An Overview to the Robust and Secure Evidence-Gathering Server for the Digital Forensic

SmitaVerma
Department of
Computer science and engineering, RITS BHOPAL,
INDIA

Anurag Jain
Department of
Computer science and engineering, RITS BHOPAL,
INDIA

ABSTRACT

Since the advent of the World Wide Web in 1990, the usage of Internet over worldwide has grown from roughly 2.6 million users (0.05% of the world population) in 1990, to roughly 2.0 billion users (30% of the world population) in 2010 (The World Bank Group, 2012). This trend is expected to continue for the foreseeable future (Cisco Internet Business Solutions Group as cited in Evans, 2011). With this, websites have become increasingly important in the lives of individuals worldwide. There security is also a big issue for the professionals.In this paper, we would like to propose a method for maintaining & managing a server called an “evidence-gathering server”. This evidence-gathering server extracts the log data from all the nodes and servers of clustered area. we are also using the concept of hashing, this hash value will always associated with every binary object of logs. In this way, we will provide a single place to get all the network level evidences. This will help the forensic analyst to analyse and reconstruct the activity, and give results in faster time.

Keywords

Digital forensics, evidences, evidence preservation, web security.

1. INTRODUCTION

Over the past decade, with the rapid growth in Internet, the arrival of blogs, virtual communities, online office, e-commerce, e-government, B2B and C2C and other emerging Web applications, the Web has become one of the core elements of human life and work. How can we enhance the value of the Web site, allowing users a better experience, and quickly find the information we need to find the user's needs? How can we improve the competitiveness of e-commerce applications and to survive in the fierce war of the Internet? These issues require answers we can find in the vast amounts of Web data. Thus, the combination of data mining technology and Internet applications constitute a very active and very important a field of study, in other words, Web mining.

Having a similar structure and content of the access log file on each Web server, Web logs automatically become an important data source for Web mining and its mining has a universal and practical significance. However, the large amount of web log data, containing a lot of noise, not suitable for Web mining, must first be pre-treated. The workload of data pre-processing accounts for more than 50% of the total web mining workload.Web forensic is the use of technology and tools to investigate and establish facts to facilitate decisive action in cyber space. The main aim of the web forensic is to find digital evidences in networking environment. The digital evidence is any data in digital storage that can be used as a proof of the criminal behavior. The Forensic analysis is a process of understanding, re-creating, and analyzing arbitrary events that have gathered from various digital sources.

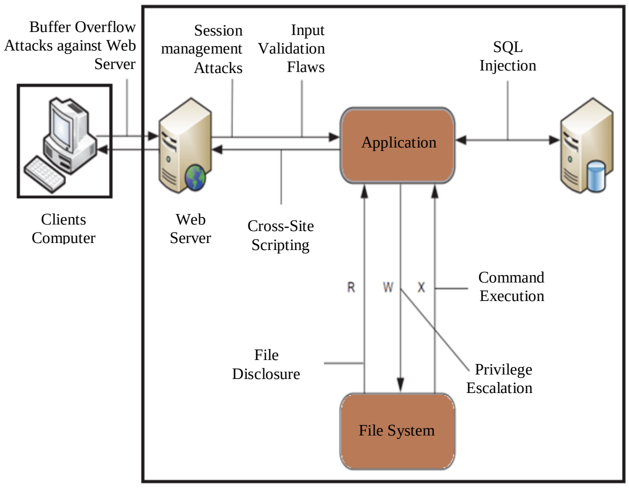


Figure 1: Web Application Architecture

The forensic analysis can be applied to anywhere in the cyber security, therefore can be applied to media: analysing physical media to have some evidence, code: Analysis of software for potentially harmful signatures and network: identify network traffic and logs to locate the activity of cyber criminals. On the other hand, the possibility of manipulation or deletion of log information (in short, log info) or log file eras ability itself is increasing. Because log files are incriminating evidence against attackers, these files are at risk of attacks. Therefore, a mechanism is needed to prevent the manipulation and deletion of log info and log files by attackers and maintain the contents of log files that are created at the time of outbreak. Log files are saved in the hard disk or RAM of each system equipment or are transmitted to log servers using a protocol (e.g., Syslog) and saved in them.

