International Journal of Engineering Research And Management (IJERM)
ISSN : 2349- 2058, Volume-03, Issue-03, March 2016

Email Threads, Vulnerabilities, Forensic Investigation Techniques and Tools

A. Sesha Rao, Prof. P. S. Avadhani

***Abstract*— E-mail is the most common mode of communication today by every individual. We are more prone to using e-mails rather than communicating on telephone or sending information through printed papers. Emails frequently contain malicious virus's threats and scams that can result in the loss of your data and confidential information, and even identity theft. Hence, Email system is inherently vulnerable to misuse. Email analysis is challenging due to not only various fields that can be forged by hackers or malicious users, but also the flexibility of composing, editing, deleting of emails using offline or email applications. So it is necessary to secure our e-mail system and also to identify criminal collect evidence against them and punish them under court of law. In light of this, Computer Forensic Specialist employ state-of-the-art tools and methodologies in the extraction and analysis of data from spammed emails received at the digital crime scene. This paper studies the comparative approach of the email forensic tools, its origins, its current position and its future directions. This paper also addresses to conduct an investigation into some of these freely available e-mail forensic tools An attempt is made to illustrate e-mail architecture from forensics perspective. It also discusses common e-mail forensic investigation techniques besides the threats and vulnerabilities that encounter in e-mail communication.**

***Index Terms*— Confidential, Forensic, Forged, Malicious, Spam emails, Vulnerable**

I. INTRODUCTION

Email is one of the most common ways people communicate, ranging from internal meeting requests to distribution of documents and general conversation. E-mail communication is a common method of correspondence. Emails are now being used for all sorts of communication including providing confidentiality, authentication, non-repudiation and data integrity. The importance of email is for corporate and private communication can be estimated by the summary presented by Radicati Group's report titled "E-mail Market, 2012-2016" that the world wide each day total emails sent in 2012 was 144.8 billion, which is increased steadily with each passing year and in 2016 approximately 192.2 billion emails will sent each day [1]. The report also states that corporate webmail clients grow from 629 million in 2012 to over one billion by the end of 2016.

As email usage increases, attackers and hackers began to use emails for malicious activities hence email system is inherently vulnerable to misuse. Over a period of years e-mail protocols have been secured through several security extensions and procedures. However, cyber criminals continue to misuse it for illegitimate purposes by sending spam, phishing emails, distributing child pornography, and hate e-mails, besides propagating viruses, worms, hoaxes and Trojan horses. Further, Internet infrastructure is suffering from denial of service, waste of storage space and computational resources are costing every Internet user directly or indirectly [2]. Hence, there is a need to identify and eliminate users and machines misusing e-mail service.

It is also good to know what security protocols are commonly used to secure our communications while using e-mail and up to what extent they are safe. Further what type of threats are facing from the cyber criminals while using e-mail as our mode of communication and how to protect ourselves or what e-mail forensic investigation tools and techniques are used by cyber forensic personnel to put these cyber criminals behind bars and prove their crime in court of law [3].

II. E-MAIL FORENSICS

Computer forensics is a systematic process to retain and analyze saved emails for the purpose of legal proceedings and other civil matters. E-mail forensic analysis is used to send the source and content of e-mail message as evidence, identifying the actual sender, recipient and date and time it was sent etc., to collect credible evidence to bring cyber criminals to justice. E-mail analysis is challenging due to not only various fields that can be forged by hackers or malicious users, but also the flexibility of composing, editing, deleting of emails using offline (e.g. MS Outlook) or online (e.g. Web mail) email applications. Towards this direction, a number of open source forensics tools have been widely used by the practitioners [4].

Forensic analysis of an email message aims at discovering the history of a message and identity of all involved entities. Besides message analysis, e-mail forensics also involves investigation of some client or server computer suspected of being used or misused for e-mail forgery. It may involve inspection of Internet favorite cookies, history, typed URL's, temporary Internet files, auto completion entries, book marks, contacts, preferences, cache, etc. This has led to the need for efficient automated tools in the hands of forensic experts. Several open source software tools have also been developed to perform e-mail header analysis to collect evidence of e-mail fraud [5].

Manuscript received March 22, 2016

A. Sesha Rao, Research Scholar, Department of Computer Science & Engineering, College of Engineering, Andhra University, Visakhapatnam, India

Prof. P. S. Avadhani. Vice- Principal, Department of Computer Science & Engineering, College of Engineering, Andhra University, Visakhapatnam, India

III. E-MAIL HEADER STRUCTURE

E-mails are made of two main parts, which are message header and message body. The header part contains routing information about the e-mail and the other information such as the source and destination of the e-mail, the IP address of the header and date and time related information. The message body contains the actual message of the e-mail i.e. message subject and body. The body may also contain attachments in the form of MIME and S/MIME [6]. Message headers are the important part for investigating e-mail messages and will be discussed in detail in this paper.

E-mail headers are organized from the bottom up. This means that the e-mail was routed 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 post marks 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 [7].

IV. E-MAIL ARCHITECTURE AND WORKING

Use either SI (MKS) or CGS as primary units. (SI units are strongly encouraged.) English units may be used as secondary units (in parentheses). **This applies to papers in data storage.** For example, write "15 Gb/cm^2 (100 Gb/in^2)." An exception is when English units are used as identifiers in trade, such as "3½ in disk drive." Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity in an equation.

The SI unit for magnetic field strength H is A/m. However, if you wish to use units of T, either refer to magnetic flux density B or magnetic field strength symbolized as $\mu_0 H$. Use the center dot to separate compound units, e.g., "A·m^2."

