An Implementation of Robust and Secure Evidence-Gathering Server for the Digital Forensic

Smita Verma, Anurag Jain

Abstract— In this paper, we implement a novel method for maintaining & managing a forensic server called an "evidence-gathering server". This evidence-gathering server stores the digital evidences in a secured way by applying hash mechanism to the log files. The concept of hashing we used, ensures the authenticity, the set of hash values are associated with every binary object of logs. With the use of evidence gathering server, we will provide a single place to get all the network level evidences. The implemented technique also ensures the integrity, privacy of the forensic evidences for security purposes. This will help the forensic analyst to analyse and reconstruct the activity, and give results in faster time.

Index Terms— Digital forensics, evidences, evidence preservation, web security, logs, log files.

1 INTRODUCTION

In this internet era the websites on the internet are a very useful information source for day-to-day activities. So there is a faster development of the World Wide Web (WWW) in its volume of traffic and the complexity of web sites. As per December 2012, according to the Web Server survey by Netcraft [12]. there are 633,706,564 sites - an increase of over 8 million since November. As per the increasing needs of the means to use the highly sophisticated applications over the Internet, recently developed large number of web applications has exploded for use across many platforms. Concern for web application security has grown dramatically as compared to the other applications.

The evidence is often in the form which is hard to find and also the investigation of cyber-crime is difficult, the collection, analysis, appreciation and preservation are the unique challenges for the Investigator. The increased in use of computer networks and the growth of the Internet have added to this complexity. Using the Internet, it is possible for a person sitting at a location to steal a computer resource at remotely located computer for his attack. To handle such challenges in attacks are not only technological, but it is also jurisdictional.

For Web-mining, Web-logs are the vital data source and this web-mining has a universal and practical significance. However, a huge amount of web log data,which contains a lot of noiseis not suitable for Web mining and must first be pre treated. The total web- mining workload for more than 50% accounts for the workload of data pre-processing. Web forensic is the use of technology and tools to investigate the crime accounts for the workload of data pre-processing. Web forensic is the use of technology and tools to investigate the crime and establish facts to facilitate decisive action in cyberspace.

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 behaviour. The Forensic analysis is a process of understanding, analysing and re-creating arbitrary events gathered from various digital sources.

Although log files were not built Specifically for forensic purposes, they are the most likely of all files resident on a system to contain the majority of evidentiary information. If log files do not contains evident information, they may provide links to the investigator to achieve an authentic representation of the virtual environment at the time of crime, known as the "cyber-crime". To show an error on applications during the run time, logs were generated by the developer. Both, normal or erroneous operation of an application can be indicated by the Events contained in these log files.

Non-standard data types, event formats, rudimentary event, and formats of file whose probity cannot be checked, are the results of this application specific log file. Managing, analyzing and organizing large files, also how to report their results in an format acceptable by court of law are some other problems faced by log analyst. Currently, the best method of practice amongst log analysis professionals or system administrators is their own empirical research which uncovers event anomalies amid the vast collection of normally logged events.

Analysis of log files is an essential process that should be done during all computer forensic investigations, but this task is the most thorough processing, due to the nature and potential size of various log files.

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.

- *Smita Verma M.Tech [CSE] Student, Department of Computer Science & Engineering, Radharaman Institute of Technology and Sciences,RGTU, Bhopal,India PH-09893679140. E-mail:smitaverma323@gmail.com*
- *Anurag Jain, Professor, Department of Computer Science & Engineering, Radharaman Institute of Technology and Sciences,RGTU, Bhopal,India E-mail:anurag.akjain@gmail.com*

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. Log-files are at a risk of intrusion, as these are the most important evidence against attackers. So, a mechanism is required to manage the data of log files at the time of out-break and to prevent the manipulation or deletion of log info and log files by attackers. 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.

