Charalambous Elisavet
Bratskas Romaios
Koutras Nikolaos
ADITESS Advanced Integrated Technology Solutions & Services, Nicosia, Cyprus
Karkas George
Anastasiades Andreas
Cyprus Police – Criminal Investigation department - Special cyber-crime unit2, Police Headquarters

Email forensic tools: A roadmap to email header analysis through a cybercrime use case

Keywords
email forensics, cybercrime, digital forensic tools, digital crime

Abstract

Email is one of the primary sources of numerous criminal activities, on the Internet, of which some threaten human lives. Email analysis is challenging due to not only various fields that can be forged by hackers or the wide range email applications in use, but also due to imposed law restrictions in the analysis of email body. Despite this being a relatively new area, a number of both open source and proprietary forensic tools, with varying possibilities and versatility, have been developed aiding use by practitioners. In this paper, we review existing email forensic tools for email header analysis, as part of email investigation, with emphasis on aspects related to online crime while still considering legal constraints. Through our analysis, we investigate a common case of cybercrime and examine the breadth of information one may gain solely through email forensics analysis. Additionally, a roadmap for email forensic analysis is presented, combining features and functionality already available, to assist the process of digital forensic analysis.

1. Introduction

We span the information age where information is maintained in digital form aiding improved efficiency and productivity [34]. The internet is a network of networks, connecting millions of computing devices [24] performing information interchange throughout the world; producing a large amount of electronic messages. As a result, e-mail has emerged as the most important application on Internet for communication of information, used not only from computers but many other electronic gadgets. Royal Pingdom, which monitors the Internet usage, stated that in 2010, 107 trillion e-mails were sent; that is 294 billion e-mails sent per day [21]. The heavy exchange of electronic messages coupled with the right of anonymity, result in increased number of criminal and terrorism activity [27] mainly due to two inherent limitations [12]. There is rarely any encryption at the sender end and/or integrity checks at the recipient end, while the widely used e-mail protocol SMTP (Simple Mail Transfer Protocol) lacks a source authentication mechanism. Although there have been many attempts into securing e-mail systems, most are still inadequately secured. Installing antivirus, filters, firewalls and scanners is simply not enough to secure e-mail communications. Some common examples of illegitimate uses of emails are spam, phishing, cyber bullying, racial abuse, disclosure of confidential information, child pornography and sexual harassment. In the vast majority of these e-mail cybercrimes the tactics used vary from simple anonymity to impersonation and identity theft. Therefore, a forensic investigator needs efficient tools and techniques to perform the
analysis with a high degree of accuracy and in a timely fashion. The anonymity factor of e-mail has made it difficult for digital forensic investigators to identify the authorship of an email, and to compound this problem further; there is no standardised procedure to follow [20]. To address this, digital forensic investigators have to follow a number of steps that are part of a process. The number of forensic models has added to the complexity of the field [23]. Therefore, this has led to a call for standardisation in the field of digital forensics as it hinders the investigation process [26]. Notwithstanding this, there are a few procedures from different authors that are known to be the “standard” procedures; with a number of discrepancies [20]. The lack of rules results in incomplete evidence collection and errors in interpretation [7].

2. Background

Computer Forensics science deals with the preservation, identification, extraction and documentation of computer evidence. The latest statistic showed that 5973 TB of data was processed for FY 2013 which is a 40% increase from FY 2011 [37]; as a side note one TB is roughly equivalent to the information in 1,000 encyclopaedias. Kara Nance et al [31] have proposed six categories of Digital forensics including Network Forensics also encompassing e-mail forensic analysis. Many studies have been carried out for analysing tools and techniques used in network forensics [5], [18], [25] and [30] which also include e-mail forensics tools and techniques.

E-mail forensic analysis is used to study the source and content of e-mail message as evidence, identifying the actual sender, recipient and timestamp to collect credible evidence to bring criminals to justice [6]. Additionally, email analysis is challenging due to not only various fields that can be forged by hackers or the wide range email applications in use, but also due to imposed law restrictions in the analysis of email body.

Forensic Analysis plays a major role by examining suspected e-mail accounts in an attempt to gather evidence to prosecute criminals in the court of law. Towards this direction, a number of forensics tools, either dedicated to or capable of performing email forensic analysis, have been widely used by the practitioners. However, these tools have been developed in an isolated manner rather than a collaborative approach [10], while despite the fact that some email forensic analysis tools start as open source and freely accessible solutions, over the years, instances of transitioning of those solutions into paid software have been noted.

