Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Comparative Analysis of Open Source Security Information & Event Management Systems (SIEMs)

Konstantinos Bezas[1], Foteini Filippidou[2]
kobezaa@mst.ihu.gr, fnfilip@cs.ihu.gr
International Hellenic University, Greece

Article Information

Submitted : 1 Apr 2023
Reviewed: 3 Apr 2023
Accepted : 27 Apr 2023

Keywords

SIEM System, Cyber Attack, Security Information Management, Security Event Management

Abstract

A Security Information and Event Management system (SIEM) is a tool used to collect, analyze, normalize and correlate data from various devices to identify potential cyber threats almost in real-time. SIEM provides a unified approach to security issues through two zones: Security Information Management (SIM) and Security Event Management (SEM). SIM deals with managing logs and reporting, while SEM deals with event management and real-time monitoring. SIEM tools collect data events in a central unit from various devices, normalize their format, analyze them, and generate reports and alerts. SIEM combines the ability of log management to generate a compliance report with the ability to manage threats. However, the central approach may present significant disadvantages, such as slowing system performance and complicating the prioritization of queries.

A. Introduction

What is a security information and event management system (SIEM)

As cyber attacks started to grow, created the requirement applying correlation mechanisms which was the beginning of SIEMs. A correlation mechanism assumes analyzing specific features of log data, and correlation with other elements, almost in real time to identify potential attacks [1].

SIEM provides a unified approach to security issues, although it can be considered as a two-zone system:

- **Security Information Management** (SIM) : this piece deals with managing logs and reporting,
- **Security Event Management** (SEM) for event management and real-time monitoring [2].

SIEM tools are used to collect, analyze, normalize and correlate data from different devices. They collect data events in a central unit, from end user devices, servers, network devices, firewalls, intrusion detection and prevention systems (IDPS), normalize its format, relate and analyse them and finally generate reports and alerts [3].

In general, a SIEM system combines the ability of log management in order to generate a compliance report with the ability to manage threats [4].

The central approach presents significant disadvantages. For example, collecting and indexing raw log information in a large central database may slow system performance and complicate the prioritization of queries.

The rest of this paper is organized as follows. In section two SIEM Architecture. In section three, we describe the Background. Next, in section four, Evaluation criteria - research. in section five Open Source Systems analysis/presentation and finally, section six closes with conclusions.

B. SIEM Architecture

1. Components of SIEM products

A typical SIEM architecture includes the following basic components [5]:

- **Data aggregation**: Collects and aggregates data from security systems and network devices.
- **Threat intelligence feeds**: Combines internal data with third-party data on threats and vulnerabilities.
- **Correlation and security monitoring**: Links events and related data into security incidents, threats or forensic findings.
- **Analytics**: uses statistical models and machine learning to identify deeper relationships between data elements.
- **Alerting**: Analyses events and sends alerts to notify security staff of immediate issues
- **Dashboards**: Creates visualizations to let staff review event data, identify patterns and anomalies.

- **Compliance**: Gathers log data for standards like HIPAA (a United States standard pertaining to organizations that transmit health information in electronic form), PCI/DSS (Payment Card Industry Data Security Standard) that created to secure credit cardholder data from theft and misuse, HITECH, SOX (Sarbanes-Oxley) a regulation that sets requirements for US public company boards, management and accounting firms and GDPR (General Data Protection Regulation) and generates reports.
- **Retention**: Stores long-term historical data, useful for compliance and forensic investigations.
- **Forensic analysis**: Enables exploration of log and event data to discover details of a security incident.
- **Threat hunting**: Enables security staff to run queries on log and event data to proactively uncover threats.
- **Incident response**: Helps security teams identify and respond to security incidents, bringing in all relevant data rapidly.
- **SOC automation**: Advanced SIEMs can automatically respond to incidents and orchestrate security systems, known as Security Orchestration and Response (SOAR).

In a physical system the above architecture is implemented as shown in figure 1 and described below[2]:

The **correlation mechanism** of a SIEM, which is responsible for gathering events from heterogeneous inputs, as well as normalizing, filtering, and reducing them, is a vital element of their **architecture** [1]. It combines existing event sources, such as agents, security services, and device logs as pieces of a puzzle, to provide fast results (i.e., security event alerts) [6]. A SIEM correlation rule specifies which sequence of events could be indicative of abnormalities that may indicate security vulnerabilities or cyber attacks. We should pay attention to the development of SIEM correlation rules because they can create false positives and divert security managers' efforts to respond to real threats and attacks. It is impossible to have zero false positives but we should try to reduce them and not missing anomalies that could indicate cyberattacks [7]. For example, a rule could be created to identify when more than x amount of requests are sent from specific IP ports within a certain amount of time [8].

A Server is responsible for collecting and processing data from the external world as well as from sources internal to the organization. All data are stored in a **Database** for analysis and runtime configuration. **FrontEnd** is the interface console between user and server. **Probes** are software or hardware elements (sensors) that collect information from the monitored infrastructure. A probe analyses the information, produces a log and in case of a security issue, creates alerts and sends informational log to the agent. The **Agents** are probes which are embedded into the server and can convert heterogeneous logs generated by different probes into logs with the same syntax and specific semantics.

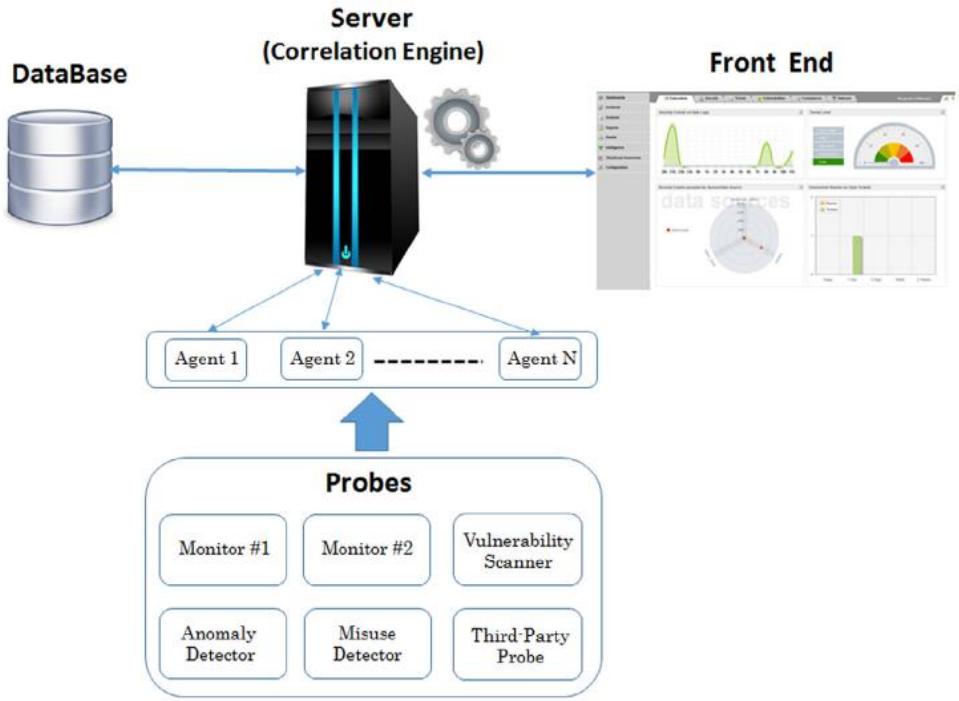

Figure 1. SIEMs Architecture [2]