A. E-Mail Components

E-mail system comprises of various hardware and software components that include sender's client and server computers and receiver's client and server computers with required software and services installed on each. It is a highly distributed service that involves several actors which play different roles to accomplish end-to-end e-mail exchange [8]. These actors fall under three groups namely User Actors, Message Handling Service (MHS) Actors, Administrative Management Domain (ADMD) Actors. There are four types of users, Authors, Recipients, Return Handlers and Mediators. The Author is responsible for creating the message, its content and its list of Recipient addresses. The MHS transfer the message from the author and delivers it to the recipients. Recipient is a consumer of the delivered message. Return Handler also called, Bounce Handler', forwards failure messages back to Author. A Mediator receives aggregates, reformulates and redistributes messages among Authors and Recipients who are principals in protected exchanges [5].

Message Handling Service (MHS) Actors performs a single end to end transfer on behalf of the author to reach the recipients addresses. They are originators, relays, gate ways and receivers which are responsible for end to end transfer of messages. These Actors can generate modify or look out only transfer data in the message. Administrative Management Domain Actors can be associated with different organizations, each with its administrative authority. There are three basic types of ADMDs.

Edge: Independent transfer services in networks at the end of the open internet mail service.

Consumer: Might be a type of edge service, as is common for web based e-mail access.

Transit: Mail Service Providers (MSPs) that offer value added capabilities for Edge ADMDs such as aggregation and filtering.

B. Sequence of e-mail message transfer:

Fig.1 shows a typical sequence of events that takes place when sender 'Alice' transmits a message to the recipient 'Bob'.

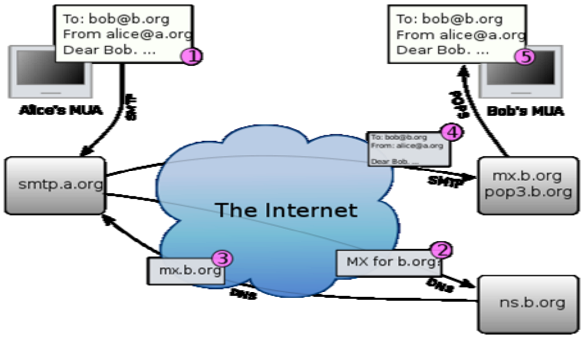

Fig.1: Sender 'Alice' transmits a message to recipient 'Bob'

i. The MUA formats the message in email format and uses the submission protocol, a profile of the Simple Mail Transfer Protocol (SMTP), to send the message to the local mail submission agent (MSA), in this case smtp.a.org.

ii. The MSA determines the destination address provided in the SMTP protocol (not from the message header), in this case bob@b.org. The part before the @ sign is the local part of the address, often the username of the recipient, and the part after the @ sign is a domain name. The MSA resolves a domain name to determine the fully qualified domain name of the mail server in the Domain Name System (DNS).

iii. The DNS server for the domain b.org (ns.b.org) responds with any MX records listing the mail exchange servers for that domain, in this case mx.b.org, a message transfer agent (MTA) server run by the recipient's ISP.

iv. 'smtp.a.org' sends the message to mx.b.org using SMTP. This server may need to forward the message to other MTAs before the message reaches the final message delivery agent (MDA).

v. The MDA delivers it to the mailbox of user bob.

vi. Bob's MUA picks up the message using either the Post Office Protocol (POP3) or the Internet Message Access Protocol (IMAP)

C. Working of E-mail:

E-mail relies on two basic communications protocols Simple Mail Transfer Protocol which is used to send messages and Post Office Protocol which is used to receive messages. A simplified version of the email life cycle is shown in fig. 1 [9].

The most important logical elements of the Internet Mail system are:

i. Mail User Agent (MUA): It is responsible for the helping the user to read and write e-mail messages. The MUA is usually implemented in software referred to as "email client". Popular email clients are Microsoft Outlook2, and Mozilla Thunderbird3, Claws mail, Zimbra Collaboration suite, etc. These programs transform a text message into the appropriate Internet format in order for the message to reach its destination.

ii. Mail Transfer Agent (MTA): It accepts a message passed to it by either an MUA or another MTA and then decides for the appropriate delivery method and the route that the mail should follow. It uses the SMTP protocol to send the message to another MTA or an MDA.

iii. Mail Delivery Agent (MDA): It receives messages from MTAs and delivers them to the users mail box in the user's mail server.

iv. Mail Retrieval Agents (MRA): It fetches mail messages from the user's mail server to the user's local in box. MRAs are often embedded in e-mail clients.[10]

D. What an e-mail header contains?

Internet e-mail has two parts header and body.

i. Header: The message header contains control information, an originator's e-mail address and one or more recipient addresses and descriptive information such as a subject header field, and a message submission date/time stamp which is structured in to fields such as From, To, CC, Subject, Date, and other information about the email.

ii. Body: Body content is unstructured text which sometimes contains a signature block at the end. The header is separated from the body by a blank line.

Each message has exactly one header, which is structured into fields. Each field has a name and a value. RFC 5322 specifies the precise syntax which contains only US-ASCII characters. The message header must include at least the following fields [9].

From: The email address and optionally the name of the author(s).

Date: The local time and date when message was written.

Message ID: Also an automatically generated fields; used to prevent multiple deliveries and for reference in In-Reply-To.

In-Reply-To: Used to link related messages together. This field only applies for reply messages.

To: The e-mail address(s), and optionally name(s) of the message's recipient(s).Indicates primary recipient.

