SIEM Network Behaviour Monitoring Framework using Deep Learning Approach for Campus Network Infrastructure

Original scientific paper

Mohd Azmi Bin Mustafa Sulaiman
National Defence University of Malaysia
Faculty of Defence Science and Technology,
Computer Science Department
Sungai Besi Camp, Kuala Lumpur, Malaysia
azmi@upnm.edu.my

Mohammad Adib Khairuddin
National Defence University of Malaysia
Faculty of Defence Science and Technology,
Computer Science Department
Sungai Besi Camp, Kuala Lumpur, Malaysia
adib@upnm.edu.my

Mohd Rizal Mohd Isa
National Defence University of Malaysia
Faculty of Defence Science and Technology,
Computer Science Department
Sungai Besi Camp, Kuala Lumpur, Malaysia
rizal@upnm.edu.my

Mohd Nazri Ismail
National Defence University of Malaysia
Faculty of Defence Science and Technology,
Computer Science Department
Sungai Besi Camp, Kuala Lumpur, Malaysia
m.nazri@upnm.edu.my

Mohd Afizi Mohd Shukran
National Defence University of Malaysia
Faculty of Defence Science and Technology,
Computer Science Department
Sungai Besi Camp, Kuala Lumpur, Malaysia
afizi@upnm.edu.my

Aznida Abu Bakar Sajak
University Kuala Lumpur,
MIIT, Computer Engineering Technology
aznida@unikl.edu.my

Abstract *– One major problem faced by network users is an attack on the security of the network especially if the network is vulnerable due to poor security policies. Network security is largely an exercise to protect not only the network itself but most importantly, the data. This exercise involves hardware and software technology. Secure and effective access management falls under the purview of network security. It focuses on threats both internally and externally, intending to protect and stop the threats from entering or spreading into the network. A specialized collection of physical devices, such as routers, firewalls, and anti-malware tools, is required to address and ensure a secure network. Almost all agencies and businesses employ highly qualified information security analysts to execute security policies and validate the policies' effectiveness on regular basis. This research paper presents a significant and flexible way of providing centralized log analysis between network devices. Moreover, this paper proposes a novel method for compiling and displaying all potential threats and alert information in a single dashboard using a deep learning approach for campus network infrastructure.*

Keywords*: SIEM, Network Behaviour Monitoring, Campus Network Infrastructure*

1. INTRODUCTION

Network security plays a critical part in Information Technology. It is still difficult for organizations to meet security standards. Identity attacks, intrusions, and hacking have been the most common security threats to the public and have also highlighted the importance of information security [6]. By focusing on threats of both internal and external of the network, network security can secure and stop the threat from entering and spreading on the network. A secure network involves a complex connection of hardware devices such as firewalls, routers, and anti-malware tools.

In the campus network, all system and server equipment depends on the admin to collect logs of network equipment and servers, and also to monitor and notify the system status to users. Therefore, it is important to have comprehensive centralized log management in a campus network. It uses to analyze events that occur from thousands of nodes to several dedicated servers where central analysis is carried out. When the analyses are obtained in a real-time process, the safety events can be identified from future events through event correlation and other advanced surveillance techniques. Moreover, it also can be an offline forensic activity, where the past events are investigated to understand the occurrence of security that has taken place.

Aggregate the data generated from multiple sources, identify specific threats and take appropriate action are the basic principles of each analysis of network and security reporting system. For instance, the system can take additional log information, generate alerts and ensure that all security controls can be monitored and prevented when such issues are detected. The networking, software, hardware, and media used to produce, distribute, store, analyze, and erase log data are referred to as log management infrastructure. Almost every organizations have one or more log management infrastructures.

Most organizations or businesses use SIEM (Security Information and Event Management) tool. This tool is designed to simplify company compliance reporting through the usage of a centralized logging solution. Each host that is in use must have a log security record included in the report and can pass log data to the SIEM server. Single SIEM servers can collect log data from as many devices as they need and can produce a detailed report and manage all security events of each log they receive. In the current situation, each system needs to be able to manually retrieve data from each device regularly and to ensure that a central configuration of configuration can be generated to produce a report.

The SIEM system server is a tool for detecting unidentified events. Almost most of the equipment used does not comply with safety regulations and cannot track events or logs more deeply. Although such tools can identify and monitor events and produce audit log entries, they cannot analyze logins to detect unacceptable activities. Best of all, tools such as personal computers and laptops can alarm users when an event occurs. SIEM equipment can also perform higher detection by linking the events or logs of the equipment used. By collecting the events or logs of the linked equipment, the SIEM system can see attacks that have different angles on each of the different devices and can therefore record events or logs to decide if the attack is of nature and if it works.

SIEM equipment is used to improve the ability to manage any future accidents that can save time and money for incident handlers. The ability to deal with accidents rapidly and effectively will speed up the delay of occurrence, thus reducing the safety risk that cannot be followed by security events. SIEM equipment can also increase performance, mainly by offering a single report and review to display all security log data from many of the devices connected to it.

This research will provide a significant and flexible way of providing centralized log analysis between the security and network devices and how to display all threats alert information in a single dashboard. The system can assist the IT administrator in collecting, analyzing, storing, investigating, and reporting on the logs and other data for forensics, incident response, and regulatory compliance purposes and as well as analyzing real-time event data to aid the early detection of advanced threats, data breaches, and targeted attacks. Hence, the proposed framework could provide an effective way of presenting the log file to the management.

This research paper is divided into four sections and begins with the introduction. The second section is on the literature review on SIEM concepts as well as the current research relevant to this research paper. It is then followed by the third section on the proposed network behavior monitoring framework based on SIEM concepts using deep learning analysis. In the fifth section, the case study evaluation is presented. The sixth and last section is on the conclusion and future works.

2. LITERATURE REVIEW

2.1 CURRENT LOG MANAGEMENT ISSUES

2.1.1 Logs are scattered

It is very hard to compile and view each event in the campus network and therefore, all logs in the campus network have been stored individually on their system. Few tools for log management, rather than performance and capabilities, are listed in a random order [1]. Although threat detection platforms such as SIEM are significantly effective based on the recent reports that Sayed [5] found.

2.1.2 High number of false positive

Based on Filkins [4], the network administrator and the company network infrastructure monitoring are facing numerous tools which are not integrated. Open standards are developed and maintained through a collaborative process that consensus-driven to facilitate interoperability and exchange of information between different products and services.

2.1.3 Logs are scattered

For analysts, a solution needs to be created. It will not be meaningful if Syslog only pulls from various data sources. While it is not difficult to preserve the data collected with traditional methods such as hacking, it is an enormous challenge in an IoT environment to preserve the scene [2].

2.1.4 Lack of support & expertise

Some logs sometimes are a massive difficulty. As a result, the agency needs to recruit dedicated staff to support the collection, analysis, correlation, and normalization of all the logs collected, or to retain time for the current team. The rapid growth of the campus network and several IT staff help with the challenging size of data. Monitoring, maintaining, and expanding IT budgets 24/7. This means that the campus network must recruit professional staff or reserve the time of the current team to support the collection of data to detect, analyze, correlate and normalize all the logs collected. Crowley and Pescatore [3] discovered that the most often reported reasons for existing SOC failures to reach excellence are a lack of competent employees, a lack of resources, and effective automation.

2.2 SECURITY AND EVENT MANAGEMENT INFORMATION (SIEM)

2.2.1 SIEM Components

Organizations need to protect themselves regularly from the daily growing number of cyber-attacks. Security and Event Management Information (SIEM) is a security system that is widely used by various organizations to protect their networks from cyber-attacks. A SIEM solution consists of several components to assist security teams to identify data violations and malicious activities through constant monitoring and analyzing network devices and events. According to Chikonga [7], the SIEM component includes a collection from various systems, network devices, and applications collected, filtering, aggregation, normalization, and correlation of event messages. These features are a significant improvement from SIEM event log collection and storage.

Sayed [5] reports that SIEM operates hierarchically by deploying different agents. SIEM also gathers security data for specific safety equipment and tools including intrusion detection systems, firewalls, and antivirus, as well as event information from end-user devices, network equipment, and systems servers. The gathered data is sent to the centralized control and administration console. On the central console, additional logs and anomalies analysis are performed. SIEM product roles are collected, consolidated, correlated, communicated, and controlled generally. Initially, log data from various devices and applications are collected. The data is then added and standardized, as a process called consolidation. Afterward, the log data is analyzed and linked. This step enhances the usefulness of contextual network information and common threats. The data is initially kept locally in an organization's network until being transferred for analysis and archiving to a central area.

