INVESTIGATING IDENTITY CONCEALING AND EMAIL TRACING TECHNIQUES

[1]Ickin Vural, [2]HS Venter

Information and Computer Security Architectures Research Group (ICSA)

Department of Computer Science, University of Pretoria

[1]ickin@tuks.co.za

[2]hventer@cs.up.ac.za

ABSTRACT

At present it is very difficult to trace the identity of spammers who use identity concealment techniques. It is difficult to determine the identity of the spammer by just analysing the electronic trail.

This paper will look at standard email tracing techniques and how email senders try and hide their electronic trail. The identity concealing techniques that that are discussed are: Spoofing, Bot-Networks, Open proxies, Open mail relays and untraceable Internet connections. The techniques used to trace spam that we discuss are: Header analysis and honeypot computers.

The paper will also Investigate advanced digital forensics techniques for email tracing namely Investigating residual data on servers and investigating network devices.

KEY WORDS

Digital Forensics, Electronic tracing, identity concealment techniques, Spoofing, Bot-Networks, Open proxies, Open mail relays, untraceable Internet connections, Header analysis, honeypot computers.

INVESTIGATING IDENTITY CONCEALING AND EMAIL TRACING TECHNIQUES

1 INTRODUCTION

Unsolicited bulk communication also known as spam is the practise of sending unwanted e-mail messages, frequently with commercial content, in large quantities to an indiscriminate set of recipients *(*Spamhaus, 2009).

The sending of unsolicited bulk communications with the intention to advertise products and generate sales is economically viable because senders have no operating costs beyond the management of their mailing lists. Because the cost of setting up a spamming operation is low spammers are numerous. Thus the volume of unsolicited bulk communications has increased dramatically over the past few years *(Messaging Anti-Abuse working group, 2007).*

The costs of spam, involve lost productivity and fraud, these costs are borne by the general public, institutions that store and retrieve mail for their employees and by Internet service providers. Institutions and Internet service providers have been forced to add extra capacity to cope with the high volumes of unsolicited bulk communications *(Europa Press Releases, 2009)*.

Anti spamming legislation has been introduced in many jurisdictions. The problem faced by law enforcement is that spammers move their operations to jurisdictions that have no or weak anti spamming laws.

At present it is very difficult to trace the identity of spammers who use identity concealment techniques. It is difficult to determine the identity of the spammer by just analysing the electronic trail using standard email tracing techniques.

This paper focuses on current email tracing techniques and how email senders try and hide their electronic trail. The objective is to present the current state of tracing the origin of unsolicited bulk communications and then suggest techniques utilising digital forensics in an attempt to trace spam.

The remainder of the paper is structured as follows. The background section defines spam in more detail and also defines its cost and causes. . The next two sections are devoted to the state of the art of spamming techniques and how to trace spammers. More specifically, these two sections contrast each other in the sense that the third section looks at techniques that spammers use to conceal their identities, whereas the fourth section looks techniques for tracing the origins of spam so that the spammers can be identified. The paper's main contribution is purported in the next section, which discusses advanced digital forensics techniques for email tracing.

2 BACKGROUND

Unsolicited bulk email otherwise known as spam is an email sent to a large number of email addresses, where the owners of those addresses have not asked for or consented to receive the mail *(Internet Service Providers' Association, 2008).* Spam is used to advertise a service or a product. An example of spam is an unsolicited email message from an unknown or forged address advertising Viagra.

Spam is one of the most significant threats to the Internet, accounting for around 60% of all email traffic (Internet Service Providers' Association, 2008). Spam costs consumers and ISPs lots of money in bandwidth charges. Despite the growing number of technological means for combating spam, the spammers somehow manage to stay one step ahead and the deluge shows little sign of abating. .

Spammers generally do not pay much for the sending of spam. They accomplish this by exploiting open mail servers to do their task for them. The spammer need only send one email message to an incorrectly configured mail server to reach thousands of email addresses, with the bulk of the transfer being handled by the mis-configured mail server. Recipients in turn need to pay access costs or telephone costs in order to receive content they didn't ask for.