Subject: A brief summary of the topic of the message.

Bcc: Blind Carbon Copy; addresses added to the SMTP delivery list but not listed in the message data, remaining invisible to other recipients.

CC: Carbon copy; generally same as 'To' field, a 'To' field specifies primary recipient.

Content Type: Information about how the message is to be displayed, usually a MIME type.

Reply To: Address that should be used to reply to the message.

Sender: Address of the actual sender acting on behalf of the author listed in the From: field.

Received: When an SMTP accepts a message it inserts this trace record at the top of the header (last to first).

Return path: Contains the address recoded by MDA from MailFrom SMTP command.

MailFrom: It is a string containing e-mail address for receiving return control information like returned messages transfer level problems.

Other header fields that are added on top of the header by the receiving server may be called trace fields, in broader sense.

Authentication results: When a server carries out authentication checks, it can save the results in this field for consumption by down-stream agents.

Received–SPF: Senders Policy Framework; It stores results of SPF checks in more detail than authentication results.

Auto-submitted: It is used to mark automatically generated messages.

References: Identifies other documents related to this message, such as other e-mail message.

Resent Message ID: Globally unique message identification string generated when it is resent.

List ID: It is globally unique mailing list identification string.

MIME Version: It describes the version of the MIME message format.

Archived-At: A direct link to the archived form of an individual email message.

V. Threats and Vulnerabilities in E-mail Communication

The need for e-mail forensic investigation is as there are cyber criminals spoof email messages to carry out various illegitimate activities through e-mail system and remain underground to evade any possible legal action against them. These include:

i. Abuses like spamming, phishing, cyber bullying, child pornography, sexual harassment, racial verification, etc.

ii. Misuse by transmitting viruses, worms, Trojan horses, hoaxes and other malicious programs with an intent to spread them over internet, and

iii. Carry out internet infrastructure crimes through denial of services and directory harvesting attacks [11]. This injudicious use of e-mail cause many technological problems like misuse of storage space, wastage of computational resources, and network conjunction. Senders can lie about their true identities in various ways by using or misusing different techniques that include:

A. E-mail Spoofing [12]:

E-mail spoofing is the forgery of an e-mail header so that the message appears to have originated from someone or somewhere other than the actual source. Spoofing can be used legitimately, distributors of spam often use spoofing in an attempt to get recipients to open, and possibly even respond to their solicitations.

B. Phishing [13]:

Phishing is a form of fraud in which the attacker tries to learn information such as login credentials or account information by masquerading as a reputable entity or person in e-mail. These e-mails entice you to click on some link present in e-mail or to open some attachment or respond to some message and that click directed to you their site in actual but it appears like your trusted website of bank and ask to fill some confidential information like passwords.

C. Email spamming [15]:

Spam email is any email that was not requested by a user but was sent to that user and many others, typically with malicious intent. Email spam is any email that meets the following three criteria. Anonymity: the address and identity of the sender are concealed. Mass Mailing: The email is sent to large group of people. Unsolicited: The mail is not requested by the recipient.

D. *Eavesdropping [14]:*

Eavesdropping refer to the unauthorized monitoring of other people communications. Email Eavesdropping usually happen if the sender has not encrypted the email message and has not used digital signature the attacker can extract security loop holes on the network to launch a man-in-the middle attack. He may intercept and deface the message before sending it to the recipient.

E. *Repudiation [3]:*

As it is known that e-mail messages can easily be forged so any one sending you some message can later on deny regarding sending of message and it is very difficult to prove it. This has implications corresponding to e-mails use as contracts in business communications.

F. *E-mail Bombing [14]:*

An email bombing is a form of net abuse consisting of sending huge volumes of emails to an address in an attempt to overflow the mailbox or overwhelm the server where the email address is hosted in a denial of service attack. There are three methods of perpetrating an email bomb: mass mailing, list linking and zip bombing. Mass mailing consists of sending memories duplicate mails to the same email address.

G. Bot-Networks [22]:

The term bot, derived from "ro-BoT Network". A Bot-Network is a large number of compromised computers that are used to generate spam relay viruses or flood a network or web server with excessive requests to launch huge numbers of denial of service attacks. Botnets today are often controlled using Internet relay Chat. The computer is compromised via a Trojan

H. *Open Mail Relays [5]:*

An open mail relay is a miss-configured mail relay that accepts mail from any computer and forwards it to another computer which otherwise should have accepted mail for and from specific computers such a relay becomes vulnerable to spammers and phishers who hide their identities behind these relays.

I. Unprotected Backups [3]:

Messages generally stored in plain text on SMPT server and also backups can be created. Even if you delete the message, they can be residing on the servers/backup servers for years. So anyone who accesses these servers can also access or read your message

J. Open Proxy [25]:

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. An open proxy server does not maintain a strict log of user activities unlike others which maintain user logs synchronized with reliable time servers. Spammers use open proxies to hide their network addresses. The recipient of a spammer's email will not see the spammer's network address on the email but the open proxy's network address.

K. *Identity Theft:*

Means someone pretend to be you on the network. If not proper security protocols are followed then someone may steal or capture your username/password and used to read your e-mail message from your account without your knowledge.

L. *Email Fraud:*

Email Fraud is the international deception made for some personal or monetary gain.

VI. E-mail Forensics Analysis and Investigation Techniques

A forensic investigation of e-mail can examine both email header and body. According to Marwan [23] an investigation should have the following:

i). Examining sender's e-mail address.
ii). Examining message initiation protocol (HTTP, SMTP).
iii).Examining message ID.
iv). Examining senders IP address.