2. Capabilities of SIEM products

A traditional siem system has the following basic capabilities [3]:

- **Real-time security tracking:** The central storage and log correlation allows real-time analysis providing alerts about live activity or attacks to take defensive measures
- **Threat Intelligence**: It provides comprehensive information and refining knowledge about the most common external threats that may endanger an organization.
- **Behavior Profiling**: learning the user activity and how an organization uses a resource, creates a regular activity profile for different event categories, so it will alert when a possible deviation from normal behavior is observed.
- **Data & User Monitoring**: checks the identity and authority of a user. After the user's authentication, checks for the authorized files in database that he can access. Any access or modification of an unauthorized file, will be considered abnormal activity and will generate an alert.
- **Application Monitoring:** targeted attacks exploit the weaknesses of an application, such as bugs or vulnerabilities. App level monitoring is the ability to analyze activity streams from applications
- **Analytics**: Includes discovery, interpretation, and communication of important patterns in data security analysis. investigates user's activity and access to detect a threat, violation or abuse of privileges.

- **Log Management and Reporting:** a SIEM system, manages, stores and analyzes large log files from different sources, such as server logs, system logs, event logs, firewalls, etc. to report an alert.

Traditional SIEM solutions are limited and don't have the flexibility to scale with security requirements while the next generation fills the gaps in the functionality and growing needs about cyberthreat. Next-generation SIEM collects a wider range of data and reduces the mean time to detect (MTTD) and response (MTTR) and monitors these metrics [9].

Above described some of the features of the next generation of SIEMs [9] :

- **Collect and manage data from all available sources:** This includes cloud service data, on-premise log data (security controls, databases, and application logs), and network data (flows, packets, etc.).
- **Well-vetted, big data architecture:** a big data architecture is needed that can scale the amount of data being collected.
- **Flat pricing for log ingestion:** the pricing is independently of the data you collect. You can ingest data from all sources and remain within your budget.
- **Automated tracking of lateral movement:** it is known that most of attackers involve lateral movement to evade detection or gain access to higher privileges by changing credentials, IP addresses, and assets. To effectively follow lateral movements from beginning to end, the SIEM must be able to tie such related events together.
- **Improved security information model:** security data stored in a useful form factor such as a timeline that contains a complete overview of each entity we are monitoring while legacy SIEM's model mostly based on discrete events. Thus when surfacing abnormal events, expert systems immediately provide their complete context
- **Prebuilt incident timelines:** using a legacy SIEM usually requires a combination of complex queries which is time consuming and requires deep security domain expertise, mastery of query languages, and the ability to interpret results. A modern SIEM can present all available context in a concise and friendly UI.
- **Enrichment of user and asset context:** advances in data science provide many insights that previously had to be correlated by experienced analysts. By using a SIEM that understands context and intent, you can look up asset ownership, user login location, peer groups, and other information that can help you discover abnormal behaviors.
- **Incident prioritization :** large companies generate hundreds of millions of log entries every day that must manage a SIEM. The ability to eliminate false positives and focus only on events with abnormal behaviors is essential for robust security..
- **User Event Behavioral Analysis (UEBA) :** advanced SIEMs use Artificial Intelligence (AI) and deep learning techniques to test human behavior patterns in order to detect threats of internal users that are the major threat in an

organization. UEBA technique can help to identify malicious activity before it leads to the theft of sensitive data from corporate networks or servers [5] [10] [9] .

- **Security Orchestration, Automation and response (SOAR)** – SIEMs integrate with enterprise systems and automate incident response before the attacker acts devastatingly [5] [9].

SIEM tools exist in many forms and differ in cost and performance. The best SIEM system for an organization may be inappropriate for another so, beyond the operational side each organization should evaluate a SIEM solution based on its own criteria and specific requirements such as functional and technical criteria. Functional criteria that determine the SIEM processes and if it does what it's supposed to do and technical criteria that includes for example the documentation about the working architecture of SIEM or help to solve the technical problems can influence the customer's decision [11]. The security team must identify the products that satisfy the organization's requirements to match the internal project and support capabilities [3].

3. Why a SIEM system

The increasing complexity of information systems combined with compliance with security regulations is a big issue even for large companies in terms of managing information safely, while for small and medium-sized businesses it is even more difficult.

SIEM systems are evolving considerably and are starting to offer solutions even for small and medium-sized businesses. SIEM solutions are designed to simplify security features by integrating the functions of individual security products into a single platform.

The scope and the focus of a SIEM system:

Audit and compliance

SIEM undertakes the collection, preservation and review of records and checks whether all rules and policies are being respected to ensure data integrity and that users have proper access to sensitive information [12].

Security

SIEM may focus on monitoring external threats and security applications by making extensive analysis of the log. Firewall, web server and Intrusion Detection Systems (IDS) log files are monitored for critical events and suspected device behavior[12].

Operations

The focus is on resource management, hardware limits and possible errors and warnings [12].

The purpose of this work is to present free and open source SIEM systems. The study was based on system's documentation or previous studies and measurements. Finally, is presented a brief review of the key features of these systems. The created corresponding table completed only for systems that was able to detect information.

The rest of this review is organized as follows. In section two describe enhancements and the latest technologies of SIEM systems and also the correlation with Critical Infrastructures (CIs). In section thee analyzed the criteria and metrics that can be used for system's evaluation. In section four introduced some of the most well-known Open Source SIEM Systems. In section five closes the review with final remarks and conclusions.

C. Background

1. Enhancements - new technologies - evolution

The rapid evolution of **big data** technologies and the large number of data sources related to cybersecurity or information security, led to the development of many SIEM systems [4].

The Attack Modeling and Security Evaluation Component (AMSEC) was proposed and studied in 2012 and could be used as a subsystem in any SIEM system. The basic elements of the AMSEC architecture (Figure 2) are the use of comprehensive security repository, effective attack graph tree taking into account known and new forms of

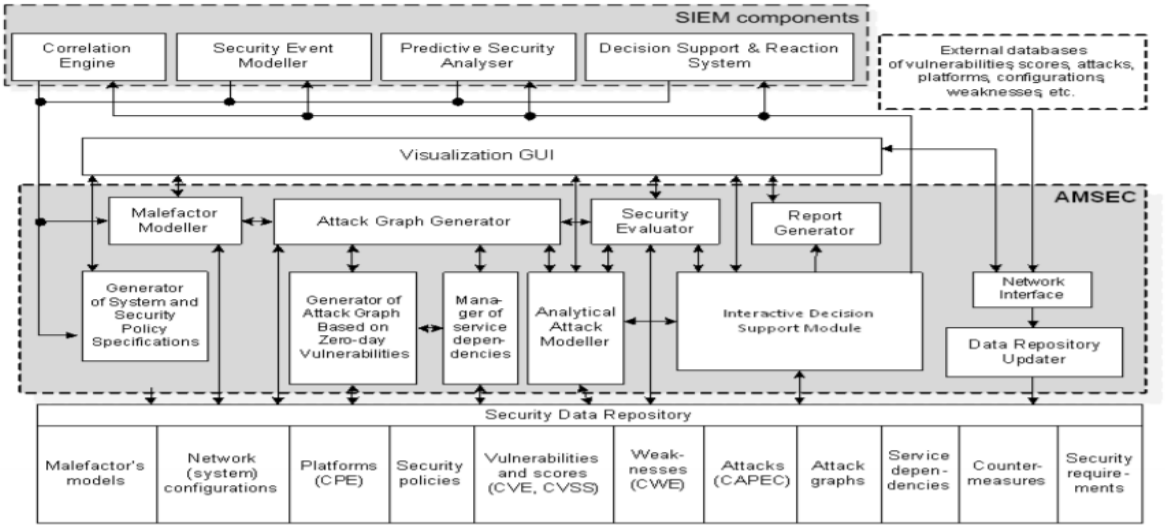

Figure 2. Generalized architecture of AMSEC [13]

threats, stochastic analytical modeling and security metric calculation [13].