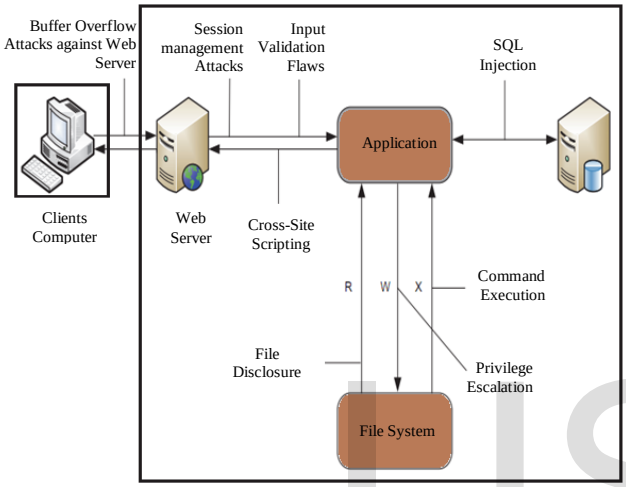

Figure 1: Web Application Architecture

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.

2.1Cyber Crime

The word Cybercrime is a term use for any illegal act that uses a computer as its primary source of commission. Any criminal activity in which computer is used either as an target, instrumentality or a means for perpetuating further more crimes includes within the ambit of cyber-crime.

The cybercrime or computer crime basically refers to any unlawful 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 execution 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 aim for unlawful and criminal acts in some 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.

2.2 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 is applies to the compilation of information in real time, as well as the examination of masked data. The objective in cyber forensics is quite straight forward. It is to identify, recover, analyse and present computer based digital material in a way such that it is useful as evidence in a court of law. Cyber forensics involves the collection, identification, validation, preservation, analysis, interpretation, documentation and presentation of digital evidences 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]. These samples are generally used for the characterization of activity in the system and to assess new methods to be used in this system. This is certainly true of HTTP traffic. On the Internet, World Wide Web (WWW), HTTP traffic logs are recorded continually as a function of most web servers as well as intermediary servers & proxies. The main function of these logs is to chronicle the operation of such systems. However, as mock-up of HTTP activity, the logs generated by these systems are also used for particularization, evaluation and usage reporting. Sometimes, the researchers capture the HTTP traffic via other sources, such as from augmented client browser. Web Server logs are independent from the server platform and are the plain text files in ASCII codes. 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 log files specified above are standard. The agent logs and referrer logs may be or may not be “turned on” at the servers or may be added to the transfer logs to frame the logs into an “extended” log file format. Each HTTP protocol transaction, weather completed or not, is recorded in the server logs and sometimes the server is tuned to record the transactions logs in more than one log.

3.1 Access Log

The data is provided by the agent log on an operating system, browser version, and user’s browser. This is the valuable information, as the type of operating system and browser determines what a simple user is able to access on a site (e.g. Java, forms).

3.2 Error Log

"Error 404 File Not Found" the average web user receive this error message several times a day. An entry is made in the Error Log whenever this message comes to any user.

3.3 Referrer Logs

The Referrer Log represents what other websites on the WWW link to a particular server. Each and every link made to a site create 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. No one can easily locate the image form of log file since this whole technique is transparent to all the users. 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-site name s-computer name s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip cs-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

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 consequential insight into the dominant 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 (Web IS) tool on an IIS web server. The authors also proposed 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 was propsed 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. And the work, described in research [2] targets on the correlation of various issues of evidences, such as sources, protection and preservation in cyber space. In this system equipment, a system is needed to defend the deletion or manipulation of log files and information by an attacker and maintaining the contents of log files are requite, because almost all the log files include an active event and an operation event in the system environment, these files are at risk of attacks such as file manipulation or deletion. In this paper, the authors propose a security management method of log files using hash values. Deletion or manipulation in the log files or information is possible to detect.

In the year 2009, Rafael Accorsi proposed "Safekeeping Digital Evidence with Secure Logging Protocols: State of the Art and Challenges" [13]. This deals with secure logging in digital forensics. In consequence, more often than not, digital evidence based on log data can be successfully challenged in the court, leading to inadmissibility or loss on probative force.

Intensive research into logging protocols is needed to advance digital evidence, in particular into the elucidation of further, fine-grained security requirements and into the verification of protocols with regard to these requirements. They intend to address these issues and investigate the trustworthiness of log data and the intricacies delegation causes to secure logging protocols, two relevant topics neglected in our analysis due to space constraints. Moreover, in the course of our analysis realized that a finer-grained view of admissibility systematically characterizing what is admissibility across different legislations is needed to analyse protocols and improve the expressiveness of our classification.