2.2.2 SIEM Architecture

Based on Figure 1, the SIEM architecture is explained as below:

- Server: It represents the core part of the entire deployment that collects and processes for the correlation engine of the logs from the external world.
- DataBase: It stores all the data to analyze the SIEM itself and to set the runtime (asset tables, taxonomies, basic modules configuration, etc.).
- FrontEnd: A console that offers a server's user interface. It provides a visual panel for the security administrator to both controls the individual component configuration and analyzes the system security under control with specific dashboards.
- Probes: A collection of sensors used within the infrastructure monitored. Probes typical examples include firewall and intrusion prevention, perimeter protection, host sensors, and security applications, for example like host IDSs.
- Agents: probes that are integrated into a server and can convert heterogeneous logs generated by various samples into logs with a syntax and a semantic.

The probe can be used to retrieve the information from IT components like routers, firewalls, web servers, anti-virus systems, and intrusion detectors so that it can generate analyzable logs. Probes usually work in two modes: active and passive. The controlled IT component cannot generate logs in the active mode, thus the sample must retrieve information actively through specific queries. For the passive mode, the component monitored is capable of generating and sending logs to the sensor in order not to require ad-hoc queries. Upon retrieval of logs, each probe may carry out preliminary security analysis by using such information. When a security issue arises, a probe alert will be generated and an information log will be sent to the respective agent representing the entry point of the SIEM architecture.

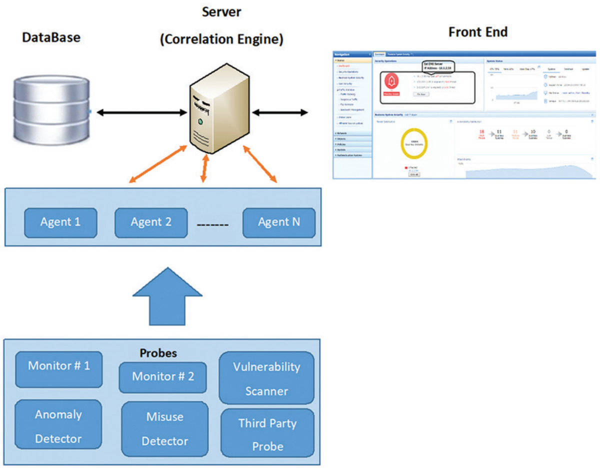


Fig. 1. A classical architecture of a SIEM system [9]

2.2.3 Benefits of SIEM

The advantage of using a SIEM tool is it serves to streamline a centralized logging solution for reporting business conformity. The host used must contain a log security record and can transfer log data to the SIEM server. Single SIEM servers can receive log data from as many devices as needed, can generate a comprehensive report, and handle every log security event it receives. It is unlikely that the efficient central logging capacity of agencies or companies without SIEM can generate fast and concise reports, as required for their reports to comply. In the current environment, the reporting of each device is essential to periodically recover data manually from each device and to ensure a central configuration for the generation of a report can be created.

The unknown events can be detected by the SIEM system server. SIEM tool can also perform better detection by connecting used equipment events or logs. The SIEM system can see attacks from various angles to each of the various equipment by collecting the events or logs of the connected equipment, and can thus record events or record logs to determine whether the attack works and its nature. SIEM tool is used to increase the ability to handle events that can save incident handlers time and resources. The ability to manage incidents fast and efficiently can accelerate the delay, thereby reducing the security risk that security events cannot be followed. The SIEM tool can also enhance efficiency mainly by reporting and analyzing each security log dataset from many connected devices.

2.2.4 SIEM Security Analysis Techniques

2.2.4.1 Event Normalization

Collecting data in the scope of SIEM is almost impossible for any person to process the data in its raw state. Through a process called event normalization, it can be performed by peculiar parsers and requires a method to modify logs and alerts delivered by probes to present homogeneous data formats to the server [9]. The process involves splitting each field of a raw event into variables and then combining them into views that are important to security administrators. This process is very important for seeking significance in often isolated and heterogeneous events.

SIEM systems can normalize logs to allow efficient analysis of data from various sources and event correlation. This normalization method includes processing logs in a readable and standardized format, extracting important data from them, and mapping the various fields that they contain.

2.2.4.2 Event Correlation

SIEM system using correlation rules to detect any potential suspicious events in the presence of encrypted real-time traffic [9]. The amount of recorded data is huge in environments. Even small to medium-sized companies are likely to send tens of GB of data every day. Sorting one-by-one authentication logs is a complicated task. Thus, using correlation rules a SIEM solution can solve the problem.

This allows administrators to see anomalies like login attempts from suspicious locations, network scans, and simultaneous user authentication attempts from various locations. The SIEM also monitors network traffic using this rule for better detection of threats and unusual activity [11]. It also can automatically extract this important information into a report or diagram that allows us to visualize activities from many sources. Events are generated using raw data to search for patterns, map them to known expressions, and assign unique identifiers. If the SIEM meets an unknown log or data type, it can define an event by the editor and allocate variables, such as name, severity, and facility.

The correlation rule is being a specific sequence of events that could be indicative of a branch in security. It combines multiple normalized events from different sources into a single correlated event [8].

2.2.4.3 Mining Process

There are millions of data and database need to be processed for useful information. In IT infrastructure, event logs provide an input of all the activities for any organization. The raw data acts as a SIEM input, providing security alerts and output reports. All raw data can be processed through a data mining technique. According to Zope et al., [12], this technique can quickly be implemented on existing software platforms and hardware platforms for enhancing the value of existing information resources. It is the process of examining data from several angles and synthesizing it into meaningful knowledge. The procedure allows people to comprehend the content of data relations. It identifies hidden patterns and trends in the data. As the data mining scope is applied to all event logs created by various networking devices, systems, and application servers, the performance of corporate security may be improved.

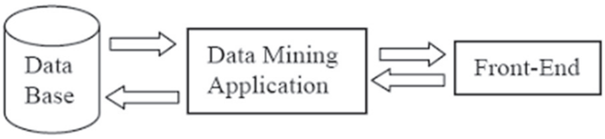


Fig. 2. Data mining architecture [12]

Based on Figure 2, the data layer might be either a database or a data storage system. An interface is a layer that connects all data sources. The findings of data mining are saved in a data layer so that they may be displayed to the user as reports or another type of visualization. To obtain database data, the application layer of the data mining is employed. There is a transformation code here that will turn the data into the necessary format. The data is subsequently processed using various data mining methods. This layer provides an easy-to-use user interface that allows end-users to engage with the data mining technology. The findings are displayed to the user via a visualization form on the front end layer.

2.2.4.4 Attack Graphs

Attack graphs represent a system's exposure. This section aims to view the approach used for calculating security measurements almost in real time. This approach should enable new security information and events in the network operating process to be taken into account and security metrics to be recalculated appropriately. To do that, it needs to develop a characterization of metrics that takes into account the following aspects. In a recent survey in the area of security metrics, modeling of attacker steps is as attack graphs. It uses known and adopted techniques to calculate security metrics. Based on these metrics, it identifies the current security situation, including attack existence, skill and position of attackers, potential previous attackers, and future attack targets.

2.2.5 How SIEM work

2.2.5.1 Collection

SIEM system collects event and logs data generated throughout a company's infrastructure from host systems, applications, and security devices such as antivirus filters and firewalls. The information that is gathered is sent to a centralized platform. The SIEM identifies and classifies data into successful and failed logins, the activity of malware, and other possibly malicious activities. So the threats can be detected and security alerts can be created. The customized dashboards and event management system of SIEM enhance investigative efficiency and reduces waste of time on false-positive elements. Real-time monitoring and incident management can be performed by SIEM for security-related events which are collected from the network, security devices, system, and applications.