As a very important element is the model used to design and implement the data repository, researchers have been studying the construction of a hybrid (ontological and relational model) data repository for the next-generation SIEM systems. According to this survey, for the representation and modeling of data, suggested the ontological approach that provides the necessary flexibility and a hybrid (ontological and relational) repository integrated with the Attack Modeling and Security Evaluation Component [14].

In 2013 was studied a **self-adaptive** SIEM system which optimizes the correlation process using AI techniques. The system was based on the idea that related data can be used for detection of a particular type of attack. The system used the pre-existing

knowledge, having identified the context of a given attack using genetic programming, generates correlation rules for the different types of multi-level attacks. Then using **artificial neural networks** trained with the prior knowledge acquired by the system itself, it classifies data obtained from the sensors according to the corresponding frame determined for each attack. Experimental measurements have demonstrated the success of the system to successfully detect different attack scenarios, such as DDoS attacks [15].

An Intrusion Prevention System (IPS) was proposed in 2015 with the extension of a commercial SIEM framework, OSSIM, to predict potential multi-stage attacks. In the proposed system, the SVM (Support Vector Machine) machine learning algorithm was used and was focused only on the distributed DDoS, Smurf and IPSweep attacks [16].

The rapid evolution of **Cloud Computing** allows the provision of cloud-based security services such as Security-as-a-Service (SECaaS). In a survey in 2017, a SIEM architecture was developed that could be deployed on the SECaaS platform to analyze and identify a smart cyber threat. The architecture was based on virtualization of security sensors, such as virtual firewalls, virtual IPS, virtual DLP (Data Loss Prevention) which means solutions to prevent data loss, virtual DPI etc and SDN (software-defined networking) / NFV (network functions virtualization) technologies that used to design, build and operate networks. The proposed SIEM architecture includes SIEM Engine for the processing of collected data, SIEM Storage for storing the collected data and the results of the analysis, and SIEM user layer to ensure the user's security service. As the correlation analytics are the most important of the various methods, a neural network was implemented for the threat identification. The neural network is trained by the collected security data, thus improving the ability of the proposed system. This method will improve SIEM ‘s intelligent threat analysis [17].

Embedding **machine learning** into SIEM systems helps organizations to analyse data before occurs the cyber attack and facilitating forensic and alert [3].

As the **Internet of Things** (IoT) is greatly increasing, attacks on these devices are increasing too, both in number and impact. It is necessary to develop a real-time data processing framework to be used to detect attacks at the time they occur, and to configure an architecture that will implement this framework. This architecture may be used by IoT vendors, third-party organizations (nonprofit organizations may seek to improve Internet security) and end users (organizations with a large number of IoTs). Some of the objectives of such a framework are to be able to quickly detect threats based on the analysis of IoT’s abnormalities and to associate the IoT’s information with other data to enhance the detection process [18].

Due to the complex operations of SIEM systems, expert knowledge is often required which may be incomplete or incorrect or costly. GraphBAD was proposed in 2018 as the solution that requires minimal end-user experience. GraphBAD parses security configurations and logs to identify anomalies that may lead to security issues and then suggest corrective steps to remove the anomaly [19].

In **smart grids** where everything is interconnected and interdependent, the new, extremely complex, cyber attacks such as Botnets, zero day, or Advanced Persistent Threats (APTs) are also emerging. It is therefore imperative to develop a platform that can respond to the new reality. SIEMs that are specialized in processing large amounts

of data, collecting and normalizing data from different sources, correlating events and alerting when a threat is detected, may be the platform that will satisfy the smart grid's requirements [20].

2. Critical Infrastructures (CIs)

Critical Infrastructures (CI) are often a combination of many organizations and domains. The rapid development of recent years creates the requirement of managing huge amounts of data coming from different sources (IT devices and applications), different levels of systems and different domains which are typically characterized by heterogeneous syntax and semantics and performs even more complex functions. Considering that SIEM reports can even be used as forensic evidence, they must guarantee data authenticity, fault and intrusion tolerance, and privacy. In many cases, SIEMs have to coexist with SCADA systems (Supervisory Control and Data Acquisition) in the same environment.

SCADA are industrial systems configured to control vital installations for the industrialized countries. The systems were not connected to the Internet, so it was difficult to be attacked by hackers. However, with the rapid increase in networking and the development of GUI command and control environment, it is now easy for hackers to penetrate the SCADA world by affecting expensive equipment and causing huge damage to the industry [21].

MASSIF *MAnagement of Security information and events in Service InFrastructures)*, is a collaborative research project co-funded under the European Commission's FP7 ICT Work Programme 2009 (FP7-ICT-2009-5). It is aligned with the objective ICT-5-1.4 - Trustworthy ICT. The MASSIF solution combines novel security technologies to provide the industry's most advanced security management solution and has successfully developed a next-generation SIEM framework for service level infrastructure [22].

In the context of the **MASSIF** project proposed enhancements in SIEM Technology**,** related to data processing (from collection to storage phase), for Critical Infrastructures protection. The basic sections of this solution are [23]:

- **GET**: collects heterogeneous data from multilevel sources, such as the older IT and SCADA components, security applications and devices (security sensors) by providing intelligence at the edge of the SIEM.
- **RS**: reliably stores the events containing useful information about the relevant security breaches.

D. Evaluation criteria - research

As the number of log files that are collected and managed by a SIEM system increases, so do the storage capacity and the computational requirements in the log repository where these files are stored. As storage costs money, it is important to count the amount of log data, to know the approximate storage that required an organization in order to operate the SIEM efficiently. The most common measurement to determining how much log data will be generated **Events Per Second (EPS)** that can

also be a performance rating for a SIEM. Considering the size (in bytes) of each event and using the formula (Average EPS x Average Bytes per Event / 1,000,000,000), we determine the average size of the logs produced per second in gigabytes. By multiplying the result by 86,400 we have the required daily capacity. Finally, by multiplying the previous result by the number of days you want to keep the log, you have approximately the total capacity requirements of an organization's repository. Because of some events can create an EPS spike that could last from minutes to days, we may have an increase in the required capacity, so it must be ensured that the storage repository is large enough to handle such unpredictable spikes [24].