Some other aspects that controls forensics step include the following properties:

i. Storage format email: Server side storage format may include maildir (each mail is kept in a separate file, for each user), mbox format (all email files are in a single text file). Server-side stores email in SQL Server databases. Reading different types of formats can be done forensics analysis by using note pad editor and applying regular expression–based searches [24]. At the client-side, an email is stored as mbox format (Thunderbird) . Client side may also store emails as .PST (MS Outlook), and NSF (Lotus Notes) files.

ii. Availability of backup copy of email: When checking from the server side, all copies are transferred to the client. This requires seizing the client computer. For Webmail, copies are always saved at the server side [24].

iii. Protocol used to transport email: Email can be initiated and transported based on SMTP or HTTP [23] depending on the email server applications. In fact, Emails frequently contain malicious viruses, threats and spam that can result in the data interception, loss of data, and confidential information theft.

E-mail forensic investigation refers to the study of source and content of e-mail as evidence, the identification of the actual sender and recipient of a message, the date/time it was sent, detailed record of email transaction, intend of sender, etc. This study involves investigating of Meta data, keyword searching, port scanning etc., for authorship attribution and identification of email scams [2]. Various approaches that are adopted for e-mail forensic investigation are briefly defined below.

A. *Header Investigation [16] [17*

It is the most common and popular technique to analyze the useful hidden personal information in e-mail. As the message travels through the communications network, an abbreviated record of the email's journey is recorded in an area of message called the header. As is the message is routed through one or more mail servers, each server adds its own information to the message header. The investigator may be able to identify Internet Protocol (IP) address from the header and use this information to determine the sender of the message. The journey of the message can usually be reconstructed by reading the e-mail header from bottom to top. This is because as the message passes through additional mail servers, the mail server will add its information on the top of the previous information in the header.

So, E-mail header is important for investigation and collection of evidence. Meta data present in the e-mail Header serves as control information. This contains information about sender/receiver and the path followed by message to reach destination. So this is very crucial information from evidence point of view. Sometimes this Meta data/control information is altered or spoofed. So in header analysis, authentication of the information present in the head2er is also checked [3].

In fact there are many fields and provide lots of information regarding e-mail, sender's IP address, return path, message-id, signature field, server transit or path follows for transmission, MIME and other security protocol information. The Received field shows date and time when e-mail is arrived at server. The From and To tells about sender and receiver. So, lots of useful information is there for analysis. E-mail header analysis is used to collect crucial evidence to prove crime in court of law.

In order to determine the authenticity of e-mail evidence,[16] says that methods to forge email are many, so it is mandatory to summarize the most common forms of forgery and analyze its implementation for identification.

B. *Server Investigation [2]:*

In this investigation, copies of delivered e-mails and server logs are investigated to identify source of an e-mail message. E-mails purged from the clients (senders or receivers) whose recovery is impossible may be requested from servers (proxy or ISP) as most of them store a copy of all e-mails after their deliveries. Further, logs are maintained by servers and they can be helpful for tracing the computer/server from where transaction takes place. When we are creating an e-mail account then we are passing some information directly like name, phone number etc., which are helpful in finding the address. If the information provided is fake then still server can store information like your location/area, etc., that can be helpful in investigation.

However, servers store the copies of e-mail and server logs only for some limited periods and some may not cooperate with the investigators. Further, SMTP servers which store data like credit card number and other data pertaining to owner of a mail box can be used to identify person behind an e-mail address.

C. *Network Device Investigation [19]:*

To operate properly network devices need to maintain information about the network traffic they process. Since network devices have limited amount of memory, they tend to collect only the bare essential information for their process and this information is discarded rather quickly when it is no longer needed. There is often a suffic

ient delay between the time a security incident occurs and the time the forensic investigation starts. And as a switch or router discards obsolete Meta data quickly you will not find forensic evidence if you react too late.

But you can improve the success rate of your forensic evidence gathering by configuring your switches and routers to collect additional data to persist this data. All professional network devices allow for the logging of events. But internal event of network device is rather small because of the memory constraints. Old events get discarded at a fast rate to make place for new events. Network devices can become compromised because their configuration gets modified or because their operating system gets Trojanised. Finding forensic evidence for these incidents can become much harder.

To detect unauthorized configuration modifications, to release management and version control process is necessary. The release management process will make sure that only approved modifications are applied to your network devices, and the version control process will make sure that these modifications are documented. Periodic review of your network device configuration will allow you to detect unauthorized configuration modifications by comparing them with the configuration kept in the version control system. This review process can be automated.

D. *Bait Tactics [2]:*

The basic aim of the bait tactics is to extract the IP address of the culprit. In this technique, an email with http:tag which has some image source at a computer that is monitored by investigators is sent to the email address that under investigation. Now the recipient is the one who originally was sender during the crime. When the email is opened, a log entry which contains the IP address of the recipient is recorded on the server which is hosting the image and the recipient is tracked. In a case when the recipient is using a proxy, server, then the IP address of the proxy server is recorded by the investigators. The log of the proxy server is used to track the culprit. In a case when the recipient is using the proxy server, then the IP address of the proxy server is recorded by the investigators. The log of the proxy server is used to track the culprit. In case, the logs of the proxy server are not available, then a tactic email is sent to the culprit. The tactic email can either be a HTML page or an embedded java applet.

