Tracing E-mail Headers

Marwan Al-Zarouni

School of Computer and Information Science
Edith Cowan University
Perth, Western Australia
E-mail: marwan@marwan.com

Abstract

This paper will discuss tracing e-mail headers and issues associated with it. It will address both HTTP & SMTP initiated e-mails. It will discuss different ways used by e-mail senders to evade tracing and workarounds used by investigators to combat them. It will also discuss advanced measures and techniques used by investigators to track emails. The paper will not however discuss any particular tools nor endorse any software products in its coverage.

Keywords

E-mail forensics, tracing e-mail headers, e-mail tracking, network forensics, fake e-mails, web mail tracking, SMTP tracing, e-mail tunnelling, e-mail anonymity, mail relay, e-mail false headers.

E-MAIL COMPONENTS

E-mails are made of two main parts; they are the message header and message body. The header part contains routing information about the e-mail and other information such as the source and destination of the e-mail, the IP address of the sender and time related information. The message body contains the actual message of the e-mail, i.e. message subject and body. The body might also contain attachments in the form of MIME or SMIME (Lewis, 2004). Message headers are the important part for investigating e-mail messages and will be discussed in detail in this paper.

E-MAIL HEADER STRUCTURE

E-mail headers are organized from the bottom up. This means that the e-mail was handed from the machines at the bottom of the e-mail header to the ones at the top of it. These machines are referred to as Message Transfer Agents (MTAs) and each of them adds a "received" section to the e-mail header, sometimes referred to as "received header". This is similar to postmarks in conventional postal systems. The order of the "received" sections will be like a stack of pancakes, with the one receiving the e-mail last at the top of the stack (Venit, 2000). In the figure below, notice that there were three received sections. This means that three MTAs were involved in the delivery of the e-mail message with the one at the bottom of the message being the one receiving the original message from the sender.

E-MAIL TRACING

E-mail tracing is conducted by examining the header information contained in e-mail messages to determine their source. Header information is included with e-mails either at the beginning or the end of e-mail messages (Jones, 2001). A typical e-mail header looks like this:

```
Received: from search.org ([64.162.18.2]) by sgiserver1.search.org with SMTP (Microsoft Exchange Internet Mail Service
Version 5.5.2650.21)
        id K9HBB4C4; Mon, 21 May 2001 09:47:01 -0700
Received: from web14506.mail.yahoo.com ([216.136.224.69]) by SEARCH.ORG
        with SMTP (IPAD 2.52) id 3579700; Mon, 21 May 2001 08:47:23 -0800
Message-ID: <20010521164640.85785.qmail@web14506.mail.yahoo.com>
Received: from [216.104.228.118] by web14506.mail.yahoo.com; Mon, 21 May 2001 09:46:40 PDT
Date: Mon, 21 May 2001 09:46:40 -0700 (PDT)
From: Can Do <can_do1@yahoo.com>
Subject: check out this e-mail header
To: todd@search.org
MIME-Version: 1.0
Content-Type: text/plain; charset=us-ascii
```

Figure 1: A Typical E-mail Header

Information contained in the header can aid investigators in tracing the sender of the e-mail. A thorough investigation of e-mail headers should include:

- Examining sender's e-mail address
- Examining message initiation protocol (HTTP vs. SMTP)
- Examining Message ID
- Examining sender's IP address

THE BOTTOM TO TOP APPROACH

To determine the source of the e-mail, investigators must first examine the received section at the bottom of the header and work their way up in a bottom to top approach (Jones, 2001).

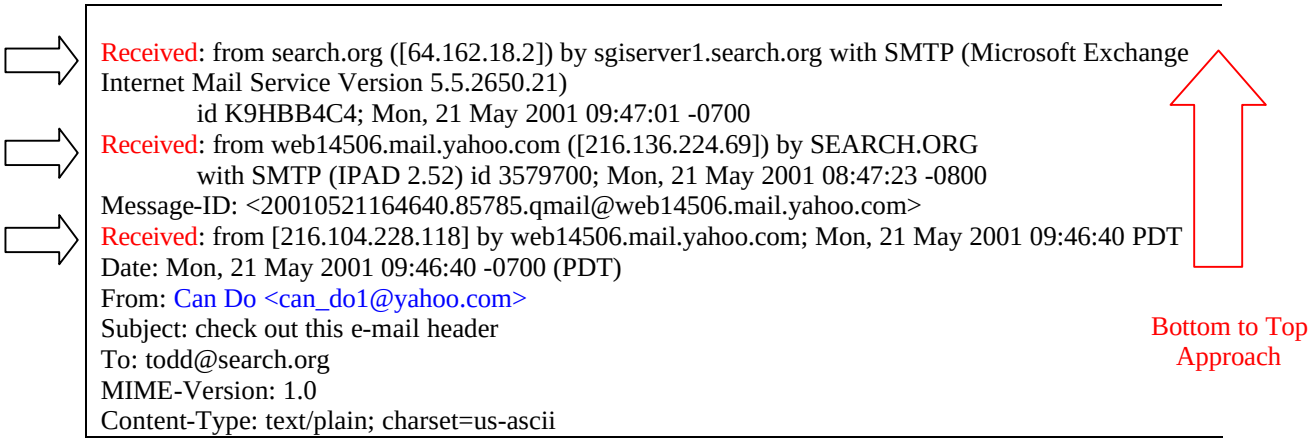

Figure 2: The Bottom to Top Approach

It is also important that e-mail cases examine the logs of all servers in the received chain as soon as possible. Time is very important in e-mail cases as HTTP and SMTP logs are archived frequently; especially by large ISPs. If a log is archived, it could take time and effort to retrieve and decompress the log files needed to trace e-mails. Some e-mails have fake headers and fake "from" e-mail addresses to fool investigators, so extreme caution and careful scrutiny should be practiced in investigating every part of the e-mail header.

