An Approach for Security Information and Event Management with Hadoop and ElasticSearch

Anand Mehta
PG Student, Dept. of Computer Engineering, GTU PG SCHOOL, Gujarat, India

Manish Kumar Abhishek,
Manager, IT Infrastructure, RailTel Corporation of India Ltd., Haryana, India

Srinivas M (Vasu),
Additional GM, Data Center, RailTel Corporation of India Ltd., Haryana, India

Abstract - Now days, technology gives more profits in the municipal sector and private sector as well as the threats and influence of the coercions also high. This is a very problematic to promise a safety in a PC and IT systems because of the swiftly expansion of IT skills and except the Information Technology structure analysis of log is very significant. Infrastructure weaknesses is revealing openly due to lake of safety. This article contains an approach for design SIEM for handle large amount of log data Hadoop prospective is a best, with the help of the HDFS file structure and ElasticSearch mechanism processing of the logs is faster and provide better search facility. So the user in a network operation room can visualize the attack or suspicious activity in real time.

Keyword: SIEM, Security, Big Data, Hadoop, Elasticsearch, HDFS, Kibana.

I. INTRODUCTION

The quantity of fake-spams or cyber-bout are intensifying every month in a huge amount. The fabricator of antivirus tools, Kaspersky Lab notifies that its resolution is noticed 23,680,646 in 2008 to 51,887,400,554 in 2013 [1, 2].

As well as with the figures of the report given by the Verizon RISK group in 2012: 55% of malware was distinguished after the long time (month) from infection. Only 59% malware attack [6] was acknowledged in a single day [3].

Now a day's research concerning text based logs to visualizing has been consistently led since past may periods. In this period technologies and its ways constantly upgraded and user friendly and the size of logs has been quickly enlarged expanded through the through the advancement of Information Technology (IT). It is integral to conceptualize the log for efficient examination or mining of item sets [7]. This is too compelling in information or network safety field. Due to the alteration in security equipment and evolution in bulk packing, there is an edge to inspect security logs/data with limited social origin. So, enlargement of Security Information and Event Management (SIEM) [8] that interrogate and visualize a few security logs has guide. The SIEM description is operational prevention and resolution in consequence of the fact that classifying Advanced Persistent Threat (APT) attack [11].

APT is a set of quiet and constant PC hacking procedure, frequently arranged by humanoid aiming a precise unit [4]. In this article, we done the examination of large data sets of real time logs with the help of open source program Hadoop [13] Hadoop-Elasticsearch is appropriate and applied in many field for Big Data [14] analysis. As Log files [15] is also one of the type of huge data which growing fast so Hadoop is the finest and suitable platform for storing

files and parallel execution of Elasticsearch [16] program for scan and provide result [9][10]. Elasticsearch is a part of ELK Stack [17] (Elasticsearch, Logstash, Kibana), but we can use each part of this stack separately.

II. ARCHITECTURE

A. Security Information and Event Management (SIEM)

SIEM associate two dissimilar field, as per the fig. 1, right side is a Security Event Management (SEM) and left side is a Security Information Management (SIM). This fields are main attention because on the analysis

and collecting of security significant data. Conversely, SEM accentuates the collection of log records in to adaptable volume of Information with the aid which Security event may be apportioned with almost 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. [12] The merger of SEM and SIM into a linked advancement of arrangement, controlling and auditing security applicable information on structure of data gathered from the various Information Security architecture is abbreviate in the term SIEM [5].

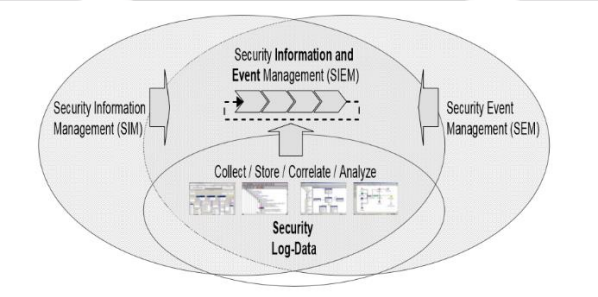

Figure 1 Introspective Architecture of SIEM

B. Hadoop

Apache Hadoop is an open-source platform, which help in readying data and analogous administration in a circulated environment. Hadoop breach the Big Data base into piece of data and set aside over the association called clusters. For a handling the massive data, Elasticsearch is available and used for side by side indexing on clusters, therefore it cut down the compilation time.