2.2.5.2 Consolidation or Normalisation and Aggregation

Consolidation happens when all the collected log data from various devices and applications are then aggregated and normalized. The SIEM consolidates logs, parsers every log, and classifies them into event types, including successful and unsuccessful logins, exploits attempts, malware activities and, port scans. Peculiar parsers are used to normalize and involve in a process for manipulating logs and alerts provided by probes, to represent homogenous data formats to the server. These types of events are then run with the outline rules to decide whether illegal traffic exists. If a rule is triggered, an alert will be created. This step enhances the usefulness of contextual information concerning a network and common threats. The collected data will be stored locally in the organization’s network first before it is transferred to the analysis and archiving central area. With this step, the data security violations can be studied as closely as necessary with the data classified.

Aggregation is described as a collection of large amounts of data in one place from different applications and databases. SIEM aggregates device data and interprets key attributes related to the identification of security incidents or problems. Devices generate event logs that are submitted for analysis to the SIEM. SIEM tools form an important part of the ecosystem for data security. It aggregates data from several systems and analyzes data to capture abnormal behavior or potential cyber-attacks. SIEM tools play an important role in the collection of events and alerts.

2.2.5.3 Correlation and Contextual Information

The core of SIEM architecture for correlation combines several normalized events from various sources into one correlated event. The log storage is used to store log volumes for retention purposes and historical queries. This correlation is done when the presence of relevant patterns of events is detected. The correlation method is used as a means for performing detection at multiple layers. The analysis of the data collected through the introduction of a set of correlation rules that detect potential suspicious events as encrypted real-time traffic. To add additional data and filter duplicate events, the correlation between data and external services is preferred. If there is a feedback mechanism, a consumer of data should use the mechanism to provide providers with information to improve the quality of their service.

Every correlation was shown to gain greater insight, eliminate false-positive effects, or detect replicates for even simple cyber incidents. The process of comparing incidents is the identification of patterns and relationships to determine events from multiple sensors and data sources that are the result of an attack or a general indicator of malicious activities. It enables an improved understanding of the nature of an event, reduces the workload required to deal with incidents, and automates the classification and forwarding of incidents that are only relevant to a specific constituency. Correlation is useful both for the processing of data on a monitored network from multiple tools and for the use of multiple external services which supply incident data.

This step is decided to make more helpful by contextual information on a network and common threats. In the network organization, the data is first stored locally before they are transferred for analysis and archiving to a central area.

2.2.5.4 Communication or Alerting/Reporting

An evaluation of three elements for the basic assessment parts of the SIEM system. Firstly, the central console, secondly, the monitoring entity, and, lastly, the process of communication between the control entity and the central console. For a SIEM to operate successfully, its design and development shall provide complete, integrated information to the central console for the supervisory entity and the communication process. The core role of the SIEM solution involving the analysis and detection of system-related incidents can be compromised through attacks on communication channels.

A proper evaluation of the SIEM solution aids in preventing an attacker from evading the system. A proper

SIEM evaluation should consist of three major steps. Firstly, to guarantee that the SIEM solution identifies the greatest number of threats while producing the fewest false events and alerts, it should first assess the entities that gather, aggregate, correlate and analyze audit log data from the monitoring entity. Secondly, the audit data collector or the Agent should be assessed independently to guarantee that all information and data acquired by the Agent is valid and truthful. Finally, all communication between SIEM entities must be evaluated and ensured that no attacks such as packet injections and packet alterations occur in the channels.

2.2.5.5 Control or Storage

The usage of SIEM in IT security has been shown to improve security professional's capacity to monitor security risks. When monitoring the threat of outgoing traffic, it may be used to examine traffic going for external locations with a high-risk rating based on their IP addresses. While this traffic was created manually and only from a few log sources, the availability of this type of event data would be useful for security data in monitoring network security risks using a centralized SIEM integrated with threat intelligence services such as IP.

In a study of botnet detection, the researchers similarly describe methods for botnet detection based on output traffic monitors to potential botnet control and control centers. The monitoring of outbound traffic to enrich threats and linked to user-related event data, for example, would allow the security professionals to see more clearly both the internal source of suspect traffic and the destination of the traffic. Insider threats are more and more recognized by the need to monitor user activity as just as damaging as external security threats. The correlation of events that might be part of a composite security threat, while reducing false positives that were an issue in security systems such as IDs, can improve threat detection in an environment with a large number of log sources [7].

As with log management solutions, SIEM technology guarantees that gathered event logs from the application, network, and system components distributed inside an IT infrastructure are consolidated. SIEM, on the other hand, offers more sophisticated features. These capabilities include advanced filtering, aggregation, normalization and correlation, alerting, and reporting. The SIEM should allow an analysis of events in real or close to real-time. This research paper focuses on the use of the SIEM framework to enhance IT security management.

2.2.6 Current Research on SIEM

Here are a few examples of SIEM research conducted by other authors that are related to this research paper. In their study, K. Agrawal and H. Makwana [1] examined a few log management tools based on criteria such as data input type, primary application area, SDK accessible for languages, dashboard design capabilities, pricing, real-time supportability, online interface, etc. On the other hand, Sayed [5], discussed on SIEMs and advanced evasion techniques. The paper examined the most frequent AETs as well as the tools used to carry out such attacks. An Adaptive learning system based on an Artificial Immune System (AIS) has been proposed by A. Majeed et al. [11]. The authors use the near-miss situation based on visual analysis for the SIEM rules. L. Coppolino, et al [8] in their research paper discussed the comparison study between OSSIM and GET Data. They concluded that data collecting is the primary role of SIEM design, Indeed, by combining information from several data sources, SIEMs may provide diverse viewpoints on security incidents that occur throughout the system.

3. THE PROPOSED FRAMEWORK

According to Agrawal and Makwana [1], the best functional log management tools must consist of the following components; Log management, Log analysis, and Event Management (Figure 3). The Campus network environment required a SIEM solution which provides all in one feature to identify, analyze, correlate, normalize and security logs from multiple data source in the campus network. The complete security features must provide essential security capabilities in a collectible platform controlled by a single management console.

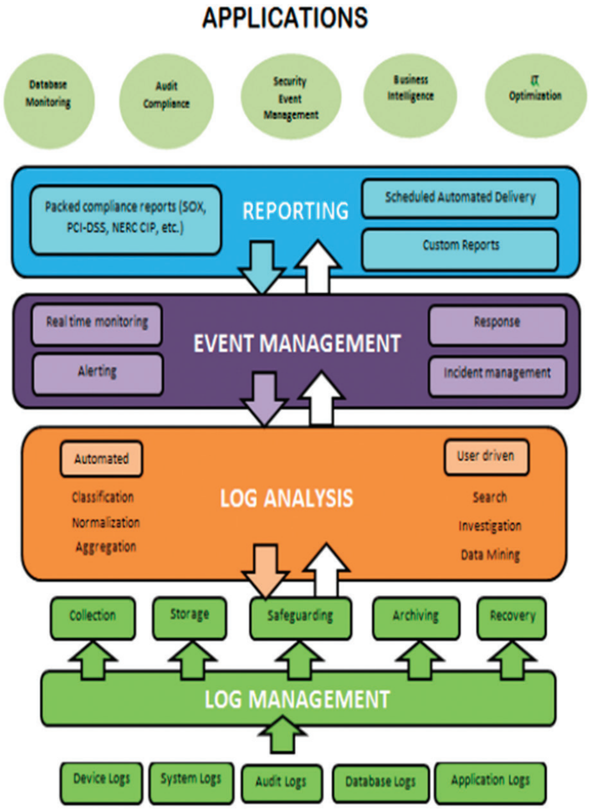


Fig. 3. Flowchart of working of log management tools [1]

Based on [1] framework, we proposed our network behavior monitoring framework to suited the campus network environment (see Figure 4). The proposed framework is divided into three components namely i. Log management, ii. Log analysis and iii Event management.