EXAMINING SENDER'S E-MAIL ADDRESS

The sender's e-mail address can usually be found after the "From" section of the header. But that is not the only place it can be found. It can also be found under other sections depending on the e-mail client used. These sections include but are not limited to, the following:

- X-Originating-E-mail
- X-Sender
- Return-Path

E-mail addresses can sometimes suggest the method used to generate the e-mail and the server that the e-mail originated from (i.e. hotmail, outlook, corporate server, ISP …etc). However, e-mail addresses should be viewed with caution by investigators as they can be easily faked.

Notice that some headers begin with an "X-", this means that they are X-Headers. X-Headers are inserted by e-mail client programs or applications that use e-mail to pass information to e-mail handling programs for processing. They may be introduced by large vendors and picked up for use by others. In this way an X-header can be considered as a de-facto standard. An example of this is the "X-Mailer" header which many e-mail clients use to define the e-mail client application and version used (Francis, 2004).

DETECTING FAKE E-MAIL ADDRESSES

The sender's e-mail address can be easily faked and can be hard to detect. If the server mentioned in the bottom "received" section does not match the server of the e-mail address, this suggests that the e-mail address is a fake one. An example of that is shown below:

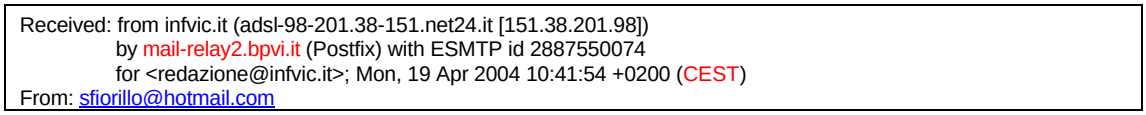

Figure 3: The Header of a Forged E-mail

Notice that the e-mail address in the "From" field has "hotmail.com" as the domain for the e-mail while in the received section of the header there is no hotmail server mentioned at all. This is clearly a forged (fake) e-mail with a fake from address. Also notice that the time on the received section is Central European Standard Time (CEST), and hotmail.com servers are not in Europe.

EXAMINING MESSAGE INITIATION PROTOCOL (HTTP VS. SMTP)

A tell-tale sign that e-mail was sent from a web based e-mail client is the use of "with HTTP" or "HTTPS" in the bottom received section of the e-mail. If "with SMTP" or "ESMTP" was used instead, this means that the e-mail was sent either manually or through SMTP client software. So, if the "From" section has a "something@hotmail.com" address but the bottom received header has with SMTP instead of HTTP or HTTPS, this means that the e-mail is forged and that the address used is fake. Using the same example as before, notice that the delivery method used is ESMTP even though the e-mail seems to come from a hotmail account:

```
Received: from infvic.it (adsl-98-201.38-151.net24.it [151.38.201.98])
        by mail-relay2.bpvi.it (Postfix) with ESMTP id 2887550074
        for <redazione@infvic.it>; Mon, 19 Apr 2004 10:41:54 +0200 (CEST)
From: sfiorillo@hotmail.com
```

Figure 4: Header of a Forged E-mail

E-MAIL INITIATION METHODS

E-mail can be generated by a number of different methods. The following is a list of the most common ones:

- Using HTTP-based e-mail services such as HotMail.com, Yahoo! Mail, and others.
- Using SMTP client applications such as Outlook, Expedia, Pine and others.
- Constructing e-mail manually. Usually used for forged e-mail by using Telnet to connect to port 25 (SMTP) or the use of other applications or tools that connect to the e-mail servers or relay servers.

EXAMINING MESSAGE IDS

Message IDs play a vital role in tracing e-mails. They facilitate searching e-mail logs and also, depending on the software used, can contain time information about the e-mail or other helpful information (Nelson, Phillips, Enfinger, Steuart, & Phillips, 2004, p460). Message IDs are created by the e-mail client software as well as the SMTP servers that receive the message and pass it along to its destination. In the example below, notice that three e-mail servers were involved with the delivery of the message and each of them assigned an ID to the message.

```
Received: from search.org ([64.162.18.2]) by sgiserver1.search.org with SMTP (Microsoft Exchange Internet Mail
        Service Version 5.5.2650.21)
        id K9HBB4C4; Mon, 21 May 2001 09:47:01 -0700
Received: from web14506.mail.yahoo.com ([216.136.224.69]) by SEARCH.ORG
        with SMTP (IPAD 2.52) id 3579700; Mon, 21 May 2001 08:47:23 -0800
Message-ID: <20010521164640.85785.qmail@web14506.mail.yahoo.com>
Received: from [216.104.228.118] by web14506.mail.yahoo.com; Mon, 21 May 2001 09:46:40 PDT
Date: Mon, 21 May 2001 09:46:40 -0700 (PDT)
From: Can Do <can_do1@yahoo.com>
Subject: check out this e-mail header
To: todd@search.org
MIME-Version: 1.0
Content-Type: text/plain; charset=us-ascii
```

Figure 5: Message IDs in a Typical E-mail Header

Notice also that the "Message-ID" from the originating server contains the time: "*Message-ID: <20010521164640.85785.qmail@web14506.mail.yahoo.com>*" which is 2001-05-21 at 16:46:40 GMT (which is equal to 09:46:40 PDT).

