On the Realization of E-mail Forensic using Meta-Data

M. Tariq Banday, University of Kashmir

Available at: https://works.bepress.com/banday/18/
On the Realization of E-mail Forensic using Meta-Data

M. Tariq Banday
Department of Electronics & Instrumentation Technology
University of Kashmir, Hazratbal, Srinagar
E-mail: sgrmtb@yahoo.com

Abstract—Cyber forensic e-mail analysis is employed to collect credible evidence to bring criminals to justice. This paper discusses various e-mail security issues and actors involved in e-mail system. It discusses the need for e-mail forensics and discusses how it can be performed using exhaustive header analysis of e-mail message. It also presents some e-mail forensic analysis software tools developed to extract credible evidences from e-mail messages.

Keywords—Header Analysis, E-mail Security, Forensics, Forensic Tools, E-mail Evidence.

I. INTRODUCTION

With the advent of World Wide Web, e-mail has emerged as one of the most widely used tool for communication over the Internet. The e-mail traffic on the Internet has increased manifold. It is used for extensive conversations between parties and for delivery of business letters and government orders. Further, it is also used for carrying out financial transactions and distribution of documents and archives of diverse natures. The primary protocols used in e-mail system lack security features for privacy of the sender, authentication of sender, integrity of e-mail message, non-repudiation by sender and consistency of e-mail envelope. These security lapses permit e-mail to be spoofed by forging its headers or by sending it anonymously. Further, its capability to transmit executable files and other similar contents using Multi-Purpose Internet Mail Extensions (MIME) and e-mail access at public places where anonymous use is easy has increased its vulnerability. Owing to e-mail’s inherent vulnerabilities, it is being misused by cybercriminals for several illegitimate purposes that include abuses like spamming, phishing, cyber bullying, child pornography, sexual harassment, racial vilification, etc. It is also misused for transmitting viruses, worms, Trojan horses, hoaxes, and other malicious programs with intent to spread them over the Internet. It is wrongly used for carrying Internet infrastructure crimes through Denial of Services and Directory Harvesting Attacks. These injudiciously use of e-mail cause many technological problems like misuse of storage space, wastage of computational resources, and network conjunction besides creating severe legal issues. Technological means like firewalls, anti-viruses, scanners and filters to prevent e-mail abuses are insufficient to make e-mail communications secure. Cyber forensic e-mail analysis is employed to collect credible evidence to bring criminals to justice. E-mail forensics refers to studying the source and content of e-mail message as evidence, identifying the actual sender, recipient and date and time it was sent, etc. An attempt is made in this paper to analysing meta-data contained in headers of e-mail messages from forensics point of view. It also presents some e-mail forensic analysis software tools developed to extract credible evidences from e-mail messages.

II. E-MAIL ACTORS AND HEADERS

E-mail is a highly distributed service involving several actors that play different roles to accomplish end-to-end mail exchange [1]. These actors fall under “User Actors”, “Message Handling Service (MHS) Actors” and “ADministrative Management Domain (ADMD) Actors” groups. User Actors are people, organizations or processes that serve as sources or sinks of messages. They can generate, modify or look at the whole message. User Actors can be of following four types. All types of Mediator user actors set HELO/EHLO, ENVId, RepTo and Received fields. Alias actors also typically change To/CC/BCC and MailFrom fields. Identities relevant to ReSender are: From, Reply-To, Sender, To/CC/BCC. Resent-From, Resent-Sender, Resent-To/CC/BCC and MailFrom fields. Identities relevant to Mailing List processor are: List-Id, List-*; From, Reply-To, Sender, To/CC and MailFrom fields. Identities relevant to Gateways are: From, Reply-To, Sender, To/CC/BCC and MailFrom fields. Message Handling Service (MHS) Actors are responsible for end-to-end transfer of messages. These Actors can generate, modify or look at only transfer data in the message. ADministrative Management Domain (ADMD) Actors are associated with different organizations which have their own administrative authority, operating policies and trust-based decision making.

E-mail Headers are included in the message by the sender or by a component of the e-mail system and also contain transit-handling trace information. Further, the message also contains special control data pertaining to Delivery Status and Message Disposition Notifications, etc. Various identities called fields are present in the message and are used in different parts of e-mail architecture called Layers. These fields serve a specific function in the system and are set by some component of the system.

