A Method Using log file analysis and reconstruction to Understand End-User classification

Ritul Saraf , Vishal Shrivastava
Computer Science & Engineering, Vedica Institute of Technology, Bhopal

Abstract — Computer crime is a crime involving computers and networks. Investigating crime in a networked environment is a tedious task. Event registration and event logs include the latest IT Critical investigation, when the last user interacts with the web environment and stores in different logs such as the client side firewall logs, gateway network logs, and server-side logs, play an important role in the system. But log files should not be stressed enough as a source of information in the system and network management. Whereas separate log files must be linked for the purpose of actively testing and collecting information. The task of parsing event log files has become inferior to continuing with the increasing size and complexity of today's logbook event. Now there is a light that one day automatically analyzes these log files. This research contains an innovative method that has been used to build a series of evidence on the basis of short and temporary series and relational algebra, and to process real generation data from logs and create Solar rules based on a number of evidence and Pre-processing classifies the actual generated data and user from the log based on the Markov model. A growing variety of data processing processes involves the audit of large log files and therefore requires processing tools and technical solutions. In the event of emergency response, human analysts have to process large amount of log data in order to detect suspicious activity and add additional evidence.

In many cases, after identifying some additional facts, the reaction of this incident is stopped. Detecting online phenomena is very difficult for many reasons. Many application-specific log formats also require deep domain-specific knowledge to correctly configure an existing rules-based event-based event engine. Secondly, provide the exact model of abuse or effective identification algorithms are required.

Keywords— Computer crime, event logs, process, event based..

I. INTRODUCTION

Event logging and event logs play an [1] important role in modern IT systems. Today, many applications, operating systems, network devices, and other system components are able to log their events to a local or remote log server. For this reason, event logs are an excellent source for determining the health status of the system, and a number of tools have been developed over the past 10-15 years for monitoring event logs in real-time. However, majority of these tools can accomplish simple tasks only, Event correlation is one of the most prominent real-time event processing techniques today. It has received a lot of attention in the context of network fault management over the past decade, and is becoming increasingly important in other domains as well, including event log monitoring. A number of approaches have been proposed for event correlation, and a number of event correlation products are available. Unfortunately, existing products are mostly expensive, platform-dependent, and heavyweight solutions that have complicated design, being therefore difficult to deploy and maintain, and requiring extensive user training. For these reasons, they are often unsuitable for employment in smaller IT systems and on network nodes with limited computing resources. So far, the rule-based approach has been frequently used for monitoring event logs – event processing tasks are specified by the human analyst as a set of rules, where each rule has the form IF condition THEN action. For example, the analyst could define a number of message patterns in the regular expression language, and configure the monitoring tool to send an SMS notification when a message that matches one of the patterns is appended to the event log. Despite its popularity, the rule-based approach has nevertheless some weaknesses – since the analyst specifies rules by hand using his/her past experience, it is impossible to develop rules for the cases that are not yet known to the analyst; also, finding an analyst with a solid amount of knowledge about the system is usually a difficult task. In order to overcome these weaknesses, various knowledge discovery techniques have been employed for event logs, with data mining methods being a common choice.

Log files are excellent sources for determining the health [2] status of a system and are used to capture the events happened within an organization’s system and networks. Logs are a collection of log entries and each entry contains information related to a specific event that has taken place within a system or network. Many logs within an association contain records associated with computer security which are generated by many sources, including operating systems on servers, workstations, networking equipment and other security software’s such as antivirus software, firewalls, intrusion detection and prevention systems and many other applications.

Reconstruction of events inside a computer requires understanding of computer functionality. Many techniques emerged for reconstructing events in specific operating systems. This dissertation classifies these techniques according to the primary object of analysis. Two major classes are identified: log file analysis and file system analysis.

Digital forensics has been defined as the use of Scientifically [10] derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation and presentation of digital evidence derived from cyber sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal or helping to anticipate the unauthorized actions shown to be disruptive to planned operations One important Element of Digital forensics is the credibility of the digital evidence.

II.OBJECTIVE OF THE DISSERTATION

Contribution of dissertation wok is to study of the field of cyber forensic, log files, role of log file in cyber forensic, evidence gathering through log file, various log files management issue and also proposes a prototype system which is based on relational algebra to build the chain of evidence. The prototype system is used to preprocess the real generated data from logs and classify the suspicious user based on decision tree.

The main approach is to correlate firewall log and web server log file for understand the end user behavior. The proposed algorithm perfume offline log files analysis by using rule based correlation and classify the suspicious user based on temporal data mining and fuzzy rule.

III.PROPOSED ARCHITECTURE & METHODOLOGY

This dissertation proposed a new approach to identify the malicious user or attacker. This can be done by analyzing the log files.