3. Email forensic investigation techniques

E-mail forensics refers to the study of email details including: source and content of e-mail, in order to identify the actual sender and recipient of a message, date/time of transmission, detailed record of e-mail transaction as well as the intent of the sender. Therefore, e-mail forensic investigation often involves analysis of metadata, keyword searching as well as port scanning, for authorship attribution and identification of cyber-crime.

It is evident that an email forensic tool may only assist the investigator during a specific stage of analysis. Various approaches that are used for e-mail forensic are described in [4]. In header analysis, metadata in the e-mail message in the form of control information containing information about the sender and/or the path along which the message has traversed become subject to in-depth analysis; understanding also the fact that some of this information might be forged to conceal the identity of the sender.

Taking it a step further, server investigation involves the analysis of email copies and logs retained on the mailing server in an effort to identify the source of a message; SMTP servers which store data pertaining to owner of a mailbox (i.e. credit card number) are of great value in revealing one’s identity. However, this type of investigation often proves to be time consuming as the logs and back up emails need to be requested either from the proxy server or the ISP (some may not co-operate with the investigators), and resource expensive due to the large amounts of processing required to restore any valuable information. Additionally, e-mail copies and server logs are only maintained for some limited periods of time (which vary according to the applicable legislation). When server investigation is not an option, investigators have the option to turn towards network device investigation, which is known as a notoriously complex type of investigation involving analysis of logs maintained by the network devices such as routers, firewalls and switches.

Other investigation techniques involve analysis of software embedded identifiers where the investigator is looking for information related to the creator or message contained data (i.e. attached files) through information incorporated by the email client/software used by the sender. This information often takes the form of custom headers or MIME (Multipurpose Internet Mail Extensions) content. Even though this type of analysis proves time consuming it may reveal some vital information about the sender’s e-mail.
preferences and options that could help client side evidence gathering. The investigation can reveal PST file names, Windows logon username, MAC address, etc. of the client computer used to send e-mail message [15]. Similarly, the received header field and email handling software at the sender side may reveal the software managing emails on the server (due to the different structure in headers). This analysis forms part of a procedure known as sender mailer fingerprints capable of describing applications and their version at the client side; useful as reveals characteristics/vulnerabilities of the bearing host machine.

At last but not least, the less time expensive but also less passive investigation method is known as bait tactics where investigators send an email, to the sender under question, containing an HTTP image source tag hosted on a server of their control. Once the email is being opened the e-mail client will request through an HTTP call the downloading of the image, revealing its, own or proxy, IP address.

It is clear that it is not viable for investigators to perform most of these analyses on a day-to-day basis (with no real evidence in hand) due to their time and resource complexity and also the risk of not being able to gather evidence of real value. Due to this, in this paper we emphasise on email header (tracing) analysis with existing tools in an effort to allow investigators to get to evidence of value in a timely manner.

4. Email structure

An e-mail system is composed of several, both, software and hardware, components providing interoperability through compliance to networking standards and communication protocols, along the path between the source and sink. E-mail is a highly distributed service involving several actors that play different roles to accomplish end-to-end mail exchange [19]. Several communicating software entities called e-mail nodes on the application layer of the TCP/IP model are involved in the process of e-mail delivery [16], while e-mail initiation is possible through HTTP and STMP, however, most e-mails are sent through the STMP protocol.

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 email message subject and body. The body might also contain attachments in the form of MIME or S/MIME (Secure/MIME) [35].

Information in email headers is organized from the bottom up, meaning that the email was handed from MTA (Message Transfer Agents) nodes at the bottom of the header to the ones at the top. Upon reception of an email by an MTA, the received header section is appended with relevant information; information related to the host who receives the email last (destination host) is enclosed at the top of the stack. The architecture of Internet mail utilizes a number of unique identities known as mailbox, domain name, message-ID, and ENVID (envelope identifier) [9].

More concretely, mailboxes are conceptual entities identified by e-mail address and receive mail. The email address has become a common identity identifier on the Internet as it consists of username and domain name separated by @ sign. A domain name is a global reference to an Internet resource like a host, network or service which maps to one or more IP address. Its structure has a hierarchical sequence of labels, separated by dots. Additionally, Message-ID and ENVID are message identifiers which respectively pertain to message content and transfer; Message-ID is used for threading, aiding identification for duplications and DNS tracking, while the ENveloped IDentifier (ENVID) is used for the purpose of message tracking; the ENVID contains transit-handling information used by the MHS. Various identities, called fields, used for analysing e-mail to determine the source (originator and the author), are present in the message and are used in different parts of email architecture called Layers, described in [6].