III. ISSUES IN E-MAIL SECURITY

The primary protocols used in e-mail system lack security features for a) privacy of the sender, b) authentication of sender, c) integrity of e-mail message, d) non-repudiation of sender, and e) consistency of e-mail envelope. These security lapses permit e-mail to be spoofed by forging its headers or by sending it anonymously. Its capability to transmit executable files and other similar contents using Multi-Purpose Internet Mail Extensions (MIME) [2] and e-mail access at public places where anonymous use is easy has increased its vulnerability. Owing to e-mail’s inherent vulnerabilities, it is being misused by cybercriminals for several illegitimate
purposes that include: a) abuses like spamming, phishing, cyber bullying, child pornography, sexual harassment, racial vilification, etc., b) transmitting viruses, worms, Trojan horses, hoaxes, and other malicious programs with intent to spread them over the Internet, and c) for carrying Internet infrastructure crimes through Denial of Services and Directory Harvesting Attacks. These injudiciously use of e-mail cause many technological problems such as misuse of storage space, wastage of computational resources, and network conjunction and severe legal issues.

Technological means like firewalls, anti-viruses, scanners and filters to prevent e-mail abuses are insufficient to make e-mail communications secure. For e-mail to remain useful and usable, besides protecting its confidentiality and integrity through cryptography, its accountability in terms of identification, authorization, attestation and retribution must be improved.

IV. NEED FOR E-MAIL FORENSICS

Cybercriminals use e-mail spoofing to misuse the trust built in SMTP to lie the recipients about their true identities by not only spoofing one or more headers in the envelope or header of the message that somehow reveals their identity but also put misleading information in these headers.

A highly technical spammer or phisher may also evade packet filters and spoof the source IP address of their packets to indicate that the message is from a trusted domain.

There are many ways in which senders can lie about their true identities, each causing different effects. These include the following: a) Unauthorized Networks, b) Open Mail Relays, c) Anomizers or re-mailers, d) Open Proxy, e) SSH Tunnel or Port-Redirector, f) Botnets, g) Untraceable Internet Connections, and h) Spoofing.

Protocols offering security and anti-spam filters capable to perform mail categorization have been developed to secure e-mail service against sender-spoofing.

Security protocols that add privacy to SMTP either create encrypted secure channel between the sender and the receiver during SMTP transactions or use end to end symmetric or asymmetric cryptographic schemes.

Further, various domain validation anti-spoofing standards using either IP addresses or digital signatures to validate sending domain have also been developed that help an e-mail system at the receiving end to detect the spoofing of addresses and as such enables it to decide how to handle incoming e-mail.

Very limited numbers of e-mail users use these protocols to secure their e-mails due to either their limited technical skill or unawareness about their existence.

Further, their use has not been made mandatory and as such unwillingness of some ESP’s also limit their use.

Furthermore, spammers constantly change spam sending techniques and its structure to evade security procedures and protocols, leaving scope for e-mail forgery and e-mail crime raising the need for e-mail forensic analysis.

V. E-MAIL FORENSICS

Network forensics is the science that deals with capture, recording, and analysis of network traffic for investigation purpose and incident response [3][4][5][6]. E-mail forensics refers to the study of source and content of e-mail as evidence to identify the actual sender and recipient of a message, data/time of transmission, detailed record of e-mail transaction, intent of the sender, etc. Forensics analysis of an e-mail message aims at discovering the history of a message and identity of all involved entities. E-mail analysis begins from the recipient’s mailbox which contains the e-mail message. The message is analyzed to determine the source (originator and the author). The analysis involves the investigation of both control information (envelope and header) and the message body. Various analysis approaches used include: a) header analysis, b) bait tactics, c) server investigations, and d) network device investigation. Besides mandatory headers, custom and MIME headers appearing in the body of the message are also analyzed for: a) sender mailer fingerprints and b) software embedded identifiers.

Once the source of the e-mail message under investigation is determined or some source is strongly suspected its computer, e-mail client software, web browser, etc. are investigated for traces of evidence.