In 2018 made a review over three of the most popular and discussed SIEM solutions: Prelude's Universal Open-Source SIEM project, AlienVault's Open Source Security Information and Event Management (OSSIM), and Elasticsearch Logstash and Kibana (ELK) by Elastic [25].

Two approaches have been developed to evaluate selected SIEM systems. The multi-step methodology for statistical evaluation of a software and the approach that defines a software selection model by introducing the **Solution Merit Index (SMI)** as the sum of the percentages of the hierarchically-ranked attributes. In a comparative study of AlienVault OSSIM, Cyberoam iView and CS Prelude, the second model was used, according to which software selection factors were divided into two groups: primary and secondary drivers. The first group includes the user-relevant features elated to the whole life cycle of a SIEM, while the second all non-essential features that are important during the development stage [20].

The Solution Merit Index for each SIEM being evaluated was calculated using the following formula:

$$\mathbf{SMI=w_pf_p+w_sf_s}$$

The symbols f_p and f_s denote scores based on a percentage base due to the primary and secondary criteria groups respectively, and w_p and w_s are their weights [20].

The results of the evaluation showed that OSSIM responded better to the defined criteria with Prelude second, while iView had much worse behavior. In the criteria of the second group, the top score was Prelude, while OSSIM and iView were similarly evaluated. The final evaluation through the MERIT index indicated that the OSSIM and Prelude systems best met the selected criteria. OSSIM is a complete SIEM system, ready to be implemented on a platform, on the other hand, the advantage of Prelude is its modular construction, which enables using various components [10].

A disadvantage of commercial products is its low flexibility, ie adaptability to the requirements of each customer, who has to apply and restrict its goals to the purchased products [26].

Splunk is a commercial software platform for large data analysis, used in banks, hospitals, communication, security, education. In 2017, ELK stack and Splunk compared to their performance. After searching for logs from a particular IP, the result was stored in a text file and in CLI (Command Line Interface) mode, the exact **execution time** was measured [26].

Splunk first saves the full text data on the hard disk and makes the mapping process later while the ELK does exactly the reverse [26].

The results showed that Splunk has a higher execution time than the ELK stack, which may be due to heavy functions built into Splunk to provide rich and generic capabilities. The ELK stack has similar or better performing when searches for specific security logs that match specific conditions. In addition, ELK provides various types of imaging tools that are useful to security administrators [26].

There are many comparable features which may be used during the evaluation of SIEM systems, such as the number of platforms supported, scalability, latency, number of built-in metrics, number of built-in dashboards, number of integration ways with third-party tools. Six SIEm systems AlienVault, Micro Focus ArcSight, Manage Engine Event Log Analyzer, Splunk, Rapid7 InsightIDR, and Solar Winds Log and Event Manager were selected for evaluation in capabilities related to the generation of custom visualizations. This paper examines predefined metrics for a selected use-case, data import options for suitable data for the use-case, the existence of built-in data structures which may be used to map the imported log files for the selected use-case, the ability to load custom data files, the ability to form custom searches, the abilities of data joining and data blending for visualizations combining multiple data sources, built-in visualization capabilities to display the selected data results [27].

The attack graphs from a SIEM system were used as input to a computer network security evaluation approach. A key feature of this approach is the implementation of the developed security metrics system and is geared to real-time safety assessment. Therefore, the technique allows you to track the attacker's current location and predict its route in the network. With a set of security metrics, their changes were identified after the occurrence of security incidents. The technique provides different algorithms according to the available input data and allows the adequate security assessment at any stage of system's operation. The case study turned out that new data from the SIEM system influences the odds and the risk values of the attack [28].

According to a proposed methodology [29] not only a technological but also a pragmatic approach is used to evaluate a tool SIEM. Firstly, a company must determine the apropriate technical and organizational requirements that a SIEM must have to meet its needs and then could follow the evaluation methodology which is divided into two phases:

- Quantify each SIEM solution requirement using a quantitative-based method that gives emphasis on numbers, measurement, experimental design, and statistical analysis.
- Measuring the applicability of the solution using a qualitative method after determining a list of indicators for evaluation. Qualitative methods evaluate parameters like the success and the eligibility of a product using nonnumerical data, such as internal discussions, interviews, comparisons to provide feedbacks, etc.

Requirements that measured are divided into 5 sections:

- Platform: Describes the technical requirements needed in the platform (e.g Log Management System capability, supporting an extended set of log sources, Method for retrieving events/flows/logs, platform computing and storage capacity, installation model, availability of both default and customizable correlation rules, ability to quickly prioritize response and

analysis, customizable and compliance reports, alerting capabilities, technical documentation and online help, ability of monitoring the platform).

- Operations: Groups the requirements needed to manage the solution (e.g Role-based access control, Accounting: log events done by operators).
- Integration: Requirements needed to integrate the SIEM solution into the Company's information system (e.g. Active Directory integration for administrative management, integration with asset management tools, integration with vulnerability management tools).
- Advanced features: Features, they could be considered as nice-to-have requirements (e.g. Threat Intelligence analysis tools support, analytics support, automatic response capabilities).
- Licensing and support: e.g. Specification of the preferred License type, delayed license activation, technical assistance support and professional services.

The goal is to select the appropriate SIEM that fits in with the environment and resources of a company.

E. Open Source Systems analysis/presentation

Open source SIEM tools open their cybersecurity design to the public, allowing IT professionals to modify and share the code of tools much more freely. Usually, businesses can obtain these Open Source InfoSec tools for free. While free SIEM tools can not provide the completeness of enterprise-level solutions, open source SIEM offers solid functionality at an affordable price [30].

In this section we provide a description of the main open source SIEM systems available. The system components, the operation and the peculiarities are presented.

a. OSSEC

The OSSEC is fully open and is provided for free use. It has extensive configuration options so that new features can be added by the user who wants to adapt it to his security needs. The user can modify the source code by writing scripts that response to his own warning rules. OSSEC is supported by a large community of developers, IT admins and users. OSSEC actively monitors the system's activity with file integrity monitoring, log monitoring, rootcheck, and process monitoring. It runs on most operating systems such as Linux, OpenBSD, FreeBSD, Mac OS X, Solaris and Windows. Customers can define for which incidents they want to be alerted and increase the priority of critical events. Thanks to the active response options, an attack can be immediately blocked. The OSSEC collects, analyzes and correlates the log files created by the operating system, the applications and the devices on the network and then informs the user if something suspicious happens (attack, misuse, bugs, etc.). It has a central manager for monitoring and receiving information from agents, syslog, databases, and from agentless devices [31].

b. OSSIM (Open Source Security Information Management)

OSSIM is the free, open source version that has been developed for more than a decade by AlientVault with also commercial releases. It performs normalization and correlation of events from different sources, such as firewalls, IDS / IPS, routers, anomaly detectors and IT assets to create security alerts [21].

Although OSSIM does not include some of the Scalability, Performance, Managed Security Service Provider (MSSP) features and forensic recording, provides all the security tools from AlienVault Professional SIEM, a technology that exists at more than the half SIEM installations worldwide [32].

