International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: 6.614
Available Online at www.journalijcar.org
Volume 7; Issue 5(D); May 2018; Page No. 12494-12499
DOI: http://dx.doi.org/10.24327/ijcar.2018.12499.2198

Research Article

DIGITAL FORENSICS – A DEEPER LOOK INTO EMAIL ANALYSIS

Deep Kumar Ganguly and Asoke Nath

Department of Computer Science, St. Xavier's College (Autonomous), 30 Park Street, Kolkata-700016, West Bengal, India

A R T I C L E I N F O

Article History:

Received 6th February, 2018
Received in revised form 20th March, 2018 Accepted 8th April, 2018
Published online 28th May, 2018

Key words:

e-mail, security, digital, forensics, tracking, cybercrime

A B S T R A C T

Computer Forensics is the fundamental study of evidences or tracing evidences of a prior committed fraud/ crime that is crucial to hold charges against the accused and be presentable in front of the eyes of law. In the present paper the authors have given a brief overview on computer forensics which will throw light upon what we understand by computer forensics and the various analysis procedures we implement to trace evidences and try to successfully recover any fragments of data without further loss. Digital forensics has a massive scope in the coming days. In the present studies the authors have tried to give basic overview of digital forensics and it application in e-mail analysis. Some experimental data also presented in the same paper.

INTRODUCTION

Computer forensics is the application of investigation and analysis techniques to gather and preserve evidence from a particular computing device in a way that is suitable for presentation in a court of law. The need for computer forensics is so essential because we are piling mountains of data every day to aid us in taking future steps and to optimize our previous conquests.

With categorizing and organizing data we also need to secure this valuable data because we don't want all our data to be accessible to all over the world. We implement various security measures to keep this data secure and prevent it from falling onto the wrong hands. Though throughout history we find evidences of security breaches and data being compromised thus resulting in various corporate failures and national security being compromised.

As for this digital age with booming cyber technologies it is necessary to enforce law and order for such crimes and this where computer forensics come into play as it identifies the nature of attack and provides a list for all the charges that can be pressed on the attacker also it tries to recover data as much as possible with minimum to no further loss.

Email Forensics – A Practical Approach. Email is a common method of communication among two parties. It is a file transfer among two servers on a specific port number.

Corresponding author:* **Deep Kumar Ganguly
Department of Computer Science, St. Xavier's College (Autonomous), 30 Park Street, Kolkata-700016, West Bengal, India

An email is usually composed using a client side application (Web Client, MS Outlook, Lotus notes) with a Sender's identity, then stored as a file, subsequently delivered to a destination user address through one or more number of servers. Although Email communication has been designed to make things simple, efficient, and powerful, the composing and communication have been under the focus of malicious intruders over the last few decades. We find it very common that email messages are the sources of transporting annoying, malicious, phishing, and spam contents. In recent years, many technologies have been developed for examining and protecting emails that include spam detection, phishing email detection, content and attachment filtering (anti-virus engines). One key aspect of designing and developing these technologies is to conduct forensics investigation on sample emails to correctly identify important information such as the recipient name or identity, the path between the sender and the recipient used for transporting the email, the client-side application used to compose the email, the timestamp when a message was generated, a unique message ID, etc.

In the literature, the examination and revealing of key information from an email is known as Email Forensics. A common example usage of forensics is to understand the key facts and rely on them for legal procedures. Over the last two decades, many email forensic tools have been developed. Further, most of the forensics tools are not intended to solve any specific cyber or computer crime related problem.

Rather, they are intended to discover or recover information. Given that a question naturally arises among stakeholders: to what extent existing tools are suitable for employing or conducting digital forensic investigation?

This paper attempts to answer this question by comparing and contrasting a number of popular email forensic tools. In particular, our focus is on email header analysis phase offered by the tools.

We examine the capability of a number of popular email forensic tools including MainXaminer, Add4Mail, Digital Forensic Framework, eMailTrackerPro, and Paraben Email Examiner. Our work is complementary to previous research work that attempts to understand the capability of other types of forensic tools such as network forensics and disk/memory forensic tools.