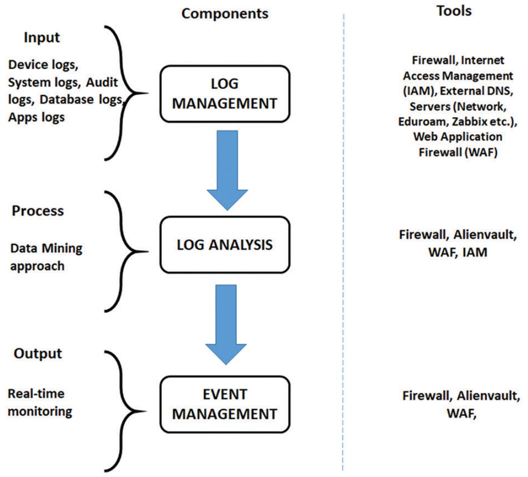

Components
Tools
Input
Device logs, System logs, Audit logs, Database logs, Apps logs
LOG MANAGEMENT
Firewall, Internet Access Management (IAM), External DNS, Servers (Network, Eduroam, Zabbix etc.), Web Application Firewall (WAF)
Process
Data Mining approach
LOG ANALYSIS
Firewall, Alienvault, WAF, IAM
Output
Real-time monitoring
EVENT MANAGEMENT
Firewall, Alienvault, WAF,

Fig. 4. The proposed framework of network behavior for campus network

3.1 COMPONENT 1 – LOG MANAGEMENT

Comprehensive centralized log management is a must to simplify the campus network. Centralized event log management is the most important component of monitoring network security and forensics, as it can analyze events that occur from thousands of nodes to several dedicated servers where central analysis is carried out. The analyses obtained can be a real-time process, where safety events can be identified from future events through event correlation and other advanced surveillance techniques; it can also be an offline forensic activity, where past events are investigated to investigate the occurrence of security that has taken place.

The underlying principles of every analysis of network & security reporting system is to aggregate data generated from different sources, identifying relevant threats, and taking appropriate action. For example, when certain problems are identified, the system will take additional log information, generate warnings and ensure other security controls can be controlled and stop them.

Log management infrastructure is a component covering networking, software, hardware, and media used to store, transmit, generate, analyze and delete log data. Almost all organizations have one or more log management infrastructure.

The main functions of a log management tool are:

- Identify and collect all logs of events involving involved software such as operating system, Syslog, flat file, database, or application
- Ensure all logs are stored integrally, scalable and secure
- All logs can be obtained quickly, fast and flexible
- Logs can be retrieved and stored for a long time
- Systems, databases, applications, databases, and devices are available in real-time.
- Normalization, aggregation, log classification, and correlation can be automated more efficiently.

3.2 COMPONENT 2 – LOG ANALYSIS

The log analysis describes how the security and traffic threats, the complete security log analysis obtained provides critical network intelligence for attempts to violate security and attacks such as viruses, trojans, service denials, and others. From the log report analysis obtained from the NGFW and WAF, security administrators and networks will be able to interpret network threatening activities and plan their strategies to protect and address threats that occur.

For servers (Eduroam and Sybase) include operating systems involving Windows, Linux, and Unix systems and other Syslog support devices, and applications such as IIS, MS SQL event log analysis is recorded for secure security. Important logs and security events are recorded and generated on equipment within the network, this SIEM system can collect important information, perform log analysis, and display all logs and events on SIEM Dashboard, in real-time and concise real-time.

Internet Access Management (IAM) is used to analyze traffic logs that provide detailed and valuable information about bandwidth usage, employee internet usage, web page confusion broadband, and smart interface traffic. From the firewall analysis report, network/security administrators will monitor the fair use of broadband to reduce existing traffic safety and data security to plan for future broadband capacity requirements.

This framework was proposed using the deep learning approach for log evaluation. Deep learning is a complex element of machine learning inspired by the function of interconnecting neurons in the human brain. The evolution of Machine Learning and an element of AI teaches itself to make more accurate and faster predictions by observing, processing and analyzing massive amounts of data. Over time, Deep Learning teaches itself every time it is executed, resulting in the identification of many previously undetected malicious domain names.

3.3 COMPONENT 3 – EVENT MANAGEMENT

In the Security Assessment System, the part that implements the proposed security assessment technique is based on attack graphs. Figure 5 shows the architecture of the component. The main component is the suite of vulnerability assessment algorithms for the computation of metrics. Mapper, another important subcomponent, allows an attacker to detect location and attack structure based on security events. The security assessment component receives data from several sources, including an attack graph generator which creates reports from network analysis. Dependency graphs showing graphical dependencies between network services are also given. Output data has several security parameters, according to the proposed categorization. Additional output data would be received from the visualization system.

4. THE PROPOSED METHODOLOGY

Figure 6 shows the methodology which is split into four main steps:

1. Data Acquisition,
2. Data Extraction and Enrichment,
3. Reporting, Alerting, and Monitoring, and
4. Dashboard, Forms, and Integration.

4.1 DATA ACQUISITION (STEP 1)

In the log management, all log sources can be found. The log sources include application logs, audit logs, device logs, system logs, and database logs. In [5] article said that the logging process should be automated, precise, and visible to have a secure custody chain. These logs collected using several tools. Examples of tools used are External Domain Name Servers (Ext DNS), Intelligent Management Center (IMC). Dynamic Host Configuration Protocol (DHCP) server and firewalls like Web Application Firewall (WAF) and Next-Generation Firewall (NGFW).

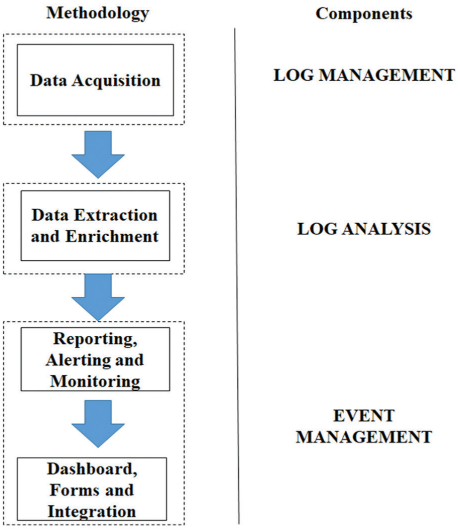

Methodology
Components
Data Acquisition
LOG MANAGEMENT
Data Extraction and Enrichment
LOG ANALYSIS
Reporting, Alerting and Monitoring
EVENT MANAGEMENT
Dashboard, Forms and Integration

Fig. 6. Proposed Methodology

There are also fewer standard data sources including Internet Access Management (IAM) systems and all other tools that can provide relevant information to better identify organizational safety cases. The proposed system collects data from different security and network devices in the form of original logs and converts the collected logs into events through normalization and parsers. As a data exchange format for event data between different hardware and equipment and regulation engines became standard format is needed. It should differentiate between formats for content and formats for series. The structure of the data to be transferred (e.g. Event data includes two fields as "event timestamp" and "information about acknowledged attacks") is descriptive to content format [14].

4.2 DATA EXTRACTION AND ENRICHMENT (STEP 2)

In the proposed framework, data extraction is the method of taking raw data input and extracting only related fields, which is known as normalization. The standard event format can be created from all sources of data that allow the consistent comparison and collection of information across the network [15]. According to [7], the context allows a more comprehensive analysis by providing further information related to the relevant event.

A firewall is a key module that makes up the proposed framework. The information from the firewall constitutes the basic source of the log and event data of the system. Currently, vendors are producing and providing very sophisticated firewalls that can detect and prevent malicious activities from attackers. The firewall filter and inspect the traffic. Monitor the process by allowing all devices to collect audit data such as system logs and firewall alerts. Such data will be sent to the central console in the proposed system, where it will be aggregated, correlated, analyzed, and reported to prevent abnormalities. Based on [16] the study found two potential sources of error in the information enrichment process. The first is the unstructured original data and the other is the accuracy of the methods used for machine learning to obtain data and information extraction. All the logs that are related to this NGFW firewall in the network will go through the outbound and inbound of the firewall.

The process of association analysis can usually be divided into three parts that are,

- filter redundant information and format safety information,
- match association rules, and
- generate security incidents.

The first two techniques can be used in the proposed system to detect anomalies. It is also known as an abnormal association rule. This rule is valid only when defining the threshold. By comparing the rules of the normal category data set with the rules of the actual traffic category data set based on the similarity measure, anomaly detection can be calculated. If the similarity result is higher than the user threshold, it means that the data set is not intrusive and vice versa. The general category data set is reference data and should not contain intrusions. To apply this technique, the data must first be converted into a data set.