2.DIGITAL FORENSICS

Lexical meaning of "forensics" is that investigate a case using the scientific and technological method in a criminal of civil trial, and it is methods that prove some truth [3]. That is, the digital forensic is a process that proves truth based on digital data built in a PC, web and mobile phone etc. For having legal force as an evidence in a law court, data collected in forensic is must handled safely.

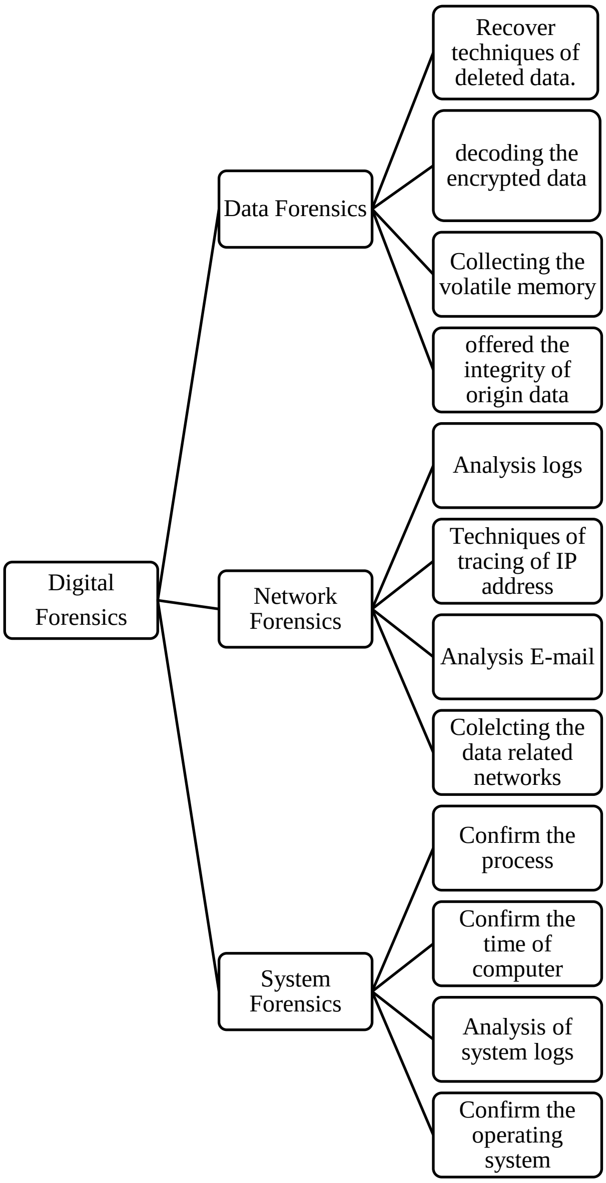

Digital Forensics
Data Forensics
Recover techniques of deleted data.
decoding the encrypted data
Collecting the volatile memory
offered the integrity of origin data
Network Forensics
Analysis logs
Techniques of tracing of IP address
Analysis E-mail
Colelcting the data related networks
System Forensics
Confirm the process
Confirm the time of computer
Analysis of system logs
Confirm the operating system

Figure 2: Forensics Analysis

Cyber Crime
Cybercrime is a term for any illegal activity that uses a computer as its primary means of commission. The cybercrime also include any illegal activity that uses a computer for the storage of evidence.

The cybercrime or computer crime basically refers to any un-lawful acts that involves a computer or any networking device and a information networks, where these inter networking devices may or may not have played an instrumental part in the commission of the crime. The Issues related with type of crime have become high-profile, particularly those surrounding copyright infringement, hacking, child grooming and child porn, intellectual property crime, online gambling, e-mail spoofing, cyber defamation, forgery and cyber stalking. The computer may however be target for criminal and unlawful acts in the certain cases like- unauthorized access/networks/computer system, theft of information contained in the electronic form, e-mail bombing, logic bombs, internet time thefts, Trojan attacks, and web jacking physically damaging the computer system and theft of computer system.