In this work there are two log files has used. One is web log and another is firewall log. But it was difficult to analyze them because of their different structure of log. Now it is necessary to convert it into the same structure. So there is a concept of temporized logging system. In this approach the log file data will convert into the database having the simple format. On this data which is selected on the basis of time, needs to apply the fuzzy rule for filtering. These rules can apply on the various logs.

There are three basic fundamentals component of the proposed work.

Temporized Logging System: This is used for the collecting data from log files with respect to time. As we know there are various log file format. This function makes the detail of log in one format in order to analysis.

Fuzzy Approach: This step is used for the applying the fuzzy association rules on the collected data. These can one or more than one rule.

Classification: Classification is a final step in order to get the suspicious users. This classification will perform on the basis of rules.

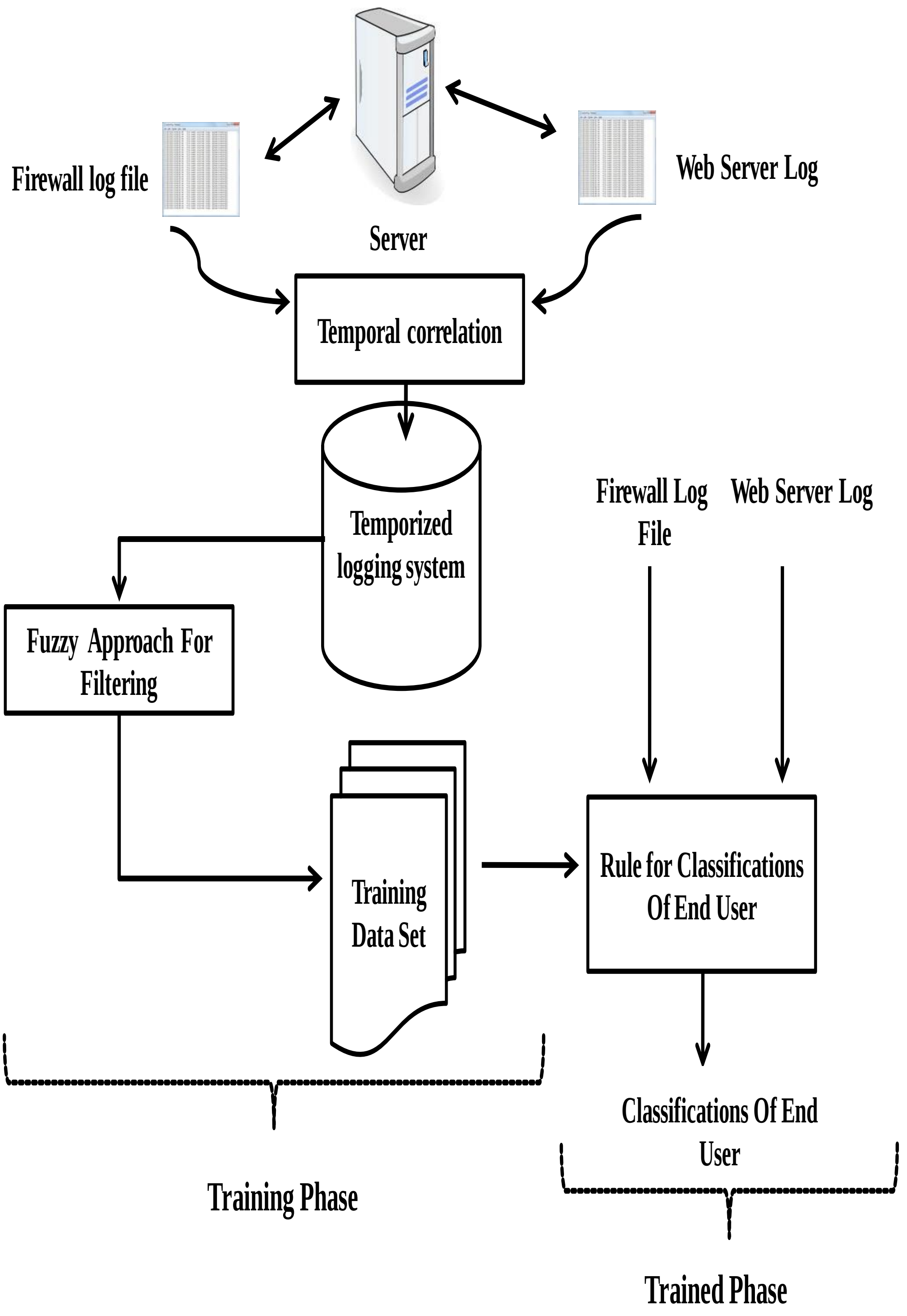

Figure 3.1 Proposed Architecture

IV. LITERATURE SURVEY

1 New Approaches for Intrusion Detection Based On Logs Correlation:

Network administrators are able to correlate log file entries manually [11] Large volume and low quality of log files justify the need for further log processing. The manual log processing is lack of flexibility. It is time consuming, and one doesn't get the general view of the log files in the network. Without this general view it is hard to correlate information between the network components. Events seemingly unessential by themselves can in reality be a piece of a larger threat. In this regard, different log co relation methods are proposed to improve alert quality and to give a comprehensive view of system security. In this paper, the authors show that how different attacks categorized in three categories with different behavior: Denial of service (DoS) attacks, user-to-root (U2R) &

remote-to-local (R2L) attacks and probing, are reflected in different logs and argue that some attacks are not evident when a single log is analyzed.