E. *Investigating residual data on servers [18]:*

SMTP servers usually keep a copy of all emails even after they are delivered. They only delete the data after a backup operation is performed. By using this information, we can trace the address of the computer that made the connection. By analyzing 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.

F. *Information Recovery* [20]:

Information recovery is one of the important processes in digital forensics. Evidence of a crime may reside in a deleted email. As mentioned by (John show et al) the solution to recover the deleted email is by expose reverse engineering on how email was deleted. The email may be manually deleted by the user or criminal to remove the evidences, or it might happen in some cases that the database is corrupted or hacked by unauthorized personal. To recover the removed email, the recovery process will be performed from vendor site that can provide a backup plan, which is called as Cache Exchange mode. The data must be synchronized to the server. The deleted email remain as deleted mail until someone tries to "up" it back, when the machines or servers get synchronize together.

The second method is using parents and orphanage method which recover the email based on the type of file or software used, for example using Microsoft Outlook. The email can be recovered by looking at history of email for a particular period. This is due to the Windows method that delete the email depends on the "flag".

G. Investigation of Anti-Forensic Approaches:[21]:

Sometimes culprits also use anti-forensic techniques which are used to counter cyber forensic investigation. As widely marketed, the forensic software may lead to defenseless state which might expose the collected information to the third party. The level of vulnerabilities is unlimited and can be exploited through software architecture, type of file, level of patching, etc. Thus, it is crucial to practice the administrative and authentication policy in IT system.

H. Sender Mailer Finger Prints [2]:

Identification of software handling e-mail at server can be revealed from the Received header field and the identification of the software handling e-mail at client can be ascertained by using different set of headers like "X-Mailer" or equivalent. These headers describe applications and their versions used at the clients to send e-mail. This information about the client computer of the sender can be used to help investigators devise an effective plan and thus prove to be very useful.

VII. E-mail Forensic Tools

Various software tools have been developed to assist e-mail forensic investigation, some are paid, free or open source. Most of the e-mail forensic tools assist in the study of source and content of email message so that an attack or the intent of the fraud email may be investigated. These tools while providing easy to use browser format, automated reports, and other features, help to identify the geographical location of the sender; trace the route traversed by the email, identify spam and phishing networks, etc This section introduces most of the available e-mail forensic tools.

A. eMailTrackerPro.

eMmailTrackerPro [26] is easy to use for analyzing email headers to disclose the original sender's or spammer's location. It takes e-mail headers and analyzes them to eventually present the sender's geographic location and identifies the network provider (or ISP) and provides contact information for further investigation. The actual path from the recipient to the sender's IP address is reported in a routing table, providing additional location information to help determine the sender's true location. Essentially the tool has four specific functions to perform.

i. Spam filtering: With a powerful spam filtering technology, the application takes control of spam e-mail that may invade your mail box, working with Hotmail, Gmail and other email clients. The spam filter scans each email as it arrives and warns the user if it is suspected spam. Essentially stopping spam e-mail before it reaches its intended recipient.

ii. E-mail origin: Just with e-mail header eMailTrackerPro can find out any email back to its true geographical location. In fact, thanks to the email header, it is possible to know key details about where an e-mail comes from.

iii. Abuse reporting: The major feature of the tool apart from filtering spam and searching for the email origin is abuse reporting. eMailTrackerPro also makes it simple to send abuse reports to the network that is responsible for phishing email and spam. ISP can then takes steps to prosecuting the account holder and help put a stop to spam.

iv. Reports: eMailTrackerPro can save all the prior traces to make it easy to check back and review a trace. The tool supports, Russian, Chinese and Japanese language spam filters besides English language.

B. Aid4Mail Forensic

Aid4Mail Forensic [27] is a fast and highly accurate mail conversion program covering three main areas of expertise, email migration, email discovery and email archiving. The tool is available in four different editions: Home, Professional, e-Discovery/Forensic and Console. All the editions can export mail to a non-proprietary highly compressed ZIP file format or index able EML files. Aid4Mail e-Discovery also offers the option to convert messages to PDF/A, an important electronic filing format used for long term preservation of case related documents.

Aid4Mail migrate mail without losing formatting, sender/recipient information, attachments, embedded contents or message status. Aid4Mail can process mail folders and files even when they are disconnected from their e-mail clients, including those stored on a CD or DVD or USB drives. It can filter mail by date range, and by keywords in the message body or in the headers. The tool easily transfers messages between email apps and web based services, for example, from Thunderbird to Outlook or Eudora to office 365, or Yahoo to Gmail. Aid4Mail supports over forty e-mail client formats and mail client programs, as well as many popular webmail services and remote accounts through IMAP.

C. EmailTracer.

EmailTracer [28] is able to analyze the email header and give the complete details of the sender including IP address, which is the key point to find the culprit. It gives the geographical location of the sender and the detected route traced by the email etc. It can also be used for retrieving information from mailbox files with extensions .dbx (Outlook Express); .pst (Microsoft Outlook); .mbx (Eudora), .cnm (Pegasus), .imm (IncrediMail), MailDir (K mail), .tbb(The Bat), .nsm (Netscope message) and .mbox (Mozilla). EmailTracer traces up to Internet Service Provider (ISP) level only. Further tracing can be done with the help of ISP and law enforcement agencies. The Message-id will be useful for analyzing the mail logs at ISP. It has the facility to extract and save the attachment(s) of mail as .eml.

It generates detailed report about the sender of the mail. The report contains: City and country name of the senders mail server; IP address of the sender; Message id of the mail, which will be unique for every mail; Path traced by the mail; Mail-id of sender; Sender's name; Sender's mail server; Subject of the mail.