The HDFS-Hadoop Distributed File System is basically a spread file system which is planned to perform on hardware. HDFS is tremendously fault-tolerant. HDFS also transports high output access to call data and is extremely fit for applications that have huge data sets.

III. PROPOSED SOLUTION

In the proposed solution, files and connections are coming from the way of internet and it will be filtered with DNS, FIREWALL and BRAS (Broadband Remote Access Server) and it will be transferred to the SYSLOG-ng. All logs are also transferred to the Hadoop Data Lake. Hadoop Data Lake is a HDFS (Hadoop Distributed File System) and large log files will be processed in this data lake like sorting and matching process. Elasticsearch plays a vital role in this system because the Elasticsearch provides a fast and compatible full text search. Elastics search also comes with its small architecture with helps of nodes. But when user search any query to dashboard i.e. Kibana to mechanism, Elasticsearch provides faster real-time results on a dashboard with helps of indexing.

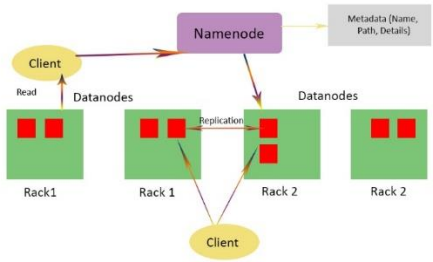

Figure 2 HDFS Architecture

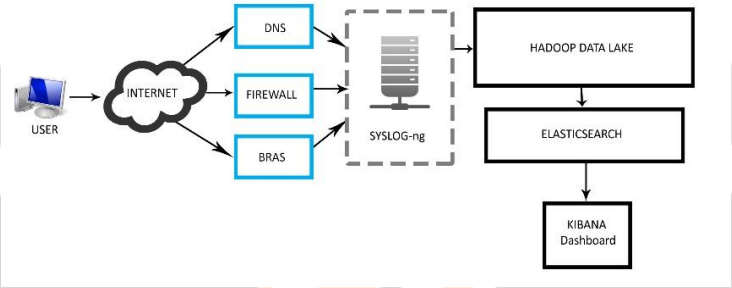

Figure 3 Proposed System Architecture

IV. RESULT

Here we are present our proposed work, for the SIEM system and log analysis we observed live log data in Data center, in this log files it contains multiple field as IP addresses like source IP address, Destination IP address, URL, timestamps. We have installed Hadoop 2.7.2 on Cent OS machine with java 1.8. Log files are circulated consistently on these nodes on the cluster, The Elasticsearch job goes on these files and get examined consequences in the graphical presentations like graph charts using Kibana visualization.