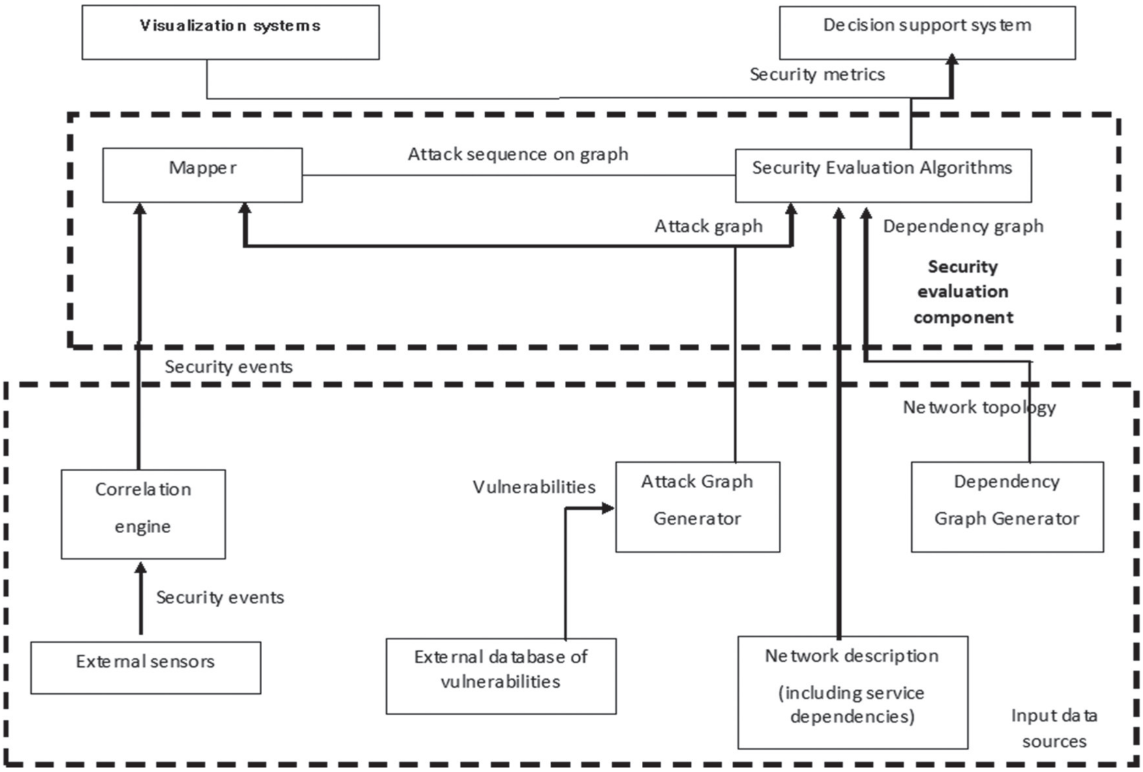


Fig. 5. Security evaluation component architecture [10]

4.3 REPORTING, ALERTING, AND MONITORING (STEP 3)

In the actual SIEM deployment environment, the monitoring requirements determined are usually mapped from specific business monitoring requirements. Monitoring requirements can be seen as business or IT security requirements for active tracking, alerting, and reporting. It is recommended to be as specific as possible when specifying monitoring requirements [7]. The proposed system will present the results and findings in a way that depends on the user's role after analyzing the log data. For all log sources, this information is displayed in a user-readable and understandable format. The reporting function allows users to configure log information files and only capture relevant information about their tasks. In addition to real-time tracking, log data analysis, and interactive reports for visual records, the proposed solution also provides compliance reporting functions, which can provide detailed and achievable audit logs.

In addition, the proposed solution has compliance reporting features, which generate a detailed and operational audit log record, in addition to real-time monitoring, log analysis, and interactive visual information reporting. During the audit of the organization, the auditor may review the records, the information reports, and other regulatory-specific content, to ensure compliance with the regulations. In short, the data will give in statistical format, such as reports, graphs.

4.3.1 Alerting threats

Network and system protections have developed as the threats grow over the years. Distributed denial of services (DDoS) and other types of attacks affect organizations worldwide. Another major network threat is unauthorized access. It occurs when a malicious user invades an account and uses it to change permissions, gain access to resources or information, and other malicious activities. This is the route for hackers to execute APT and is a common problem in large organizations, which usually keep confidential information and other data with high business value. To secure the network, AlienVault generated alerts that provide response procedures. Of course, the data itself will be passed on when received for real-time analysis and monitoring. Analysis of the data from the devices used can show patterns and how conducive activities are to the fingerprints of individuals or groups of threats, they are detected throughout the sensor network through a trace of scattered data [17].

4.3.2 Monitoring logs

The proposed systems will identify and warn a network in real-time if an incident and an important protection problem are detected. SIEM's key functions within the network of an organization are to track, capture and archive log data in a central console. The log data must be analyzed, warnings filtered and correlation rules developed. Both file origins are included in the log management. The log sources include network-based applications, systems, and computers. Based on

[18], effective log monitoring requires active data log analysis due to the size and quantity of the log file.

4.4 DASHBOARD, FORMS, AND INTEGRATION (STEP 4)

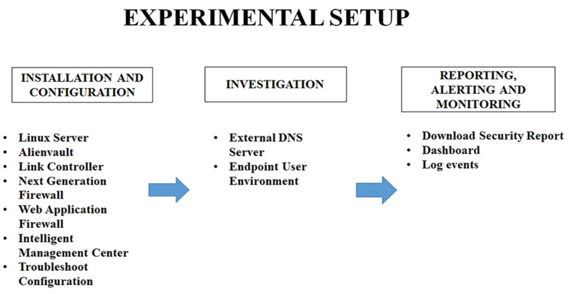
The infographic of network security provides for the visual depiction of security data, which may help users understand complicated technical information and security aspects [19]. The visual representation also enables users to search for specific types of graphical chart methods, appropriate visualization methods, classification of security data visualizations, etc., allowing us to meet the growing demand for network security monitoring.

5. CASE STUDY INVESTIGATION – EVALUATION OF THE PROPOSED FRAMEWORK

Unauthorized access introduces serious security problems. The SIEM system is of great significance in dealing with the security issues of critical infrastructure. The proposed solution based on the SIEM framework can monitor the network at any time to detect and issue alerts when incidents and serious security issues are discovered. Figure 7 shows the experimental setup for this study.

5.1 UNAUTHORIZED ACCESS – AN EXTERNAL DNS SERVER (HACKED)

Figure 8 and Figure 9 shows an example for unauthorized access from an external DNS server from NGFW where it can detect the security operations found three IP that threatened the external DNS Server and the graph and other analysis.

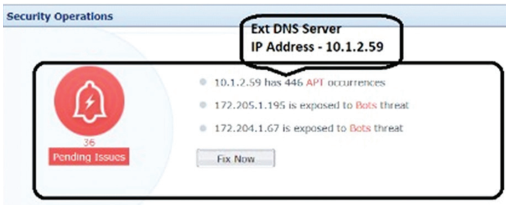

EXPERIMENTAL SETUP

INSTALLATION AND CONFIGURATION
- Linux Server
- Alienvault
- Link Controller
- Next Generation Firewall
- Web Application Firewall
- Intelligent Management Center
- Troubleshoot Configuration

INVESTIGATION
- External DNS Server
- Endpoint User Environment

REPORTING, ALERTING AND MONITORING
- Download Security Report
- Dashboard
- Log events

Fig. 7. Experimental Setup for this project



- 10.1.2.59 has 446 APT occurrences
- 172.205.1.195 is exposed to Bots threat
- 172.204.1.67 is exposed to Bots threat

Fig. 8. The security operations found three IP that threatened the external DNS Server (Ext DNS server)