Mail server log analysis is for evidence collection, creating report after the analysis. Able to process mail header from Yahoo, Hotmail, etc. It helps on how to extract email header from different email clients also.

D. Abuse pipe

Abuse pipe [29] works on the platform windows 8 and windows XP. Abuse pipe identifies abuse complaint emails (typically sent by irate users to abuse@yourISP.com) and determines which of your customers is sending spam. It automatically generates reports showing which customers have violated ESP's acceptable user policy, and you can verify quickly and take action to shut them down. Abuse requires no technical knowledge to operate it. It takes the place of the technical staff that you currently have investigating spam complaints. Abuse pipe is only system that deals with abuse complaints, and helps ISPs in their fight against spam from within their own networks. It can assist in meeting legal obligations such as reporting on the customers connected to a given IP address at a given date and time.

E. Adcomplain

Adcomplain [30] is a tool for reporting appropriate commercial email and use net postings, as well as chain letters and "make money fast" postings. It runs under UNIX, Windows NT, and Windows 95 and does not currently run on Macintosh systems. It automatically analyses the message, composes an abuse report and mails the report to the offender's Internet Service Provider (ISP) by performing a valid header analysis. A third party forwarding service called Abuse.net is used for complaints to the offender's service provider. Adcomplain may detect mail forgeries or other problems and asks the user to confirm by pressing return. It will then compose the message and display for approval before mailing. The user has opportunities to edit or list the message, or just to abort. Adcomplain can be invoked from the command line or automatically from many news and mail readers.

Syntax: Ad complain [-b] [-c] [-f template] [-m] [-O out name] [-P in name] [-q] [-s] [-u] [-v] [news group]

F. AccessData's FTK

AccessData's FTK [31] is a court cited .digital investigations platform built for speed, stability and ease of use. FTK's database-driven, enterprise class architecture allows the user to handle massive data sets, as it provides stability and processing speeds not possible with other tools. Further more because of its architecture, FTK can be set up for distributed processing and incorporate web based case management and collaborative analysis. FTK scans a hard drive looking for various types of information. It can for example locate deleted emails and scan a disk for text strings to use them as password dictionary to crack encryption. The tool kit also includes a standalone disk imaging called FTK imager. The FTK imager is simple but concise tool. It saves an image of a hard disk in one file or in segments that may be later on reconstructed. It calculates MD5 hash values and confirms the integrity of the data before closing the files. The result is an image file (s) that can be saved in several formats including DD raw. It supports popular encryption technologies, such as Credant, Safe Guard Enterprise and Easy EFS, PGP, Guardian Edge, Point Sec, S/MIME. To that matter FTK is able to decrypt almost as many file formats as PRTK. Users can also import password lists to decrypt files during the processing phase. Its current supported email types are Lotus Notes NSF, Outlook PST/OST, Exchange EOB, Outlook Express DBX, Eudora, EML (Microsoft Internet mail, Earth link, Thunderbird, Quick mail, etc), Netscape, AOL and RFC 833.

It also process and analyze MG (compressed and uncompressed) Ext4, exFAT, VxFS (Veritas file system) Microsoft VHD (Microsoft Virtual hard disk) Black berry, IPD back up files, Android YAFFS/YAFFS2 and many more. Creates and process Advanced Forensic Format (AFF) images also.

G. EnCase Forensic

EnCase Forensic [32] is the shared technology within a suite of digital investigations products by guidance software. The software comes in several products designed for forensic, cyber security analytics, and e-discovery use. EnCase contains tools for several areas of the digital forensic process, acquisition, analysis and reporting. EnCase contains functionality to create forensic images of suspect media. Images are stored in proprietary EnCase evidence file format. The tool acquire data from disk or RAM, documents, images, e-mail, Webmail, Internet artifacts, web history and cache, HTML page reconstruction, Chat sessions, compressed files, back up files, encrypted files, RAIDs, Workstations, servers, and with version 7: smart phones and tablets. Encase forensic produces an exact binary duplicate of the original drive or media, then verifies it by generating MD5 hash values for related image files and assigning CRC values to the data. These checks and balances reveal when evidence has been tampered with or altered, helping to keep all digital evidence forensically sound for use in Court Proceedings or internal investigations. Advanced analysis recover files and partitions, detect deleted files, by parsing event logs, file signature analysis, and hash analysis, even within a computer files or unallocated disk space.

The EnCase forensic tool includes Instant Messenger tool kit for Microsoft Internet Explorer, Mozilla Firefox, Opera and Apple Safari. The e-mail support includes for Outlook PSTs/OSTs, Outlook Express DBXs, Microsoft Exchange EDB parser, Lotus Notes, AOL, Yahoo, Hotmail, Netscape Mail and MBox archives'.

H. . FINALeMAIL

FINALeMAIL [33] can recover the email database file and locates lost emails that do not have data location information associated with them. It has the capability of restoring lost emails to their original state, recover full email database files even when such files are attacked by viruses or damaged by accidental formatting. It can recover email messages and attachments emptied from the 'Deleted Items Folder' in Microsoft Outlook Express, Netscape Mail, and Eudora. It has the capability of restoring lost emails to their original state.

I. Sawmill – GroupWise.

Sawmill-GroupWise [34] is a Group Wise Post Office Agent log analyzer which can process log files in group wise Post Office Agent format, and generate dynamic statistics from them, analyzing and reporting events. Sawmill is a powerful hierarchical analysis tool that runs on every major platform. It can process almost any type of data. The reports that Sawmill generates are hierarchical, alternative and heavily cross-linked for easy navigation. Sawmill stores log statistics in an optimized database, which can be Sawmill's own built in high performance database, or MySQL database.