Email Forensics Analysis Steps. A forensic investigation of e-mail can examine both email header and body. This paper will look at header examination. According to Marwan an investigation should have the following:

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

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

1. Storage format of email: Server side storage format may include maildir (each email is kept separate in a 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 for forensics analysis by using notepad editor and applying regular expression-based searches. At the client-side, an email is stored as mbox format (Thunderbird). Client side may also store emails as .PST (MSOutlook), and NSF (Lotus Notes) files.
2. Availability of backup copy of email: When checking from the serve side, all copies are transferred to the client. This requires seizing the client computer. For webmail, copies are always saved at the server side.
3. Protocol used to transport email: Email can be initiated and transported based on SMTP or HTTP depending on the email server applications.

Comparison Criteria of Email Forensic Tools. We compare email forensic tools based on a set of desired attributes required by forensic tool that focuses on issues to support forensics investigation where most cases cannot be generalized. At the same time, expected criteria should be relevant to the reverse engineering capabilities, as well as scaling up data processing, considering the changes of technologies for storage, data transportation, and in between environment between devices processing and storing inputs. We identify nine criteria that may be useful and present in forensic tools listed below:

1. Requirement of input file in the hard disk,
2. Search option
3. Information extracted or provided by the tool
4. Recovery capability
5. Email file format supported
6. Visualization support
7. Operating system (OS) supported
8. Extended device supported
9. And export format supported.

LITERATURE REVIEW

Definition and Importance. Forensics is the process of using scientific knowledge for collecting, analyzing, and presenting evidence to the courts. (The word forensics means "to bring to the court.") Forensics deals primarily with the recovery and analysis of latent evidence. Latent evidence can take many forms, from fingerprints left on a window to DNA evidence recovered from blood stains to the files on a hard drive.

Because computer forensics is a new discipline, there is little standardization and consistency across the courts and industry. As a result, it is not yet recognized as a formal "scientific" discipline. We define computer forensics as the discipline that combines elements of law and computer science to collect and analyze data from computer systems, networks, wireless communications, and storage devices in a way that is admissible as evidence in a court of law.

Adding the ability to practice sound computer forensics will help you ensure the overall integrity and survivability of your network infrastructure. You can help your organization if you consider computer forensics as a new basic element in what is known as a "defense-in-depth" approach to network and computer security. For instance, understanding the legal and technical aspects of computer forensics will help you capture vital information if your network is compromised and will help you prosecute the case if the intruder is caught.

Failure of Practicing Forensics. What happens if you ignore computer forensics or practice badly? You risk destroying vital evidence or having forensic evidence ruled inadmissible in a court of law. Also, you or your organization may run afoul of new laws that mandate regulatory compliance and assign liability if certain types of data are not adequately protected. Recent legislation makes it possible to hold organizations liable in civil or criminal court if they fail to protect customer data.

From a technical standpoint, the main goal of computer forensics is to identify, collect, preserve, and analyze data in a way that preserves the integrity of the evidence collected so it can be used effectively in a legal case.

Forensics and Investigation. First, those who investigate computers have to understand the kind of potential evidence they are looking for in order to structure their search.

Crimes involving a computer can range across the spectrum of criminal activity, from child pornography to theft of personal data to destruction of intellectual property. Second, the investigator must pick the appropriate tools to use. Files may have been deleted, damaged, or encrypted, and the investigator must be familiar with an array of methods and software to prevent further damage in the recovery process.

Cybercrime–Introduction, Scope and Types. The proliferation of technology devices and other equipment; their pervasive use across age, gender, socioeconomic and geographic boundaries; and, for many, a false sense of information security have merged to create a perfect storm for cybercriminal activity. In fact, cybercrimes have risen so dramatically in recent years that they have seemingly replaced old-fashioned, organized crime.

A cybercrime is defined as an intended act involving the use of computers or other technologies, and the criminal activity must take place in a virtual setting, such as the Internet.

Cybercrimes share three elements

1. Tools and techniques to perpetrate a crime
2. Approach or methodology for executing the criminal plan-known as a vector
3. Crime itself that is the end result of those plans and activities (a cybercrime is the ultimate objective of the criminal's activities)

Virtual environments have become fertile territory for cybercrime, with the number of crimes escalating each year along with the severity of losses. In 2011, online revenue losses resulting from fraudulent transactions were estimated to be $3.4 billion, up from $2.7 billion in 2010. Revenue losses are based only on fraud associated with e-commerce and exclude fraud involving theft/loss of mobile devices and other forms of cybercrimes.

