INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING & APPLIED SCIENCES
ISSN: 2322-0821(O)
ISSN: 2394-9910(P)
VOLUME 9 ISSUE 2
Apr 2021 - Jun 2021
www.irjeas.org

AN ASSESSMENT OF FUZZY TEMPORAL EVENT CORRELATION TOWARDS CYBER CRIME INVESTIGATION

James Murray[1]

[1] *Electrical and Computer Engineering Department, Virginia Tech, Arlington, VA 22203, USA*

**Corr. Author - mrry.james87@hotmail.com*

Abstract - Event logging and event logs play an important role in modern IT systems criminal investigation which is generated when end user with each other in web environment and stored in various logs like firewall log file at side, network log file at gateway and web log file at server side. But log file is not to be over emphasized as a source of information in systems and network management. Whereas conduct efficient investigation and gathering of use full information need to correlate different log file. Task of analysing event log files with the ever-increasing size and complexity of today's event logs has become cumbrous to carry out manually. Nowadays latest spotlighted is automatic analysis of these logs files. . This paper presents a bird eye on two basic concepts one is temporal data mining and another is fuzzy association rules. Using log files it is possible to classify the attacker from the normal user.

Keywords— **Event Logging, Fuzzy Logic, Temporal Correlation**

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, e.g., raise an alarm immediately after a fault message has been appended to a log file. On the other hand, quite many essential event processing tasks involve event correlation a conceptual interpretation procedure where new meaning is assigned to a set of events that happen within a predefined time interval. 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[2]. 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 [3]. It should be noted that while event log monitoring tools conduct on-line (real-time) analysis of event log data, data mining methods are designed for off-line analysis an existing event log data set is processed for discovering new knowledge, with the data set remaining constant throughout the discovery process.

HTTP & TCP METHOD

Both protocols are responsible for the data communication from source node to destination node. In browser the http protocol works as application layer protocol. Computer store all the information regarding

access the server store in log file. At that time which protocol has used during the communication will note in the log files.

LOG AND FUZZY LOGIC

Log file that have all the entry related to incoming user and outgoing user. These file are generated by the process of installation. It can maintain by server machine, firewall, web servers, and routers etc. Generally the log files are in the text format can be read by notepad or simple text editor. Due to the plain text the size of log file will also reduces group of items having similar sort of properly is known as set. These set items are the elements of set. This is a traditional approach to represent the set theory. There are some problems with the traditional approach of set theory. The set theory is used for the specific data set. It always gives the answer is Yes or No. some time it seems to be that it is not practically suitable. To remove such types of complexity fuzzy set theory comes in existence.

TEMPORAL DATA MINING

Time and space are the two basic category of data mining. The temporal [4] data mining is a newly data mining approach with respect to time. Some time it used the temporal data base. There are basically three types of time used in this approach these are valid time, transaction time and Bi-temporal. The valid time shows the duration in which the event was performed in province of real world. The transaction time give the detail about the duration in which the event was saved in the records. The combination of both time periods is known as a Bi-Temporal. Extraction of Temporal Association Patterns for Temporal data mining has been productively applied in number of fields including trading, marketing, social analysis, medical, fraud detection, robotics and assisted design Because of that explorer's number of efficient algorithms for temporal data model like symbolic time series, symbolic time sequences, symbolic interval series, numeric time series, item set sequences, etc have been proposed.

II. LITERATURE REVIEW

In this research paper, the authors [13] 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. Authors have emphasized the importance of logs, attacks and have discussed how different logs are affected by three categories of attacks. Authors applied logs correlation table with 44 common attacks and 15 important logs that represented beneficial and valuable information to researchers in this area. Authors also performed experiments to successfully correlate data from multiple logs for cases of anomaly detection and misuse detection.

In this paper [14] authors use the log-correlation distance method analyze the complete genome of the 124 large dsDNA 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 CVTree method and the taxonomy of the VIIIth report of ICTV.

This paper introduces a novel algorithm for performing policy-based security monitoring. Authors [8] 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.

In this paper [15] authors 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.

This paper [16] 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.

III. PROPOSED ARCHITECTURE FRAMEWORK

In this work 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.

International Research Journal of Engineering & Applied Sciences, IRJEAS
www.irjeas.org, ISSN(O): 2322-0821, Volume 9 Issue 2, Apr 2021-Jun 2021, Page 10-14