TRACING E-MAIL BASED ON INITIATION METHOD

Investigating e-mail to determine its origin can differ depending on its initiation method. In the following sections we will discuss each type of origin and the means that can make tracing the source of e-mail harder or easier.

TRACING HTTP-BASED E-MAIL

Web based e-mail is one of the common ways of sending nuisance e-mails due to many factors. These factors include:

- The convenience of being able to send and receive e-mail through a web page.
- Free web based e-mail accounts are easily obtained and can take minutes to set-up.
- Fake details can be supplied when setting up accounts.
- E-mail can be sent from virtually anywhere including Internet Cafés and public places without suspicion.
- "Anonymous" web based e-mail services offer added level of security to senders of nuisance e-mails.

The following diagram shows how HTTP-based e-mail can be made harder or easier to trace:

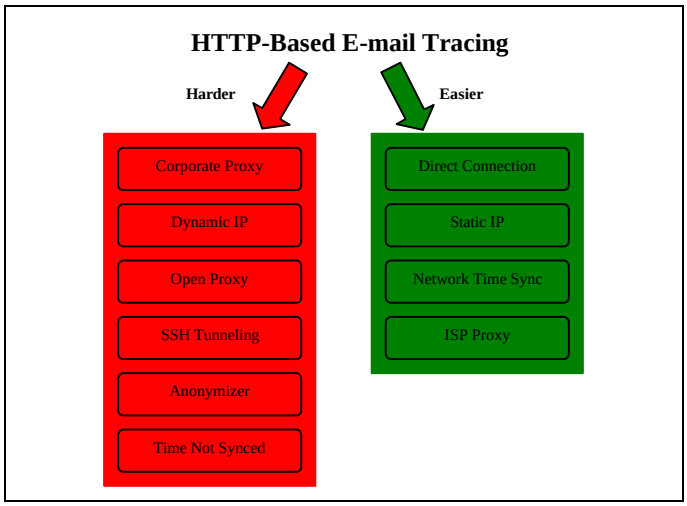

Figure 6: Issues Effecting HTTP-Based E-mail Tracing

Factors Effecting HTTP E-Mail Tracing

As shown in the above diagram, each factor listed effects e-mail tracing and can make it either harder or easier to investigate, track or prove in court. In the following section of the paper, each factor will be discussed and its effects will be explained.

Direct Connection

When an e-mail user connects directly to an HTTP-based e-mail client, the IP address of the sender shows in the header section of the e-mail under the last "received" section of the header as shown below:

```
MIME-Version: 1.0
X-Originating-IP: [195.252.34.107]
X-Originating-E-mail: [jane_doe@hotmail.com]
X-Sender: jane_doe@hotmail.com
Received: from 195.252.34.107 by by1fd.bay1.hotmail.msn.com with HTTP;Sat, 01 May 2004 14:24:48 GMT
```

Figure 7: HTTP-based E-mail Header

Notice that the e-mail and the IP address of the sender are clearly shown and there is only one "Received" section in this e-mail header. This is because the sender used Hotmail.com to send his e-mail to another hotmail user. This means that no servers exchanged the e-mail outside of Hotmail.com. This makes tracing the e-mail a matter of finding out the ISP that owns the IP block that includes the IP: 195.252.34.107 and getting the billing contact information for the user of that IP Address. Using the basic command "nslookup (IP Address)" the IP resolves to: "h107n1fls25o1067.bredband.comhem.se" which looks like a broadband user in Sweden.

Static vs. Dynamic IP

The ease of determining the user who sent the e-mail depends on the IP address policy of the ISP. The IP address could be either Static (usually for dedicated, high speed broadband users) or Dynamic (most dial-up users). If the IP address is Dynamic, retrieving the billing contact information can be done by getting the time and date when the e-mail was sent (01 May 2004 14:24:48 GMT in the example given above), and tying that to the time on the billing server to determine which user was allocated the IP that was used to send the e-mail (Venit, 2000).

The billing server (or access server) usually has a "Session Start" and "Session Stop" record that contains time and IP allocation for each account.

Network Time Sync
If the time on the ISP's billing server is not perfectly synced to network time, it could complicate determining the sender of the e-mail. This is especially true for large ISPs or if the time widely of sync or has unpredictable deviations. Time has very substantial value in a court of law. If time sync information can not be verified, it will be hard to prove the case in court.
Time synchronization is essential for all servers involved in tracing e-mail. The way each server is synchronized should be questioned by investigators. Secure protocols must be used versus insecure ones. An example of a secure time sync protocol is STIME and an example of insecure protocol is NTP.
Synchronization via GPS is also considered a reliable source for time sync (GPSClock, 2003; Valli, 2004).