Growth in cybercrimes and their attendant costs are documented in a 2012 Ponemon Institute study. In the study, 56 large U.S. businesses surveyed reported an average annual cost of $8.9 million for cybercrimes, with costs reaching $46 million for one company. The average annual cost climbed six percent from the 2011 study. Companies participating in the study suffered an average of 102 successful attacks per week, up from 72 in 2011.

The report concluded that cybercrime appears to be worsening, and that 51 percent of CEOs reported that their companies have been attacked either daily or hourly.

The top five types of cybercrimes:

1. Tax-refund fraud
2. Corporate account takeover
3. Identity theft
4. Theft of sensitive data
5. Theft of intellectual property

Digital Evidence. The definition proposed for Digital Evidence by the International Organization for Computer Evidence (IOCE) is – all information that is generated, stored or transmitted using electronic devices that may be relied upon in court during trail is regarded as digital evidence.

Even though it is commonplace for the terms "electronic evidence" and "digital evidence" to be used interchangeably, the latter is, in fact, merely a subset of the broader category of "electronic evidence" which, he posits, also includes evidence in the form of analog data, such as video and audio tape recordings, photographic film and tape-recorded intercepts of fixed-line phone conversations.

Whilst all these types of data may well be "digitalized", they do not originate in digital form. More elaborately electronic evidence is data (comprising the output of analogue evidence devices or data in digital format) that is created, manipulated, stored or communicated by any device, computer or computer system or transmitted over a communication system, that is relevant to the process of adjudication.

Being such a modern term many legal and penal systems are still unaware and are oblivious to the fact of computer generated digital evidence or the forensics part in play that is necessary to process this evidence and give a valid legal report which can be used in prosecution.

Modern Trends and Tools in Forensic Analysis. Currently it has become a very important measure for the treatment and analysis of data for securing future data and to prevent future attacks and minimize risks. The various tools at our disposal conducting forensic analysis to name a few are:

- Image Forensic
- Memory Forensic
- Smartphone Forensic
- Personnel identification
- File Forensic
- Information Recovery
- Artificial Intelligence
- Vectored Image processing
- Pattern Matching/ Analysis
- E-mail Analysis
- Data Integrity
- USB forensic
- Location Analysis

These are few of the various methods and techniques that we have in hand to try to process evidence and generate meaningful reports from these analyses which are of importance to the public for court hearing.

RESULTS AND DISCUSSION

The following email session is conducted via sending a dummy email and studying the analytics produced by a modern email forensics tool called – "aid4Mail".

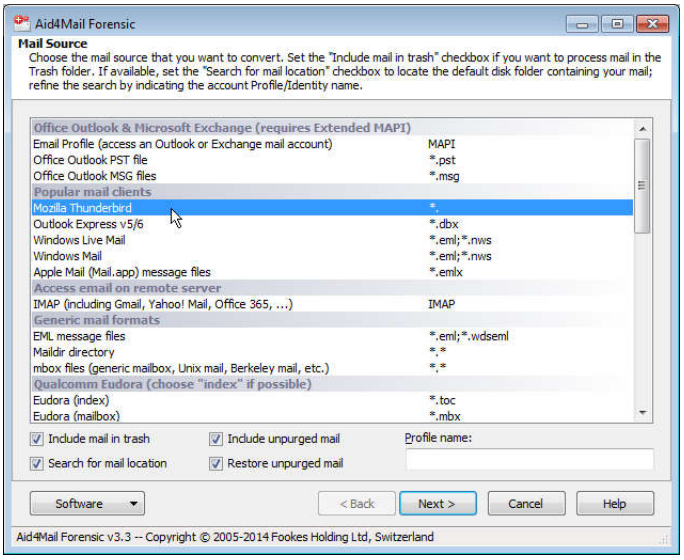

Fig 1 First we begin by establishing the email service that we shall be using, Mozilla Thunderbird (for this particular example)

Common Terminologies associated with email

SMTP Server. Stands for Simple Mail Transfer Protocol and the primary task that it performs are:

- Receives emails from the sender.
- Validate source and destination addresses.
- Sends and Receives Emails to and from other SMTP Server.
- User/Email Client/ Webmail use SMTP to SEND the intended email.

POP3 Server: It is an incoming mail server that helps the user to RECEIVE the email residing in its e-mailbox.