ISPs have to bear the bulk of the cost for bandwidth overuse by spammers, this cost is often passed onto the consumer through increased Internet access fees or a degraded service level.

With the introduction of the "Electronic Communications and Transactions Act, 2002" unsolicited emails now have a legal definition and the sending of spam is illegal *(Acts Online, 2002)*. Spammers if identified are liable for a fine and prosecution. Thus spammers will attempt to cover their trail to prevent identification.

Spammers are able to send email and cover their trail because Emails use Standard Mail Transfer Protocol (SMTP) which is not a secure protocol and can be tampered with *(Tzerefos, P. Smythe, C. Stergiou, I. Cvetkovic, S. 1997)*.

Emails consist of two main parts a message header and a message body. The message header contains information about the destination network address and the source network address as well as routing information. The email headers are not secure and can be easily forged to add false routing data and to hide the source network address. This paper will discuss both email concealment and email tracking techniques respectively in the following two sections.

The following section will describe in detail techniques used by spammers to conceal their identities from persons who would attempt to identify the source of spam mail.

3 HOW SPAMMERS CONCEAL THEIR IDENTITIES

Spammers conceal their identities for a number of reasons. If they are based in a jurisdiction which has strict anti-spamming laws they do not want to be traced for fear of prosecution. If they are based in a jurisdiction which has weak anti-spamming laws then the primary motive is not to be traced and blacklisted. As many ISP's will block any mail from blacklisted sites. The techniques studied are Spoofing, Bot-Networks, Open proxies and untraceable Internet connections.

3.1 Spoofing

Spoofing is the process whereby a spammer would insert fictitious headers into the email address to hide the network address of their computer. The

spammer will usually insert fake “From” and “Reply-To” headers into the email, these headers would point to a non-existent network address or more commonly an innocent third parties’ network address *(Boneh, Dan, 2004)*.

3.2 Bot- Networks

A Bot-Network consists of a set of machines that have been taken over by a spammer using Bot software sent over the Internet. This Bot software hides itself on its host machine and periodically checks for instructions from its human Bot-Network administrator. Botnets today are often controlled using Internet Relay Chat *(Evan Cooke, Farnam Jahanian, and Danny McPherson. 2005)*. The owner of the computer usually has no idea that his machine has been compromised until its Internet connection is shut down by an ISP. As most ISP’s block bulk mail if they suspect it is spam the spammers who control these Bot-Networks typically send low volumes of mail at any one time so as not to arouse suspicions. Thus the spam mail can be traced to an innocent individuals network address and not the spammers network address.

While the number of Botnets appears to be increasing, the number of bots in each Botnet is actually dropping. In the past Botnets with over 80 000 machines were common *(Evan Cooke, Farnam Jahanian, and Danny McPherson. 2005)*. Currently Botnets with a few hundred to a few thousands infected machines are common. One reason for this is that smaller Botnets are more difficult to detect and may be easier to sell or rent.

3.3 Open Proxies

An open proxy is a machine that allows computers to connect through it to other computers on the Internet. Open proxies exist because they enable unhindered Internet usage in countries that restrict access to certain sites for political or social reasons. An Internet user in a country that restricts Internet access can access blocked sites by using an open proxy in a country that does not restrict Internet access.

Spammers use open proxies to hide their network addresses. The recipient of a spammers email will not see the spammers’ network address

on the email but the open proxy's network address. It is estimated that sixty percent of all spam is sent using an open proxy *(Boneh, Dan, 2004)*.

3.4 Open mail relays

Emails sent over the Internet pass through a number of gateways on their way from the sender to the receiver, these gateways are called mail relays. Each time an email passes through a mail relay it has a Received header inserted, this will have the network address of the computer that connected to the mail relay.