HTTP Proxies and their Effect on HTTP-Based E-mail Tracing
An e-mail sent from a person accessing the Internet via an http proxy will NOT have his/her IP address in the last "received" section; rather, it will have the Proxy's IP address. This means that the proxy server's logs must be examined to find the original sender's IP address.

Corporate Proxies
An example of an e-mail sent from a client using a corporate proxy server is shown below:

```
MIME-Version: 1.0
X-Originating-IP: [213.132.32.130]
X-Originating-E-mail: [jd@hotmail.com]
X-Sender: jd@hotmail.com
Received: from 213.132.32.130 by by2fd.bay2.hotmail.msn.com with HTTP;Tue, 20 Jan 2004 06:09:55 GMT
```

Figure 8: E-mail Sent Via a Corporate Proxy Server

Notice that the IP shown can be traced back to Proxy server (213.132.32.130) which resolves to (eth1.cache2.dubaiinternetcity.net) which is a corporate HTTP Proxy server. It is not the sender's IP address. The sender may or may not have an Internet-publishable IP address. To trace the e-mail to its source requires searching the http proxy logs for connections to the specific hotmail server that the e-mail was sent from (by2fd.bay2.hotmail.msn.com) from a user within that organization at the exact same time it was sent from (Tue, 20 Jan 2004 06:09:55 GMT). Note that it is crucial that the time on the proxy server is synced down to the tenth of a second.

ISP Based HTTP Proxies
Some Internet Service providers (ISPs) have HTTP proxy servers available for their customers. Other ISPs use transparent proxies for purposes such censorship. If a user sends an e-mail through a proxy server of an ISP, tracing that e-mail will result in the ISP proxy server's IP address showing in the received section of the header. Unlike Corporate Proxy Servers, ISP proxies are usually time synced to reliable time servers and usually maintain strict logs in regards to time updates. This is also true for financial firms (Valli, 2004). This makes tracing e-mails originating from ISP proxies and financial firms relatively easier and increases the likelihood that they can be considered as evidence in a court of law.
However, if the IP assigned to the customer is dynamic, it will require more time and effort from the ISP to trace the user as explained in the "Static vs. Dynamic IP" section of this paper.
There is also another avenue to pursue in order to trace the sender of the e-mail. If the HTTP-based e-mail site uses SSL to log-in users, the logs for that particular e-mail account can be retrieved from the e-mail service provider. Unlike HTTP, SSL connections are made directly between the server and the client. Therefore, the real IP of the client would appear in the SSL connection logs rather than the IP of the proxy server.

Open Proxies
Open proxies can vary in type and anonymity. When an e-mail is traced back to an IP of an open or so called anonymous proxy, it can be difficult to get the IP of the e-mail sender. Heavy demand on servers, their location outside the areas of jurisdiction and lack of co-operation from administrators, who guard user anonymity, can all complicate investigations. However, if SSL was used for log-in to the HTTP-based mail server, open proxies would not be an issue.

Tunnelling
Tunnelling can be used by e-mail senders to evade being traced by law enforcement. Tunnelling can be done in many ways depending on the software and techniques used. SSH tunnelling is a straight forward approach and is usually used by home users who have full access to the Internet. It requires logging into a remote server and running the SSH daemon on it with a command such as: ***ssh -R 12340:webmailsite.com:80 localhost*** This forwards all requests on port 12340 coming to the machine running the daemon to port 80 of the webmail.com,

the HTTP-based e-mail server. When Tunnelling is used, only the IP of the Tunnelling server shows up on received section of the e-mail header.
Tracing the source of a tunnelled e-mail is complex. This is due to the following: The usage of an SSH daemon as a tunnel is not usually logged. The only way to prove that a user used SSH tunnelling is to determine who logged to the server via SSH at the time the e-mail was sent. There might be a number of users logged in to the same server via SSH at the same time the e-mail was sent which makes it a hard to prove in court (Akin, 2003). The diagram below shows how SSH tunnelling is performed:

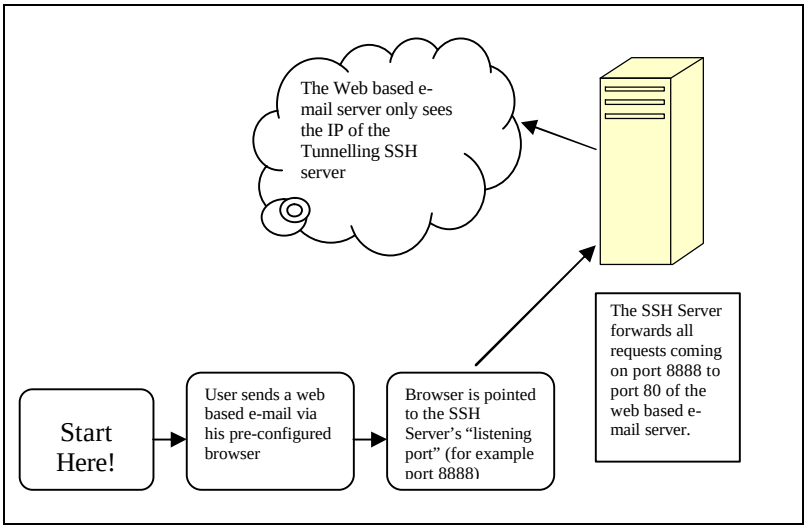

Figure 9: How SSH tunnelling works

There are also other ways of tunnelling by using client-server based tunnelling software. This technique is usually used by HTTP-based e-mail abusers from within companies or users who are forced to go through a proxy to access the Internet. The diagram below shows an example of such set up with a Windows based program called JAP (JAP, 2003):