2 Constructing Genome Phylogenetic Tree of Large dsdna Viruses Using Log Correlation Distance:

The taxonomy of the large ds DNA viruses [12] has been provided in the VIIIth report of ICTV. The phylogenetic tree of large ds DNA viruses has been constructed using CV Tree method (Gao and Qi, BMC Evol. Biol.7(2007)41). In this paper, we use the log-correlation distance method analyze the complete genome of the 124 large ds DNA viruses and construct phylogenetic trees based on compositional vectors of DNA sequences or protein sequences. The phylogenetic trees show the large dsDNA virus genomes are separated into nine families. The structures of the trees based on log-correlation distance are mostly consistent with the result of CV Tree method and the taxonomy of the VIIIth report of ICTV.

3 Confidentiality of event data in policy-based monitoring:

Monitoring systems observe [13] important information that could be a valuable resource to malicious users: attackers can use the knowledge of topology information, application logs, or configuration data to target attacks and make them hard to detect. The increasing need for correlating information across distributed systems to better detect potential attacks and to meet regulatory requirements can potentially exacerbate the problem if the monitoring is centralized. A single zero-day vulnerability would permit an attacker to access all information. This paper introduces a novel algorithm for performing policy-based security monitoring. We use policies to distribute information across several hosts, so that any host compromise has limited impact on the confidentiality of the data about the overall system. Experiments show that our solution spreads information uniformly across distributed monitoring hosts and forces attackers to perform multiple actions to acquire important data.

4 A Log Correlation Model To Support The Evidence Search Process In A Forensic Investigation:

Computer forensics [14] searches for evidence to reassemble the actions that led the system from a secure state to the moment an intrusion was detected. The main source of data for a forensic investigation is the information provided by log files. Log files are generated by applications to keep a register of the actions occurred on the system. However, the massive amount of recorded events complicates the forensic investigation. A model composed by a set of agents in order to collect, filter, normalize, and to correlate events coming from diverse log files is proposed in this paper. The purpose of the model is to assist the analyst in the evidence search process of a forensic investigation.

5 LEC: Log Event Correlation Architecture Based on Continuous Query:

The rapidly evolving society, every [15] corporation is trying to improve its competitiveness by refactoring and improving some if not all of its industrial software infrastructure. This goes from mainframe applications that actually handle the company's profit generating material, to the internal desktop applications used to manage these application servers. These applications often have extended activity logging features that notify the administrators of every event encounter at runtime. Unfortunately, the standalone nature of the event logging sources renders the correlation of log event infrastructure prone to continuous queries. This paper described an approach that adapts and employs continues queries for distributed log event correlation with the aim to solve problems that face the present log event management systems. It will present LEC architecture that analyze a set of distributed log events that follow a set of correlation rules; then the main output is a stream of correlated log events.

6 Log Master: Mining Event Correlations in Logs of Large-Scale Cluster Systems:

This paper [16] presented a set of innovative algorithms and a system, named Log Master, for mining correlations of events that have multiple attributions, i.e., node ID, application ID, event type, and event severity, in logs of large-scale cloud and HPC systems. Different from traditional transactional data, e.g., supermarket purchases, system logs have their unique characteristics, and hence the authors proposed several innovative approaches to mining their correlations. The authors parsed logs into an

n-ary sequence where each event is identified by an informative nine-tuple. The authors proposed a set of enhanced Apriori-like algorithms for improving sequence mining efficiency, the authors proposed an innovative abstraction event correlation graphs (ECGs) to represent event correlations, and present a ECGs-based algorithm for fast predicting events. The experimental results on three logs of production cloud and HPC systems, varying from 433490 entries to 4747963 entries, show that the author's method can predict failures with a high precision and an acceptable recall rates.

7 A Correlation Analysis Method of Network Security Events Based on Rough Set Theory:

Network security event correlation [17] can find real threat through correlating security events and logs generated by different security devices and can be aware of the network security situation accurately. This paper has proposed a network security events correlation scheme based on rough set, build database of network security events and knowledge base, gives rule generation method and rule matcher. This method has solved the simplification and correlation of massive security events through combining data discretization, attribute reduction, value reduction and rule generation.