An open mail relay is a mis-configured mail relay that accepts mail from any computer on the Internet and forwards it to any other computer on the Internet as opposed to a normal mail relay that accepts mail from a limited number of computers on the Internet and forwards it to a limited number of computers *(Flavio D. Garcia , Jaap-Henk Hoepman and Jeroen van Nieuwenhuizen. 2004).*

This helps the spammer conceal his identity as it appears that the mail is from the open relay and not from the spammer. However as the spammers network address is still found in the emails headers the spammer would insert fake headers into the email. Open mail relays are usually used together with open proxies to conceal the network address of the spammer.

3.5 Untraceable Internet connections

Spammers can also conceal their identities by accessing the Internet from Internet cafes, university computer labs and by using stolen 3G cards. There is thus no way of tracing spammers who access the Internet using these methods. Even if the network address of the computer used is identified this cannot be connected to the identity of the spammer.

4 HOW DIGITAL FORENSIC INVESTIGATORS CAN TRACE THE IDENTITY OF SPAMMERS

The two primary methods for tracing the origins of spam are header analysis and honeypot computers. The following section studies methods for email tracing and their limitations. The methods studied are header analysis and honeypot computers.

4.1 Header analysis

By studying the email headers in a spam email we should be able to identify the senders' network address. Spammers know this and try and divert us from the trail by inserting fake headers. In addition as mentioned previously by using open proxies or Bot-networks the network address of the spammer is not even on the headers. Thus the use of header analysis to trace spammers is highly unlikely.

The only header that cannot be easily forged is the first received header, as all the others may be faked. Spammers will fake their headers to conceal their network addresses. This means that header analysis is not a time and cost effective method to use when tracing spam. The following figure shows an email with the various header tags.

```
Microsoft Mail Internet Headers Version 2.0

Received:       from       s058eml004004.ds1.ad.absa.co.za[1]
([10.6.50.91])  by  V058EMLFFF004.ds1.ad.absa.co.za[2]  with
Microsoft SMTPSVC(6.0.3790.3959);

        Thu, 5 Mar 2009 11:47:49 +0200

Received:    from    S200INT006001    ([169.202.65.146])    by
s058eml004004.ds1.ad.absa.co.za        with        Microsoft
SMTPSVC(6.0.3790.3959);

        Thu, 5 Mar 2009 11:47:49 +0200

Received: from relayin-at1.absa.co.za ([169.202.65.20]) by
S200INT006001 with InterScan Message Security Suite; Thu, 05
Mar 2009 12:03:31 +0200

Received:   from   kendy.up.ac.za   ([137.215.101.101])   by
relayin-at1.absa.co.za            with            Microsoft
SMTPSVC(6.0.3790.3959);

        Thu, 5 Mar 2009 11:45:59 +0200

Received:   from   b040pc181.up.ac.za[3]   ([137.215.40.181]
helo=notebook)

       by kendy.up.ac.za with esmtp (Exim 4.63)

       (envelope-from <hventer@cs.up.ac.za>)

       id 1LfAB1-0001RS-AW

       for   Ickin.Vural@absa.co.za;   Thu,   05   Mar   2009
11:47:39 +0200
```

```
From: "Prof. Hein Venter" <hventer@cs.up.ac.za>
To: <Ickin.Vural@absa.co.za>
Subject: Meeting next week
Date: Thu, 5 Mar 2009 11:45:02 +0200
Message-ID: <FC414216270A45DEAEB887C3BF3C8A18@UP>
MIME-Version: 1.0
Content-Type: text/plain;
       charset="us-ascii"
Content-Transfer-Encoding: 7bit
X-Mailer: Microsoft Office Outlook 11
Thread-Index: Acmddw5GmQGAim2STqiIW8+jHkG6AQ==
X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2900.5579
X-Scan-Signature: 1241d9d45ce102941afa91f8ab9dc533
Return-Path: hventer@cs.up.ac.za
X-OriginalArrivalTime: 05 Mar 2009 09:46:03.0937 (UTC)
FILETIME=[33047910:01C99D77]
```

Figure 4.1 An Email Header

The above email is sent by the University of Pretoria's email sever b040pc181.up.ac.za as highlighted at number [3] in the figure. This is sent to the relay server relayin-at1.absa.co.za [2] which, in turn, sends it to Absa's email server s058eml004004.ds1.ad.absa.co.za [1]. This shows us how we can find the identity of the email sender by tracing the route the email took by analysing the header. But as mentioned earlier a spammer can tamper with these headers so as to confuse an investigator.