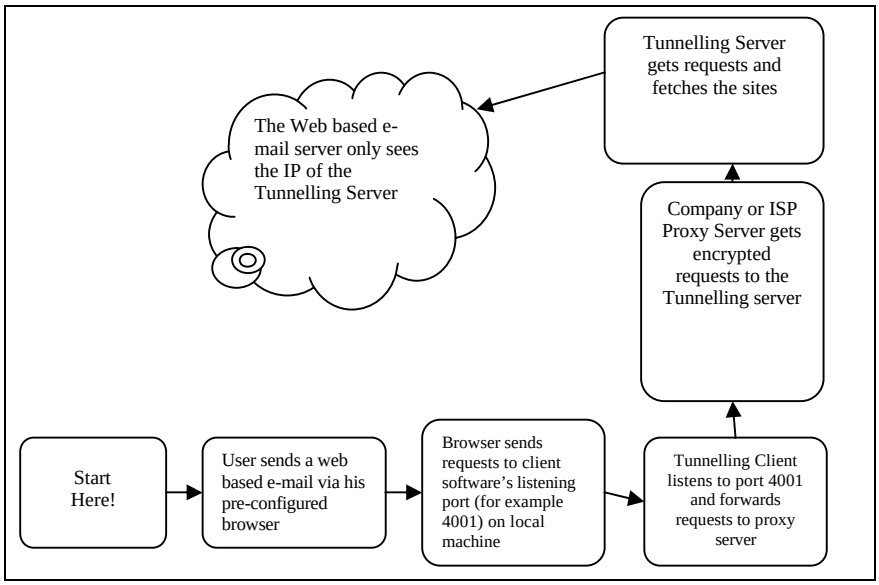

Figure 10: How JAP tunnelling works

The IP Address in the bottom received header will contain the IP Address of the tunnelling proxy server used. An example of an HTTP-based e-mail sent via JAP is below:

```
X-Received: 12 May 2004 15:41:14 GMT
Received: (cpmta 19530 invoked from network); 12 May 2004 08:41:14 -0700
Received: from 69.93.239.130 (HELO ns.hostsearchindia.com)
by smtp.inbound.c001.snv.cp.net (209.228.32.108) with SMTP; 12 May 2004 08:41:14 -0700
Received: from marwan by ns.hostsearchindia.com with local (Exim 4.24)
id 1BNvr6-0000Bq-3z; Wed, 12 May 2004 21:11:12 +0530
```

```
Received: from 141.76.1.121 ([141.76.1.121])
(SquirrelMail authenticated user marwan)
by www.marwan.com with HTTP;
Wed, 12 May 2004 21:11:12 +0530 (IST)
MIME-Version: 1.0
X-Antiabuse: This header was added to track abuse, please include it with any abuse report
X-Antiabuse: Primary Hostname - ns.hostsearchindia.com
X-Antiabuse: Original Domain - marwan.net
X-Antiabuse: Originator/Caller UID/GID - [33052 33052] / [47 12]
X-Antiabuse: Sender Address Domain - ns.hostsearchindia.com
User-Agent: SquirrelMail/1.4.0
```

Figure 11: Header of E-mail Sent Via JAP Tunnelling Software

Notice that the bottom IP Address is 141.76.1.121. This resolves to *proxy1.anon-online.org* which is the proxy server used by JAP to tunnel the connection.
The ISP or company proxy logs in this case are useless, as all data to and from the server and client is encrypted. The logs of HTTP-based e-mail server are also useless as the IP address of the JAP server will only show in the logs. The line below shows the corresponding entry in the web server log:

```
141.76.1.122 - - [12/May/2004:21:09:00 +0530] "GET /webmail HTTP/1.0" 302 277 "-" "Mozilla/4.0 (compatible; MSIE
6.0; Windows NT 5.1) Opera 7.50 [en]"
```

Figure 12: Web Server Log Entry Related to the E-mail Sent Via JAP

Notice that the IP Address in the log is 141.76.1.122. This resolves to *proxy2.anon-online.org* which is another proxy server used by JAP to tunnel the connection. The only way to trace e-mail sent this way is to access the logs of servers running the tunnelling service. In this case, both 141.76.1.121 and 141.76.1.122 should be checked. Obtaining these logs depends on the country in which the server resides and also the type of case in which the e-mail is involved. Paedophilia and terrorism cases get high attention and level of cooperation. Harassment and threat e-mails usually are not responded to as quickly and sometimes not at all. Nevertheless, an increasing number of anonymity servers maintain logs and provide them to law enforcement on request.