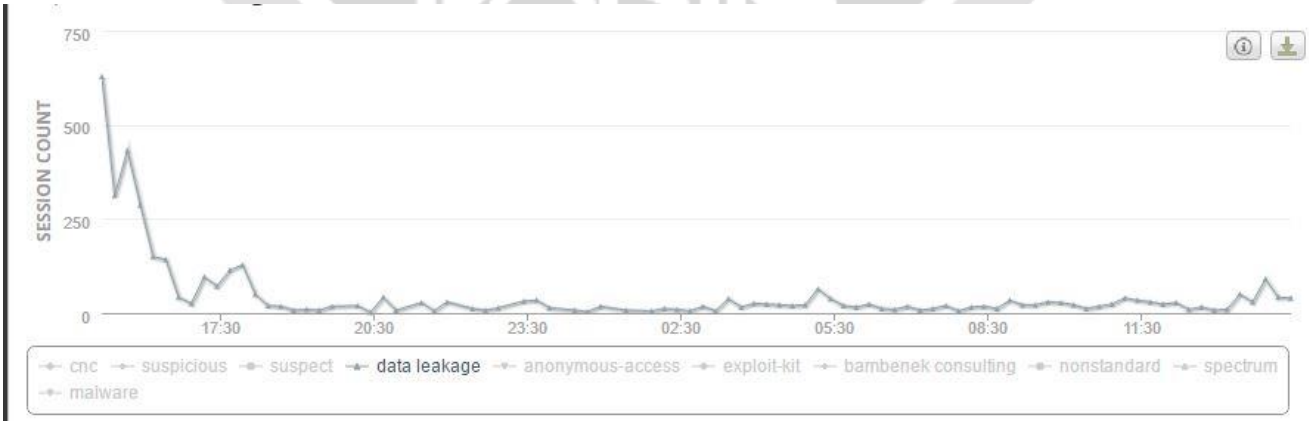

Figure 4 Kibana Dashboard with Real Result

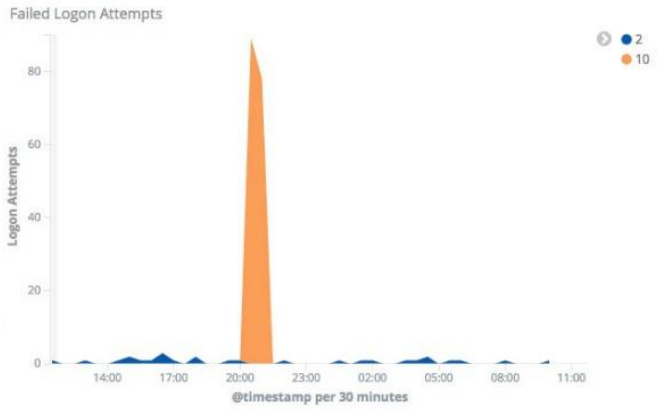

Figure 5 Multiple Log-in Failure

V. CONCLUSION

Log investigation supports to develop the business tactics and helps to make statistical reports. Hadoop and Elasticsearch established log file investigation mechanism will provide us graphical reports with the help of Kibana Dashboard showing hits for pages, activity, in part of users are interested in sites, consistence login failures, traffic, attack etc. From these intelligences industry societies can estimate which portions of the site essential to be upgraded on behalf. Using of Hadoop HDFS framework delivers parallel distributed and steadfast data storing by duplicating data for huge log files. At first stage, files get stored block wise in on a number of nodes in a cluster so that time obligatory can be reduced that save more execution time and give better presentation. Here hadoop main representative of improve response time. And Elasticsearch positively workings scattered for huge data set and providing the more well-organized results with help of indexing and full text search support.

VI. ACKNOWLEDGEMENT

The author[1] is highly thankful to his respected guide Mr. Manish Kumar Abhishek for marvelous guidance and support to complete this paper. Author[1] is also thankful to RailTel Corporation of India Limited (Ministry of Railways undertaking) and thankful to Mr. Srinivas Vasu (ADGM) of Railtel Corporation of India Ltd. For encouraging and support. The author[1] would like to thank parents for financial and moral supports throughout their technical education.

VII. REFERENCES

[1]. Gostev, A. Kaspersky Security Bulletin: Statistics 2008, https://securelist.com/analysis/kaspersky-security-bulletin36241/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/2016/. (visited on Dec 2016)

[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, "The 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.
[12]. Anand Mehta, Manish Kumar Abhishek, "A survey on log corelation in security information and event management with hadoop" published in IJARIIE Vol-2 Issue-6 2016.
[13]. Apache Hadoop, http://hadoop.apache.org/ (visited on 21 Dec 2016)
[14]. Big Data, https://www.sas.com/en_us/insights/big-data/what-is-big-data.html (visited on 19 Dec 2016)
[15]. Vangie Beal, "Log Files" by, http://www.webopedia.com/TERM/L/log_file.html (visited on 15 Oct 2016)
[16]. Elasticsearch, https://www.elastic.co/products/elasticsearch (visited on 2 Jan 2017)
[17]. ELK Stack, https://logz.io/category/blog/elk-stack/ (visited on 1 Jan 22017).