OSSIM consists of 4 major **components** [32] [33] [34] :

- **Server**: The most basic that controls all other components. Its role is to store, process and correlate events that are received from the sensor through TCP port 40001, manage the entire OSSIM core functionality and communicate with any other element.
- **Sensor**: is responsible for collecting logs from network devices and then make log analysis and normalization to send that to Server. May have privileged access certificates to recover logs and information from systems.
- **Framework**: is the user interface and management console, needs the least hardware resources and is usually installed together with the Server component.
- **Database**: is a MySQL server instance that contains OSSIM's runtime configuration and assembled events.

OSSIM is a good solution for small businesses. Generally can not respond to all types of computer threats [33].

It has relatively easy configuration. The user can integrate OSSIM into multiple network / security equipment and in most cases has to manually solve the parsing error [11].

OSSIM maintains a large record of events in its database that can help to detect vulnerabilities over a long term. As it increases the interest in protecting SCADA / ICS systems, OSSIM can be considered as a Unified Security Management (USM) solution for critical infrastructures [21].

The security officer can set policies and instructions in OSSIM to improve the processing of events in accordance with the organization's requirements. Instructions are the OSSIM mechanism for correlating events taken from sensors [21].

Further extending the OSSIM may take into account the following [26]:

- applying ELK stack for log retention,
- usage of machine learning for intrusion recognition,
- using evolutionary computation methods in automatic rule generation,
- feeding the system with information from honeypot systems,
- horizontal scaling of the core to distribute correlation-related processing .

c. The ELK Stack

The ELK stack is released as open source under Apache License 2.0 [35].

It consists of three components, the ElasticSearch, Logstash and Kibana (Figure 3) and can be run in a virtual environment.

- **ElasticSearch** is a search engine based on Apache Lucene [26].
- **Logstash** collects the necessary logs, filters and sends them to ElasticSearch after converting them to JSON format [26] .
- **Kibana** is a screen for visualization the results in various forms such as graphs, tables and maps [26].

A comparative study shows that ELK Stack is as good as the commercial product Splunk. Generally, this tool is recommended for use in small and medium-sized organisations [26].

Generally ELK is an effective, easy-to-use log management tool. An important observation noticed that if only the three elastic components (ElasicSearch, Logstash and Kibana) are used, we get an efficient centralization tool without alerting management and other important SIEM procedures as they are offered instead in the Elastic X-PACK (Elastic's proprietary softwares) [11].

It is the basic component of most of the available open source SIEM solutions. Taking care of the collection, parsing, storage, and analysis, ELK is part of the architecture for OSSEC Wazuh, SIEMonster, and Apache Metron [36].

ELK has the ability to integrate with many other products and can also receive many input sources. It can generate and export reports as a PDF file. [25].

Architecture

Figure 3. Architecture for a small-sized development environment [36]

d. Apache Metron

Apache Metron was created by the Hadoop community evolving from Cisco OpenSOC. It is a framework that enables businesses to detect cyber anomalies and respond quickly [37].

Metron's framework provides 4 key capabilities [37],[38]:

- **Safety Data Lake / Vault** : a platform that provides a cost-effective way of storing and combining rich telemetry data for long periods of time.
- **Pluggable Framework** : The platform not only provides a large number of parameters for common security data sources, but also enables the addition of new custom parameters. New enrichment services and custom functionality also characterize it.

- **Security application :** It has the typical capabilities of a SIEM system, but also has utilities commonly used by SOC analysts.
- **Threat Intelligence Platform** : Metron will provide advanced defense techniques and implement machine learning algorithms in real time as events are streaming in.

Logical Architecture

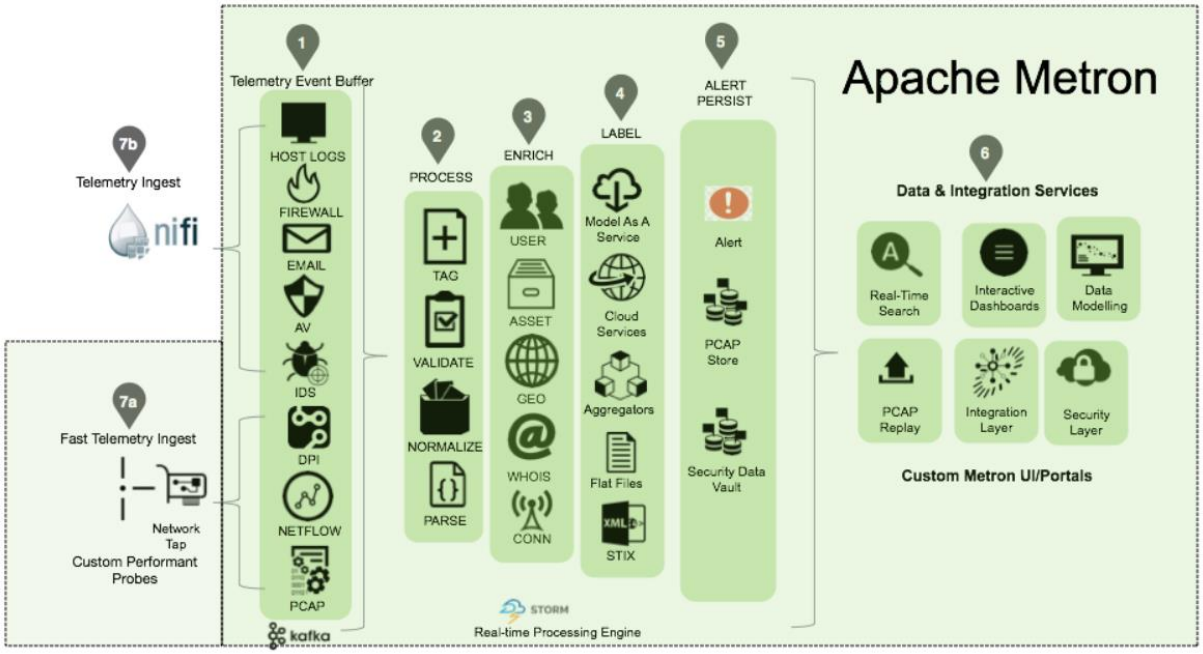

Figure 4. The logical architecture of the Metron Platform [39]

Figure 4 shows the logical architecture of the Metron Platform which consists of the following components [39]:

1. Telemetry Event Buffer

All raw events from each telemetry security data source captured by Apache Nifi (supports powerful and scalable directed graphs of data routing, transformation, and system mediation logic)[40] or custom Metron probe will be pushed into its own Kafka topic, which is a publish-subscribe messaging software where topics-categories can be defined to where applications can add, process and reprocess data-messages [41]. Metron processing begins when a telemetry event arrivals into the ingest buffer.

2. Process (Parse, Normalize, Validate and Tag)

Each raw event will be parsed and normalized into a standardized flat JSON structure, so the topology correlation engine further downstream can correlate messages from different topologies.

3. Enrich

Enrich different data elements of the normalized event from step2. For example with GEO enrichment an external IP address is enriched with GeoIP information (lat/long coordinates + City/State/Country).

4. Label

After enrichment, the telemetry event goes through the labeling process.