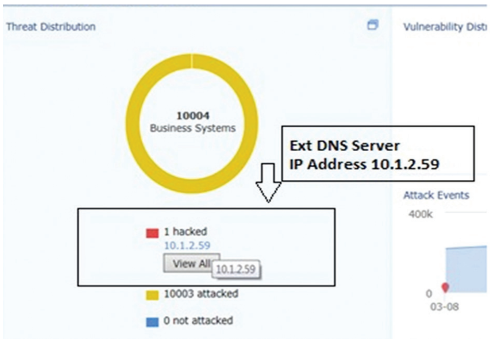
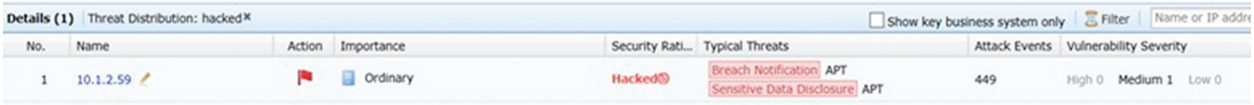
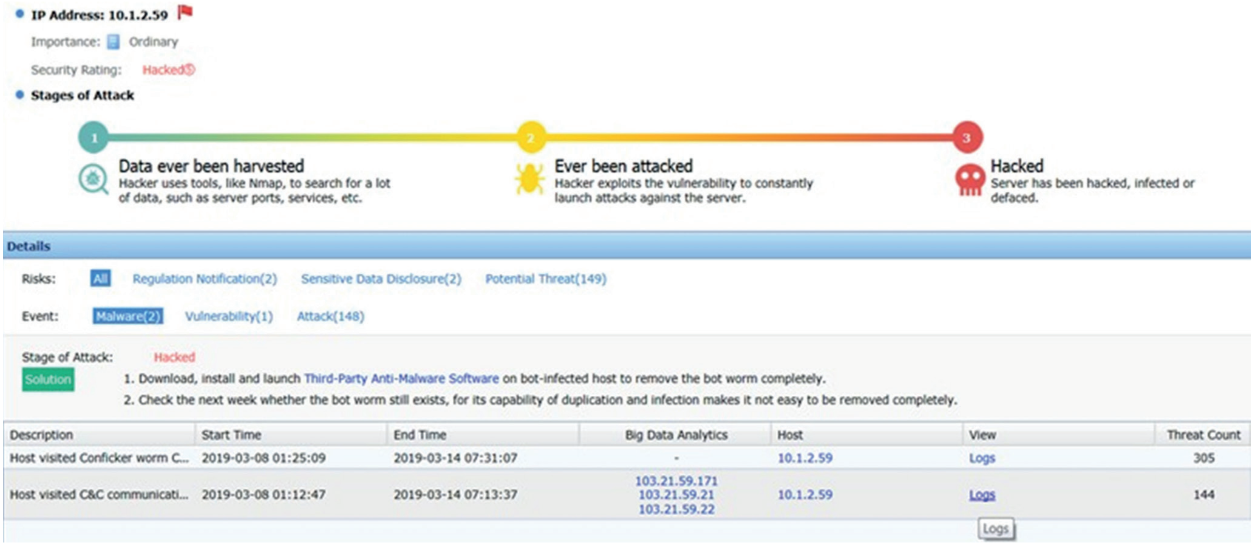


Fig. 9. The graph and other analysis



No.	Name	Action	Importance	Security Rati...	Typical Threats	Attack Events	Vulnerability Severity
1	10.1.2.59		Ordinary	Hacked	Breach Notification APT Sensitive Data Disclosure APT	449	High 0 Medium 1 Low 0

Fig. 10. The detail of the threat


IP Address: 10.1.2.59

Importance: Ordinary

Security Rating: Hacked

Stages of Attack

1 Data ever been harvested
Hacker uses tools, like Nmap, to search for a lot of data, such as server ports, services, etc.

2 Ever been attacked
Hacker exploits the vulnerability to constantly launch attacks against the server.

3 Hacked
Server has been hacked, infected or defaced.

Details

Risks: All Regulation Notification(2) Sensitive Data Disclosure(2) Potential Threat(149)

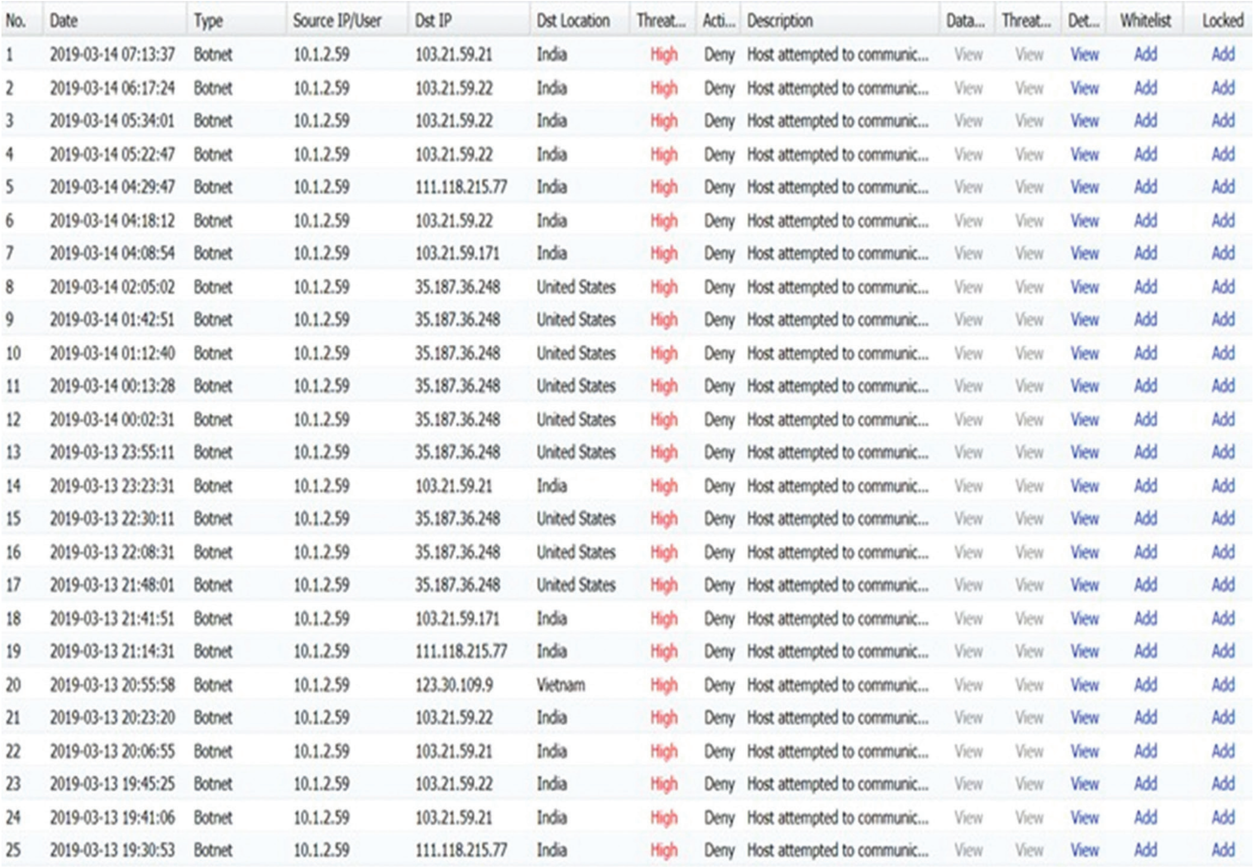
Event: Malware(2) Vulnerability(1) Attack(148)

Stage of Attack: Hacked

Solution
1. Download, install and launch Third-Party Anti-Malware Software on bot-infected host to remove the bot worm completely.
2. Check the next week whether the bot worm still exists, for its capability of duplication and infection makes it not easy to be removed completely.

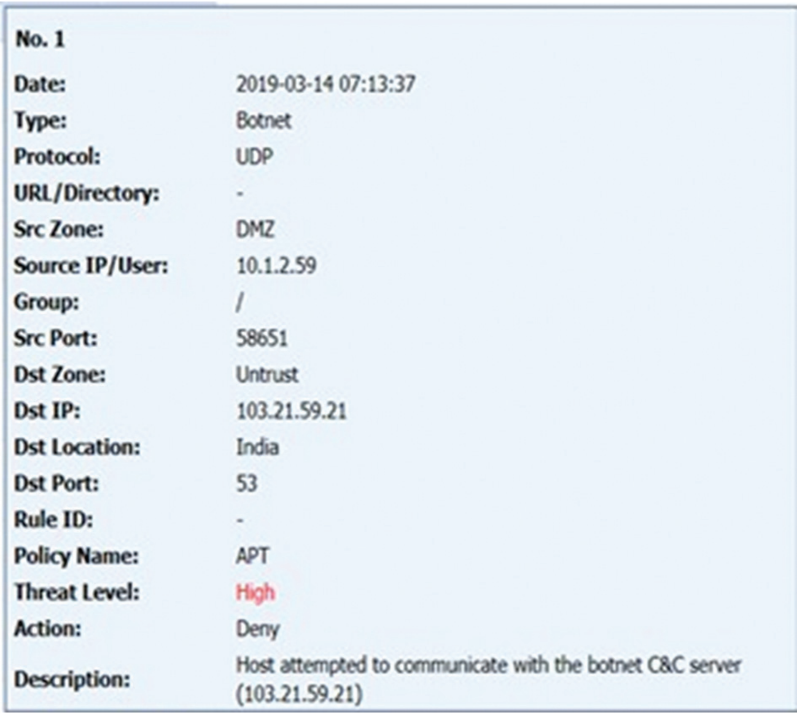
Description	Start Time	End Time	Big Data Analytics	Host	View	Threat Count
Host visited Conficker worm C...	2019-03-08 01:25:09	2019-03-14 07:31:07	-	10.1.2.59	Logs	305
Host visited C&C communicati...	2019-03-08 01:12:47	2019-03-14 07:13:37	103.21.59.171 103.21.59.21 103.21.59.22	10.1.2.59	Logs	144

