient. The outliers are relevant, as the phishing statistics are real-life samples of phishing attacks. When investigating these outliers and which months they occurred in, one could understand and appreciate the realistic statistics. These outliers, or increase in phishing attacks, occurred in the same month where the South African financial firms were affected by a certain event. These events include a new rewards programme being launched, a launch of a new payments method and internet access, the retrenchment of staff , the launch of a new mobile application, a glitch in the system which caused major downtime and sale of a partner firm.

5.2 Findings

The research goals, as stated in Chapter 1, were to investigate the growth rates of phishing globally and compare the results to South African phishing targets to find if any correlation exists. This was achieved by analysing the data set and then using statistical analysis to determine correlation, which was deemed as significant, from a statistical point of view. Table 23 describes the findings from the statistics that indicated which South African phishing targets had a correlation with either the global statistics or the global top 15 financial firms.

In terms of Themes 1 and 2, the research first analysed the monthly statistics over the 28 month period which indicated that SA firm #4 had a positive relationship with the global trend of phishing attacks. The monthly statistics when looking at a 12 month period in Theme 2, indicated that SA #5 had a significant negative correlation with the global trend, while SA #3, SA #5 and SA #6 had a negative correlation with the global top 15 financial phishing targets.

In Themes 3 and 4, the weekly statistics were investigated. Over the 28 month period, it was found that SA #4 had a significant positive relationship with the global trend. Theme 4 investigated the weekly statistics within the calendar year, the correlation calculated that SA #5 had a negative linear relationship with the global trend, while SA #5 and SA #6 had a negative linear relationship with the global top 15 financial phishing targets.

Table 23: Summary of findings from the analysis

Theme Number	**Comparison by**	**Time Period**	**Correlation with Global trend**	**Correlation with Global top 15 financial phishing targets**
1	month	sample size time period	SA #4	-
2	month	month of the year	SA #5	SA #3, SA #5, SA #6
3	week	sample size time period	SA #4	-
4	week	week of the year	SA #5	SA #5, SA #6
5	day	day of the year	SA combined 6, SA #2	SA #2
6	day	day of the month	-	-
7	day	day of the week	SA combined 6, SA #1, SA #2, SA #3, SA #5, SA #6	SA combined 6, SA #1, SA #2, SA #3, SA #5, SA #6
8	hosting country	sample size time period	SA #1, SA #2, SA #3, SA #4, SA #5, SA #6	SA #1, SA #2, SA #3, SA #4, SA #5, SA #6

In Themes 5, 6 and 7, the daily statistics were analysed. Over the period of a year, it was found that the SA combined 6 and SA#2 had a positive linear relationship with the global trend, while SA #2 also had a positive linear relationship with the global top 15 financial phishing targets. For the day of the month, no significant relationship was found. For day of the week, the SA combined 6, SA #1, SA #2, SA #3, SA #5 and SA #6 was found to have a positive linear relationship with both the global trend and the global top 15 financial phishing targets.

It was observed that the South African phishing targets' URL hosting countries have a similar trend to the global statistics from the data set. In observing the trends and statistics, it was found that at least 80% of phishing sites are hosted within the top 20 hosting countries for the respective phishing targets.

5.3 Reliability and Validity of the Research

Reliability is defined as the reproducibility of the same results. It was noticed that there were a few outliers in the data set, which reduced the correlation coefficient significantly. However, this research is based on actual phishing statistics, and outliers will occur. The reliability is based on the accuracy of the phishing data set and users who contribute the knowledge base thereof.

The validity of the research represents the extent to which the research study actually measures as compared to what it claims to measure. There are two forms of validity, namely internal validity which looks at the measurements of data obtained from the analysis, and external validity which describes the generalised findings. The internal validity appears to be high, as the sample of data used is real phishing statistics of targets around the world. The external validity is also high, as the observations are based on the correlation coefficient and significant values.

5.4 Future Research

There are multiple areas where future research can be conducted.

A study which investigates the impact of laws, when they were implemented vs increase/decrease in phishing trend can be performed to understand the impact of enforcing laws against phishing. This can be researched both globally and within South Africa.

A comparison of phishing statistics can be compared to internet usage growth statistics to find correlation. As more developing nations' users start using the internet, they will not be too familiar with modern day advanced phishing. It would be interesting to see if phishers target these developing nations where internet penetration has been relatively low.