Anonymous Surfing Sites
If a user sends an HTTP-based e-mail via a websites offering anonymous surfing only the anonymous server's IP will show in the "received" part of the header. The logs from the anonymity server can be used in this case and the exact time and HTTP-based e-mail server should be used to track down the IP of the sender of e-mail. If anonymity is a payed service, user account and credit card details should be obtained (Akin, 2003).

TRACING SMTP-BASED E-MAIL

Tracing SMTP-based e-mail is somewhat different from that of HTTP-based e-mail. Therefore, issues that are associated with it are different too as shown below:

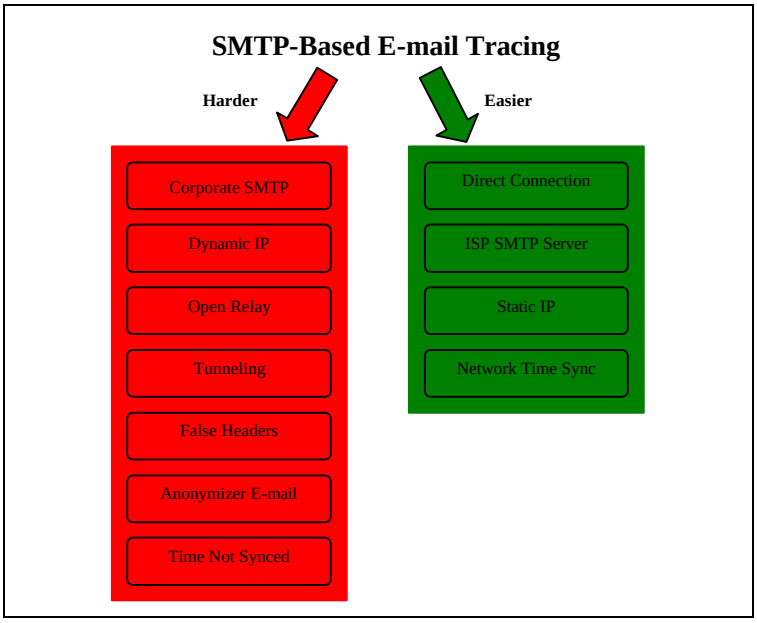

Figure 13: Issues Effecting SMTP-Based E-mail Tracing

Factors Effecting Tracing SMTP Initiated E-mail
Tracing e-mail from initiated with SMTP differs slightly from that of HTTP-based e-mail. Some factors are shared such as static vs. dynamic IPs, the importance of time, and tunnelling. Some issues such as open relay and false headers are more relevant to SMTP e-mail.

Direct Connection to SMTP Servers
The example below shows a header of an SMTP initiated e-mail:

```
X-Received: 28 Apr 2004 16:23:05 GMT
Organization: TheTrainingCo.
X-Mailer: Version 5.0
Return-Path: <jack@thetrainingco.com>
Received: (cpmta 21033 invoked from network); 28 Apr 2004 09:23:05 -0700
Received: from 216.189.100.151 (HELO thetrainingco.com)
by smtp.inbound.c001.snv.cp.net (209.228.32.109) with SMTP; 28 Apr 2004 09:23:05 -0700
```

Figure 14: E-mail Initiated with SMTP

In this case the IP resolves to "*pppoe-151.rockhill.cetlink.net*" PPPoE (Point-to-Point Protocol over Ethernet), a protocol used by many ADSL Internet Service Providers. This hints that the sender is either using an ADSL connection from home or a corporate ADSL connection. This makes tracing this e-mail a straight forward task. Logs in this case are obtained from the ISP, the company (if it was sent from an IP belonging to a company) and the SMTP server.

ISP SMTP Server
If the SMTP server belongs to an ISP, prompt action must be taken by investigators to obtain a court order and the server logs from the ISP. If the ISP uses NAT for its clients, this does not negatively effect investigations, as long as logs are kept and time sync is maintained properly.

Corporate SMTP Server
If a user uses a company's SMTP server, NAT will most likely be used and a non-publishable IP will be obtained from the SMTP Server's logs. If DHCP is used, IP allocation logs should also be obtained.

Open Relay SMTP Server
Open relay SMTP servers are servers that forward e-mail messages regardless of sender and receiver. If a user uses an open relay SMTP server, his IP will still be present in the headers. His IP will also be present in the open relay server's logs. The use of open relay servers is often used to send spoofed messages. They can only fool unaware receivers as they can present the "from" address to be anything (such as w@whitehouse.gov) for example. Investigators can easily detect such false "from "addresses as discussed earlier in this paper. Only when combined with other techniques can the use of open relay hinder an investigation (Akin, 2003). There are many ways to test if a server allows for open relays (SecWiz, 2004). The simplest of which is to connect to it via Telnet and try to send e-mail from one non-existent domain to another.

Tunnelling and Port Redirecting
As with HTTP-based tunnelling, SMTP can also be tunnelled, redirected and encrypted. The received headers in this case will also contain the IP of the SSH server and it will be as hard to determine the source of the e-mail as in the HTTP-based e-mail (Akin, 2003).