In the year 2010 Wu & Wang proposed "Tamper Resistance Protection of Logs Based on Forward-Secure" [14], it proposes the forward digital signature scheme as the protection for damage resistance of the server logs and deals with digital forensic towards protected log files. The attacker cannot recover the plain message of the sign- encryption of this and past time, even though the private key of the signer is revealed and thus he cannot damage the logs that are encrypted before the key exposure.

To maintain the system security, monitor system activity and to keep the system healthy, Integrity and security of log files are the essential aspects. This article describes and analyses the traditional protection method of logs, and points out the disadvantages. Then present an asymmetric forward-secure digital signature scheme to make up the defect.

In [10], the authors present a technique that allows for securely storing a temporal sequence of event records (log lines) in a file. Each log records are checked and signed by an authority, and therefore they are unalterable with no detection. The Data is also encrypted (secured) in the file, and may be accessed by the granularity of a single log line with possession of the 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.

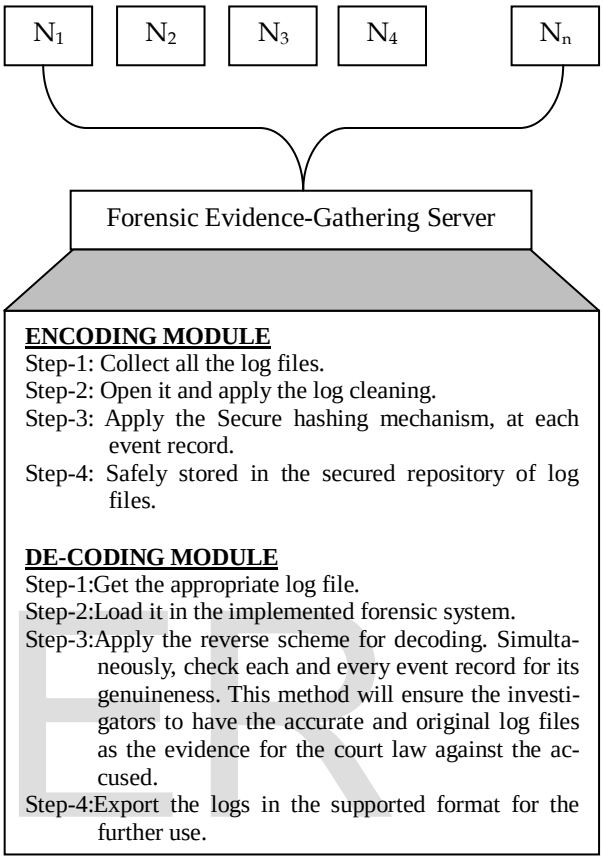

Figure 2: Algorithm of Proposed evidence-gathering server

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 Fumiharu et. al. [2], according to Fumiharu et. al. log files include an active event and an operation event in the system equipment, log files are at the risk of being deleted or manipulated. So, we create the copy of logs at different places, using dispersion method.

Whereas in the proposed research [11], 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 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 [15].

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. IMPLEMENTATION

The proposed scheme implemented by developing an application, using Microsoft .Net Framework 4.0, Visual Studio 2008 which is tested on windows environment with IIS 6.0. As the nature of both the work is synergetic, there is always a lot of backing from mailing lists and documentation. Expeditiously the bugs are fixed, any requests made for features are always heard, evaluated, and if practicable, it is implemented. The implemented system works according to the proposed approach [11], designed in two modules:

1- Encoding Module

With computer data on magnetic media, investigation of cyber forensic is not very simple process. Physical stresses such as magnetism, extremes of heat and cold, and, even, physical or electrical shock are encountered by magnetic media. Additionally, if one can gain access over it, it is very simple to alter logical evidence. The evidence has been altered can be very difficult to prove, because the alteration may leave no indication that it ever occurred.

To assure that the evidence used by the investigators in court is the same evidence they collected, we need to be able to mark it logically and seal it in such a manner that it simply is not accessible to anyone, except administrators and investigators. So that nobody can create their own, slightly different (presumably, to their benefit) version of the evidence, and present it as identical to actual, Therefore, there is a need of establishing a method, that all evidence meets the best evidence rule and that it is all identical to the original.