- Recipient Receives the email
- POP3 Server Deletes the email from its server once user's email client downloads the particular email.
- Usually work on port 110

IMAP Server: Incoming mail Server exhibits same functionality as of POP based server but retain copy of email even after user downloads the email.

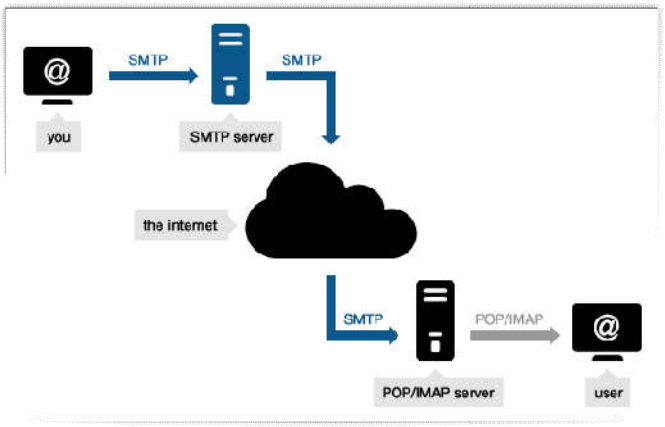

Fig 2 Basic framework of an email server

SMTP Protocol Commands (Client-Server interaction). SMTP protocol governs the email system language [as specified in RFC2821]. The process of SMTP client [application (MS Outlook)/webmail (Gmail)] making a request to SMTP Server (e.g. MS Exchange) and the server responding back to the request with acknowledgment code. SMTP Server becomes SMTP client when it transmits email to other SMTP server.

The protocol commands for REQUEST are:

- **EHLO or HELO:** SMTP Client identifies itself to SMTP server with this command.
- **MAIL FROM:** This command tells the Server the source of the email message (sender).
- **RESET:** SMTP Client asks the SMTP server to abandon the current transaction.
- **VERIFY:** SMTP client ask the Server to verify a user/mailbox.
- **EXPN:** SMTP client ask the Server to confirm the mailing list and after confirmation should return membership of the list.
- **HELP:** The client asks the server to send helpful information.
- **NOOP:** this SMTP protocol commands request the server to send an "ok" reply.
- **QUIT:** this command tells the server to send an "ok" command and thereafter terminate the transmission channel.

Aid4Mail rebuilds an email to a format almost as close to RFC such that further analysis can be done on the particular email.

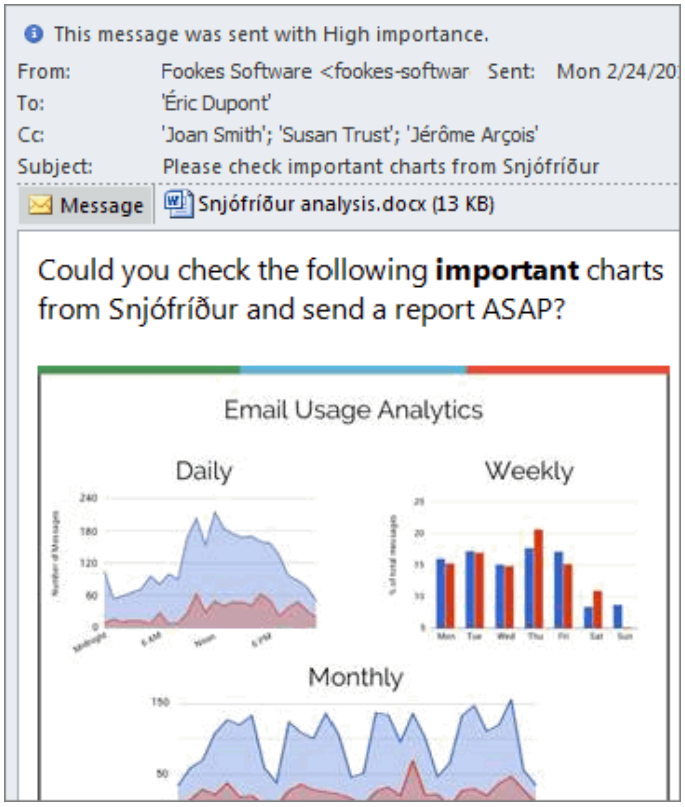

Fig 3 Rebuilding with Aid4Mail

Client	Server
	220 mail.example.com SMTP Foo Mailserver
HELO mail.example.org	
	250 Hello mail.example.org,
MAIL FROM:<hans.muster@example.org>	
	250 Sender OK
RCPT TO:<foo@example.com>	
	250 Recipient OK
DATA	
	354 End data with <CR><LF>.<CR><LF>
From: <hans.muster@example.org> To: <foo@example.com> Subject: Testmail Date: Thu, 26 Oct 2006 13:10:50 +0200 Testmail .	
	250 Message accepted for delivery
QUIT	
	221 See you later

Fig 4 The communication between the client and the server.