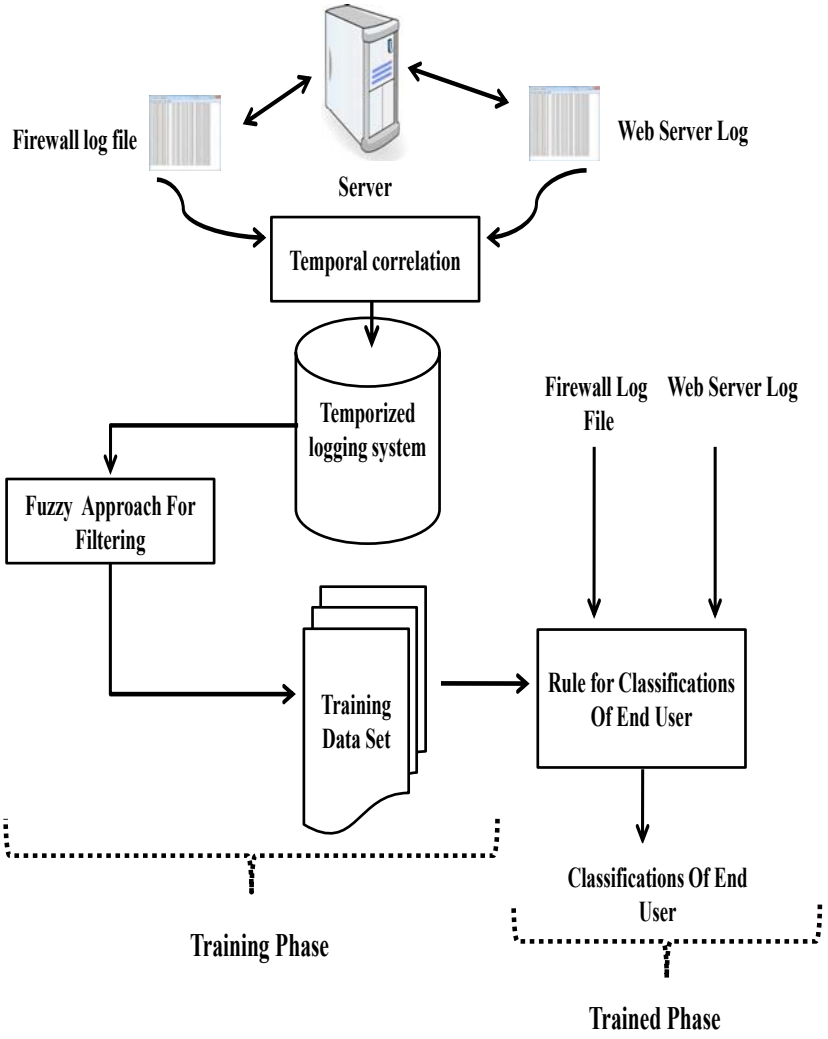

Figure 3.1: Proposed Architectural Framework

IV. EXPERIMENTAL ENVIRONMENT

The implementation of the proposed work has done in MATLAB 10.0 with the help of MySQL database. In this work use Intel core i3 CPU with 2.53 GHz, having 4 GB of RAM and 500 GB HDD. 32 bit OS (Windows 7) is used in this experiment. A client server architecture having many clients and one server is taken as a scenario for verification of proposed work whole verification is done over MATLAB 7.10. MySQL is also used in many high profiles, large-scale World Wide Web products including Wikipedia, Google and Facebook. Figure shows the login window of MySQL [43] in command mode. Generally we use the Command prompt in order to access the sql database.

V. RESULT ANALYSIS

In this paper, the MATLAB simulated experiments are performed to verify the accuracy of proposed model. Log format synchronization is one of great challenge in log management issue, recently researcher focus on that problem. Proposed model in [1, 2] is log format dependent where as Proposed model is not format dependent. Along with that log rotation (size of log file) and clock synchronization is another most challenging issue in log management .Proposed model is time dependent and result set is independent form size of log file.

SN	Year	Log Format	Time	Size
1	IEEE,2012 [44]	Dependent	Dependent	Independent
2	IEEE,2011 [45]	Dependent	Independent	Independent
3	IEEE,2011 [46]	Independent	Dependent	Dependent
Proposed Methodology		Independent	Dependent	Independent

Table 5.1 Comparison table

VI. CONCLUSION AND FUTURE WORK

Computer crime rate has increased a lot in these days. To detect the culprit there is a need to improve the investigation mechanism. Web server logs are commonly captured the behavior of machine not the behavior of end user. Log file provide troubleshooting, security and pro-active system administration that provide significant help in caching suspicious end user and in process of cyber forensic. 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. The proposed methodology gives the solution to identify the attacker. This approach uses the temporal data mining and fuzzy association rules. As the results show that the proposed methodology gives the improved results which are better from the previous work. Proposed frame work encourages the web investigator to navigate the end user behavior and assist to enforce the effective security policy. Although the initial results are encouraging, there is still a lot of work to do for improving the evidence gathering efficiency. A major issue's for future work are related to log consistency, log integrity and log rotation to do more extensive test with large volume of log data. Future work also includes clock synchronization problem arises in evidence correlation.

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]. Guillame-Bert M., Crowley J.L., "New Approach on Temporal Data Mining for Symbolic Time Sequences: Temporal Tree Associate Rules" IEEE Conference Publications 2011 pp 748 – 752

[5]. Ye Changguo, Wei Nianzhong , Wang Tailei , Zhang Qin, Zhu Xiaorong The Research on the Application of Association Rules Mining Algorithm in Network Intrusion Detection IEEE Conference Publications 2009 pp 849 – 852

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

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

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

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

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

[11]. 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.

[12]. 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.

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

[14] L.Q.Zhou and J.M.Bai, "Constructing Genome Phylogenetic Tree of Large dsdna Viruses Using Log-CorrelationDistance", IEEE 2010, pp 2182-2185.

[15] 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.

[16] A Correlation Analysis Method of Network Security Events Based on Rough Set Theory.