A SURVEY ON LOG CORELATION IN SECURITY INFORMATION AND EVENT MANAGEMENT WITH HADOOP

Anand Mehta[1], Manish Kumar Abhishek[2]

[1] *Dept. of Computer Engineering, GTU PG SCHOOL, Gujarat, India*
[2] *Manager, IT Infrastructure, RailTel Corporation of India Ltd., Haryana, India*

ABSTRACT

In a today's era, technology gives more and more benefits in the public sector as well as the threats and impact of the threats also rises high. It's a very difficult to assure a security in a computer systems because of the rapidly development of IT technologies and to save the IT infrastructure analysis of log is very important. Infrastructure vulnerabilities is exposing to the worlds openly due to lake of security and more APT attacks are targeting the large business and organization's important and secret information. This article contains some information and surveys of different papers regarding the security information and management system and log analysis and log correlation. Right now many SIEM systems are available in the market and this SIEM system has ability to identify the correlation. If we are matching the logs efficiently then we can identify the problem and its origin. This survey taken from the various papers from journals and articles. It is helps for developed a good open source SIEM system with a high capacity processing framework like Hadoop because of the log data generated from the various sets of devices to the SIEM came with the large amount.

Keyword: *SIEM, Security, Log Correlation, Hadoop*

1. INTRODUCTION

The amount of cyber-attacks are rising each months in a very large size. The producer of antivirus tools, Kaspersky Lab informs that its solution is detected 23,680,646 in 2008 to 51887400554 in 2013 [1, 2]. As a further matter, with the statistics of the report provided by the Verizon RISK team in 2012: 54% of malware was detected after the months from infection. Only 13% malware was identified in a single day [3].

The Investigation concerning text based log to imagining has been always led since past many decades. In that size of log has been speedily enlarged through the growth of Information Technology. It is vital to imagine the log for well-organized examination. This is too significant in information safety field. Due to modification of security paraphernalia and growth of bulk packing, there is a boundary to examine security logs with partial social source. So, expansion of Security Information and Event Management (SIEM) which investigate and visualize several security logs has accompanied. SIEM explanation is operative protection clarification for identifying Advanced Persistent Threat (APT) attack. APT is a set of quiet and constant PC hacking procedure, frequently arranged by humanoid aiming a precise unit [4].

This report is directed to show Information technology management environment interest on turnout security events, their analysis and taking suitable choice to alleviate negative consequence of attacks embattled in contradiction of observed system. The IT Security monitoring tool is developed with the use of open source tool because the open source helps to cut the cost and it will help more beneficially to small scale organization.

Therefore, we want to focus on distinguishing advance persistent attack more specifically. If we sense APT occurrence in some week, we, may save a lots of economic destruction to the organization and drip of important evidence. For SIEM which we established is created on open source platform, it is economical.

1.1 SECURITY INFORMATION AND EVENT MANAGEMENT (SIEM)

As per the figure 1, SIEM combines two different field, first is security event management (SEM) and security information management (SIM). In this fields main attention is on the examination and gathering of security significant data. Conversely, SEM accentuates the collection of log records in to adaptable volume of info with the aid which security event may be apportioned with proximately though security info management (SIM) basically attentions on investigation of previous data in instruction to expand the extensive term usefulness and proficiency of infrastructure in information security structure.