It supports Windows, Linux, Free BSD, Open BSD, Mac OS, Solaris, other UNIX and several other platforms. It also supports Role Based Authentication Control.

J. Forensics Investigation Tool Kit (FIT)

Forensics Investigation Tool Kit (FIT) [35] is windows-based content Forensics Tool Kit to read and analyze the content of the Internet raw data in Packet CAPture (PCAP) format. FIT comes with Graphic User Interface. All protocols and services analyzed and reconstructed are displayed in a readable format for the users. The other unique feature of FIT is that imported raw data files will be immediately parsed and reconstructed. Raw data file (PCAP) captured from sources like LAN and WLAN networks can be imported to the FIT accordingly. Imported raw data files will be parsed and the output content will be displayed in intended Service Categories. It supports case management functions, detailed information including Date-Time, Source IP, Destination IP, Source MAC etc. It fully supports search function (Full Text Search) and Book Mark functions. Another product feature is analyzing and reconstruction of various Internet traffic types which includes email (POP3, SMTP, IMAP), Webmail (read and Sent) IM or Chat (MSN, ICQ, Yahoo, QQ, Skype Voice Call, Log, UT chat ROOM, G-talk, IRC Chat Room), File Transfer (FTP, DZP), Telnet, HTTP (content, upload/download, Video Streaming, Request and others (SSL).

K. Paraben (Network) E-mail Examiner

Paraben(Network) E-mail Examiner [36] is an advanced personal email archive analysis and conversion tool. It forensically examines hundreds of email formats including Outlook (PST and OST), Thunderbird, Outlook Express, Windows mail and more. Email Examiner allows you to analyze message headers, bodies and attachments. Email Examiner does not just recover email in the deleted folders; it recovers email deleted from deleted items. Supports advanced searching, reporting and exporting to PST and other formats.

The most important supported emails are America on-Line (AOL), Microsoft Outlook (PST), and Microsoft Outlook Offline Storage (OST). The Bat (Version 3.x and higher), Thunderbird, Outlook Express, Eudora, E-mail file (EML), Windows main databases, support for more than 750 MIME types and related file extensions, and plain text mail.

L. Forensic E-mail Analysis Tool

Forensic E-mail Analysis Tool has been designed by highly trained and technically sound professional. Hence, tool offers all the required techniques and accuracy of analyzing an email message with preciseness. The other features are:

- Recover PST, EDB, OST, OLM, MBOX, TBB, MBX and other email forensic investigation.
- Performs deep scanning of corrupted email.
- Examines e-mail evidence in Hex Code.
- Provide email and attachment forensic examination platform.
- Shows complete email header information if an email.
- Many modules of investigation can be divided in order to do the group analysis.

It exports email evidence into HTML, EML, MSG, or PDF file type

CONCLUSION

Overall contribution of this paper is an exhaustive survey of the several tools and techniques available to conduct email forensics. All the tools surveyed in this paper are free to use, at least available for trials.

Email Forensics deals with the investigation of content of email messages to identify the actual sender, recipient, date/time when it was sent, etc. Various software tools which essentially perform automated header analysis and network device inspections have been developed to assist speedy investigations.

According to Simson L. Garfinkle [8] current forensic tools are designed to help examiners in finding specific pieces of evidence and are not assisting in investigations.

Today's tools are created for solving crimes committed against people where the evidence resides on a computer, they were not created. Our analysis among the tools compared shows that Aid4Mail can analyze emails stored both in hard disk (Offline analysis) and on remote email servers (Online analysis). It has highest amount of capability to gather information than other tools. The recovery capability of Aid4Mail and Paraben E-mail Examiner appears better than the other tools since then can recover emails from delete folder. Of course E-Mail Examiner supports most of the familiar email formats and 750 MIME content types. In general the more coverage a tool has, the better it can be suitable to address various types of forensic activities and legal procedures.

It is a known fact that the security and forensic personnel need to keep up pace with the latest attack tools and techniques adopted by the attackers. With freely available tools, one can enforce the security mechanisms and analyze attack traffic only to a certain extent. To detect all kinds of attacks and conduct a comprehensive email forensic analysis, one would have to deploy and analyze the effectiveness of commercial tools.

REFERENCES

[1] Radicati Group, Inc., "Email Market, 2012-2016," http://www.radicati.com/wp/wp-content/uploads/2012/04/Email-Statistics-Report-2012-2016-Executive-Summary.pdf, October 2012.

[2] M. Tariq Banday, "Technique and Tools for Forensic Investigation of E-Mail", International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.6, November 2011

[3] Dilpreet Singh Bajwa et al , "Review of E-mail System, Security Protocols and Email Forensics", International Journal of Computer Science & Communication Networks, 2015, Vol 5(3),201-211

[4] V. K. Devendran, Hossain Shahriar, and Victor Clincy, "A Comparative Study of Email Forensic Tools", Journal of Information Security, April 2015, 6, pp. 111-117

[5] M. Tariq Banday, "Technology Corner Analysing E-Mail Headers for Forensic Investigation", *Journal of Digital Forensics, Security and Law, 2012, Vol. 6(2)*

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

*[7]*Venit, A. J. (2000), *The Key to Unlocking E-Mail Headers*. Retrieved from: http://ncfs.ucf.edu/email%20tracing%20SA%20Venit.ppt Simsom L. Garfinkel,(2010),"Digital Forensics Research:The next 10 years ", Digital Invetigation,Vol 7, pp 64-73; available at www.sciencedirect.com.

