Detection of Spoofed Mails

Aparna Jayan
ER &DCI-IT
Centre for Development of Advanced Computing(C-DAC)
Trivandrum, India
aparnajayancs@gmail.com

Dija S
Resource Centre for Cyber Forensics
Centre for Development of Advanced Computing(C-DAC)
Trivandrum, India
dija@cdac.in

***Abstract*— Cyber forensics deals with the collection of digital artifacts from digital devices that can be presented in the court as evidence. There are many areas of cyber forensics. Some of these are Disk Forensics, Network Forensics, Mobile Forensics, Database Forensics, etc. This paper mainly focuses on E-mail Forensics coming under network forensics. Emails are highly vulnerable to attacks. Email spoofing is the major one among many forms of email-based attacks. Detecting email spoofing is an important challenge in email forensic investigation. In this paper, different methods to detect spoofing by analyzing the mail header are discussed.**

Keywords—cyber forensics; email spoofing; email headers

I. INTRODUCTION

Cyber forensics involves techniques to collect, analyze and preserve digital evidence that can be presented in the court. The main aim of cyber forensics is to perform an investigation by event reconstruction. Event reconstruction involves finding the events happened in a digital device and documenting these events as evidence to be presented in court. Some of the important areas of cyber forensics include disk forensics, mobile forensics, network forensics etc. Disk Forensics involves extraction of evidence from storage media like hard disks, USB devices, CDs, etc. Mobile Forensics involves extraction of deleted mobile phone data such as contact information, call information, messages, pictures, emails, video recordings, etc that can be used as evidence. Network forensics deals with capture, recording and analysis of network traffic for investigation purpose and incident response[3].

Email forensics deals with the extraction of forensic evidence from an email message. Email provides a rapid means of communication, but at the same time it has many negative impacts. Email is vulnerable to attacks. Some of the common attacks include Spoofing, Denial of service, Replay, Release of message contents, Traffic analysis, etc. Of these spoofing is one of the main forms of attack occurring in email messages. Email spoofing is the sending of email messages with forged sender's address. In this paper, the focus will be on analyzing email header to detect spoofed mails. By analyzing email messages, information about the sender, receiver, transmission time, the date at which mail is sent, etc. can be obtained.

A. Need for Email Forensic Investigation

Email spamming, phishing, drug trafficking, cyber bullying, racial vilification, child pornography, and sexual harassment are some common email mediated cyber crimes etc.[5]. Email messages can also be used to spread viruses, worms and other malicious attacks. SMTP is used for sending email from the sender's client to the sender's server and it is also used for its transmission from the sender's SMTP server to the recipient's SMTP server[4]. Email is transmitted via SMTP protocol. Extraction of evidence forensically from email message can be done using various tools. However, these tools are not designed to detect spoofing of mails. The development of new methods and increase in cybercriminal activities makes it necessary to analyze email messages to detect attacks.

II. ANALYSIS OF AN EMAIL

Email follows a uniform format defined by RFC822[1]. Analysis of an email message is done by analyzing the header of the mail. The email header is the envelope of the mail[2]. Every received email is associated with an email header. Analysis of this email header is the main means of email-related attacks. identification. There is only one header for each email. Basic header fields[2], which have been defined in RFCs include:

- From: contains the email address of the sender.
- To: contains email addresses of the recipient.
- Subject: contains information about the topic of the message.
- Date: contains time and date at which mail was sent.
- Reply-To: contains mail address that is used to reply back by the recipient.
- Message-ID: an automatically generated field, it uniquely identifies the message.
- Received: contains information about mail servers involved in mail transmission which can be used to trace the path of message transmission.

Of these fields, Received and Message-ID[1] fields are of great importance for the identification of the integrity and authenticity.

1) Received: The received field is the most important field in the email header and is usually the most reliable field as it is automatically added by the email servers during email transmission from the sender to the receiver. This provides information about all the servers or computers

978-1-4799-7849-6/15/$31.00 ©2015 IEEE

through which the email traveled from the sender to reach the destination. There are many Received: fields in an email header. These Received: fields are analyzed in the order from bottom to the top. That is, the Received: line at the top of the header contains information about the receiver of the mail. And the Received: line at the bottom contains information about the sender of email[1], domain name, IP address, timestamps etc. at which mail is sent.

2) Message-ID: It is a globally unique identifier used in a mail which helps to prove the authenticity of a mail. The Message-ID is required to have a specific format defined by RFC 2822.The common format is like a valid mail address which is included within a pair of angle brackets. It contains 2 parts connected by @ sign. The left part of @ sign contains the date and time stamps and the right part contains the domain name of the local host. Message-ID is usually generated and added to the email header by the sender's mail transfer agent. Message –ID is unique and cannot be reused[1].

