Towards an Enhanced Security Data Analytic Platform

Gustavo Gonzalez-Granadillo, Susana Gonzalez-Zarzosa and Mario Faiella
Atos Research and Innovation, Cyber Security Department, Spain
{*gustavo.gonzalez, susana.gzarzosa, mario-ferdinando.faiella*}*@atos.net*

Keywords: XL-SIEM, SIEM Enhancements, Security Data Analytic Platforms, SIEM Analysis

Abstract: We present in this paper a Cross-Layer Security Information and Event Management tool (herein after denoted as XL-SIEM) as an enhanced security data analytics platform with added high-performance correlation engine able to raise alarms from a business perspective considering different events collected at different layers. The platform is composed of a set of distributed agents, responsible for the event collection, normalization and transfer of data; an engine, responsible for the filtering, aggregation, and correlation of the events collected by the agents, as well as the generation of alarms; a database, responsible of the data storage; and a dashboard, responsible for the data visualization in the web graphical interface. The proposed platform has been deployed on top of the open-source SIEM OSSIM (AlienVault) providing enhanced features compared to current open-source solutions, in particular associated to data sources, correlation engine, visualization, and reaction capabilities. A testbed implementation is described to show the integration and applicability of the tool over a security infrastructure.

1 INTRODUCTION

Several companies have developed SIEM software products in order to detect network attacks and anomalies in an IT system. Gartner reports (Kavanagh et al., 2016) analyze the SIEM tools available in the market provided by the top 14 leading SIEM vendors. According to these reports, products are classified into four groups based on the ability to execute and the completeness of vision.

Among them, we find classical IT companies such as IBM, LogRhythm, Splunk, HP, and Intel as the *leaders*, that provide products with good requirements behavior and have the foresight for future requirements; others as *challengers* (i.e., EMC - RSA Security, providing products that do not comply general market requirements, execute well at present or may control a large segment, but do not demonstrate knowledge of future requirements; others as *visionaries* (i.e., AlienVault, that forecast trending of market, but do not yet execute well; and others as *niche players* (i.e., MicroFocus, Trustwave, SolarWinds, Fortinet, EventTracker, ManageEngine, and BlackStratus), that execute well in a particular segment of market but are unfocused and do not out-innovate or outperform others.

Besides the great variety of commercial and open-source SIEMs, current tools are unable to cope with the new and complex attack patterns. Most of them do not provide high-level security risk metrics and their correlation rules are basic and weak (Barros, 2017). Current solutions lack on user and entity behavior analytic (UEBA) features, and their storage, visualization and reaction capabilities are very limited (Scarfone, 2015), (Kavanagh et al., 2016), (Sheridan, 2017), (Caccia et al., 2017).

Based on the aforementioned limitations, we introduce a Cross-Layer SIEM (XL-SIEM) as an enhanced security data analytic platform deployed on top of Alienvault Open Source SIEM (OSSIM)[1], that overcomes limitations detected in current solutions. In particular, XL-SIEM enhances the performance and scalability of current open source SIEMs, allowing the processing of increasing amounts of data and adding the possibility of event correlation at different layers with more complex rules.

Paper Organization: The remainder of the paper is structured as follows: Section 2 describes the architecture of the XL-SIEM. Section 2 defines architecture and the different modules that compose the XL-SIEM platform. Section 3 details the extended capabilities of the proposed platform. Section 4 shows the implementation and testbed of the XL-SIEM platform. Finally, conclusions and perspective for future work are presented in Section 5.

[1] https://www.alienvault.com

2 XL-SIEM Architecture

Figure 1 depicts the XL-SIEM architecture with its main components. The collection of data is done on the monitored infrastructure by SIEM Agents, and the events are sent to the XL-SIEM engine core running on Apache Storm[2], where they are processed and correlated. The events gathered as well as the alarms generated and the configuration used are integrated with the OSSIM deployment in the XL-SIEM for its storage and visualization.

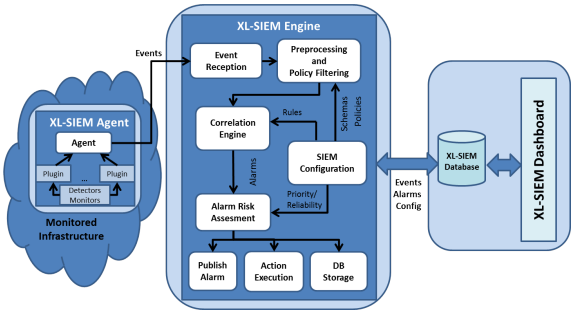

Figure 1: XL-SIEM Architecture

2.1 XL-SIEM Agent

The XL-SIEM agent is the resource layer component responsible for the event collection, normalization and transfer to the XL-SIEM Engine for its processing. Events are generated by different sensors (e.g., network traffic, honepypot, intrusion detection systems) deployed on the customer's monitored infrastructure.