Meta data in the e-mail message in the form of control information i.e. envelope and headers including headers in the message body contain information about the sender and/or the path along which the message has traversed. Some of these may be spoofed to conceal the identity of the sender. The e-mail address of sender responsible for submitting the message to the transfer service is specified in the Sender header field. This field is optional and needs to be specified only if the author of the message is different from the sender. The e-mail address of the author is contained in the mandatory From header field. Various other addresses related to author or sender of the message are addresses specified in Reply-To, MailFrom and Return-Path. Address specified in Reply-To header overrides the From address for responses from recipients if specified. MailFrom specifies the address for receiving return control information like MDN and DSN. This address need not be same as that of author or sender responsible for submitting the message. Return-Path address is the address recorded by MDA from MailFrom control identifier. E-mail client programs and webmail interfaces add useful headers called X-Headers which besides other information contain useful information about the sender or author of the e-mail. A comparison of these addresses in all cases cannot be used to determine the genuineness of a sender or author because it is possible to spoof all of these addresses and thus not only hide one’s address but also pretend to be somebody else by using somebody else’s valid e-mail address in these fields. Trace information in the form of Received header field recorded by originator, relay, mediator or destination may be used to validate only the domain part of the sender’s or author’s e-mail address. Received-SPF also cannot validate the e-mail address of a sender or author as it can only validate domain of address specified in MailFrom parameter.

DKIM-Signature security field if present may be used to validate the domain part of the sender’s or author’s e-mail address if the sending and receiving servers are following DKIM signing protocol. Thus, on the basis of inconsistencies
of sender’s identity as may be revealed by various header fields, forensic experts can only make a wider guess about the genuineness of the sender’s e-mail address and not a final decision. Originator, Relay including MTA, and Receiver add trace information at the beginning of the message which is in the form of Received and Return-Path header fields. Received header field specifies the address used in MailFrom parameter and may not be much useful for forensic investigation. The Received field contains vital information including names and IP addresses of originating host, relays or MTA’s, Mediators and MSA’s.

Besides header analysis, e-mail forensic also involves investigation of some client or server computer suspected of being used or misused for e-mail forgery.

It can involve inspection of: a) Internet favorites, b) Cookies, c) History, d) Typed URL’s, e) Temporary Internet Files, f) Auto-completion Entries, g) Bookmarks, h) Contacts, and i) Preferences, Cache, etc. Several Open Source software tools have also been developed to perform e-mail header analysis to collect evidence of e-mail fraud.

VI. E-MAIL FORENSIC TOOLS

There are many tools which may assist in the study of source and content of e-mail message so that an attack or the malicious intent of the intrusions may be investigated [7]. These tools while providing easy to use browser format, automated reports, and other features, help to identify the origin and destination of the message, trace the path traversed by the message; identify spam and phishing networks, etc. These include following:

- eMailTrackerPro (http://www.emailtrackerpro.com/)
- EmailTracer (http://www.cyberforensics.in)
- Adcomplain(http://www.rdrop.com/users/billme/adcomplain.html)
- Aid4Mail Forensic (http://www.aid4mail.com/)
- AbusePipe (http://www.datamystic.com/abusepipe.html)
- AccessData’s FTK (www.accessdata.com/)
- EnCase Forensic (http://www.guidancesoftware.com)
- FINALeMAIL (http://finaldata2.com)
- Sawmill-GroupWise (http://www.sawmill.net)
- Forensics Investigation Toolkit (FIT) (http://www.edecision4u.com/FIT.html)
- Paraben (Network) E-mail Examiner (http://www.paraben.com/email-examiner.html)

The forensic tools have included features like e-mail tracing using the headers, Email tracing using the email address, World map location, Network Whois Data, Domain Whois Data, Abuse Reporting, Sender IP address, Misdirection detection, Outlook Plugin, Spam Filtering, Whitelists and blacklists, Supports POP access [8], Filter System, DNS Blacklist, etc.

Current forensic tools are designed to help examiners in finding specific pieces of evidence and are not assisting in investigations. Further, these tools were created for solving crimes committed against people where the evidence resides on a computer; they were not created to assist in solving typical crimes committed with computers or against computers. Current tools must be re-imagined to facilitate investigation and exploration. This is especially important when the tools are used outside of the law enforcement context for activities such as cyber-defense and intelligence. Construction of a modular forensic processing framework for digital forensics that implements the “Visibility, Filter and Report” model would be the first logical step in this direction.

CONCLUSION

E-mail being a highly distributed application involves several actors that play different roles in its transmission process. These actors include hardware and software components, services and protocols which provide interoperability between users along the path of its transfer. Cybercriminals forge e-mail headers or send it anonymously for illegitimate purposes which lead to several crimes and thus make e-mail forensic investigation crucial. With the help of various header fields, a detailed forensics of the e-mail message can be carried out which can lead to various clues useful to identify criminals. Many forensic tools are available that not only permit to make tracing of e-mail message from its source but also have other unique features such as filtering, blacklisting, etc.

REFERENCES