III. DETECTING SPOOFED MAILS

Email is one of the important services provided by the internet and it is an essential communication medium today. At the same time, email is one of the most troublesome services in terms of security. Email is vulnerable to both active and passive attacks. Spoofing is one of the important forms of attacks among them. Sending of a spoofed mail is usually carried out for the purposes of causing embarrassment or the misinterpretation of the individual or organization. Spoofed emails can also be used to spread malware on your system. Normally emails are spoofed by forging sender's mail address given in the 'From:' field of the mail header. But there are chances for other fields in the header to be forged. This paper discusses different ways to detect spoofing in the mail by analyzing the mail header.

Email address spoofing is done by changing the address of the sender of the mail. Email spoofing can be identified by analyzing the date and time stamps available inside the email header. This type of analysis is important when spoofing is done by configuring new mail servers for sending spoofed emails. Another scenario is when, the email header content is spoofed by adding junk Received: fields and other fields into the header. A noticeable difference in time between the system times in different mail servers involved in the sending of an email indicates a spoofing. The proposed technique calculates the time taken to receive the mail[2] from the sender. The time obtained from the calculation is compared with the usual range of time required for receiving a mail from the sending machine. The deviation from this range may denote spoofing.

In order to calculate the time taken by the email, all the timestamps in the mail header are converted to a standard time format, usually to UTC(Universal Time Coordinated) before comparing their difference with the time range. The time difference between sender's time and intermediate servers and the time difference between sender's time and receiver's time is calculated. The time difference calculation between intermediate mail servers helps to know the actual time the mail was sent if the sender's machine's local time was wrongly set. If their differences are out of this range, then there are chances for the mail to be spoofed. The Received: and Date: fields of a legitimate mail header is shown in Fig.1. Table I. shows the date and timestamps of the sender, first server, last server and receiver from mail header shown in Fig.1. Table I. shows no time difference calculated between sender's time and intermediate mail servers and time difference between sender's time and receiver's time from the usual time range required for sending an email.

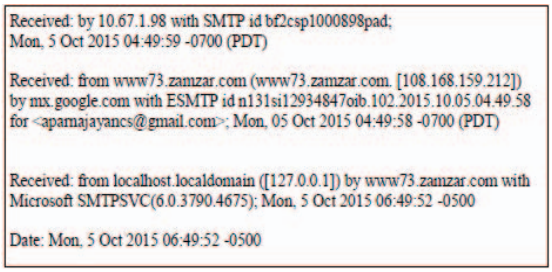

Fig. 1. Legitimate mail header

TABLE I. DATE AND TIME FOR LEGITIMATE HEADER

	Date	Time
Sender	Mon,5 Oct 2015	11:49:52
First Server	Mon,5 Oct 2015	11:49:52
Last Server	Mon,5 Oct 2015	11:49:52
Receiver	Mon,5 Oct 2015	11:49:52

Fig.2 shows the Received: and Date: fields of a time spoofed mail header. Table II. shows the date and timestamps of the sender, first server, last server and the receiver from mail header shown in Fig.2. The date and timestamps given in Table II. show greater time deviation between the intermediate mail servers in sending the mail from the usual time range. Hence, there are chances that spoofing had happened and the email header is needed to be further analyzed to detect spoofing.

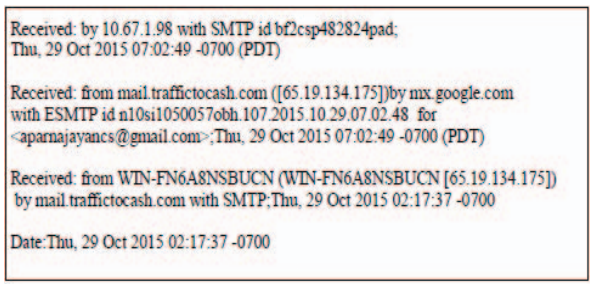

Fig. 2. Time spoofed mail header

Email headers contain domain names inside the Received: field. These domain names can also be used to detect spoofing by doing DNS or Reverse DNS lookup. DNS stands for Domain Name System. DNS provides domain names to any devices or resources connected to the Internet. DNS lookup is the process of looking up the IP associated with a domain name, and Reverse DNS lookup is just the reverse process, finds out the domain name corresponding to an IP address. A Received: field may contain the domain name and IP address corresponding to that domain name. If DNS Lookup is performed with domain name present in the Received field, the IP address obtained should be same as the corresponding IP address in the Received: field. And if a Reverse DNS lookup is performed with the IP Address available in Received: field, domain name obtained should be same as the corresponding domain name available in the Received: field. If there is a mismatch in results, then there are chances that mail is spoofed. Fig.3 shows the domain name selected to perform DNS lookup process.