False Headers
False headers are additional headers added to the e-mail when sending e-mail manually. The figure below shows how they are created:

```
220 mailgw1.ecu.edu.au ESMTP Mirapoint 3.4.5-GR; Thu, 13 May 2004 01:53:38 +0800
 (WST)
mail from:malzarou@student.ecu.edu.au
250 malzarou@student.ecu.edu.au... Sender ok
rcpt to: webmaster@marwan.net
250 webmaster@marwan.net... Recipient ok
data
354 Enter mail, end with "." on a line by itself
Received: from 216.189.100.151 (HELO thetrainingco.com) by smtp.inbound.c001.snv.cp.net (209.228.32.109)
with SMTP; 28 Apr 2004 09:23:05 -0700
```

```
Hello
.
250 ALM01196 Message accepted for delivery
```

Figure 15: How to add false headers to an e-mail

Notice that only one header was added in the "data" field of the message. A whole list of them can also be added to confuse the investigators. The investigator in this case has to pay close attention to time as it could be the only way to determine the false headers from the genuine ones. The figure below shows the received e-mail containing the false headers:

```
Message-Id: <200405121754.ALM01196@mailgw1.ecu.edu.au>
Received: (cpmta 18042 invoked from network); 12 May 2004 10:55:22 -0700
Received: from 139.230.225.11 (HELO mailgw1.ecu.edu.au)
by smtp.inbound.c001.snv.cp.net (209.228.32.110) with SMTP; 12 May 2004 10:55:22 -0700
Received: from [10.77.31.114] ([10.77.31.114]) by mailgw1.ecu.edu.au (MOS 3.4.5-GR)
with SMTP id ALM01196; Thu, 13 May 2004 01:54:15 +0800 (WST)
Received: from 216.189.100.151 (HELO thetrainingco.com) by smtp.inbound.c001.snv.cp.net (209.228.32.109) with
SMTP; 28 Apr 2004 09:23:05 -0700
```

Figure 16: E-mail Containing a False Header

Notice that the method of delivery for the e-mail is SMTP which means that the sender did not use a HTTP-Based client to send the e-mail (but he sent it from an HTTP-based address: malzarou@student.ecu.edu.au) which suggests that the e-mail is forged. Furthermore, the time on the bottom most received header is off by a wide margin (28 Apr 2004). This leads to one of two conclusions. Either the header is false or the time on that SMTP server is out unsynchronised. Simply connecting to the corresponding SMTP server on port 25 and checking the time on it rules out the second conclusion.
Only determined and skilful e-mail abusers will be able to match the time on the servers in the chain of e-mail, in which case, the investigators will have to investigate all headers and request logs corresponding to each one. This could be an enormous task depending on the number of headers and the location of servers involved. Common sense and careful inspection of each header in this case can be used to rule out false headers.

Note: In the example above, the internal IP address of the sender is contained in the header (10.77.31.114). Therefore, e-mails that contain false headers do actually contain the IP of the sender somewhere in the chain of received headers. This means that cases involving false headers are not completely hopeless as long as every received header is thoroughly investigated.

Anonymizer E-mail
There are many companies offering free or SMTP relay services with anonymity to the sender. What these servers basically do is strip out the IP address of the e-mail originator from the header and then forward the e-mail to its destination. Headers on the received end will only contain the IP of the Anonymous SMTP server that sent the e-mail on behalf of the sender. Unlike HTTP-based anonymizers, SMTP-based anonymizers scrub all information about the sender of the e-mail. This includes the client SMTP program used (outlook for example) and other information. They also delay the time the e-mail was sent by up to ten minutes.
One way to determine the sender of e-mail in this case is to contact the administrators of the anonymity site and get the logs from the anonymous SMTP server corresponding to the time the e-mail was sent AND the content of the e-mail. The servers usually keep track of these details for auditing and legal purposes (LuxSci, 2004).

COMBINING TECHNIQUES

Combining e-mail trace evasion techniques can complicate matters for investigators. This is true for both HTTP and SMTP based e-mail. The possibilities and combinations are limitless. Tracing back logs in this case can be a cumbersome task. The IP chase can lead from an SSH server to an Open Proxy and then from there to a port redirector and so on until it gets to the IP of the originator. Failure in obtaining the logs of any server along the way can lead to the failure of the investigation. Also, the process of tracing e-mail that combines such techniques can become time consuming and process intensive for the investigators involved. Time factors alone can bring the investigation to a halt.

ADVANCED MEASURES TO TRACE E-MAILS

In some cases, investigating headers can lead to a dead end. This could be due to many factors. They include:

- Lack of international cooperation.
- Inability to prove secure time on any of the servers in the chain of investigation.
- Unavailability of logs on some corporate servers.
- Failure to maintain "chain of evidence" by investigators

This however does not always close a case. Other measures and techniques can be used by investigators to establish the sender of the e-mail. These techniques include: investigating network devices, investigating residual data on servers, bait tactics and software embedded identifiers.

Investigating Network Devices
If no logs were obtainable from any server in the delivery chain, a network device such as a router or firewall's log can be used instead of a corresponding server's log to identify the source of e-mail (Nelson et al., 2004,p462-463). Although this avenue of investigation is not likely to be pursued by investigators, it is still a valid one.

Investigating Residual Data on Servers
SMTP Servers usually keep a copy of all e-mails even after they are the delivered. They only delete such data after a backup operation is performed. It is essential however that an investigator requests such data and log files as soon as possible to minimize the possibility that data is erased or overwritten if it is not backed up. Recovering erased files is also possible but would take much more time and effort and might disrupt operations. E-mail saved on a server can be used in cases where the receiver or the sender erases the e-mail intentionally or unintentionally (Nelson et al., 2004, p465).