4.2 Honeypot computers

A honeypo*t* is a closely monitored computing resource that is intended to be compromised *(Niels Provos. 2004)*. A honeypot computer can be applied to Bot-networks, open proxies and open relays. Thus by setting up a computer to imitate an open proxy or a Bot-network, investigators can attempt to trap the spammers into revealing their network addresses.

4.2.1 Honeypots on Bot-Networks

One way of identifying spammers is to set up a computer to pretend that it is part of a Bot-network *(Boneh, Dan, 2004)*. By allowing the honeypot computer to become part of the Bot-network we can obtain the Bot-network software used by the spammer. Once this has been done the honeypot waits for the spammer to send new instructions and then identifies the network address of the sender. The problem with this approach is that spammers could send the instructions to the Bot-networks under their control over open relays and open proxies thus it may be impossible to discover the identity of the spammer's network address.

4.2.2 Honeypots on open proxies

By setting up a honeypot on an open proxy and waiting for spammers to use it in order to send their spam, we can attempt to identify the spammer's network address. This could be done by keeping records of all connections made by the proxy to locate the source of the spam.

The fake open proxies emulate a subset of the HTTP protocol. Requests made with methods other than GET and CONNECT are answered with an error message. GET requests are answered with a randomly generated page. CONNECT requests to port 25 are internally redirected to an emulated open relay. The reason for this redirection is that the spammer may think nothing went wrong and he is connected to the SMTP server he requested, while he actually is connected to a honeypot. CONNECT requests to ports other than 25 are served with a "Request timeout" message (*Mauro Andreolini,Alessandro Bulgarelli, Michele Colajanni and Francesca Mazzoni. 2005)*.

To identify spammers, it is necessary to encourage them to use honeypot services to their advantages. This is done through the deployment of fake servers, such as open proxies. To ensure traceability of their actions, logging must be enabled for the honey pot open proxy *(Mauro Andreolini,Alessandro Bulgarelli, Michele Colajanni and Francesca Mazzoni. 2005)*.

Spammers try and get around this by using a proxy chain. A proxy chain is when a spammer sends spam mail through a chain of open proxies

and, thus, reducing the chances of their network addresses being compromised *(Boneh, Dan, 2004)*.

4.2.3 Honeypots on open relays

This works by setting up a honeypot on an open relay and waiting for spammers to use it. We would then be able to trace the network address of the spammer using the open relay to send spam.

The fake open relays emulate a SMTP server. All the main commands of the SMTP protocol are implemented, so that spammers cannot notice the difference with a real server. When an e-mail is sent through the open relay, it actually does not reach destination, since all messaged are logged but not forwarded, except the very first one. This is done in order to fool a spammer who sends a first probe message to himself to see if the service is properly running *(Mauro Andreolini,Alessandro Bulgarelli, Michele Colajanni and Francesca Mazzoni. 2005)*.

This technique is similar to that used on open proxies. Thus the spammer would attempt to get around this in the same manner by using a relay chain. The following section describes some advanced techniques used in identifying spammers.

5 ADVANCED EMAIL TRACING TECHNIQUES USING DIGITAL FORENSICS

As discussed previously it is very difficult to obtain the network addresses of spammers. This paper discussed techniques such as header analysis and honeypot computers used to discover the network addresses of spammers. This section discusses digital forensic techniques used to determine the network addresses of spammers. The techniques studied are investigating network devices such as routers, investigating residual data on servers and using bait tactics to identify spammers.

5.1 Investigating Network devices

If logs are unattainable from the servers used by the spammer a routers log can be used instead to obtain information about the spammers network address. *(Patryk Szewczyk, 2007)* If say no access was given to the server logs of the ISP or proxy server that sent the email, the investigator can

analyse the log files of the router or switch that routed the email. This should enable an investigator to determine where the email was sent from.