5. Alert and Persist

During this phase, certain telemetry events can initiate alerts and then indexed in an alert index store. The alerts are triggered either because the raw telemetry event is an alert itself or it has a threat intel hit that will be marked as an alert. These alerts are stored and maintained for a long time to enable next generation analytics to be performed

6. UI Portal and Data & Integration Services

Steps 1 through 6 provide the mechanism to ingest, parse, normalize, enrich, label, index and store all security telemetry data across a diverse set of data sources in the enterprise into a single security data vault. Metron platform provide a set of services for different types of security users to perform their jobs more effectively.

7a. Fast Telemetry Ingest

For high volume network telemetry data like packet capture (PCAP), Netflow/YAF, and Bro/DPI, custom Metron probes will be available to ingest data directly from a network tap.

7b. Telemetry Ingest

For most security telemetry data sources that use transports and protocols like file, syslog, REST, HTTP, custom API, etc.

In analyst/investigator workflow Apache Metron includes three steps not found in traditional security tools, Looking through Alerts, Collecting Contextual data and Investigate [42].

e. SIEMonster

SIEMonster is based on open source technology and is available for free and as a paid solution. It was created by a team of professional hackers with experience of more than 20 years. Using this experience, SIEMonster has built advanced SIEM security tools for organizations that want to stop real-time attacks. Using machine learning, human-based analytics watch SIEMonster Deep Learning kill the attacks automatically [43].

SIEMonster uses Kubernetes for all builds, which is a portable, extensible, open-source platform for managing containerized workloads and services, that facilitates both declarative configuration and automation. It provides a framework to run distributed systems resiliently. [44]. Each application within the framework is represented by a Pod [45].

The open source Open Distro for Elasticsearch Security, provides Elasticsearch cluster protection & Role based access control, including fine grained role-based access control to indices, documents and fields as well as compliance logging for GDPR, HIPAA, PCI, SOX and ISO compliance [45].

It has two instances of Logstash in use [45],

- Logstash-Collector : for data ingestion in JSON format to the Apache Kafka Message Broker.
- Logstash-Indexer : uses Apache Kafka as an input, parses and correlates data before forwarding to Elasticsearch.

All event data including syslog is initially received and processed by the Logstash-Collector before being sent on to the Apache Kafka message broker queue. The data is then input to the Logstash-Indexer where it is normalized and formatted for suitability for indexing into Elasticsearch [45].

SIEMonster uses agent for collecting logs from Unix hosts, typically Apache web logs. Network devices with remote syslog settings should be set to the SIEMonster appliance IP address. Syslogs are accepted on the ports 1516 TCP, 1516 UDP, 1517 TCP, 1518 TCP, any unused TCP ports between 3520-3529 [45].

The primary alerting system, configured with many index interfaces to provide a feature rich method of event log analysis and outputs for email, Slack, Pager Duty, and webhooks to send to custom applications [45].

Product features

Enables access to all logs through a gateway, provides the ability to see all live alerts and vulnerability scan data, enables all events and notifications from SIEM and other commercial products such as McAfee AV and Cisco to be transferred to the same product.

Feeding the “Monster” with what's going on outside the world with bad IP addresses, bad e-mail addresses and other information, can stop the attacks before they affect the organization [43].

f. Prelude (Open Source Siem - OSS)

Prelude OSS is the open source edition of Prelude SIEM which collects, normalizes, sorts, aggregates, correlates and reports all security-related events independently of the product brand or license giving rise to such events. It is **"agentless"** [46].

It can be installed on many different Linux distributions. It provides a centralized database to store logs and allows users to create and send reports in pdf format via email [25].

Components

Prelude Manager: is a high availability server capable of handling large number of connections, and processing large amounts of alerts

Libprelude : a library that guarantees secure connections between all sensors and the Prelude Manager.

LibpreludeDB : allows developers to use the Prelude IDMEF database

Prelude-Correlator : allows conducting multistream correlations for writing correlation rules

Prewikka : the official web Graphical User Interface (GUI) [47].

Architecture (Figure 5)

Sensors detect intrusions and report alerts using a TLS connection to a "prelude-manager" server. The server can then process these alerts and deliver them to a user-defined medium. Notifications are displayed on the Prelude console [48].

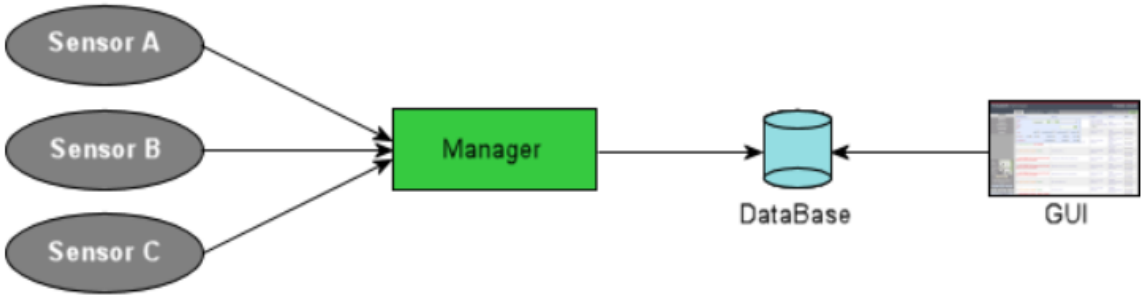

Figure 5. Prelude Architecture [48].

g. LOGalyze

LOGalyze is an open source, centralized log management and network monitoring software. It supports Linux / Unix servers, network devices, Windows hosts and provides real-time event detection and correlation and extensive search capabilities. Can extract reports into CSV, XLS, PDF or HTML formats. It is compatible with syslog, rsyslog, syslog-ng, lasso and Snare SOAP API service [49].

The following table summarizes the key features of SIEM systems.

Table 1. SIEM's Features

	OSSIM	**ELK**	**Prelude**	**SIEMonster**	**Apache Metron**
Correlation and Alerts	it supports three types of correlations: logical correlation : where events are associated with others within a window of time after an activation event. cross-correlation : in which an event is associated with a vulnerability identified by a vulnerabilities database, which periodically updated by the Internet inventory correlation : is used to relate events with known vulnerabilities from the assets of the organization	does not come with built-in correlation rules, and so it is up to the analyst to use Kibana queries, based on the parsing and processing performed using Logstash, to correlate between events [36].	supports rule-based correlation in real-time. It comes with rules for correlation out of the box which can be customized. support fully customizable statistical correlation in real time. Alerts are sent to the concentrator [50].	uses MineMeld [51] UI for threat intelligence, Alerts for creating and managing event-based notifications [52]	Real-time alerts, anomaly detection, data correlation - Rules and reports, predictive modeling via an interface for centralized view of data and alerts. Capability to tag a message as an alert via a custom rule language.