This module helps the administrator to make a safe and secure repository of log files, which helps the forensic investigators during the investigation of any cyber-crime scene. Although the concept of evidence gathering server is simple but very effective, as there is a secured copy of log files are stored in an organised manner. Hence, the forensic investigator does not need to check the individual nodes of the network because, all it gets from the proposed evidence gathering server.

2- Decoding Module

Log examination is probably the single most productive part of your investigation if the logs are kept properly. It is also very tedious, especially when the logs are from multiple machines and are thousands of lines long.

The decoding module allows the administrator to make available the resources to the forensic investigators to their investigation by ensuring that the evidences stored at evidence gathering server are safe, secure and unaltered. So, that the investigators can use it for investigation and can be use as prime evidence for the judicial system against the accused.

Figure 3: GUI of proposed software

Figure 4: Encoding Module-Takes the log file and apply the security

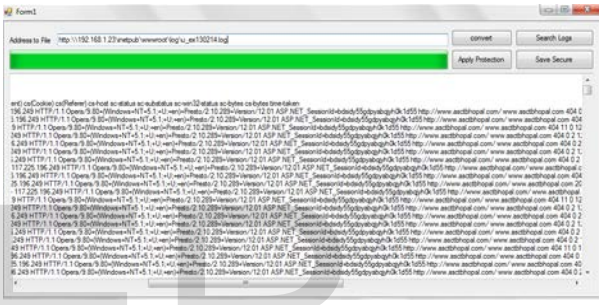

Figure 5: Encoding Module - Reading of log file

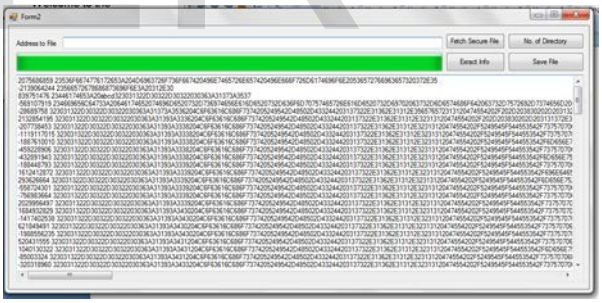

Figure 6: Decoding Module - Reading of Secure file

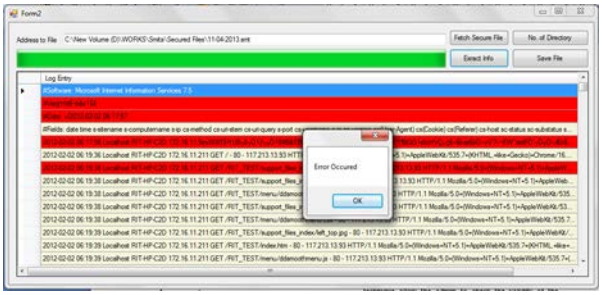

Figure 7: Decoding Module - Extraction process - Altered entry

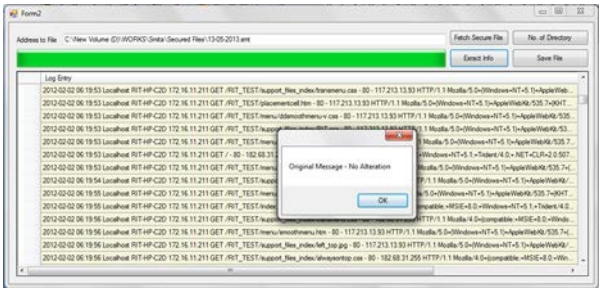

Figure 8: Decoding Module - Extraction process - No Altered content

Table 1: Results Obtained from the proposed approach

Log Files	Encoding Time	Size	Decoding Time	Size
u_ex130415.log	2 ms	105 kb	4 ms	95 kb
u_ex130115.log	2 ms	135 kb	5 ms	90 kb
u_ex130315.log	3 ms	301 kb	7 ms	290 kb
u_ex130615.log	4 ms	537 kb	7 ms	480 kb
u_ex130515.log	4 ms	685 kb	9 ms	610 kb

Table 2: Comparative Study with other approaches

Features	**Proposed**	**Log Dispersion Method [2]**	**Sudheer Reddy [15]**	**Tamper Resistance [14]**
Log Cleaning	Yes	-	Yes	Yes
Security	Yes	Yes	-	Yes
Security With hash Code	Yes	Yes	-	Yes
Zero Redundancy	Yes	-	-	-