5.2 Investigating Residual data on servers

SMTP servers keep a copy of emails even after they have been delivered. By using this information we can trace the address of the computer that made the connection. By analysing these in a proxy server the identity of the computer making the connection could be obtained. This would require access to the servers which might not always be given as the proxy server might be located in a jurisdiction that does not have anti spamming laws *(Al-Zarouni Marwan, 2004)*.

5.3 Using Bait tactics to identify spammers

If the email address of the spam message is genuine, forensic investigators can e-mail a message to the sender containing an http "<img src>" tag where the source of the picture is placed on an http server. As soon as the person receiving the e-mail opens it, a log entry with his IP address is recorded on the http server holding the image. This tracks down the sender of the e-mail and establishes his ownership of the e-mail account (Al-Zarouni Marwan, 2004). However this technique is not always useful as some browsers automatically block the downloading of images by default (Microsoft Office online, 2009).

If the person receiving the e-mail is using a proxy server, his IP address will not show in the HTTP logs but rather, the IP of the Proxy server he/she used. In this case the proxy logs can be checked for persons accessing that picture at that time.

If the person in question is using an open proxy server that does not cooperate with law enforcement, one of the following two tactics can be used to track him/her down:

1. Java Applet: The investigator sends an e-mail with an "embedded" Java applet that runs on the receiver's machine and extracts his IP address and e-mails it to the investigator.

2. Active X Control: The investigator sends an e-mail address containing an HTML page with Active X that extracts the receiver's IP address and other information from his machine and sends it to the investigator.

6 SPAMMER IDENTIFICATION

One of the issues with identifying spammers is that SMTP is not a secure protocol and can be tampered with. Some researchers have advocated the adoption of a secure email protocol. *(A. Herzberg, 2005)* But until such time that this technology is widely adopted, and its usefulness would be limited if spammers make use of bot-networks and open proxies to send spam, other means of discouraging spammers must be found.

Spammers on the other hand are in the business of spamming because they want to make a profit. Spammers send spam advertising a product that they hope to sell and a bank account number to which payment should be made. By tracing the information in the email message body an investigator should be able to identify the source of some spam. This however is not enough as the enterprise can deny having sent the spam mail and the investigators may not be able to conclusively prove ownership.

Thus a proposal to identify spammers would be to create an implementation that identifies computers which are sending spam. These computers should then be added to a spam list that could be blocked by an ISP. This framework would need to identify bot-networks as well as open proxies sending spam. Once on a spam list, it is up to the individual or organisation concerned to have their network address removed from the spam list.

The detection of spamming computers can be done by analysing the network layer traffic and determining patterns that match bot-networks and open proxies sending spam.

The implementation would detect spam by analysing data provided by an ISP to identify abnormal behaviour on the network to identify spammers. This system will enable ISP's to proactively locate open proxies and bot-networks.

This would require analysing large data sets which would require a large amount of computing power. However the computing power of computers has increased and computers such as blade servers that can be programmed to analyse data using parallel computing techniques are now

available. Thus it is possible for ISP's to analyse large datasets to detect abnormal behaviour in a way that was not possible a number of years ago.

7 DISCUSSION

The implementation of a system to detect spammers by analysing network traffic for abnormal behaviour has some shortcomings, mainly that spammers do not usually send out mail in bulk but in smaller packets so as to avoid detection.

The implementation would have to take into account spam email sending patterns to effectively identify spammers. The implementation could either make use of artificial intelligence to learn behaviour by feeding it training patterns or by using graph analysis which is perhaps better suited for large scale data analysis.

The authors of this paper, however, still need to explore their ideas mentioned in this paper in future work so as to produce a proof-of-concept prototype.

8 CONCLUSION

This paper outlines the challenges facing digital forensic investigators when attempting to identify spammers. Servers that contain forensic data such as log files showing the network addresses of the computers that have connected to it, that would enable a digital forensic investigation, are not made available by the servers' owners for various reasons. The usual reason for the refusal is that court orders requesting this data to be made available, may not apply to that jurisdiction.