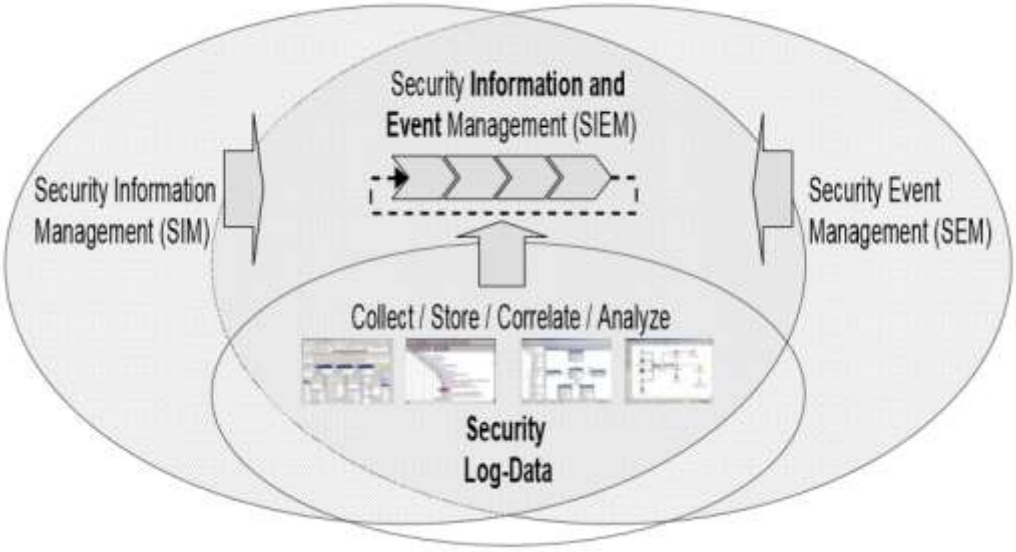

Figure 1: Visionary Architecture of SIEM

The incorporation of SEM and SIM into a combined development of preparation, directing and monitoring security applicable info on the foundation of statistics gathered from the Information Security architecture is shortened below the term SIEM [5].

Through classifying information and subtracting information from the current capacity of records SIEM endeavors to assurance the security of the info and the information system advantage standards of an institute. In the direction of accomplish this aim it is indispensable to behavior SIEM as a combined and uninterrupted supervision procedure. In turn, this progression is reliant on information related to choice creating which is take out from the data lake. It is consequently critical to found apposite performs and apparatuses which provision the use of data in administration development as successfully and resourcefully as possible.

As an outcome of the frequent mechanisms fitted in an information security (IS) architecture, the volume of set of procedures as well as the volume of records (data) produced is massive. Dependent on the scheme and the accomplishment done, log data can cover info about frequencies or threats [6].

1.2 HADOOP

Apache Hadoop is an open-source stage, which assistances in loading data and corresponding handling in a circulated (distributed) environment. Hadoop ruptures the big data-base into slabs of data and allocates over the groups called clusters. For a processing the data, MapReduce is available and used for parallel-processing on clusters, therefore it reduces the compilation time.

The HDFS (Hadoop Distributed File System) is fundamentally a circulated file system which is intended to execute on commodity hardware. HDFS is extremely fault-tolerant and it is planned in a way that it must be organized on low-cost hardware. HDFS also delivers high throughput (output) access to request data and is highly fit for applications that have huge data sets [7].

2. LITERATURE REVIEW

Some papers were published in the area of log correlation and visualization. In this section we will discuss about various techniques that proposed for the correlation of logs by authors in last few years.

In 2014, authors presented and discussed on the operational role of SIEM in a security operation center [8], which is a very important tool in the operation center for a infrastructure security. This tool will help to collecting, analyzing and normalizing security events from various sources. Its primary goal is to monitor the event which is a related to a security including the IT Network and other devices like a firewall, server, accounts, IDS and IPS. Each of this device producing the log files and SOC receives an information in a form of log files and alerts in a huge amount. Might be this alerts and log files indicates the malevolent performance.

In this article, authors also discussed about the challenges in a Rule Creation and Management of logs for better awareness. In this a cost for transforming the data management to a big data management comes with the more challenges. One aspect of this problem is the systems inability to efficiently execute complex queries. Another challenge is a lake of contextual information due to unnecessary overhead from alerts which created by SIEM. Authors also faced challenges on Event collection, Storage, Correlation and analysis, because the correlation is the time consuming.

In a 2014, authors Igor Anastasov, Danco Davcev presented a paper about the Implementation of SIEM for Global and Distributed Environments. In this article they proposed a model for SIEM based on the hierarchical manager. They proposed an extension of architecture called, "Hierarchical Manager Archtecture", in this the model work as a parent child manager relationship. Server act as a parent and communicate with child, child servers is a SIEM server. Advantage of this proposed way is extendable in a big organization but for a small scale organization and institutions cannot afford this architecture. [10]

In a 2015, author Damian Hermanowski presented a paper with the open source solution, Open Source Security Information Management (OSSIM), this tool is an integrated and designed to help a network administrator for a security reasons. This tool performs SIEM functionalities using other well-known open source components. OSSIM is developed by AlienVault Company and main goal is correlation engine development.