Cyber Forensics
It can be defined as the collection and analysis of data from computer systems, networks, communication streams (wireless) and storage media in a manner that is admissible in a court of law (International Judicial system). It is a combination of the streams of computer science and the law. This definition applies to the collection of information in real time, as well as the examination of latent data. The objective in cyber forensics is quite straight forward. It is to recover, analyze and present computer based material in such a way that it is useable as evidence in a court of law. Cyber forensics involves the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of computer evidence stored on a computer.

3. LOG FILES

In dynamic systems such as the Internet (Wide Area Network), this is a common practice to keep record samples of activity in a periodic manner [3]. Those samples are then used to characterize the activity in the system and to evaluate new mechanisms to be used in this system. This is certainly true of HTTP traffic. On the World Wide Web (WWW), logs of HTTP traffic are recorded continuously as a function of most origin web servers as well as intermediate servers and proxies. The main function of these logs is to chronicle the operation of such systems. However, as mock-up of HTTP activity, logs generated by these systems are also used for characterization, evaluation and usage reporting. Sometimes, the researchers capture the HTTP traffic via other sources, such as from augmented client browser. Web Server logs are the plain text files in the ascii codes and independent from the server platform. There are some distinctions between server software, but conventionally there exist, four types of server logs:

1. Transfer Logs
2. Agent Logs
3. Error Logs
4. Referrer Logs

In the context of server logs, the first two types of log files specified above are standard. The referrer and agent logs may or may not be “turned on” at the server or may be added to the transfer log file to create an “extended” log file format. Each HTTP protocol transaction, weather completed or not, is recorded in the server logs and some of the transactions are recorded in more than one log.

Access Log
The Agent Log provides data on a user's browser, browser version, and operating system. This is the significant information, as the type of browser and operating system determines what a user is able to access on a site (e.g. Java, forms).

Error Log
The average Web user will receive an "Error 404 File Not Found" message several times a day. When a user encounters this message, an entry is made in the Error Log. Below is a sample Error Log entry

Referrer Logs
The Referrer Log indicates what other sites on the Web link to a particular server. Each link made to a site generates a Referrer Log entry.As the log files are managed by the web servers, we will protect web log files by converting them into an image and then encrypt that image. So that, it becomes more secure from any tampering. This whole technique is transparent to all the users. Hence, no one can easily locate the image form of log file. This process converts the log file periodically into the image as per the web server owner policy.In addition, we can detect the tampering errors in image, if occurs and try to reconstruct the original image.

For example:
date time s-sitename s-computername s-ipcs-method cs-uri-stem cs-uri-query s-port cs-username c-ipcs-version cs(User-Agent) cs(Cookie) css(Referer) cs-host sc-status sc-sub status sc-win32-status sc-bytes cs-bytes time-taken

2013-02-22 00:09:35 W3SVC9613 H11-SECUREHOST 209.59.179.25 GET /RIT_NEW/d_cse/images/amster6.jpg - 80 - 66.249.73.248 HTTP/1.1 Googlebot-Image/1.0 - - www.radharaman.com 200 0 0 228405 262 1403

Considering the IIS server of Microsoft whose log file contents are as follows:

The log files are stored in a customizable ASCII text-based format. With the help of IIS Manager, we can select which fields to include in the log file, it allows us to keep log files as small as possible. Because HTTP.sys manages the W3C Extended log file format, this format records HTTP.sys kernel-mode cache hits.

Field	**Appears As**	**Description**
Date	Date	The date on which the activity occurred.
Time	Time	The time, in coordinated universal time (UTC), at which the activity occurred.
Client IP Address	c-ip	The IP address of the client that made the request.
User Name	cs-username	The name of the authenticated user who accessed your server. Anonymous users are indicated by a hyphen.
Server IP Address	s-ip	The IP address of the server on which the log file entry was generated.
Server Port	s-port	The server port number that is configured for the service.
Method	cs-method	The requested action, for example, a GET method.
URI Stem	cs-uri-stem	The target of the action, for example, Default.htm.
URI Query	cs-uri-query	The query, if any, that the client was trying to perform. A Universal Resource Identifier (URI) query is necessary only for dynamic pages.
HTTP Status	sc-status	The HTTP status code.
User Agent	cs(User-Agent)	The browser type that the client used.
Protocol Sub status	sc-sub status	The sub status error code.