Phishing is continuously evolving, and moving further away from traditional email-based attacks and more into the social media and mobile landscapes. The campaigns include similar types of lures that are evident in phishing and spam email. A scam could be advertised as a new application to try out, or offer to download of a song from a user's favourite music artist. If a user clicks on it, the scam often asks the user to enter his or her social media login details. A phishing study can be conducted to understand the increase in mobile adoption and social media against phishing on these platforms.

References

Aaron, G., & Rasmussen, R. (2013a). Global Phishing Survey: Trends and Domain Name Use in 1H2013. *Anti-Phishing Working Group*. Retrieved from `http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_1H2013.pdf` (Last accessed: 2015-05-30)

Aaron, G., & Rasmussen, R. (2013b). Global Phishing Survey: Trends and Domain Name Use in 2H2012. *Anti-Phishing Working Group*. Retrieved from `http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_2H2012.pdf` (Last accessed: 2015-05-30)

Aaron, G., & Rasmussen, R. (2014). Global Phishing Survey: Trends and Domain Name Use in 2H2013. *Anti-Phishing Working Group*. Retrieved from `http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_2H2013.pdf` (Last accessed: 2015-05-30)

Almomani, A., Gupta, B. B., Atawneh, S., Meulenberg, A., & Almomani, E. (2013). A survey of phishing email filtering techniques. *IEEE Communications Surveys and Tutorials*, *15*(4), 2070-2090.

Anderson, R. (2007). Closing the phishing hole - fraud, risk and nonbanks. In *Federal Reserve Bank of Kansas City, Conference on Nonbanks in the Payments System* (p. 1-16).

Bose, I., & Leung, A. C. M. (2014). Do phishing alerts impact global corporations? A firm value analysis. *Decision Support Systems*, *64*(8), 67-78.

Butler, R. (2005). Investigation of phishing to develop guidelines to protect the Internet consumer's identity against attacks by phishers. *South African Journal of Information Management*, *7*(9), 1-15.

Butler, R. (2007). A framework of anti-phishing measures aimed at protecting the online consumer's identity. *The Electronic Library*, *25*(5), 517-533.

Chaudhary, G. K. (2014). Development Review on Phishing: A Computer Security Threat. *International Journal of Advance Research in Computer Science and Management Studies*, *2*(8), 55-64.

Chhabra, S. (2005). *Fighting spam, Phishing and Email Fraud* (Doctoral dissertation). University of California Riverside.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (4th Edition).* Thousand Oaks, California, CA, USA: Sage publications.

Dagada, R. (2011). Digital Banking Security, Risk and Credibility concerns in South Africa. *The Second International Conference on Cyber Security, Cyber Peacefare and Digital Forensic (CyberSec2013)*, 1-17. Retrieved from `http://rabelanidagada.co.za/wp-content/uploads/2014/01/134.pdf` (Last accessed: 2015-05-30)

Dlamini, M., Eloff, J., & Eloff, M. (2009). Information security: The moving target. *Computers and Security*, *28*(3-4), 189-198.

Fister, S. E. (2009). *Consumers' shopping value and their responses to visual merchandise displays in an in-store retail setting* (Masters dissertation). Oregon State University.

Frauenstein, E. D., & Von Solms, R. (2013). Using Theories and Best Practices to Bridge the Phishing Gap. In *European information security multi-conference* (p. 69-78).

Goldstuck, A., & Dagada, R. (2009). Information Security for South Africa. In *ISSA 2009 Conference* (p. 117-135). University of Johannesburg, johannesburg, South Africa. doi: 10.1145/2030376.2030396

Grobler, M. (2010). Strategic Information Security: Facing the Cyber Impact. *Council for Scientific and Industrial Research*, 12-21. Retrieved from `http://researchspace.csir.co.za/dspace/handle/10204/4507` (Last accessed: 2015-05-30)

Grobler, M., & Dlamini, Z. (2012). Global Cyber Trends a South African Reality. In *Ist-africa* (p. 1-8).

Gudkova, D., & Demidova, N. (2014). *Spam and phishing in the Q2 2014* (Tech. Rep.). Retrieved from `https://securelist.com/analysis/quarterly-spam-reports/65755/spam-and-phishing-in-q2-2014/` (Last accessed: 2015-09-26)

Gupta, N. (2014). Analysis of Issues in Phishing Attacks and Development of Prevention

Mechanism. *Journal of Global Research in Computer Science*, *5*(6), 22–25.

Halevi, T., Memon, N., & Nov, O. (2015). Spear-Phishing in the Wild: A Real-World Study of Personality, Phishing Self-Efficacy and Vulnerability to Spear-Phishing Attacks. *Social Science Research Network*. doi: 10.2139/ssrn.2544742

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, *30*(2), 87–93.