The main advantages of the XL-SIEM compared to other open source solutions are as follows:

- support the usage of TLS (Transport Layer Security) certificates in the communication between agents and XL-SIEM engine;
- support for anonymization/encryption of normalized fields before transmission. The user has the possibility of defining which fields in an event type are sensitive and need to be transmitted and stored anonymized;
- new output to RabbitMQ (RabbitMQ, 2008) messaging system in JSON (Bray, 2014) format for sharing events collected;
- generation of heartbeats to monitor the status of the agents.

[2]Open source distributed real-time computation system for processing large volumes of data http://storm.apache.org/

2.2 XL-SIEM Engine

The XL-SIEM Engine is the component on the back-end layer of the monitoring architecture responsible for the analysis and processing of the events collected by the XL-SIEM Agents, and the generation of alarms based on a predefined set of correlation rules or security directives. XL-SIEM is implemented in a Storm[3] topology running in an Apache Storm cluster[4] which allows to take advantage of the benefits of this architecture.

Previous to the correlation of the collected events, there is a phase of *Pre-processing and Policy Filtering*, in which the system verifies if the user has specified some conditions to filter the incoming events before they arrive to the correlation engine (e.g., source/destination IP, port, time/date range, type of event, or the SIEM agent where the event is collected).

The *Correlation Engine* is the core of the XL-SIEM engine and integrates the open source high performance correlation engine Esper[5]. The correlator uses Event Processing Language[6] (EPL), which allows a flexible and complex definition of the correlation rules. It is a SQL-like language that includes for example the detection of patterns, the definition of data windows or the aggregation and filtering of incoming events into more complex events.

2.3 XL-SIEM Database

The XL-SIEM takes advantage of the OSSIM database and storage capabilities. The format used by the XL-SIEM for the events and alarms storage is consequently the same defined in OSSIM. The main aspects of this module are as follows:

- Data is stored in MySQL relational databases;
- There is a separate database for historical data;
- It does not support integration with cloud storage services;
- The data storage can be in a different machine from the one where the event processing takes place to improve the performance or in case it is required more storage capacity.

[3]http://storm.apache.org/index.html

[4]http://storm.apache.org/releases/2.0.0-SNAPSHOT/Setting-up-a-Storm-cluster.html

[5]http://www.espertech.com/esper/

[6]https://docs.oracle.com/cd/E13157_01/wlevs/docs30/epl_guide/overview.html

2.4 XL-SIEM Dashboard

XL-SIEM web graphical interface is deployed on top of OSSIM dashboard and consequently, it integrates the following visualization capabilities.

- different graphical charts are shown to the user as an overview of the monitored system status;
- alarms, security events and raw logs can be visualized by the user;
- additional information provided by open source tools (e.g., Netflow traffic detected by Fprobe, vulnerabilities detected by Nessus or OpenVAS) is integrated in the graphical interface;
- it is possible to generate PDF reports with a summary of the SIEM analysis;

3 XL-SIEM Extended Capabilities

This section details the additional features integrated in the XL-SIEM that enhance their capabilities in terms of data sources, correlation, visualization, and reaction.

3.1 Data Sources

Besides the data sources supported by open source SIEMs (e.g., database, log, remote logs, security device event exchange, windows management infrastructure), the following data sources have been added to the XL-SIEM:

- **Structured Threat Information eXpression(STIX) format data** (Barnum, 2014): Cyber-threat observations, represented using this type of structured language for cyber threat intelligence, are supported through a STIX plug-in that parses the STIX data and generates its representation in the OSSIM normalized event format used in the XL-SIEM.
- **JavaScript Object Notation (JSON) format data** (Bray, 2014): It supports data received from the open source message broker RabbitMQ[7] that implements the Advanced Message Queuing Protocol (AMQP).

3.2 Correlation Engine

One of the advantages of XL-SIEM architecture is the use of a high-performance correlation engine running in an Apache Storm cluster for the processing of the incoming security events.

XL-SIEM makes use of Esper[8] for event correlation and generation of alarms. This complex event processing engine is able of processing 500,000 events per second with latency below 10 microseconds average with more than 99% predictability. For more complex queries, these values are slightly reduced to a throughput of 120,000 events per second (Mathew, 2014), keeping good performance capabilities for processing large volume of data.