Fig. 11. The summary of the attacks



No.	Date	Type	Source IP/User	Dst IP	Dst Location	Threat...	Acti...	Description	Data...	Threat...	Det...	Whitelist	Locked
1	2019-03-14 07:13:37	Botnet	10.1.2.59	103.21.59.21	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
2	2019-03-14 06:17:24	Botnet	10.1.2.59	103.21.59.22	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
3	2019-03-14 05:34:01	Botnet	10.1.2.59	103.21.59.22	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
4	2019-03-14 05:22:47	Botnet	10.1.2.59	103.21.59.22	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
5	2019-03-14 04:29:47	Botnet	10.1.2.59	111.118.215.77	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
6	2019-03-14 04:18:12	Botnet	10.1.2.59	103.21.59.22	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
7	2019-03-14 04:08:54	Botnet	10.1.2.59	103.21.59.171	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
8	2019-03-14 02:05:02	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
9	2019-03-14 01:42:51	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
10	2019-03-14 01:12:40	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
11	2019-03-14 00:13:28	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
12	2019-03-14 00:02:31	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
13	2019-03-13 23:55:11	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
14	2019-03-13 23:23:31	Botnet	10.1.2.59	103.21.59.21	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
15	2019-03-13 22:30:11	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
16	2019-03-13 22:08:31	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
17	2019-03-13 21:48:01	Botnet	10.1.2.59	35.187.36.248	United States	High	Deny	Host attempted to communic...	View	View	View	Add	Add
18	2019-03-13 21:41:51	Botnet	10.1.2.59	103.21.59.171	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
19	2019-03-13 21:14:31	Botnet	10.1.2.59	111.118.215.77	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
20	2019-03-13 20:55:58	Botnet	10.1.2.59	123.30.109.9	Vietnam	High	Deny	Host attempted to communic...	View	View	View	Add	Add
21	2019-03-13 20:23:20	Botnet	10.1.2.59	103.21.59.22	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
22	2019-03-13 20:06:55	Botnet	10.1.2.59	103.21.59.21	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
23	2019-03-13 19:45:25	Botnet	10.1.2.59	103.21.59.22	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
24	2019-03-13 19:41:06	Botnet	10.1.2.59	103.21.59.21	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add
25	2019-03-13 19:30:53	Botnet	10.1.2.59	111.118.215.77	India	High	Deny	Host attempted to communic...	View	View	View	Add	Add

Fig. 12. APT logs 1



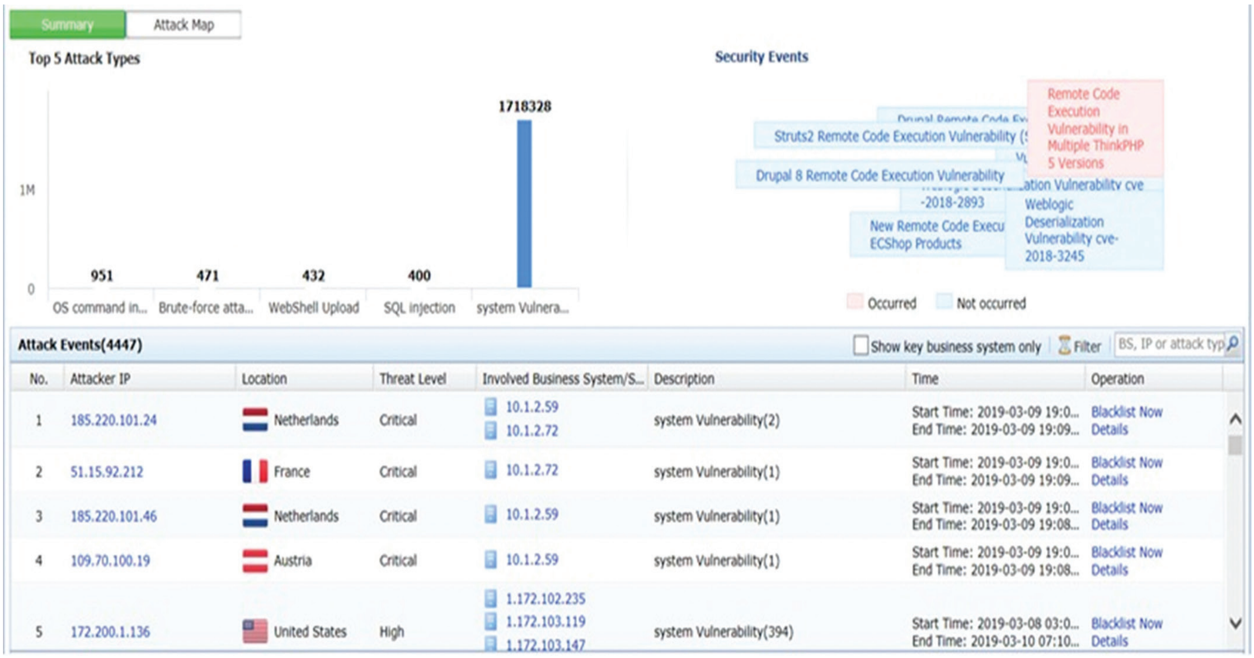
No. 1

Date:	2019-03-14 07:13:37
Type:	Botnet
Protocol:	UDP
URL/Directory:	-
Src Zone:	DMZ
Source IP/User:	10.1.2.59
Group:	/
Src Port:	58651
Dst Zone:	Untrust
Dst IP:	103.21.59.21
Dst Location:	India
Dst Port:	53
Rule ID:	-
Policy Name:	APT
Threat Level:	High
Action:	Deny
Description:	Host attempted to communicate with the botnet C&C server (103.21.59.21)

Fig. 13. APT logs 2

While, Figure 10, Figure 11, Figure 12, and Figure 13 show an example of unauthorized access from an external DNS Server from NGFW where it can detect the detail of the threat, the summary of the attacks come from, APT logs 1 and logs 2 actions.

Lastly, Figure 14 and Figure 15 shows the sample attack and map event of the attacker location.