V.CONCLUSION

These days the level of computer crime has increased dramatically. We need to improve the investigative system to identify the culprit. Web server logs are usually fixed on the behavior of the machine, and not on the behavior of the end user. The log file provides troubleshooting, security, and proactive system administration, which provides significant support for suspicious users in the caching and cyber expertise process. In this dissertation, implemented system extracts the evidence from log file and correlates these generated logs on the basis of relational algebra and classifies end user .v model. We have proposed a novel log analysis method using TL based on reconstruction. The proposed method provides a solution to identify an attacker. This approach uses time-based data analysis and fuzzy communication rules. As the results show, the proposed thesis provides better results which are better than the previous work. The proposed baseline work encourages a web researcher to navigate the end-user behavior and ensure an effective security policy.

REFERENCES

[1]. Risto Vaarandi "Tools and Techniques for Event Log Analysis", Faculty of Information Technology, Department of Computer Engineering, Chair of System Programming, Tallinn University of technology,2005

[2]. Muhammad Kamran Ahmed, Mukhtar Hussain and Asad Raza "An Automated User Transparent Approach to log Web URLs for Forensic Analysis" Fifth International Conference on IT Security Incident Management and IT Forensics 2009.

[3]. Pavel Gladyshev "Formalising Event Reconstruction in Digital Investigations" Ph.D. dissertation Department of Computer Science, University College Dublin, 2004.

[4]. Carrier, B.D., Spafford, E.H "Defining Digital Crime Scene Event Reconstruction" Journal of Forensic Sciences, 49(6). Paper ID JFS2004127,2004

[5]. Stephenson. P, "Application Of Formal Methods To Root Cause Analysis of Digital Incidents", International Journal of Digital Evidence, 3(1) ,2004

[6]. Stevens, M.W. "Unification of relative time frames for digital forensics", Digital Investigation journal, 1(3), pp. 255-239, 2004

[7]. Weil, M.C. “Dynamic Time & Date Stamp Analysis”, International Journal of Digital Evidence, 1(2),2002

[8]. Willassen, S.Y., Mjølsnes, S.F . “Digital Forensics Research”, Teletronikk journal, 1, pp. 92-97. 2005

[9]. Gary L Palmer “A Road Map for Digital Forensic Research”. Technical ReportDTR-T0010-01, DFRWS. Report for the First Digital Forensic Research Workshop (DFRWS), 2001

[10]. Bennie Kar Leung Fei, “ Data Visualisation In Digital Forensics” University of Pretoria etd-Fei,BKL,2007

[11]. S.O. Azarkasb, S.S. Ghidary, “New Approaches for Intrusion Detection Based On Logs Correlation” IEEE 2009, pp 234.

[12]. L.Q.Zhou and J.M.Bai, “Constructing Genome Phylogenetic Tree of Large dsdna Viruses Using Log-Correlation Distance”, IEEE 2010, pp 2182-2185.

[13]. Montanari, M.; Campbell, R.H. “Confidentiality of event data in policy-based monitoring” IEEE 2012 Page(s): 1 – 12

[14]. Jorge Herrerias and Roberto Gomez, “A Log Correlation Model To Support The Evidence Search Process In A Forensic Investigation”, IEEE 2007, pp 31-42.

[15]. N.Hammoud, “LEC: Log Event Correlation Architecture Based on Continuous Query” IEEE 2009, pp 422-429.

[16]. Xiaoyu Fu, Rui Ren, Jianfeng Zhan, Wei Zhou, Zhen Jia and Gang Lu, “Log Master: Mining Event Correlations in Logs of Large-Scale Cluster Systems”, IEEE 2012, pp 71 – 80.

[17]. Jing Liu, Lize Gu, Guosheng Xu and Xinxin Niu, “A Correlation Analysis Method of Network Security Events Based on Rough Set Theory ” IEEE 2012, pp 517 - 520.

[18]. Chu-Hsing Lin, Jung-Chun Liu and Ching-Ru Chen, “Access Log Generator for Analyzing Malicious Website Browsing Behaviors", IEEE 2009, pp 126-129

[19]. D.S. Sisodia and S. Verma, "Web usage pattern analysis through web logs: A review", IEEE 2012, 49-53.

[20]. Log file format from “http://www.w3.org” accessed on 14/03/2013.

[21]. Wei Peng ; Tao Li ; Sheng Ma “Mining Logs Files for Computing System Management” IEEE 2005,P 309-310

[22]. Tomono, A. ; Uehara, M. ; Shimada, Y. Improvement and Evaluation of a Method to Manage Multiple Types of Logs IEEE 2011, P 601-607

[23]. Karen Kent and Murugiah Souppaya, “Guide to Computer Security Log Management”, Computer Security Division Information Technology Laboratory National Institute of Standards and Technology Gaithersburg, 2006.