Security rules in the XL-SIEM are expressed using the Event Processing Language[9] (EPL). This latter is a declarative programming language that allows expressing security directives with rich event conditions and patterns in a simple way. The usage of EPL adds a business perspective to the definition of security directives.

In addition, for the definition of security directives, XL-SIEM supports two means: (i) Preconfigured categories of rules: where the user selects one or several directive categories (e.g., scans behaviours, malware detection, denial of service attacks, brute force attacks, network attacks) for which a pre-configured set of rules or security directives is included; and (ii) User custom rules: where users have the possibility to define their own rules or security directives and select them in the configuration of each correlation process.

3.3 Visualization Capabilities

Besides the visualization capabilities inherited by OSSIM, the XL-SIEM includes high-level charts and diagrams in different dashboards to provide valuable information about incidents to non-security expert users.

These diagrams can be adapted based on the client needs and requirements. Examples of available dashboards are **Executive Dashboard:** showing a high-level information relevant for a C-level administrator, e.g., the current threat level of the monitored system with a color code (green, yellow and red); **Operational Dashboard:** showing information relevant for system administrators to be able of taking decisions, e.g., the top five detected incidents, the hosts identified as source of security incidents or alarms, or the destination TCP/UDP ports of the attacks; and **Situational Awareness Dashboard:** showing a graph with the monitored network topology including the number of events detected in each component as well as

[7] https://www.rabbitmq.com/

[8] http://www.espertech.com/esper/

[9] https://docs.oracle.com/cd/E13157_01/wlevs/docs30/epl_guide/overview.html

representing with a color code (green, yellow, red) the risk level of each node.

3.4 Reaction Capabilities

The processes running in the XL-SIEM topology include support to provide the same reaction capabilities offered by the open source version of the AlienVault SIEM (e.g., execute a script, send an email or create a ticket) but with an enhanced performance.

On the one hand, the different actions can be associated to each correlation process, which allows its execution for specific alarms and not only associated to a defined policy. On the other hand, since there is a specific process defined in the Storm topology for the application of these reactions, it is possible to increase its parallelism to reduce the reaction time or even decide in which node of the cluster this process will be launched. This is useful in situations where the script to be invoked or the email server is only available in a specific node.

4 XL-SIEM Implementation and Test-bed

The XL-SIEM has been deployed in a security infrastructure to correlate logs and detect abnormal situations. The test-bed is composed of two servers, each one holding a web application (i.e., ownCloud[10], and Github[11]); the XL-SIEM Engine, holding the core of the SIEM; the XL-SIEM Agent deployed together with two IDSs (i.e., Snort[12] and Suricata[13]; the XL-SIEM Database and Dashboard, as depicted in Figure 2. In addition, an internal sub-net composed of several users with different privileges and access right over the applications has been integrated into the platform.

The purpose of this test-bed is to detect unusual connections against the web applications that could lead to the execution of attacks (e.g., brute-force, DoS, SQL injection). The Agent sends the suspicious events to the engine for correlation and treatment. Events are stored in the XL-SIEM Database and displayed in the Dashboard for further analysis.

[10]https://owncloud.org/
[11]https://github.com/
[12]https://www.snort.org/
[13]https://suricata-ids.org/

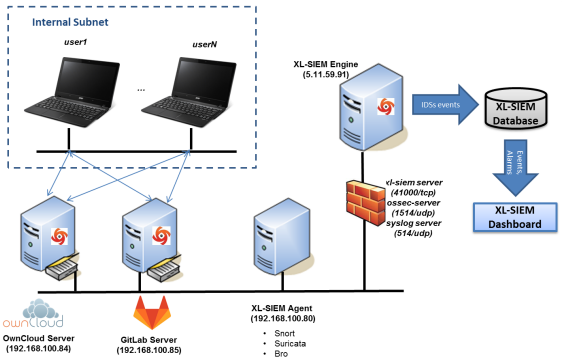

Figure 2: XL-SIEM Testbed

4.1 Attack Description

A possible Brute-force has been detected via correlating events seen on the network . Brute Force attempts are one of the few events in security that are identifiable by their volume, not by their type. While a system can be exploited with as little as a single packet of data, brute-force intrusions require greater numbers of packets to be achieved.

All hosts accessible from the Internet are continually been scanned and attacked from malicious hosts, searching for vulnerable systems to compromise and re-use. While this brute-force attempt may therefore be legitimate, it is not necessarily targeted directly against the organization. We must cross-reference against other activities from this system 222.186.52.199 to be aware if this host is engaging in further activities beyond an unsuccessful brute-force access attempt.