Hsu, C. H., Wang, P., & Pu, S. (2011). Identify Fixed-path Phishing Attack by STC. In *Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference* (pp. 172–175). New York, NY, USA: ACM. doi: 10.1145/2030376.2030396

Illowsky, B. (2015). *Testing the Significance of the Correlation Coefficient.* Retrieved from `http://cnx.org/contents/1fc300c9-4d54-4117-94ee-f9ff260a8a56@2.49:36/Collaborative-Statistics-for-M` (Last accessed: 2015-10-04)

Kaspersky Lab. (2013). *Financial cyber threats in 2014* (Tech. Rep.). Retrieved from `https://securelist.com/files/2015/02/KSN_Financial_Threats_Report_2014_eng.pdf` (Last accessed: 2015-09-26)

Kaspersky Lab. (2014). *Financial cyber threats in 2013* (Tech. Rep.). Retrieved from `http://media.kaspersky.com/en/Kaspersky-Lab-KSN-report-Financial-cyber-threats-in-2013-eng-final.pdf` (Last accessed: 2015-09-26)

Klensin, J. (2008). *Simple Mail Transfer Protocol* (RFC No. 5321). RFC Editor. Internet Requests for Comments. Retrieved from `http://www.rfc-editor.org/rfc/rfc5321.txt` (Last accessed: 2015-10-04)

Kopsovich, R. D. (2001). *A study of correlations between learning styles of students and their mathematics scores on the texas assessment of academic skills test* (Doctoral dissertation). University of North Texas.

Longe, O., Ngwa, O., Wada, F., Mbarika, V., & Kvasny, L. (2009). Criminal Uses of Information and Communication Technologies in Sub-Saharan Africa: Trends, Concerns and Perspectives . *Journal of Information Technology Impact*, *9*(3), 155-172.

Marforio, C., Jayaram Masti, R., Soriente, C., Kostiainen, K., & Capkun, S. (2015). Personalized Security Indicators to Detect Application Phishing Attacks in Mobile Platforms. *ArXiv e-prints*. Retrieved from `http://adsabs.harvard.edu/abs/2015arXiv150206824M` (Last accessed: 2015-10-04)

Megaw, G., & Flowerday, S. V. (2010, August). Phishing within e-commerce: A trust and confidence game. In *Information Security for South Africa (ISSA)* (pp. 1–8). Sandton, Johannesburg, South Africa.

Milletary, J. (2005). Technical trends in phishing attacks. *CERT Coordination Center*, *1*, 1-17. Retrieved from `https://www.us-cert.gov/sites/default/files/publications/phishing_trends0511.pdf` (Last accessed: 2015-05-30)

Nenty, H. J. (2009). Writing a quantitative research thesis. *International Journal of Educational Sciences*, *1*(1), 19-32.

Perez, T. (2012). *Google Safe Browsing Program 5 Years Old: Been Blacklisted Lately?* Retrieved from `http://blog.sucuri.net/2012/06/google-safe-browsing-program-5-years-old-been-blacklisted-lately.html` (Last accessed: 2015-09-27)

Pfleeger, C. P., & Pfleeger, S. L. (2006). *Security in Computing (4th Edition).* Upper Saddle River, NJ, USA: Prentice Hall PTR.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rajalingam, M., Alomari, S. A., & Sumari, P. (2012). Prevention of Phishing Attacks Based on Discriminative Key Point Features of WebPages. *International Journal of Computer Science and Security*, *6*(1), 1-18.

Ramzan, Z., & Wuest, C. (2007, August). Phishing Attacks: Analyzing Trends in 2006. In *Conference on Email and Anti-Spam (CEAS).* Microsoft Research Silicon Valley, California, USA.

Rasmussen, R., & Aaron, G. (2012). Global Phishing Survey: Trends and Domain Name Use in 1H2012. *Anti-Phishing Working Group*. Retrieved from `http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_1H2013.pdf` (Last accessed: 2015-05-30)

Rasmussen, R., & Aaron, G. (2014). Global Phishing Survey: Trends and Domain Name Use in 1H2014. *Anti-Phishing Working Group*. Retrieved from http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_1H2014.pdf (Last accessed: 2015-05-30)

Resnick, P. W. (2008). *Internet Message Format* (RFC No. 5322). RFC Editor. Internet Requests for Comments. Retrieved from http://www.rfc-editor.org/rfc/rfc5322.txt (Last accessed: 2015-10-04)