	[21].				
EPS	200	100			
Standard			Intrusion Detection Message Exchange Format (IDMEF)[53].		
Compatibility			provides a C, C++, Python, Ruby, Lua and Perl framework so that you can convert existing security applications to use the Prelude system (Native compatibility) is capable of analyzing any type of log (system logs, syslog, flat files, etc.) (Log Compatibility) [54]		
Installation Requirements	2 CPU cores 4-8GB RAM 250GB HDD E1000 compatible network cards [55]		GnuTLS,Python, PCRE,MySQL, PostgreSQL, SQLite [56].		
Last Version	5.5.1 (26 February 2018) [57]		Prelude OSS 5.0.0 (12/23/2018) [58].	Version 4 (10 June 2019)	0.4.3 -2018
Documentation	complete analytical understandable	complete analytical understandable	Well-detailed documentation is available on the Prelude	complete analytical understandable	complete analytical understandable

			website https://www.prelude-siem.org/projects/prelude/wiki/ManualUser		
Operating Systems supported	Linux, Windows	Linux, Windows and Mac [59]	Linux, OpenBSD, FreeBSD, NetBSD, Sun/Solaris, MacOSX, Tru64, and more generally most UNIXes systems [50]	Mac, Ubuntu, CentOS, and Debian	Mac and Windows
License	GNU General Public License (GPL) [20]	Apache License 2.0	GNU General Public License (GPL) [20]		Apache License 2.0

From the above features, correlation and alerts, documentation, supported Operating Systems, EPS and Last Version are among the most important for the evaluation of free open source SIEM systems. Event correlation, the connection of signals coming from different data sources into a pattern that could be indicative of a breach in security, is an important feature. The correlation rules connected with alerts that created when a possible attack pattern is identified. An organization that decide to adopt an open source solution based on the documentation help and detailed information about the system to solve problems. The EPS metric is useful to count storage requirement or to evaluate a system. Finally the date of Last Version in combination with the quality of the documentation is a criteria that shows if the system is update with the last security trend and requirements.

These are features about which we found information on internet. The features used for the evaluation do not require experimental measurements. The objectives of the present work are limited to the study of SIEM systems based on features known by the bibliography and the experimental measurements carried out by researchers. Carefully and analytical study of one of the tools is the target for next research. According to table1, OSSIM provides the most completed correlation process. About EPS metric we haven't got measurements for all the systems. All the systems are recently updated which means that they can cover the needs of a modern organization. They also have complete analytical and understandable documentation. Finally they supported the most widely used operating systems.

F. Conclusion

SIEM Systems have become an essential part of a company's defense system to detect an attack and to response immediate. In this work, a comparative study of the most widely used open-source SIEM tools has been performed. As commercial SIEM solutions tend to be complex and expensive the open source tools meet the needs of many organizations and businesses, especially small ones. The Ossim system seems to be the most researched. Its features make it the most complete of the open source that has been studied and equal to many commercial ones. According to our comparison (table1) this system presents many correlation capabilities and its documentation is complete analytical and understandable which means that the user could find help easy. It has a big EPS (= 200) and the last update made in 2018 which means that it has not been abandoned by its creators and possible new updates will improve the system . An in-depth and experimental study of OSSIM system is the next goal of our research.

G. References

[1] J. Inns, "The evolution and application of SIEM systems," *Netw. Secur.*, vol. 2014, no. 5, pp. 16–17, May 2014.

[2] M. Di Mauro and C. Di Sarno, "Improving SIEM capabilities through an enhanced probe for encrypted Skype traffic detection," *J. Inf. Secur. Appl.*, vol. 38, pp. 85–95, Feb. 2018.

[3] S. S. Sekharan and K. Kandasamy, "Profiling SIEM tools and correlation engines for security analytics," in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017, pp. 717–721.

[4] J. Kaskade, "Magic Quadrant for Security Information and Event Management," p. 32.

[5] "Exabeam - Smarter SIEM, UEBA and SOAR," *Exabeam*. [Online]. Available: https://www.exabeam.com/. [Accessed: 29-Jun-2019].

[6] "A Comparative Study of Correlation Engines for Security Event Management | Planet JBoss Developer." [Online]. Available: https://planet.jboss.org/post/a_comparative_study_of_correlation_engines_for_security_event_management. [Accessed: 07-Oct-2019].

[7] "How SIEM Correlation Rules Work." [Online]. Available: https://www.alienvault.com/blogs/security-essentials/how-siem-correlation-rules-work. [Accessed: 08-Oct-2019].

[8] "Using the ELK Stack for SIEM," *Logz.io*, 20-Jun-2018. [Online]. Available: https://logz.io/blog/elk-siem/. [Accessed: 13-Oct-2019].

[9] "Traditional SIEM versus NextGen SIEM | Threat Management." [Online]. Available: https://threatmanagement.info/traditional_siem_versus_nextgen_siem/. [Accessed: 12-Oct-2019].

[10] "Securonix Next-Generation SIEM: A Modern SIEM," *Securonix*. [Online]. Available: https://www.securonix.com/products/securonix-next-generation-siem/. [Accessed: 10-Oct-2019].

[11] M. Nabil, S. Soukainat, A. Lakbabi, and O. Ghizlane, "SIEM selection criteria for an efficient contextual security," in *2017 International Symposium on Networks, Computers and Communications (ISNCC)*, 2017, pp. 1–6.

[12] J. C. Gomez, "Radboud University Nijmegen Security Information and Event Management Master Thesis."

[13] I. Kotenko and A. Chechulin, "Attack Modeling and Security Evaluation in SIEM Systems," vol. 8, p. 20, 2012.

[14] I. Kotenko, O. Polubelova, A. Chechulin, and I. Saenko, "Design and Implementation of a Hybrid Ontological-Relational Data Repository for SIEM Systems," *Future Internet*, vol. 5, no. 3, pp. 355–375, Sep. 2013.

[15] G. Suarez-Tangil, E. Palomar, A. Ribagorda, and I. Sanz, "Providing SIEM systems with self-adaptation," *Inf. Fusion*, vol. 21, pp. 145–158, Jan. 2015.

[16] E. T. Anumol, "Use of Machine Learning Algorithms with SIEM for Attack Prediction," in *Intelligent Computing, Communication and Devices*, 2015, pp. 231–235.

[17] J. Lee, Y. S. Kim, J. H. Kim, and I. K. Kim, "Toward the SIEM architecture for cloud-based security services," in *2017 IEEE Conference on Communications and Network Security (CNS)*, 2017, pp. 398–399.

[18] R. I. Bonilla and C. L. Abad, "Towards a Real Time Framework for Monitoring IoT Devices for Attack Detection: Vision Paper," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2017, pp. 699–703.

[19] S. Parkinson, M. Vallati, A. Crampton, and S. Sohrabi, "GraphBAD: A general technique for anomaly detection in security information and event management," *Concurr. Comput. Pract. Exp.*, vol. 30, no. 16, p. e4433, 2018.

[20] R. Leszczyna and M. R. Wróbel, "Evaluation of open source SIEM for situation awareness platform in the smart grid environment," in *2015 IEEE World Conference on Factory Communication Systems (WFCS)*, 2015, pp. 1–4.

[21] A. Mahboob and J. A. Zubairi, "Securing SCADA systems with open source software," in *2013 High Capacity Optical Networks and Emerging/Enabling Technologies*, 2013, pp. 193–198.

[22] Newsroom, "MASSIF creates next-generation framework for Security Information and Event Management (SIEM)," *Digital Single Market - European Commission*, 25-Nov-2013. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/massif-creates-next-generation-framework-security-information-and-event-management-siem. [Accessed: 30-Jun-2019].

[23] L. Coppolino, S. D'Antonio, V. Formicola, and L. Romano, "Enhancing SIEM Technology to Protect Critical Infrastructures," in *Critical Information Infrastructures Security*, vol. 7722, B. M. Hämmerli, N. Kalstad Svendsen, and J. Lopez, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 10–21.