[8] Mrs. Pranjal S. Bogawar, Dr. Kishor. K. Bhoyar,"Email Mining: A Review" , IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012

[9] Ioannis Katakis, Grigorios Tsoumakas, Ioannis Vlahavas,"Email Mining: Emerging Techniques for Email Management", 2006, pp- 1-32

[10] Hastings, N. E, McLean, P. A., (1996), "TCP/IP spoofing fundamentals", In Proceedings of the IEEE 15th Annual International Phoenix Conference; 1996 .pp. 218-224.

[11] Radvanovsky, B. (2006),"Analyzing spoofed e-mail header", Journal of Digital Forensic Practice, 1(3), 2006, pp. 231-243.

[12] Gori Mohamed .J, M. Mohammed Mohideen, Mrs.Shahira Banu. N, "E-Mail Phishing-An Open Threat to Everyone", International Journal of Scientific and Research Publications, Vol-4, No.-2, Feb-2014.

[13] Olalekan Adeyinka, "Internet Attack Methods and Internet Security Technology", Second Asia International Conference on Modeling & Simulation, May-2008.

[14] Jitender Nath Srivastva, Maringati Hima Bindu, "E-Mail Spam Filtering using Adaptive Genetic Algorithm", I. J Intelligent System and Applications, pp-54-60, January 2014.

[15] Hong Guo, Bo Jin, Wei Qian, "Analysis of Email Header for Forensic purpose",International Conference on Communication Systems and Network Technologies", Computer Society, IEEE, 2013.

[16] Satheesaan Pasupatheeswaran, "Email Message-IDs' helpful for forensic analysis", Proceedings of the 6th Australian Digital Forensics Conference, Edith Cowan

[17] University,Perth Western Australia, December3rd2008, http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1048&context=adf.

[18] Nelson, B. et.al,(2004),"Guide to Computer Forensics and Investigation", Boston, Massachusetts,U.S.A: Course Tecnology. Society, IEEE, 2009.

[19] Wang WenQi, Liu WeiGuang, "The Research on Email Forensics Based Network", Ist International Conference on Information Science and Enginnering" , Computer Society, IEEE, 2009.

[20] Farhood Norouzizadeh Dezfoli, et al, " Digital Forensic Trends and Future", International Journal of Cyber-Security and Digital Forensics (IJCSDF), 2013,2(2): 48-76, The Society of Digital Information and Wireless Communications.

[21] Anu Jain, Gurpal Singh Chhabra, "Anti Forensics Techniques: An analytical Review", Seventh International Conference on \Contemporary Computing (IC3), IEEE, Aug – 2014.

[22] Ickin Vural, HS Venter, "Investigating Identity Concealing and Email Tracing Techniques", Information and Computer Security Architecture Research Group (ICSA), Dept. of Computer Science ,University of Pretoria.

[23] Marwan A.Z. (2004) Tracing E-mail Headers. *Proceedings of Australian Computer, Network & Information Forensics Conference*, November 2004, School of Computer and Information Science, Edith Cowan University Western Australia,16-30.

[24] Conan Albrecht, Email Analysis. http://www.gsaig.gov/assets/File/other-documents/Forensics- EmailAnalysis.pptx.pdf

[25] Boneh, Dan. 2004. " The Difficulties of Tracing Spam Email," Department of Computer Science Stanford University.Forensic E-mail Investigation tools:

[26] eMailTrackerPro, http://www.emailtrackerpro.com/

[27] Aid4Mail Forensic, http://www.aid4mail.com/

[28] EmailTracer, http://www.cyberforensics.in

[29] AbusePipe, http://www.datamystic.com/abusepipe.html

[30] dcomplain, http://www.rdrop.com/users/billmc/adcomplain.html

[31] AccessData's FTK, http://www.accessdata.com/

[32] EnCase Forensic, http://www.guidancesoftware.com

[33] FINALeMAIL, http://finaldata2.com

[34] Sawmill-GroupWise, http://www.sawmill.net

[35] Forensics Investigation Toolkit (FIT), http://www.edecision4u.com/FIT.html

[36] Paraben (Network) E-mail Examiner, http://www.paraben.com/email- examiner.html

A.Sesha Rao, is a research scholar in the department of Computer Science and Engineering, College of Engineering, Andhra University, Visakhapatnam. His research interests are related but not limited to cryptography, security, privacy and email forensics. In 1972, he received M.Sc. (Applied Maths) from Andhra University, Waltair, M.Tech and in Computer Science & Engineering, from IIT, Mumbai, 1985. After that he worked as Scientist for 25 years in Naval Science and Technological Lab, DRDO, Visakhapatnam, presently he is working as Professor in Vignan's Institute of Engineering for Women, Visakhpapatnam.

Dr.P.S. Avadhani is a Professor in the department of Computer Science and Engineering of Andhra University. He has guided 11 Ph.D. students, 3 students already submitted and right now he is guiding 12 Ph.D. Scholars from various institutes. He has guided more than 100 M.Tech Projects. He received many honors and he has been the member for many expert committees, member of Board of Studies for various universities, Resource persons for various organizations. He has co-authored 4 books. He is a Life Member in CSI, AMTI, ISIAM, ISTE, YHAI and in the International Society on Education Technology. He is also a Member of IEEE, and a Member in AICTE.