RSA. (2013). *Phishing Kits: the Same Wolf, Just a Different Sheep's Clothing* (Tech. Rep.). Retrieved from http://www.emc.com/collateral/fraud-report/rsa-online-fraud-report-012013.pdf (Last accessed: 2015-05-30)

RSA. (2014a). *2013 A year in review* (Tech. Rep.). Retrieved from http://uk.emc.com/collateral/fraud-report/rsa-online-fraud-report-012014.pdf (Last accessed: 2015-05-30)

RSA. (2014b). *The Current State of Cybercrim 2014: An Inside Look at the Changing Threat Landscape* (Tech. Rep.). Retrieved from http://www.emc.com/collateral/white-paper/rsa-cyber-crime-report-0414.pdf (Last accessed: 2015-10-03)

RSA. (2015). *Cybercrime 2015: An Inside Look at the Changing Threat Landscape* (Tech. Rep.). Retrieved from http://www.emc.com/collateral/white-paper/rsa-white-paper-cybercrime-trends-2015.pdf (Last accessed: 2015-10-03)

Saad, Z. S., Glen, D. R., Chen, G., Beauchamp, M. S., Desai, R., & Cox, R. W. (2009). A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *NeuroImage*, *44*(3), 839 - 848. doi: 10.1016/j.neuroimage.2008.09.037

Salar, K. (1989). *Sample Size for Correlation Estimates* (Tech. Rep.). Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA219810 (Last accessed: 2015-10-11)

Shcherbakova, T., Vergelis, M., & Demidova, N. (2014). *Spam and phishing in the Q3 of 2014* (Tech. Rep.). Retrieved from https://securelist.com/analysis/quarterly-spam-reports/67851/spam-and-phishing-in-the-q3-of-2014/ (Last accessed: 2015-09-26)

Shcherbakova, T., Vergelis, M., & Demidova, N. (2015a). *Spam and phishing in Q2 2015*

(Tech. Rep.). Retrieved from https://securelist.com/analysis/quarterly-spam-reports/71759/spam-and-phishing-in-q2-of-2015/ (Last accessed: 2015-09-26)

Shcherbakova, T., Vergelis, M., & Demidova, N. (2015b). *Spam and Phishing in the First Quarter of 2015* (Tech. Rep.). Retrieved from https://securelist.com/analysis/quarterly-spam-reports/69932/spam-and-phishing-in-the-first-quarter-of-2015/ (Last accessed: 2015-09-26)

Shumway, R. H., & Stoffer, D. S. (2013). *Time Series Analysis and its Applications - with R examples.* Secaucus, NJ, USA: Springer Science & Business Media.

Singh, A. C., Somas, K. P., & Tambre, K. G. (2013). Phishing: A Computer Security Threat. *International Journal of Advance Research in Computer Science and Management Studies*, 64-71.

Symantec Corporation. (2013). *Internet Security Threat Report* (Vol. 18; Tech. Rep.). Retrieved from http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v18_2012_21291018.en-us.pdf (Last accessed: 2015-09-28)

Symantec Corporation. (2014). *Internet Security Threat Report* (Vol. 19; Tech. Rep.). Retrieved from https://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v19_21291018.en-us.pdf (Last accessed: 2015-09-26)

Symantec Corporation. (2015). *Internet Security Threat Report* (Vol. 20; Tech. Rep.). Retrieved from https://www4.symantec.com/mktginfo/whitepaper/ISTR/21347932_GA-internet-security-threat-report-volume-20-2015-social_v2.pdf (Last accessed: 2015-09-26)

Trend Micro Incorporated. (2012). *Spear-Phishing Email: Most Favored APT Attack Bait* (Tech. Rep.). Retrieved from http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-spear-phishing-email-most-favored-apt-attack-bait.pdf (Last accessed: 2015-09-28)

van der Merwe, A., & Seker, R. (2004). Mobile Phishing. *South Africa Computer Jounal*, *33*, 111-112.

van der Merwe, A., Seker, R., & Gerber, A. (2005). Phishing in the system of systems settings: mobile technology. In *Systems, man and cybernetics, 2005 ieee international conference* (Vol. 1, pp. 492–498).

Von Mises, R. (2014). *Mathematical theory of probability and statistics.* Berkeley, London, United Kingdom: Academic Press.