4.2 XL-SIEM Detection

Data are collected through the XL-SIEM Agents, directly from the sensors deployed in the monitored infrastructure. These data are then normalized and sent to the XL-SIEM Engine, for further analysis and correlation. For this purpose we use the open source high performance correlation engine Esper[14], which allows defining specific security rules using the Event Processing Language[15] (EPL), a flexible SQL-like declarative language.

As explained in section 2.2, the incoming event, before being correlated, is filtered considering specific policies. The filtering operation is performed taking into account the source and destination IP addresses, the source and destination port numbers, the

[14]http://www.espertech.com/esper/
[15]https://docs.oracle.com/cd/E13157_01/wlevs/docs30/epl_guide/overview.html

event type (e.g., ANY; DS Groups (i.e., Executable files, Suspicious DNS, Network anomalies); Taxonomy (Product Type, Category, Subcategory)), and the name of the XL-SIEM agent involved.

The following example shows a policy created for any IP source and destination, any source and destination port, an event type of DS Groups detected by any agent: $\langle ANY, ANY, ANY, ANY, DSGroups, ANY \rangle$

After this pre-processing and filtering operation, the normalized events are correlated, using specific correlation rules. The definition of these latter is a process that requires (i) the definition of statements and (ii) the definition of directives. A statement indicates the sources from which the information will be evaluated. They do not fire an alarm but the generated output is used by other services. A statement is composed of a unique name and a value that indicates the variables and the sources of information. For instance, plugin_id=1001 refers to data source SNORT.

Directives contain the pattern used in the evaluation of the rule. They are defined using Event Processing Language (EPL). They fire an alarm for every matching pattern. For every EPL directive, we can assign a category (e.g., alarm , antivirus, authentication, application, etc.), a subcategory (e.g., Attacks, Bruteforce, DoS, Malware, Network, etc.), a reliability score (from 1 to 10), and a priority score (from 1 to 5).

A concrete example of a correlation rule, associated to the attack described in section 4.1, used by the XL-SIEM engine is shown in Table 1.

Table 1: XL-SIEM Correlation Rule Example

Name: BruteForce Microsoft SQL Server authentication attack against SRC_IP
EPL Statement
Insert into BruteForce_Microsoft_SQL_Server_authentication_attempt_failed_detected select * from ossimSchema_default where (plugin_id = 1001) and (plugin_sid = 50051_2) and (dst_port=1433)
EPL Directive
Insert into directive_sls_1 select * from pattern [every-distinct(a.src_ip, 15 seconds) a = BruteForce_Microsoft_SQL_Server_authentication_attempt_failed_detected → ([3] b = BruteForce_Microsoft_SQL_Server_authentication_attempt_failed_detected ((b.src_ip=a.src_ip) and (b.dst_ip=a.dst_ip)))]

The previous rule is defined to raise an alarm after 3 failed Microsoft SQL Server authentication attempts. The EPL statement will increase the value of the related variable from the default OSSIM scheme, where the plugin_id corresponds to Snort (value = 1001), and the plugin_sid correspond to a failed Microsoft SQL Server authentication attempt (value = 50051_2).

The EPL directive shows the correlation rule used to fire an alarm if there is a match in the pattern defined within the brackets. In this case: every sequence that considers distinctively the IP source and 15 seconds of timeout will generate an alarm. For this example we have defined a failed Microsoft SQL Server authentication attempt as the event to check, if the defined condition matches (i.e., the IP source and destination are the same) and the same event occurs more than 3 times, then we will generate a correlated alarm.

4.3 XL-SIEM Analysis

The XL-SIEM dashboard shows a list of both events (e.g. Suspicious TCP traffic, Potential SSH Scan attempts) and alarms (e.g., Brute-force attack, Network Scan). The former should not be considered as real alarms, they constitute the output produced through the EPL statements. Then, this output is further analyzed through the EPL directives for raising alarms, if specific conditions are met, as explained in Section 4.2.

Regarding the alarms, they are triggered from a correlation rule having Snort and/or Suricata as input devices. Two or more conditions must be met (e.g., several particular log events in the same time period, or an event from a security control that matches against a particular host's current condition). We will focus on an alarm with a high risk level: Brute-force attack, Microsoft SQL server authentication attack against 222.186.52.199.