4.1. E-mail header tracing

Since in this paper, we are focusing on header analysis, it is necessary to review common processes involved in this process. 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. A typical e-mail header looks like this:

```plaintext
Received: from mail-lb0-f178.domain.com (mail-lb0-f178.domain.com [209.85.217.178])
    by linux247.grserver.gr (Postfix) with ESMTPS id 7D5794604B3
    for <address@domain.com>; Wed, 30 Mar 2016 10:43:32 +0300 (EEST)
Received-SPF: pass (linux247.grserver.gr: domain of
gmail.com designates 209.85.217.178 as permitted sender) client-ip=209.85.217.178; envelope-from=sender@mail.com; helo=mail-lb0-f178.domain.com;
```

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 examination of the sender’s e-mail address and IP address, examination of the message ID as well as the messaging initiation protocol (HTTP or SMTP). 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.

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/forged headers in order to deceive investigators, so extreme caution and careful scrutiny should be practiced in investigating every part of the e-mail header.

4.2. Review of existing email 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. 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. This section introduces some of these tools.

According to [17] 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.

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 [6]. The following forensic tools offer email forensic analysis, it is noted that the available solutions do not only vary in the type of offered functionality but also in the amount of available documentation for their use and deployment; the list involves both open source and closed source solutions.

Aid4Mail [3] is a proprietary solution that can analyse emails stored in hard disk. Further, it supports, both online and offline, email analysis (i.e. either directly from webmail services that use IMAP access or from a local storage unit). Aid4Mail supports email filtering based on text, time, date, keywords, logical operators, and regular expressions whilst also performing searches by date, header content, and by message body content. Aid4Mail also offers the ability to process unpurged (deleted) e-mail from mbox files and can restore unpurged e-mail during exportation.

Digital Forensic Framework (DFF) [11] is an Open Source computer forensics platform built on top of a dedicated Application Programming Interface (API). DFF proposes an alternative to the aging digital forensics solutions used today. Designed for simple use and automation, the DFF interface guides the user through the main steps of a digital investigation so it can be used by both professional and non-expert to quickly and easily conduct digital investigations and perform incident response. Digital Forensic Framework can analyse emails stored in hard disk while it can also perform some features like virtual machine disk reconstruction. For each messages header contained information can be displayed and their filtering is possible with the use of regular expressions and based on the email content, tags, and time-line. The digital forensic framework can support various formats allowing interoperability with other tools.

eMailTrackerPro [14] is a proprietary email forensic solution that analyses email files stored in local disk and supports automatic email analysis for the identification of spamming incidents. eMailTrackerPro is capable of recovering the IP address that sends the message along with its associated geographical location (city) to determine the threat level or validity of an e-mail message. It can find the network service provider (ISP) of the sender. A routing table is provided to identify the path between the sender and receiver of an email. It also can check a suspected email against Domain Name Server blacklists to safeguard against spam
and malicious emails. It also displays whether any port is open in any of the HTTP or FTP server in the tracked IP addresses. Paraben E-Mail Examiner [33] is another proprietary solution capable of processing emails found on a local hard disk and supports comprehensive analysis features, bookmarking as well as advanced searching; including searching within attachments. The tool can examine email headers and bodies, provides information based on the search (including contents from attachments). Paraben E-mail Examiner can recover deleted emails from Exchange (EDB), Lotus Notes (NSF), and Group-Wise email even though they may be deleted from the deleted items folder.

EmailTracer [13] is an Indian effort in cyber forensics by the Resource Centre for Cyber Forensics (RCCF) which is a premier centre for cyber forensics in India. It develops cyber forensic tools based on the requirements of law enforcement agencies. This tool traces the originating IP address and other details from e-mail header, generates detailed HTML report of email header analysis, finds the city level details of the sender, plots route traced by the mail and display the originating geographic location of the e-mail. Besides these, it has keyword searching facility on e-mail content including attachment for its classification.

Adcomplain [2] is a command line tool which however does not only take into consideration the header of the message but also the body. It is presented as a tool for reporting inappropriate commercial e-mail and usenet postings, as well as chain letters and "make money fast" postings. It automatically analyses the message, composes an abuse report, and mails the report to the offender's internet service provider by performing a valid header analysis; the report is submitted to the U.S. Federal Trade Commission.