4. LITERATURE SURVEY

Data pre-processing is an important activity for discovering behavioural patterns [8]. The analysis of web logs is a significant task for the System Administrators to provide the safety to adequate bandwidth and to maintain capacity of servers on their organization websites. The web Log file represents activities of users occurring over a time span. These web log files offer valuable insight into the effective usage of the web applications. It helps to maintain an account of the actual usage in a real world working system as compared to the virtual setting of the usability lab. This research paper emphasizes on the pre-processing techniques developed at a specially designed Web Sift (WebIS) tool on an IIS web server. The authors also proposes some other efficient heuristics and techniques.

Another research in [9], shows the analysis of Web log servers are needed in many applications and can be useful for analysts and designers of computers networks, and are also for some interesting research problem. Conventionally, some statistics are computed & used for design and analytic purposes. In this research, the authors presents the uses of results of Web server logs and their data analysis/mining through linguistic data summaries, which is based on fuzzy logic with linguistic quantifiers. The Linguistic summaries of both static and dynamic analyses are presented in this paper, with an emphasis on the latter.

In Web environment, a major challenge facing by the law-enforcing agency is to collect effective & accurate evidences from the huge volumes of crime data. In the field of cyberspace multistep attack, it involve group of action where some of these may be lawful but when combine together constitute malicious activity. The Code injection attack [1] is a type of multistep attack, which may be carried out by potentially malicious (unlawful) invaders by inserting script code and SQL statement into available textboxes (data supplier to database) on vulnerable web site.

In Network Digital Forensic, the log maintenance method is a method in which a reliable third party (TTP: Trusted Third Party) stores and keep the log safely is proposed in [6]. Further, the related guarantee method that considers the time order of the information in each log entry when plural entities exist is proposed in [7]. Time-stamp service as the method using HTTP, has been proposed to ensure compliance for law such as SOX. However, there are some problems related to the cost of data capacity, increase in traffic, and security in the case that all log files from all system equipment of a single organization are saved. Further, a distributed file saving system was proposed in [5]. However, there is a risk that a log file may be manipulated or deleted before the information from this file is saved into the distributed files.

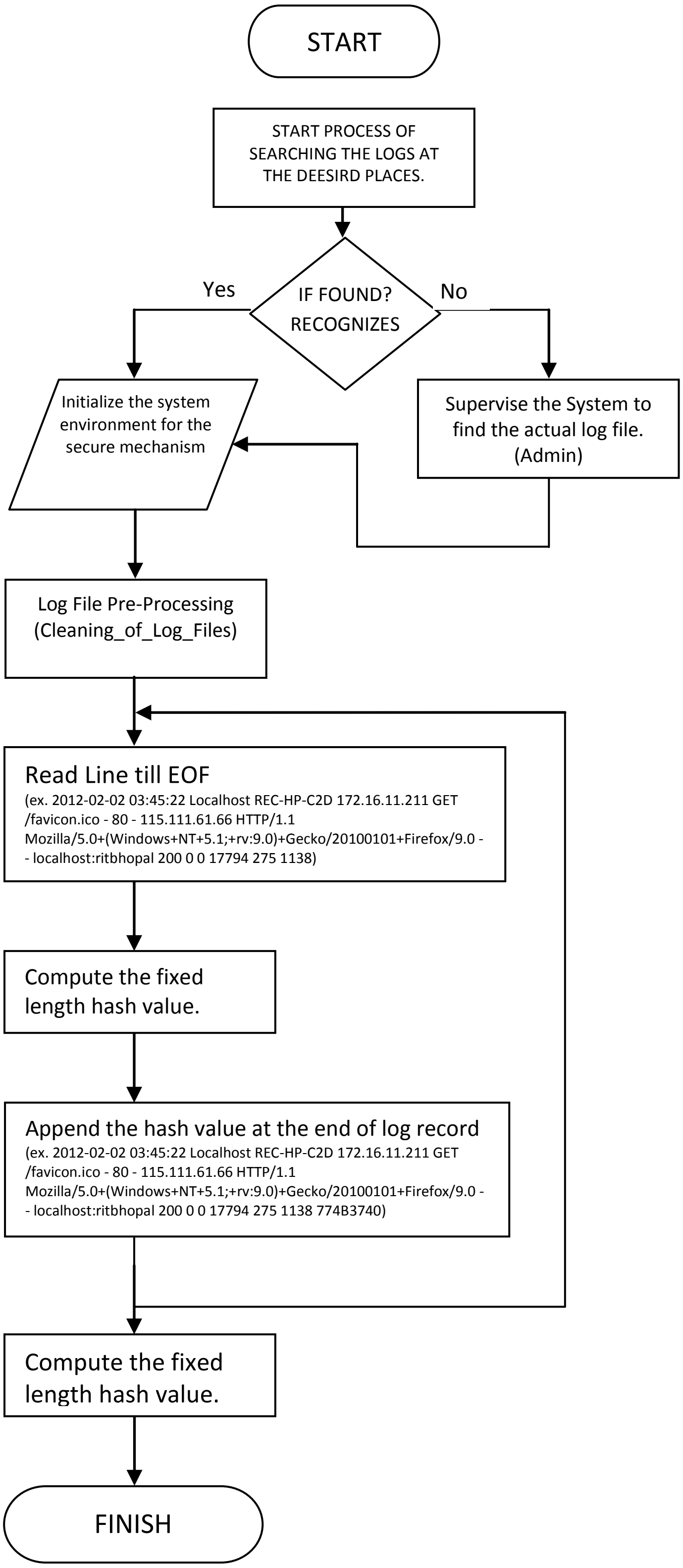
In the research [1] architecture for gathering evidence subjected to code injection attack is proposed. On the other hand, the work, described [2] focuses upon the correlation of various sources of evidences, protection of evidences and preservation of evidences in cyber space. In system equipment, a mechanism to prevent the manipulation or deletion of log information and log files by an attacker and to maintain the contents of log files are essential. Because log files include an active event and an operation event in the system equipment, these files are at risk of attacks such as file manipulation or deletion. In this paper, the authors propose a log file security management method using hash values. It is possible to detect the manipulation or deletion of log information and log files.