By clicking on the alarm name, we obtained more information about the events, that triggered such an alarm. In this case, we identified that four individual events with the same source and destination IP addresses within a period of 12 hours have triggered this alarm. Moreover, by clicking on the tabs "Source" and "Destination", specific information related to the source and destination IP addresses, respectively, can be consulted. For example, the geographic location of the IP address can be visualized through a map. For a further analysis, information associated to both the alarm and each single event, which contributed to raise the former, can be visualized.

5 Conclusion

This paper presents a Cross-Layer Security Information and Event Management tool (XL-SIEM) as a security data analytic platform that enhances current open-source SIEMs in terms of data sources, correlation, visualization and reaction capabilities.

The XL-SIEM provides several advantages compared to current open source solutions: e.g., the support of STIX and JSON data formats, secure communications between the XL-SIEM engine and agents, Data anonymization, high processing performance, advanced visualization options, and enhanced correlation capabilities that reduce the reaction time and allow customizing the reaction options.

The main drawbacks of the XL-SIEM are the following: (i) No cloud storage service supported, XL-SIEM uses MySQL relational databases for data storage; (ii) No User and Entity Behavior Analysis (UEBA) or machine learning capabilities, although it is possible to include behavioral analysis at the application level through the implementation of specific plug-ins to normalize the data; (iii) No built-in network forensics capabilities; (iv) very basic ad-hoc importing capabilities to deal with a huge number of unstructured data as inputs; (v) no support for dealing with new and specific targeted attacks (e.g., zero-day attacks).

Future work will concentrate in improving the aforementioned limitations and providing a platform able to evaluate defense in depth by analyzing multiple and diverse security devices (e.g., IDSs, firewalls, honeypots). As a result, accuracy on the detection will be improved, and false alarm rates reported back to the system will be reduced. In addition, limitations of actual SIEMs based on the last two drawbacks, as stated in (Gonzalez-Zarzosa, 2017) and (ThreatConnect, 2018), will be studied more in details.

The final objective of the security data analytic platform is to integrate the XL-SIEM with a threat intelligence platform, able to gather, process, and normalize unstructured information from external sources (e.g., OSINT sources). In this way, the XL-SIEM drawback in terms of ad-hoc importing capabilities will be overcome and the new detection rules could be dynamically created and injected directly into both the XL-SIEM and the sensors that interact with it, making it possible to recognize zero-days attacks that were not previously identifiable. Moreover, integrating the XL-SIEM with a threat intelligence platform, as highly recommended in (ThreatConnect, 2016), will also improve information sharing capabilities, considering that the platform itself will handle both the input and the output information flows. This will have a positive impact, because threat intelligence sharing is considered as a critical activity against new generation threats by governments and companies, for empowering incident response and defense capabilities (Ring, 2014) (US-CERT, 2013).

ACKNOWLEDGEMENTS

This work is supported by the European Commission, as part of the H2020 Diversity enhancement for SIEMs (DiSIEM) project (under grant agreement 700692) and H2020 European Network for Cybersecurity (NeCS) project (under grant agreement 675320).

REFERENCES

Barnum, S. (2014). Standardizing cyber threat intelligence information with the structure threat information expression (stix). Whitepaper .

Barros, A. (2017). Siem correlation is overrated. Gartner Blog Network.

Bray, T. (2014). The javascript object notation (json) data interchange format. RFC7159 available online at: https://tools.ietf.org/html/rfc7159.

Caccia, R., Cassetto, O., and Shteiman, B. (2017). The future of siem. International Information Systems Security Certification Consortium (ISC^2) Webminar available at https://www.brighttalk.com/.

Gonzalez-Zarzosa, S. (2017). In-depth analysis of siems extensibility. DiSIEM Project Technical Report D2.1.

Kavanagh, K. M., Rochford, O., and Bussa, T. (2016). 2016 magic quadrant for siem. Gartner Technical Report G00290113.

Mathew, A. (2014). Benchmarking of complex event processing engine esper. Technical Report.

RabbitMQ (2008). Amqp advanced message queuing protocol, protocol specification, version 0-9-1. A General-Purpose Messaging Standard, Technical Report.

Ring, T. (2014). Threat intelligence: why people dont share. 3:5–9.

Scarfone, K. (2015). Comparing the best siem systems on the market. Online Research.

Sheridan, K. (2017). Future of the SIEM. Dark Reading, threat intelligence article.

ThreatConnect (2016). SIEM + Threat Intelligence: Quickly Identify the Threats that Matter to You. White Paper.

ThreatConnect (2018). Threat intelligence platforms everything youve ever wanted to know but didnt know to ask. Technical Report.

US-CERT (2013). Cybersecurity questions for ceos. White Paper.