Email header plays a crucial role in identifying the sender of an email. Many fields can be forged within the header part but it still gives enough information about the sender. The investigator upon performing the email header forensics will able to identify the following:

Other information in the email header that indirectly will help you during the forensics process:

- Sender of the email
- Network path it traversed and path of origination
- SMTP Servers it went through
- Time Stamp Detail
- Email Client information
- Encoding information

Header contains several lines of header information also known as fields. Each field itself is divided into three components.

- Field Label
- Followed by semicolon ":"
- Field Body

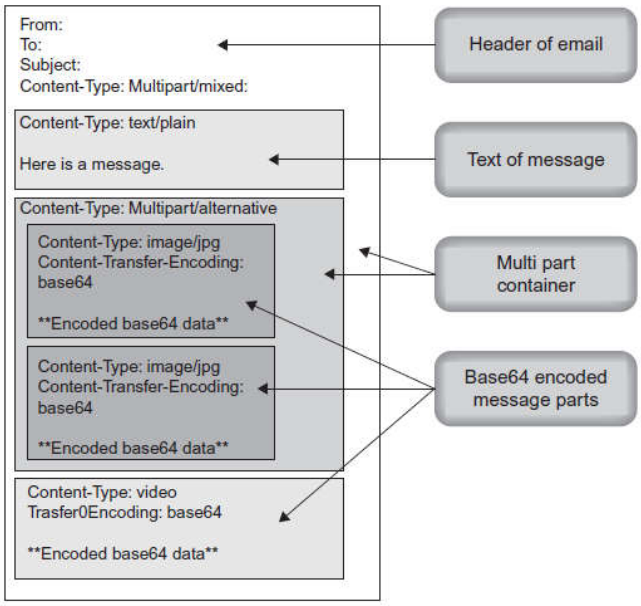

Fig 5 Email header schematics

Forensics of Email Metadata information

Header Line	Meaning
Microsoft Mail Internet Headers Version 2.0	Email header title
Received: from EXIC1.lse.ac.uk ([132.148.290.111]) by Exf2.lse.ac.uk with Microsoft SMTPSVC(5.0.2195.5329); Tue, 2 Nov 2010 12:20:40 +0100	Email passed from Exchange gateway servers to staff mailbox Server
Received: from EXAV2.pc.lse.ac.uk ([132.148.290.123]) by EXIC1.lse.ac.uk with Microsoft SMTPSVC(5.0.2195.5329); Tue, 2 Nov 2010 12:20:39 +0100	Email passed from anti-virus servers to Exchange gateway Server
Received: From exas1.lse.ac.uk ([132.148.290.127]) by EXAV2.pc.lse.ac.uk (WebShield SMTP v4.5 MR1a); id 1058267813844; Tue, 2 Nov 2010 12:20:37 +0100 Content-Class: urn:content-classes:message X-MimeOLE: Produced By Microsoft MimeOLE V5.50.4910.0300	Email passed from anti-spam servers to anti-virus servers
Received: from web60003.mail.yahoo.com ([120.204.138.115]) by exas1.lse.ac.uk with Microsoft SMTPSVC(5.0.2195.5329); Tue, 2 Nov 2010 12:18:20 +0100	Email received by LSE anti-spam server from originator's email server
Message-ID: <200305689256.6388.qmail@web60003.mail.yahoo.com>	
Received: from [132.148.290.30] by web60003.mail.yahoo.com via HTTP; Tue, 2 Nov 2010 12:18:20 BST	Date and time received by the originator's email servers
Date: Tue, 2 Nov 2010 12:18:20+0100 (BST)	Date and time sent
From: "Daniel Simpson" <djs@yahoo.co.uk>	Originator's email address
Reply-To: <d.simpson@hotmail.com>	Reply to address
Subject: Test Email	Subject
To: <f.muir@lse.ac.uk>	Recipients email address
MIME-Version: 1.0 Content-Type: multipart/alternative; boundary="0-714528637-1235867664=:1234" Content-Transfer-Encoding: 8bit	
Return-Path: <djs@yahoo.co.uk> X-OriginalArrivalTime: 2 Nov 2010 11:20:20.0570 (UTC) FILETIME=[3F7D6840:01C34AC2]--00-714528637-1235867664=:1234 Content-Type: text/plain; charset="iso-8859-1" Content-Transfer-Encoding: 7bit--0-714528637-1235867664=:1234 Content-Type: text/html; charset="iso-8859-1" Content-Transfer-Encoding: 7bit--0-714528637-1235867664=:1234--	Originator's email address Arrival time at originators email servers Start of the email message