In [10], the authors present a method that allows for securely saving a temporal sequence of data (log lines) in a file. Log lines are signed by an authority, and are thus unalterable with no detection. The Data is also encrypted (secured) in the file, and may be accessed with the granularity of a single log line with the possession of a decryption key. Also, it is possible that for some lines data must be accessed by a group of cooperating users. In this paper, the authors deals with the problem of keeping the content of a log file both unalterable without detection) and private.

5. PROPOSED WORK

The proposed research is partially based on the work described in [1][2], according to J.L.Rana et. Al. [1] the code injection attack is a Multi-step attack which generate the data into multiple locations. This is necessary to reconstruct the activity later during the forensic analysis. Therefore, rather than store the log data, activity records at different places we proposed a methodology that make the clusters of network and in each cluster, we will manage a server called an “evidence-gathering server”.

This evidence-gathering server extracts the log data from all the nodes and servers of clustered area. This action repeated again after a fixed periodic span. This approach will reduces the time to extracts the information from the remote computer, as all the necessary information already replicated at the evidence-gathering server in the form of binary objects.


START
START PROCESS OF SEARCHING THE LOGS AT THE DEESIRD PLACES.
IF FOUND? RECOGNIZES
Yes
No
Initialize the system environment for the secure mechanism
Supervise the System to find the actual log file. (Admin)
Log File Pre-Processing (Cleaning_of_Log_Files)
Read Line till EOF
Compute the fixed length hash value.
Append the hash value at the end of log record
Compute the fixed length hash value.
FINISH

```
(ex. 2012-02-02 03:45:22 Localhost REC-HP-C2D 172.16.11.211 GET
/favicon.ico - 80 - 115.111.61.66 HTTP/1.1
Mozilla/5.0+(Windows+NT+5.1;+rv:9.0)+Gecko/20100101+Firefox/9.0 -
- localhost:ritbhopal 200 0 0 17794 275 1138)
```

```
(ex. 2012-02-02 03:45:22 Localhost REC-HP-C2D 172.16.11.211 GET
/favicon.ico - 80 - 115.111.61.66 HTTP/1.1
Mozilla/5.0+(Windows+NT+5.1;+rv:9.0)+Gecko/20100101+Firefox/9.0 -
- localhost:ritbhopal 200 0 0 17794 275 1138 774B3740)
```

Figure 3: Proposed Algorithm

On the other hand, after storing the evidences at the "evidence-gathering server", the main question arises for the security measures. Therefore, we would like to apply the hash based security method of Fumiharuet. al. [2], according to Fumiharuet. al. log files include an active event and an operation event in the system equipment, these files are at risk of attacks such as file manipulation or deletion. Therefore, they use a dispersion method, that creates the replica of logs at different places.