MailXaminer [28] by SysTools Software is a digital forensic program built to allow the examination of messages from both web and application based email clients. MailXaminer loads messages from the chosen email storage source and arranges them hierarchically for the purpose of evidence analysis and extraction. The programming of the application provides carving out of deleted evidence or evidence from damaged sources in cases of evidence spoliation [32]. Post analysis, the software serves output generation in court admissible digital formats. AbusePipe [1] analyses abuse complaint e-mails and determines which of ESP’s customers is sending spam based on the information in e-mailed complaints. It automatically generates reports reporting customers violating ESP’s acceptable user policy so that action to shut them down can be taken immediately. AbusePipe can be configured to automatically reply to people reporting abuse. 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.

Internet Evidence Finder (IEF) [29] by Magnet Forensics allows the definition of specific profiles for the recovery and detection of emails contained on physical drives. Once a search is being completed the results are shown and for each recovered email the investigator may observe details related to the artefact, header information, as well as the email body and attachments. IEF is capable of recovering a range of files including PST, OST files as well as ‘mbox’ emails from which may extract emails, contacts, appointments, notes and tasks as well as any attachments. Additionally, IEF may recover instant messaging chats and file transfers.

NUIX [22] continues to lead the industry in breaking open data formats and forensic artifacts. It performs complex processes efficiently and consistently. Amongst its main features are email threading which groups email messages together so that the investigator can review them in context and make bulk decisions quickly.

AccessData’s FTK [15] is standard court-validated digital investigations platform computer forensics software delivering computer forensic analysis, decryption and password cracking within an intuitive and customizable interface. It has speed, analytics and enterprise class scalability. It is known for its intuitive interface, e-mail analysis, customizable data views and stability. It supports popular encryption technologies, such as Credant, SafeBoot, Uitmaco, EFS, PGP, Guardian Edge, Sophos Enterprise and S/MIME. Its current supported e-mail types are: Lotus Notes NSF, Outlook PST/OST, Exchange EDB, Outlook Express DBX, Eudora, EML (Microsoft Internet Mail, Earthlink, Thunderbird, Quickmail, etc.), Netscape, AOL and RFC 833. EnCase Forensic [19] is computer forensic application that provides investigators the ability to image a drive and preserve it in a forensic manner using the EnCase evidence file format (LEF or E01), a digital evidence container vetted by courts worldwide. It contains a full suite of analysis, bookmarking and reporting features. Guidance Software and third party vendors provide support for expanded capabilities to ensure that forensic examiners have the most comprehensive set of utilities. EnCase Forensic includes e-mail threading and related conversations to enable investigators to understand the context of e-mails, which is critically important in an investigation. Including many other network forensics investigations, it also supports Internet and e-mail investigation. It included Instant
Email Forensic Tools: A roadmap to email header analysis through a cybercrime use case

5. Exploring the boundaries of Email Forensic tools through a Use Case

A realistic scenario for a DFE examiner is, provided a number of artifacts, question the authenticity of the material, verify the capturing source and then to reconstruct the depicted scene so as to be able to draw conclusions [8]. A common forensic case for authorities in Cyprus is the investigation of child pornographic cases for which the utilization of images is rather critical. The use of the email forensic analysis could reveal possible connections of the offender as well as identification of cases of victimization.

Email header analysis may allow the investigators in identifying links to criminal activity and the start of an investigation. Upon tracing of sufficient suspicious activity, investigators shall build the confidence in seizing the offender’s rights on data protection and request a warrant for further analysis. It is seen that there is overlapping between operations performed by the different existing solutions. The use of one solution over the other is dependent on the investigator’s preference and technical knowledge as some solutions are more user friendly and intuitive than the others. Additionally, in the case of command line solutions deployment of the tool is highly related also to the available documentation. It is noticed that freeware or open-source solutions are not characterized by good documentation and supporting material, therefore their use as part of an investigator’s standard procedures is not very likely. However, some freeware or low cost tools might be used as a means of double checking email contained information.

6. Conclusion

Digital forensic analysis is a complex and time consuming process which involves the analysis of digital evidence. Emails might contain valuable information that could lead investigators to the identity and/or location of the offender. Additionally, email forensic tools through email header analysis may even reveal information related to the host machine used during the composition of the message. In this paper, we have discussed key information related to email forensic analysis as well as important aspects of header tracing. Finally, we listed the available tools that can be utilised for email analysis emphasising on their key features in an effort to assist investigators in the selection of the appropriate tools.

Acknowledgements

The work presented in this paper received funding from the European Union ISEC programme (HOME/2013/ISEC) entitled Cybercrime Center of Excellence Network under grant agreement number HOME/2013/ISEC/AG/INT/400005229.
References


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


[36] Sawmill - Universal log file analysis and reporting, [available at: https://www.sawmill.net/].