A Phishing statistics from the South African APWG data set

Table 24: Phishing statistics per week for 2012

2012 Week #	Global	Global Fin Top 15	SA combined 6	SA #1	SA #2	SA #3	SA #4	SA #5	SA #6
1	2252	1255	6	0	0	0	0	6	0
2	1926	911	2	0	0	1	0	1	0
3	2973	1386	7	1	0	0	0	6	0
4	2453	1217	6	4	1	0	0	1	0
5	3020	1684	14	4	0	4	0	6	0
6	3463	2174	12	7	0	4	0	1	0
7	2249	1226	18	1	3	12	1	1	0
8	1822	1133	8	3	0	4	0	1	0
9	2057	974	11	5	0	0	0	4	2
10	2444	1438	17	7	0	4	0	1	5
11	2152	1241	49	5	0	10	2	16	16
12	1494	555	25	3	0	5	0	10	7
13	2231	1209	33	6	0	0	0	0	27
14	2391	1450	49	5	0	4	0	19	21
15	2057	1303	29	6	0	0	0	13	10
16	2977	1664	42	3	0	6	0	12	21
17	3209	2054	41	2	0	5	0	11	23
18	2423	1455	77	5	0	11	0	33	28
19	2971	1595	92	1	0	46	0	27	18
20	3370	1978	101	0	0	19	0	35	47
21	2949	1601	69	5	0	1	0	24	39
22	3084	1890	49	3	0	0	1	12	33
23	3590	1952	68	3	0	5	0	42	18
24	4525	2984	47	6	0	1	0	10	30
25	8396	6856	51	1	0	9	2	8	31
26	5815	4343	72	2	0	18	0	12	40
27	6313	5060	24	2	0	11	0	6	5
28	3628	2231	23	4	0	2	0	4	13
29	4864	3012	27	12	0	12	0	1	2
30	5814	3887	19	1	0	6	0	3	9
31	5341	3309	14	0	0	6	0	5	3
32	6921	5412	20	2	1	2	0	8	7
33	4593	2730	13	1	2	0	0	10	0
34	3865	2177	19	3	0	7	1	7	1
35	4453	2192	14	2	1	1	1	9	0
36	5447	2967	17	7	0	0	0	8	2
37	5023	3240	14	4	0	0	1	4	5
38	4151	2466	18	2	4	1	0	11	0
39	5002	3194	20	3	3	2	0	5	7
40	4697	2566	22	2	0	2	0	11	7
41	4510	2428	17	4	0	6	0	4	3
42	4006	1909	23	6	1	2	3	5	6
43	5174	2479	22	6	0	3	2	2	9
44	3897	2009	15	5	0	5	0	5	0
45	4854	2475	17	4	0	2	0	10	1
46	2788	986	26	1	0	3	1	7	14
47	2585	484	20	1	0	8	1	8	2
48	6404	829	30	1	0	17	4	7	1
49	2934	552	22	0	0	0	6	3	13
50	6000	2287	35	1	0	10	3	15	6
51	4755	2824	13	1	0	0	2	4	6
52	4105	2489	8	2	0	1	1	4	0

Table 25: Phishing statistics per week for 2013

2013 Week #	**Global**	**Global Fin Top 15**	**SA combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
1	5403	3028	4	0	1	0	0	2	1
2	5713	3800	13	1	1	4	0	7	0
3	4236	2205	11	1	1	0	0	9	0
4	6654	3979	44	3	4	8	7	22	0
5	5469	2933	56	0	2	10	10	31	3
6	4807	2250	53	0	3	4	5	24	17
7	6251	3965	32	2	0	1	0	15	14
8	5050	2815	40	1	0	5	7	17	10
9	4761	2680	16	2	2	2	2	3	5
10	3078	1585	42	3	12	5	3	14	5
11	5382	2598	125	2	33	20	16	19	35
12	5203	2848	96	2	23	12	11	18	30
13	3550	2081	39	0	7	4	9	7	12
14	1490	665	14	1	4	1	2	2	4
15	3434	1682	58	4	15	4	11	17	7
16	4062	2223	97	7	30	14	13	17	16
17	4382	1792	135	6	46	23	10	27	23
18	6629	2591	36	5	13	3	4	9	2
19	7989	3374	46	2	25	2	7	4	6
20	5298	1917	52	4	21	12	2	7	6
21	5924	2373	66	5	28	2	6	20	5
22	3622	1660	17	1	11	0	0	5	0
23	4934	2032	19	1	10	1	1	4	2
24	2794	879	24	0	11	1	2	8	2
25	5072	1993	50	5	28	3	4	5	5
26	4581	2179	73	2	43	6	3	10	9
27	5953	2604	47	4	21	3	4	10	5
28	4260	1657	37	1	21	3	6	2	4
29	3460	1488	36	1	19	1	0	6	9
30	6997	3139	32	1	15	1	4	9	2
31	5681	2545	18	8	5	1	0	4	0
32	5158	2648	27	2	19	0	0	3	3
33	5126	3065	18	2	11	0	3	0	2
34	6324	3239	24	6	8	0	4	3	3
35	8957	3576	13	1	4	3	4	1	0
36	10983	6997	42	2	15	4	12	8	1
37	2929	1536	29	5	9	5	2	7	1
38	5375	2887	43	1	18	7	6	4	7
39	4810	2301	49	4	14	2	1	23	5
40	3990	2885	22	1	12	1	1	3	4
41	3213	1686	49	2	28	1	7	5	6
42	6404	3274	32	5	9	2	4	10	2
43	9090	4772	30	5	4	6	4	8	3
44	7791	3290	36	3	10	2	2	7	12
45	5830	2537	35	0	14	2	7	5	7
46	4828	1806	69	5	12	0	6	9	37
47	6146	2052	78	5	19	2	9	11	32
48	8552	2331	33	5	3	0	2	3	20
49	13090	8806	25	2	8	1	3	0	11
50	7244	3109	25	5	8	2	4	0	6
51	7085	2530	56	4	9	3	26	9	5
52	7072	3781	39	6	4	1	18	2	8