[24] B. Hale, "Estimating Log Generation for Security Information Event and Log Management," p. 6.

[25] F. E. Christopher and K. J. Myers, "Siem-Enabled Cyber Event Correlation (What

And How)," p. 133.

[26] S. J. Son and Y. Kwon, "Performance of ELK stack and commercial system in security log analysis," in *2017 IEEE 13th Malaysia International Conference on Communications (MICC)*, 2017, pp. 187–190.

[27] F. Ö. Sönmez and B. Günel, "Evaluation of Security Information and Event Management Systems for Custom Security Visualization Generation," in *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*, 2018, pp. 38–44.

[28] I. V. Kotenko and E. Doynikova, "Evaluation of Computer Network Security based on Attack Graphs and Security Event Processing," *JoWUA*, vol. 5, pp. 14–29, 2014.

[29] H. Mokalled, R. Catelli, V. Casola, D. Debertol, E. Meda, and R. Zunino, "The Applicability of a SIEM Solution: Requirements and Evaluation," in *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 2019, pp. 132–137.

[30] B. Canner, "The 10 Best Open Source SIEM Tools for Businesses," *Top SIEM Vendors, News & Reviews for Security Information and Event Management*, 06-May-2019. [Online]. Available: https://solutionsreview.com/security-information-event-management/the-10-best-open-source-siem-tools-for-businesses/. [Accessed: 01-Jul-2019].

[31] "Home — OSSEC." [Online]. Available: http://www.ossec.net/. [Accessed: 30-Apr-2019].

[32] P. Shivhare and P. Savaridassan, "Addressing Security Issues of Small and Medium Enterprises Through Enhanced SIEM Technology," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2592463, Apr. 2015.

[33] D. Hermanowski, "Open Source Security Information Management system supporting IT security audit," in *2015 IEEE 2nd International Conference on Cybernetics (CYBCONF)*, 2015, pp. 336–341.

[34] L. Coppolino, S. D'Antonio, V. Formicola, and L. Romano, "A framework for mastering heterogeneity in multi-layer security information and event correlation," *J. Syst. Archit.*, vol. 62, pp. 78–88, Jan. 2016.

[35] "production." [Online]. Available: http://repo.bg.pw.edu.pl/index.php/en/r#/info/bachelor/WUT61430bfcf45a4cff91a01ccceece67bc/. [Accessed: 09-May-2019].

[36] "Using the ELK Stack for SIEM," *Logz.io*, 20-Jun-2018. .

[37] "Apache Metron Big Data Security." [Online]. Available: http://metron.apache.org/. [Accessed: 02-May-2019].

[38] © 2019 Cloudera, I. A. rights reserved Terms, C. | P. Policy, D. P. A. Hadoop, associated open source project names are trademarks of the A. S. F. F. a complete list of trademarks, and C. Here, "Apache Metron," *Cloudera*. [Online]. Available: https://www.cloudera.com/content/www/en-us/products/open-source/apache-hadoop/apache-metron.html. [Accessed: 13-Oct-2019].

[39] "Metron Architecture - Metron - Apache Software Foundation." [Online]. Available: https://cwiki.apache.org/confluence/display/METRON/Metron+Architecture#Metr

onArchitecture-MetronComponents. [Accessed: 03-Jun-2019].
[40] "Apache NiFi." [Online]. Available: https://nifi.apache.org/. [Accessed: 13-Oct-2019].
[41] "Part 1: Apache Kafka for beginners - What is Apache Kafka? - CloudKarafka, Apache Kafka Message streaming as a Service." [Online]. Available: https://www.cloudkarafka.com/blog/2016-11-30-part1-kafka-for-beginners-what-is-apache-kafka.html. [Accessed: 15-Oct-2019].
[42] "Metron Benefits - Metron - Apache Software Foundation." [Online]. Available: https://cwiki.apache.org/confluence/display/METRON/Metron+Benefits#MetronBenefits-MetronBenefits.SOCAnalyst&InvestigatorPerspective. [Accessed: 04-Jun-2019].
[43] "SIEMonster | Affordable Security Monitoring Software Solution," *SIEMonster*. [Online]. Available: https://siemonster.com/. [Accessed: 02-May-2019].
[44] "What is Kubernetes." [Online]. Available: https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/. [Accessed: 09-Oct-2019].
[45] "siemonster-v4-starter-edition-operations-guide-v10.pdf." .
[46] "Επισκόπηση - PRELUDE SIEM." [Online]. Available: https://www.prelude-siem.org/. [Accessed: 02-May-2019].
[47] "PreludeComponents - PRELUDE SIEM." [Online]. Available: https://www.prelude-siem.org/projects/prelude/wiki/PreludeComponents. [Accessed: 25-May-2019].
[48] "PreludeArchitecture - PRELUDE SIEM." [Online]. Available: https://www.prelude-siem.org/projects/prelude/wiki/PreludeArchitecture. [Accessed: 25-May-2019].
[49] "LOGalyze - Open Source Log Management Tool, SIEM, Log Analyzer." [Online]. Available: http://www.logalyze.com/. [Accessed: 05-May-2019].
[50] "PreludeSpecifications - PRELUDE SIEM." [Online]. Available: https://www.prelude-siem.org/projects/prelude/wiki/PreludeSpecifications. [Accessed: 25-May-2019].
[51] "MineMeld Threat Intelligence Sharing - Palo Alto Networks." [Online]. Available: https://www.paloaltonetworks.com/products/secure-the-network/subscriptions/minemeld. [Accessed: 13-Oct-2019].
[52] "6 of the Leading Open Source SIEM Tools," *Logz.io*, 07-May-2018. [Online]. Available: https://logz.io/blog/open-source-siem-tools/. [Accessed: 13-Oct-2019].
[53] "PreludeStandards - PRELUDE SIEM." [Online]. Available: https://www.prelude-siem.org/projects/prelude/wiki/PreludeStandards. [Accessed: 25-May-2019].
[54] "PreludeCompatibility - PRELUDE SIEM." [Online]. Available: https://www.prelude-siem.org/projects/prelude/wiki/PreludeCompatibility. [Accessed: 25-May-2019].
[55] "Installation of AlienVault OSSIM®." [Online]. Available: https://www.alienvault.com/documentation/usm-appliance/initial-setup/ossim-installation.htm?tocpath=DOCUMENTATION%7CAlienVault%C2%AE%20USM%20Appliance%E2%84%A2%7CDeployment%20Guide%7CUSM%C2%A0Appliance%20Deployments%7C_____8. [Accessed: 25-May-2019].

[56] "InstallingPreludeRequirement - PRELUDE SIEM." [Online]. Available: https://www.prelude-siem.org/projects/prelude/wiki/InstallingPreludeRequirement. [Accessed: 25-May-2019].
[57] "OSSIM," *Wikipedia*. 18-Mar-2019.
[58] "Αρχεία - PRELUDE SIEM." [Online]. Available: https://www.prelude-siem.org/projects/prelude/files. [Accessed: 25-May-2019].
[59] "The Complete Guide to the ELK Stack," *Logz.io*, 23-May-2016. .