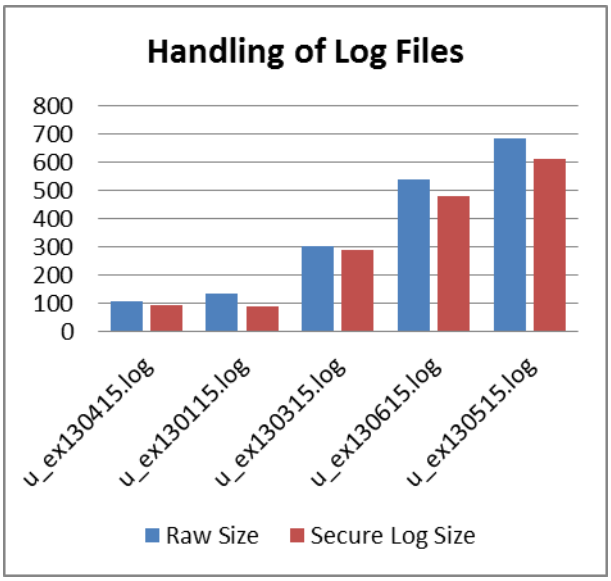

Figure 9: Handling of log files at evidence gathering server

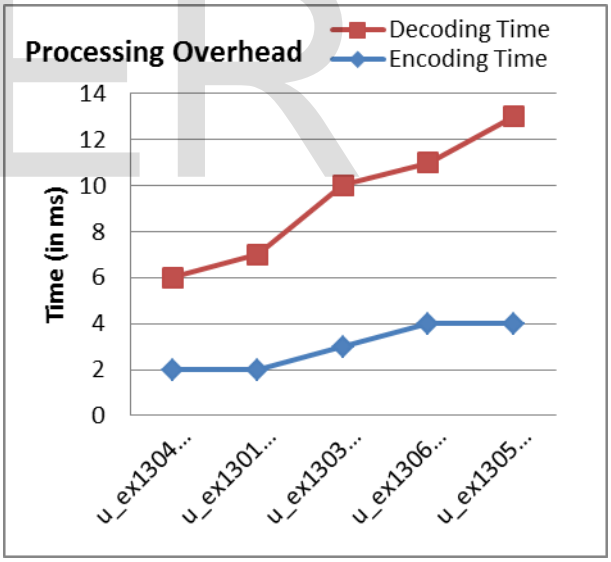

Figure 10: Time overhead in processing of log files

Experimental results are carried out on log files taken from web server of different time-periods. The table 1, shows the actual experimental scenario, where one can easily find out the performance of proposed system. While in table 2, it is clearly seen the novelty of the proposed approach, as it provides security with very less processing overhead and zero redundancy.

7. 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.

8. 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. Fumiharu Etoh, Kenichi Takahashi, oshiaki Hori, Kouichi Sakurai, *"Study of log file dispersion management method"*, in 10th Annual International Symposium on Applications and the Internet, IEEE, 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. Partha Saradhi Varma, "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, Davide Cavagnino, 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.
11. Smita Verma, Anurag Jain, "An Overview to the Robust and Secure Evidence-Gathering Server for the Digital Forensic", International Journal of Computer Applications (0975 –8887) Volume 71–No.15, May 2013.
12. December 2012 Web Server Survey, http://news.netcraft.com/archives/2012/12/04/december-2012-web-server-survey.html, on 18/07/2013 23:08.
13. R. Accorsi, Safe keeping digital evidence with secure logging protocols : state of the art and challenges, in: O. Goebel, R. Ehlert, S. Frings, D. Günther, H. Morgenstern, D. Schadt (Eds.), Proceedings the IEEE Conference on Incident Management and Forensics, IEEE Computer Society, 2009, pp. 94–110
14. Zhiyongwu, Bin Zhuge Weiming Wang "Tamper Resistance Protection Of Logs Based On Forward-Secure" Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on Volume:8, Date 9-11 July, 2010 Pp 90-94.
15. K. Sudheer Reddy, G. Partha Saradhi Varma, I. Ramesh Babu, Preprocessing the web server logs: an illustrative approach for effective usage mining, ACM SIGSOFT Software, Volume 37 Issue 3, May 2012, Pages 1-5.

IJSER