Table 26: Phishing statistics per week (combined) for 2012 and 2013

Week #	**Global**	**Global Fin Top 15**	**SA Combined 6**	**SA #1**	**SA #2**	**SA #3**	**SA #4**	**SA #5**	**SA #6**
1	7655	4283	10	0	1	0	0	8	1
2	7639	4711	15	1	1	5	0	8	0
3	7209	3591	18	2	1	0	0	15	0
4	9107	5196	50	7	5	8	7	23	0
5	8489	4617	70	4	2	14	10	37	3
6	8270	4424	65	7	3	8	5	25	17
7	8500	5191	50	3	3	13	1	16	14
8	6872	3948	48	4	0	9	7	18	10
9	6818	3654	27	7	2	2	2	7	7
10	5522	3023	59	10	12	9	3	15	10
11	7534	3839	174	7	33	30	18	35	51
12	6697	3403	121	5	23	17	11	28	37
13	5781	3290	72	6	7	4	9	7	39
14	3881	2115	63	6	4	5	2	21	25
15	5491	2985	87	10	15	4	11	30	17
16	7039	3887	139	10	30	20	13	29	37
17	7591	3846	176	8	46	28	10	38	46
18	9052	4046	113	10	13	14	4	42	30
19	10960	4969	138	3	25	48	7	31	24
20	8668	3895	153	4	21	31	2	42	53
21	8873	3974	135	10	28	3	6	44	44
22	6706	3550	66	4	11	0	1	17	33
23	8524	3984	87	4	10	6	1	46	20
24	7319	3863	71	6	11	2	2	18	32
25	13468	8849	101	6	28	12	6	13	36
26	10396	6522	145	4	43	24	3	22	49
27	12266	7664	71	6	21	14	4	16	10
28	7888	3888	60	5	21	5	6	6	17
29	8324	4500	63	13	19	13	0	7	11
30	12811	7026	51	2	15	7	4	12	11
31	11022	5854	32	8	5	7	0	9	3
32	12079	8060	47	4	20	2	0	11	10
33	9719	5795	31	3	13	0	3	10	2
34	10189	5416	43	9	8	7	5	10	4
35	13410	5768	27	3	5	4	5	10	0
36	16430	9964	59	9	15	4	12	16	3
37	7952	4776	43	9	9	5	3	11	6
38	9526	5353	61	3	22	8	6	15	7
39	9812	5495	69	7	17	4	1	28	12
40	8687	5451	44	3	12	3	1	14	11
41	7723	4114	66	6	28	7	7	9	9
42	10410	5183	55	11	10	4	7	15	8
43	14264	7251	52	11	4	9	6	10	12
44	11688	5299	51	8	10	7	2	12	12
45	10684	5012	52	4	14	4	7	15	8
46	7616	2792	95	6	12	3	7	16	51
47	8731	2536	98	6	19	10	10	19	34
48	14956	3160	63	6	3	17	6	10	21
49	16024	9358	47	2	8	1	9	3	24
50	13244	5396	60	6	8	12	7	15	12
51	11840	5354	69	5	9	3	28	13	11
52	11177	6270	47	8	4	2	19	6	8