But in this solution, main challenge about the integration and imperfections, lack of a raw log storage mechanism and OSSIM SIEM in overall not a mixture for all types of threats and attacks [9].

In a 2016, author presented a paper on visualization of a security log efficiently, in this paper they discussed and implemented a SIEM system based on open-source package such like D3 component. The main result is decrease the cost of package. For analyzing the attacks, analysis of a correlated security logs in necessary and visualize them for accepting those connections. Therefore, we need to research the security-log visualization constantly. This research is also focused on visualizing security logs efficiently but with the large logs the efficiency is less in this also [11].

3. CONCLUSION

As per survey in area of log correlation in a security information and event management we find many approaches is available like an enterprise solution as well as the open-source tool also but in all the approaches cannot work properly due to lake of the robust framework. When the log data is available in an amount of TB's in a day to the administrator at that time the correlation of the logs must be accurate. It will save the time and cost both for an organization, with the help of the Apache Hadoop framework this lake must be covered because the Hadoop has the ability to process a large amount of a data at a time. So with this way the deployment of an Open source SIEM system with Hadoop framework provide a extra ordinary result in a log correlation and visualization in a field of computer security.

4. ACKNOWLEDGEMENT

The author is highly thankful to his respected guide Mr. Manish Kumar Abhishek for marvelous guidance and support to complete this survey paper. Author is also thankful to RailTel Corporation of India Limited and GTU PGSCHOOL, Gandhinagar for providing essential facilities to complete this manuscript in present nature. The

author would like to thank Prof. Bhadreshsinh Gohil (GTU PGSCHOOL) and parents for financial and moral supports throughout their technical education.

5. REFERENCES

[1]. Gostev, A. Kaspersky Security Bulletin: Statistics 2008, https://securelist.com/analysis/kaspersky-security-bulletin/36241/kaspersky-security-bulletin-statistics-2008/

[2]. Funk, C.; Garnaeva, M. Kaspersky Security Bulletin. The Overall Statistics for 2013. Available online: https://securelist.com/analysis/kaspersky-security-bulletin/58265/kaspersky-security-bulletin-2013-overall-statistics-for-2013/.

[3]. Verizon RISK Team. Verizone 2012 Data Breach Investigations Report. Available online: http://www.verizonenterprise.com/DBIR/2012/.

[4]. Binde, Beth, Russ McRee, and Terrence J. O'Connor. "Assessing outbound traffic to uncover advanced persistent threat." SANS Institute. Whitepaper (2011).

[5]. Tobias Hoppe, Alexander Pastwa, Sebastian Sowa, "Business Intelligence Based Malware Log Data Analysis as an Instrument for Security Information and Event Management" International Journal on Advances in Security, vol 2 no 2&3, year 2009.

[6]. Roland Gabriel, Tobias Hoppe, Alexander Pastwa, Sebastian Sowa, "Analyzing Malware Log Data to Support Security Information and Event Management: Some Research Results" published in IEEE conference, year 2009

[7]. Tushar M. Chaure, Kavita R. Singh, "Frequent Itemset Mining Techniques – A Technical Review" published in IEEE WCFTR year 2016.

[8]. Sandeep Bhatt, Pratyusa K. Manadhata and Loai Zomlot, "Th e Operational Role of Security Information and Event Management Systems" published in IEEE Computer and Reliability Societies, year 2014.

[9]. Damian Hermanowski, "Open Source Security Information Management System Supporting IT Security Audit" published in IEEE, year 2015

[10]. Igor Anastasov, Danco Davcev, "SIEM Implementation for Global and Distributed Environments" published in IEEE year 2014.

[11]. Jaehee Lee, Changyeob Lee, Jaebin Cho, "A Study on Efficient Log Visualization Using D3 Component Against APT How to visualize security logs efficiently?" published in IEEE year 2016.