The use of bot-networks means that even if the source of the machine sending the spam is identified the person owning the machine is not the one responsible for sending spam. The use of untraceable Internet connections and open proxies to communicate instructions to bot-networks makes the use of Honeypots unlikely to succeed.

Thus any success in tracing spammers will be matched by spammers using increasingly sophisticated techniques to evade detection. Greater responsibility will have to fall to ISP's in monitoring connections to open

proxies as well as attempting to shut down open relays. Nevertheless an arms race between spammers and forensic investigators will continue for the foreseeable future.

9 REFERENCES

Al-Zarouni Marwan. 2004. Tracing E-mail Headers. We-B Centre & Edith Cowan University.

Boneh, Dan. 2004. The Difficulties of Tracing Spam Email. Department of Computer Science Stanford University.

A. Herzberg, Controlling Spam by Secure Internet Content Selection, Proceedings of Secure Communication Networks (SCN) 2004, LNCS vol. 3352, Springer-Verlag.

Acts Online, 2002. Electronic Communications and Transactions Act, 2002. Available: http://www.acts.co.za/ect_act/. [April 2009]

Europa Press Releases, 2009. Data protection: "Junk" e-mail costs Internet users 10 billion a year worldwide. Available: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/01/154&format=HTML&aged=0&language=EN&guiLanguage=en. [April 2009]

Messaging Anti-Abuse Working Group, 2007.Email Metrics Program: The Network Operators' Perspective. Available: http://www.maawg.org/about/MAAWG20072Q_Metrics_Report.pdf . [April 2009]

Evan Cooke, Farnam Jahanian, Danny McPherson. 2005 . The advanced computing systems association. [Online] The Zombie Roundup: Understanding, Detecting, and Disrupting Botnets. Available: http://www.usenix.org/events/sruti05/tech/full_papers/cooke/cooke_html/ [April 2009]

Flavio D. Garcia , Jaap-Henk Hoepman and Jeroen van Nieuwenhuizen, 2004. Spam Filter Analysis: IFIP International Federation for Information Processing. Springer Boston

Internet Service Providers' Association, 2008. 'What is Spam?' Available: http://www.ispa.org.za/spam/whatisspam.shtml. [April 2009]

Microsoft Research. S-GPS: Spammer Global Positioning System. Available: http://research.microsoft.com/en-us/projects/S-GPS/. [April 2009]

Mauro Andreolini, Alessandro Bulgarelli, Michele Colajanni and Francesca Mazzoni. 2005. The advanced computing systems association. [Online] HoneySpam: Honeypots fighting spam at the source. Available: http://www.usenix.org/event/sruti05/tech/full_papers/andreolini/andreolini_html/. [April 2009]

Microsoft Office online, 2009. About protecting your privacy by blocking automatic picture downloads. [Online] Available: http://office.microsoft.com/en-us/outlook/HP010440221033.aspx. [April 2009]

Niels Provos. 2004. The advanced computing systems association. [Online]. A Virtual Honeypot Framework. Available: http://www.usenix.org/event/sec04/tech/full_papers/provos/provos_html/. [April 2009]

Patryk Szewczyk, 2007. ADSL Router Forensics Part 1: An introduction to a new source of electronic evidence We-B Centre & Edith Cowan University.

Spamhaus, 2009. The Definition of spam. Available: http://www.spamhaus.org/definition.html [April 2009]

Tzerefos, P. Smythe, C. Stergiou, I. Cvetkovic, S. 1997. A comparative study of Simple Mail Transfer Protocol (SMTP), PostOffice

Protocol (POP) and X.400 Electronic Mail Protocols. Proceedings., 22nd Annual Conference on Publication Date: 2-5 Nov1997.

Yao Zhaoy , Yinglian Xie, Fang Yu, Qifa Ke, Yuan Yu, Yan Cheny, and Eliot Gillumz BotGraph: Large Scale Spamming Botnet Detection. Microsoft Research