TABLE II. DATE AND TIME OF A SPOOFED HEADER

	Date	Time
Sender	Thu,29 Oct 2015	09:17:37
First Server	Thu,29 Oct 2015	09:17:37
Last Server	Thu,29 Oct 2015	14:02:49
Receiver	Thu,29 Oct 2015	14:02:49

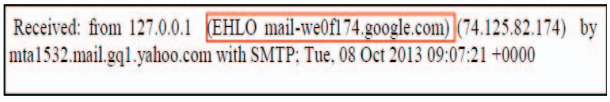

Fig. 3. DNS lookup

The other field used for detecting spoofing is Message-ID: field. It is a globally unique identifier. This field is generated either by the sender or the sender's mail transfer agent. In order to detect spoofing domain name in the message-ID field is compared with the domain name given in the Received line just above it. If they are different, spoofing can happen. Fig. 4 shows the domain name checking to detect spoofing by looking the domain names in the Message-ID: and Received: fields.

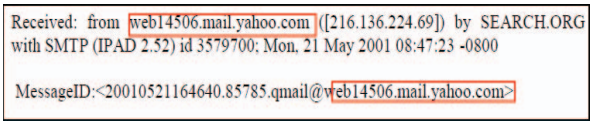

Fig. 4. Domain name checking

The Received: lines in a mail header can be used to detect spoofing. This is done for looking matching of mail servers involved in mail transmission. The mail transfer agent at the sender transfers the mail to mail transfer agent at the receiver's mail transfer agent. The transfer agent at the receiver's side should be responsible for the transfer of mail to the receiver. The mail server that receives the mail should be responsible for the transfer of mail to next server. Fig.5 shows the sender mail server in last Received: line receives the mail and transfers the mail to receiver's server.

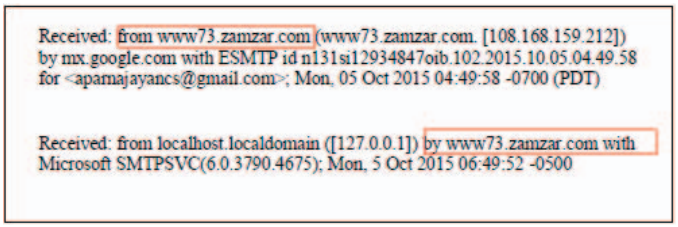

Fig. 5. Mail server matching

A. *Algorithm for detection of spoofed mails*

Step 1: Read the mail header from bottom to top line by line.

Step 2: Identify the different fields in the header to detect spoofing.

Step 3: Analyze the Date: and Received: field to obtain timestamps and convert them to UTC.

Step 4: Compare these timestamps with the standard timestamp needed for a mail communication.

Step 5: Perform DNS or Reverse DNS lookup using the IP addresses and domain names available in the Received: fields.

Step 6: Compare domain name in the Message-ID field with the domain name in the Received: field added by sender's mail server.

Step 7: Look for mail server matching in Received: field.

Step 8: If more than two methods show a mismatch, then the mail can be considered as spoofed.

IV. CHALLENGES

Email is important as it is one of the main forms of communication. It brings great convenience to the people, at the same time it has become a potential carrier of crime related data. Emails can be used to spread spam, viruses and other malware through spoofing. So it is important to identify spoofed mails. The proposed technique in this paper enables to detect spoofing. The major challenge in the proposed technique is that it is applicable only if email header is available for the forensic analysis.

V. CONCLUSION AND FUTURE WORK

Although email is a convenient means of communication, it also has negative impacts. Email has become a potential carrier of criminal evidence. Using of email to spam, and to spread pornography, fraud and other criminal activities has

become increasingly rampant. One of the important forms of attacks on email is Spoofing. Analysis of date and time stamps present in different fields in an email header gives information related to spoofing. In addition to this, the proposed technique uses DNS lookup also to identify spoofed mails by using domain names and IP addresses present in the header. Thus, this technique provides a good solution for the identification of email spoofing. However, when spoofing is not reflected on the mail header, or if the header is not available, this methodology cannot be adopted. Advanced research is needed to find novel methodologies for identifying other types of spoofing.

ACKNOWLEDGMENT

We would like to express our gratitude towards the Resource Centre for Cyber Forensics for giving us the opportunity to work on this engaging project and guiding us by providing the necessary support. We would also like to thank the ER&DCI-IT for being instrumental in the work done by us.

REFERENCES

[1] Hong Guo, Bo Jin, and Wei Qian, "Analysis of Email Header for Forensics Purpose", International Conference on Communication Systems and Network Technologies, 2013.

[2] Preeti Mishra, Emmanuel S. Pilli and R. C. Joshi "Forensic Analysis of E-mail Date and Time Spoofing", Third International Conference on Computer and Communication Technology, 2012.

[3] M. T. Banday, "Analysing E-Mail Headers for Forensic Investigation", Journal of Digital Forensics Security and Law, vol. 6, pp. 49-64, 2011.

[4] M. T. Banday, "Algorithm for detection and prevention of E-mail date spoofing ", International Journal of Computer Applications, vol. 06, pp. 7-11, 2011.

[5] R. Hadjidj, M. Debbabi, H. Lounis, F. Iqbal, A. Szporer, and D. Benredjem,"Towards an integrated E-mail forensic analysis framework", Digital Investigation, vol. 5, pp. 124-137, 2009.