Whereas in the proposed research, we are using the concept of hashing, this hash value will always associated with every binary object of logs. In this way, we will provide a single place to get all the network level evidences. This will help the forensic analyst to analyze and reconstruct the activity, and give results in faster time.

According to the proposed work, the algorithm of the robust and secure evidence-gathering server is represented in the figure 3. In which it is easily seen that, the system start searching for the log files in the server, if found the necessary log it simple provide the log files to the data pre-processing section. From where, the operation of data cleaning is to be done.

After cleaning of the data, the remaining that is a quality data, that doesn't contain the unwanted or irrelevant information. This cleaned log file are to be read by the security process, that start reading the log file line by the line (record by record). Each line is processing by the hashing technique that generated a fixed length hash code. This hash code is appended at the end of the line. This is how the process of providing security to the logs are monitored.

6. CONCLUSION

During the course of our research, several log files that could be valuable to the development of an encompassing forensic server-evidence gathering log records large discovered. With the use of proposed system, one can manages the log files with the security and anti-alteration protection, the proposed technique allow the admin to check the validity of the evidences for the next phases of investigation. The advantage of the proposed technique is to detect the tampering of digital evidence, this technique allow the user only to check the data at the tampered location not for the whole log file. Although this method needs more refinement for the attributes like processing overhead, time consuming and storage album of all log files.

7. REFERENCES

[1] Deepak Singh Tomar, J.L.Rana, S.C. Shrivastava, *"Web Forensics System on the Basis of Evidence Gathering with Code Injection Attack"*, in International Journal of Computer Science & Communication, Vol. 1, No. 2, July-December 2010, pp. 313-315.

[2] FumiharuEtoh, Kenichi Takahashi, oshiaki Hori, Kouichi Sakurai, *"Study of log file dispersion management method"*, in 10th Annual International Symposium on Applications and the Internet, 2010, pp. 371-374.

[3] Warren G.Kruse II, Jay G.Heiser. "COMPUTER FORENSICS:Incident Response Essentials", Addison Wesley.

[4] Robert Rinnan "*Benefits of Centralized Log file Correlation*" Master‟s Thesis, Master of Science in Information Security30 ECTS, Department of Computer Science and Media Technology Gjøvik University College, 2005.

[5] H. Tomori, S. Tezuka and R. Uda, "A proposal of a distributed file backup system for digital forensics [in Japanese]," Computer Security Symposium (CSS 2008), Oct. 2008.

[6] B. Schneier and J.Kelsey, "Cryptographic Support for Secure Logs on Untrusted Machine," Proc. of the 7th USENIX Security Symposium, Jan. 1998, pp.53–62.

[7] M. Ando, K. Matsuura and A. Baba, "An analysis of ensuring order of log entries in distributed environment [in Japanese]," Computer Security Symposium (CSS 2002), Oct. 2002.

[8] K. Sudheer Reddy, G. ParthaSaradhiVarma, "Preprocessing the web server logs: an illustrative approach for effective usage mining", ACM SIGSOFT Software Engineering Notes archive Volume 37 Issue 3, May 2012 Pp 1-5.

[9] Zadrożny, S., Kacprzyk, J.: From a static to dynamic analysis of weblogs via linguistic summaries. In: Proc. of 2011 IFSA World Congress, pp. 110–119 (2011).

[10] Francesco Bergadano, DavideCavagnino, Paolo Dal Checco, Pasquale Andrea Nesta, Michele Miraglia, and Pier Luigi Zaccone, "Secure Logging for Irrefutable Administration", International Journal of Network Security, Vol.4, No.3, PP.340–347, Mar. 2007.