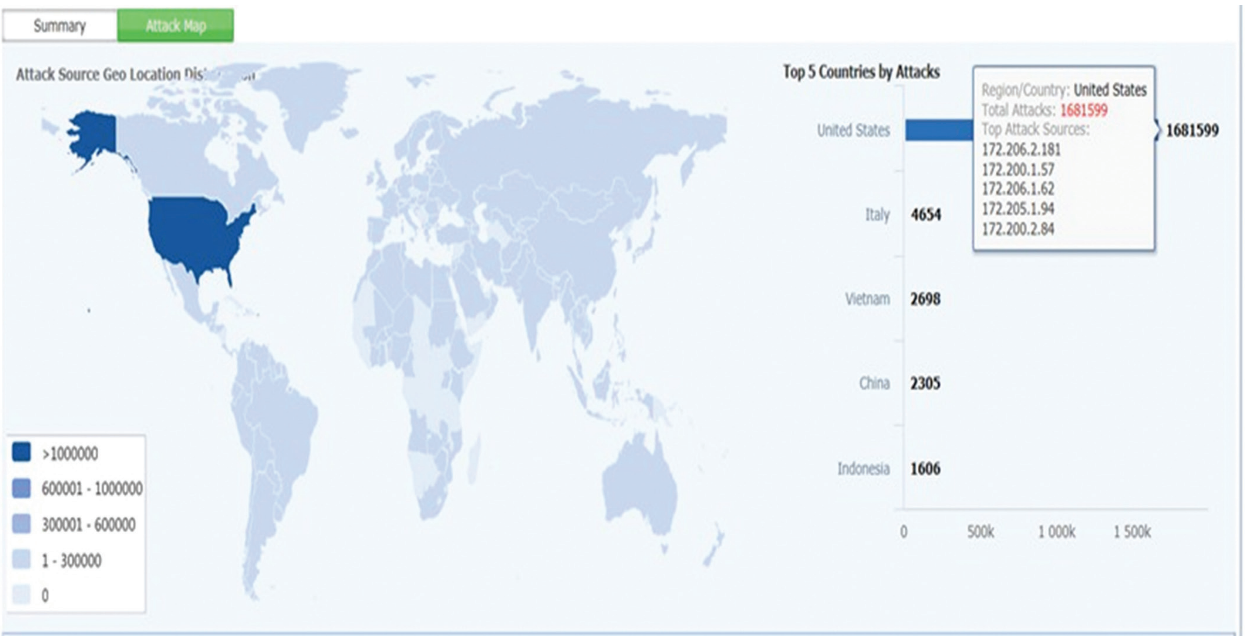

Summary
Attack Map
Top 5 Attack Types
1718328
1M
0
951
471
432
400
OS command in...
Brute-force atta...
WebShell Upload
SQL injection
system Vulnera...
Security Events
Remote Code Execution Vulnerability in Multiple ThinkPHP 5 Versions
Drupal Remote Code Ex
Struts2 Remote Code Execution Vulnerability (S
Drupal 8 Remote Code Execution Vulnerability
...ation Vulnerability cve -2018-2893
New Remote Code Execu ECShop Products
Weblogic Deserialization Vulnerability cve-2018-3245
Occurred
Not occurred

Attack Events(4447)

□ Show key business system only Filter BS, IP or attack typ

No.	Attacker IP	Location	Threat Level	Involved Business System/S...	Description	Time	Operation
1	185.220.101.24	Netherlands	Critical	10.1.2.59 10.1.2.72	system Vulnerability(2)	Start Time: 2019-03-09 19:0... End Time: 2019-03-09 19:09...	Blacklist Now Details
2	51.15.92.212	France	Critical	10.1.2.72	system Vulnerability(1)	Start Time: 2019-03-09 19:0... End Time: 2019-03-09 19:09...	Blacklist Now Details
3	185.220.101.46	Netherlands	Critical	10.1.2.59	system Vulnerability(1)	Start Time: 2019-03-09 19:0... End Time: 2019-03-09 19:08...	Blacklist Now Details
4	109.70.100.19	Austria	Critical	10.1.2.59	system Vulnerability(1)	Start Time: 2019-03-09 19:0... End Time: 2019-03-09 19:08...	Blacklist Now Details
5	172.200.1.136	United States	High	1.172.102.235 1.172.103.119 1.172.103.147	system Vulnerability(394)	Start Time: 2019-03-08 03:0... End Time: 2019-03-10 07:10...	Blacklist Now Details

Fig. 14. Attack events



Region/Country: United States
Total Attacks: 1681599
Top Attack Sources:
172.206.2.181
172.200.1.57
172.206.1.62
172.205.1.94
172.200.2.84

Fig. 15. The map of attack events

6. CONCLUSION AND FUTURE WORKS

Nowadays, High Education Institution (HEI) is working towards the deployment of SIEM as a solution to provide control across the whole network by gathering logs from both security and network equipment. Throughout the framework, logs may be correlated to include reliable and appropriate threat warnings. As a result, this research paper outlines the proposed network behavior monitoring architecture based on SIEM principles using a deep learning analysis. The case study to evaluate the proposed framework is also presented. Future studies will concentrate on evaluating a deep learning approach to DDOS attacks to determine the detection accuracy rate.

7. REFERENCES:

[1] K. Agrawal, H. Makwana, "Data Analysis and Reporting using Different Log Management Tools", International Journal of Computer Science and Mobile Computing, Vol. 47, No. 7, 2015, pp. 224-229.

[2] M. Conti, A. Dehghantanha, K. Franke, S. Watson, "Internet of Things security and forensics: Challenges and opportunities", Future Generation Computer Systems, Vol. 78, 2018, pp. 544-546.

[3] C. Crowley, J. Pescatore. "Common and Best Practices for Security Operations Centers: Results of the 2019 SOC Survey", https://www.dflabs.com/lp/thank-you-for-downloading-the-sans-2019-soc-surveyreport/?__s=aivfmwwer9oqny8sxwyf&drip_email=jack.whitter-jones%40southwales.ac.uk&drip_subscriber_id=aivfmwwer9oqny8sxwyf (accessed: 2021)

[4] B. Filkins, "2019 SANS Automation & Integration Survey", https://www.sans.org/media/vendor/Automation-and-Integration-Survey.pdf (accessed: 2021)

[5] M. Z. Seyed, "Analysis of Security Information and Event Management (SIEM) – Evasion and Detection Methods", Tallinn University of Technology, Faculty of Information Technology, Tallinn, Estonia, Master Thesis, 2016.

[6] A. Khan, R. Khan, F. Nisar, "Novice threat model using SIEM System for Threat Assessment", Proceedings of the 2th International Conference on Communication Technologies, Rawalpindi, Pakistan, 19-21 April 2017, pp. 72-77.

[7] M. Chikonga, "Exploring the Applicability of SIEM Technology in IT Security", Auckland University of Technology, Auckland, New Zealand, Master Thesis, 2014.

[8] L. Coppolino, S. D'Antonio, V. Formicola, L. Romano, "A framework for mastering heterogeneity in multi-layer security information and event correlation", Journal of Systems Architecture, Vol. 62, 2016, pp. 78-88.

[9] M. Di Mauro, C. Di Sarno, "Improving SIEM capabilities through an enhanced probe for encrypted Skype traffic detection", Journal of Information Security and Applications, Vol. 38, 2018, pp. 85-95.

[10] I. Kotenko, E. Doynikova, "Security assessment of computer networks based on attack graphs and security events", Lecture Notes in Computer Science, 8407, 2014, pp. 462-471.

[11] A. Majeed, R. ur Rasool, F. Ahmad, M. Alam, N. Javaid, "Near-miss situation based visual analysis of SIEM rules for real time network security monitoring", Journal of Ambient Intelligence and Humanized Computing, Vol. 10, No. 4, 2019, pp. 1509-1526.

[12] A. R. Zope, A. Vidhate, N. Harale, "Data Mining Approach in Security Information and Event Management", International Journal of Future Computer and Communication, Vol. 2, No. 2, 2013, pp. 80-84.

[13] K. O. Detken, M. Jahnke, C. Kleiner, M. Rohde, "Combining Network Access Control (NAC) and SIEM functionality based on open source", Proceedings of the IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Bucharest, Romania, 21-23 September 2017, pp. 300–305.

[14] H. A. Khan, "Advancing Security Information and Event Management Frameworks in Managed Enterprises using GeoLocation", University of Cape Town, Faculty of Science, Department of Computer Science, Master Thesis, 2014.

[15] P. Andruszkiewicz, H. Rybinski (2018), "Data Acquisition and Information Extraction for Scientific Knowledge Base Building". Proceedings of the 12th IEEE International Conference on Semantic Computing, Laguna Hills, CA, USA, 31 January - 2 February 2018, pp. 256–259.

[16] B. D. Bryant, & H. Saiedian, "Improving SIEM alert metadata aggregation with a novel kill-chain based classification model", Computers and Security, Vol. 94, 2020, p. 101817.

[17] I. Yagoub, M.A. Khan, L. Jiyun, "IT Equipment Monitoring and Analyzing System for Forecasting and Detecting Anomalies in Log Files Utilizing Machine Learning Techniques", 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems, Durban, South Africa, 6-7 August 2018, pp. 1-6.

[18] A. Majeed, R. ur Rasool, F. Ahmad, M. Alam, N. Javaid, "Near-miss situation based visual analysis of SIEM rules for real time network security monitoring", Journal of Ambient Intelligence and Humanized Computing, Vol. 10, No. 4, 2019, pp. 1509–1526.