Fig 6 Full data extracted from the email header and its corresponding significance

The Email forensic investigator can use several header fields to trace the email but it can be broadly categorized into the following area of interest the investigator should look into:

- Sender's SMTP Server (OUTGOING Mail Server)
- Encrypted mail header
- Typical To, From, Subject, and Date Lines
- Mail transfer email client information
- Various X-header information added by different SMTP server and email clients during the whole email sending process.

Fig 7 Extracted metadata information collected from header and it's relevant mapping of the IP address.

CONCLUSION AND FUTURE SCOPE

More and more tools are available or developed to facilitate the digital forensic investigators to acquire the digital evidence from the devices. Some of the tools are very powerful to extract the information from and reduce the duration of evidence analysis. Besides the advancement in the digital forensic investigation tools, the methodologies or techniques developed to obtain the information also become more advanced.

One of the key factors of the situation is contributed by the way computing technology evolves. The rapid development of computing devices requires new methods or tools to be used by the digital forensic investigators to obtain the evidences as a legally acquired evidence to be presented in the court.

The advancement in communication device also contributes to the following situation. Nowadays, mobile phones do not only transfer voice and text message, they have become a multipurpose device that can transfer multimedia files, perform video streaming, internet browsing and other operation that relates to data transfer. Thanks to the advancement in networking speed, the user can transfer their data easily with their mobile device. Even though this is a great situation for the user, it may lead them to the become a target of privacy invasion. Their personal data that reside in their mobile are valuable and might attract unauthorized attacker to gain their information for illegal purposes. As the computing technology evolves, the way computer user use or transfer the data in their environment also different from traditional computing system. The digital evidence acquisition methodologies need to adapt the new environment like cloud computing and peer-to-peer networking environment.

It differs from traditional computing system, where normally a single user uses the device and the application and user's data reside in their devices only. Unlike traditional computing environment the evidences or the data might not reside on single device but may be scattered around several devices. This requires the investigator to be extra careful with the data acquisition process because they might invade other users' private information that resides in that type of network. So with the complexity of networking, computing environment and the advancement of mobile devices, the digital forensic investigators also need to be advanced in their tools and

methodologies to obtain the evidences legally without affecting the user's privacy to the court.

Apart from the above issue there is a scope to do research on effective method on privacy education. As an initial step to reduce the privacy issue, it is crucial to combat the problems at the root level. The root level solution is in our mind. Educating the human mind to become an ethical person in their work is one of the key factors that we think will help to reduce the issues in privacy. It is crucial to educate different level of person not to invade into other person's private information and to educate on what to do if they accidentally found that type of information. The method to educate people on privacy need to be effective enough, as we are human tends to explore something new to us. So, regardless how powerful the above mentioned tools might evolve, in the hand of ethical person, the privacy of related parties can be preserved if we have successfully educated ourselves to not interfere with the information which is not for our eyes to see.

References

1. International Journal of Cyber-Security and Digital Forensics (IJCSDF) 2(2): 48-76 The Society of Digital Information and Wireless Communications, 2013 (ISSN: 2305-0012)
2. S. MASON, Electronic Evidence. Discovery & Admissibility, LexisNexis Butterworths, London, 2007, paragraph 2.03.
3. *Journal of Digital Forensics, Security and Law*, Vol. 5(2)

How to cite this article:

Deep Kumar Ganguly and Asoke Nath (2018) 'Digital Forensics – A Deeper Look into Email Analysis', *International Journal of Current Advanced Research,* 07(5), pp. 12494-12499. DOI: http://dx.doi.org/10.24327/ijcar.2018.12499.2198

****