Bait Tactics
Bait tactics are nothing new to investigators. But in this case, they are used with a technical twist. If the e-mail address of the sender is a real one, 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.
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.

In the case where an e-mail address of the sender was faked, there are other ways to track the source. One is to use specific user details embedded in software. An example of this is explained in the following section.

Software Embedded Identifiers
Software used to compose e-mail and documents and files attached to e-mails can hold vital information about the creator of the e-mail, file or document. Microsoft Office 97 products for example have an embedded Globally Unique Identifier (GUID) which is a unique number that matches e-mails written using them (e-mail written in Word 97 or Outlook 97) with the unique MAC Address of the originator of the document as well as the name and login ID of the creator of the document (Wyble, 2002).

CLIENT-SIDE EVIDENCE GATHERING

It is important to mention that once the e-mail sender's identity has been confirmed, investigation must begin on the client side (i.e. e-mail sender). Investigators must establish who was at the keyboard of the computer that the e-mail was sent from at the time the e-mail was sent. Also, they must look for evidence on the computer that links it to the crime. This could be data remanence in web browser files (in case of HTTP-based e-mail) or in the e-mail client program (in case of SMTP based e-mail). The location of files related to SMTP e-mail can be different depending on the program used (Mandia, Prosise, & Pepe, 2003, p308). Message ID from the client side is a strong piece of evidence that can tie the machine used to compose the e-mail to the received e-mail headers.
It is essential to use specialized forensics tools when investigating on the client side, as these tools will be able to find deleted files and will also look for files depending on the structure of the file rather than its name or extension (Nelson et al., 2004, p460-461).

CONCLUSION AND FURTHER RESEARCH

Tracing e-mails is a vast and complex subject. This paper has only covered how e-mail evasion can occur and ways to determine the source of the e-mail when such evasion is attempted.

This paper has not discussed forensic examination of servers. This area alone is complex as operating systems, SMTP daemons, and HTTP daemons store files and logs differently. Custom installation and configuration of servers can also complicate the forensic examination of such servers.
Another interesting area for research is packet level inspection of e-mails. This area of research can reveal "fingerprints" or watermarks that specific e-mail client software can embed into e-mails which might help establish the sender of the e-mail.
In depth "Message ID" examinations is also a good area for research. Such research can uncover fake Message IDs as well as uncover patterns that could identify the source of e-mail or at least place an e-mail within a sequence of events on a server that could help determine the exact time it was sent.
This paper has only touched on advanced measures to trace emails as each measure is a suitable topic for research. Such measures include investigating network devices, investigating residual data on servers, bait tactics and software embedded identifiers.

REFERENCES

Akin, T. (2003). *Webmail Forensics*. Retrieved 12/5/2004, from http://opensores.thebunker.net/pub/mirrors/blackhat/presentations/bh-usa-03/bh-us-03-akin.pdf

Francis, N. (2004). *Anatomy of an E-mail Message*. Retrieved 16/5/2004, from http://www.bath.ac.uk/bucs/e-mail/anatomy.shtml

GPSClock. (2003). *GPS Clocks for Computer Time Synchronization*. Retrieved 6/5/2004, from http://www.gpsclock.com/gps/g4.html

JAP. (2003). *JAP Anonymity & Privacy: Anonymity is Not a Crime*. Retrieved 12/5/2004, from http://anon.inf.tu-dresden.de/index_en.html

Jones, H. (2001). *Removing the Mystery from E-mail Tracing*. Retrieved 6/5/2004, from http://ncfs.ucf.edu/E-mail%20Tracing2.ppt

Lewis, E. (2004). *E-mail Attachments 101*. Retrieved 16/5/2004, from http://perl.about.com/library/weekly/aa032302a.htm

LuxSci. (2004). *Additional Privacy with Anonymous SMTP*. Retrieved 13/5/2004, from http://luxsci.com/extranet/info/e-mail-smtp-anon.html

Mandia, K., Prosise, C., & Pepe, M. (2003). *Incident Response and Computer Forensics, Second Edition*. Emeryville, California, U.S.A.: McGraw-Hill Osborne Media.

Nelson, B., Phillips, A., Enfinger, F., Steuart, C., & Phillips, A. (2004). *Guide to Computer Forensics and Investigation*. Boston, Massachusetts, U.S.A.: Course Technology.

SecWiz. (2004). *How to Test Your SMTP Server for Open Relay*. Retrieved 13/5/2004, from http://www.secwiz.com/Default.aspx?tabid=46

Valli, C. (2004). GPS Synchronized Time. In M. Al-Zarouni (Ed.). Perth.

Venit, A. J. (2000). *The Key to Unlocking E-Mail Headers*. Retrieved 6/5/2004, from http://ncfs.ucf.edu/e-mail%20tracing%20SA%20Venit.ppt

Wyble, C. (2002). *Microsoft Office 97 & 2000 Have A Dirty Little Secret*. Retrieved 23 March 2004, from http://advogato.org/article/440.html

COPYRIGHT