This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2017 conference.

Profiling SIEM Tools and Correlation Engines for Security Analytics

S. Sandeep Sekharan[1] and Kamalanathan Kandasamy[2]
Amrita Center for Cybersecurity Systems & Networks, Amrita School of Engineering, Amritapuri,
Amrita Vishwa Vidyapeetham, Amrita University, India
Email: [1]sandeepsekharans@gmail.com [2]kamalanathan@am.amrita.edu

***Abstract*—Nowadays, IT organizations generate colossal amounts of data. Handling these chunks of data itself is critical in the IT world. Hence centralizing the log management system improves security thereby enhances data protection in an organization. Such enterprises require a high profiling tool that helps in managing the information and events data to improve the level of security. Security Information and Event Management (SIEM) is a procedure for security analysis that prominence an overview of security in an organization. SIEM tools collect, analyze, normalize and correlates all files and analyze data coming from the various device and give a centralized view of logs. This paper articulates an abstraction of SIEM tools and event correlation engines, furnishing a description of their technical comparative study, focusing on most popular SIEM tools and open source rule-based correlation engines and profiles them.**

***Index Terms*—Information Management, Event Management, Log Analysis, Log management, Correlation, SIEM Products, IBM QRadar, HP ArcSight, Splunk, LogRhythm.**

I. Introduction

The exponential growth in the world of Internet & Technology leads to the connection of heterogeneous machines by a distributed networked system and protecting them against malicious activity is an endless task. Enterprises must protect their data from cyber-attacks by not only monitoring logs and network flow data. They require an advanced high profiling tool to gain a much more inclusive view of the organization's security infrastructure. Analytics tools help in real time analysis of servers, endpoints and network traffic, application and network logs, which collect and correlate events and performs forensic analytics to understand the attack model and system vulnerabilities. Security analytic tools provide services such as malware detection, incident detection and data loss reporting by continuous monitoring. When a security breach or threat is happened or detected, the analytics software relieves the cause. Security Analytics tools are classified as Three generation of series. Three generation series of security analytics tools are Intrusion Detection and Prevention System (IDPS), Security Information and Event Management (SIEM) and Big Data.

A. Intrusion Detection and Prevention System (IDPS)

An Intrusion Detection System is a software which automates the process of incident detection by observing the events occurring in a system, network and scrutinize them for signs of possible intrusion (incidents) [1]. An Intrusion Prevention System (IPS) is software/hardware which have the capabilities of an Intrusion Detection which help to mitigate possible incidents. IDPS focuses on monitoring possible incidents, log information, attempting to terminate and reporting information to administrators. Different types of IDPS (a) Signature based detection (b) Anomaly based detection (c) Protocol state analysis.

B. Security Information and Event Management (SIEM)

Security Information and Event Management (SIEM) are to focus on organization's Information technology (IT) security which provides a holistic view of the security management. SIEM system [2] collects relevant data produced in an organization from multiple locations thereby makes it easier to find profiled attacks by matching patterns that are not legitimate. SIEM combines Security Information Management (SIM), which accumulates data to a central repository for analysis and gives automated & centralized reporting, and Security Event Management (SEM), centralizes the storage management and correlates log files and allows near real-time monitoring, functions into one security management system [3].

C. Big Data Analytics

Information security mainly aims at analyzing data events on servers, networks, and other end devices. Big Data analytics are used for advanced security monitoring that furnishes a broader and in-depth analysis [4]. Big data security analytics and analysis is an extension of SIEM. However, the quantitative difference in volumes keeps them apart. Hadoop [4] map reduce function enables to collect and analyze Big Data fast enough to perform threat mitigation.

This paper describes the popular SIEM products used across the industry in Section II and their critical capabilities are described in Section III. Section IV gives a description of types of correlation process and different correlation engine used in Security Analytics. The conclusion is drawn in Section V.

II. SIEM Tools

SIEM system uses logs and relevant data for analysis. Event data analysis and collection of log data events from end user devices, servers, network devices, firewalls and intrusion detection and prevention systems (IDPS). The collector application forwards data to a centralized unit, which performs normalization and correlation of data for anomaly behavior of

978-1-5090-4442-9/17/$31.00

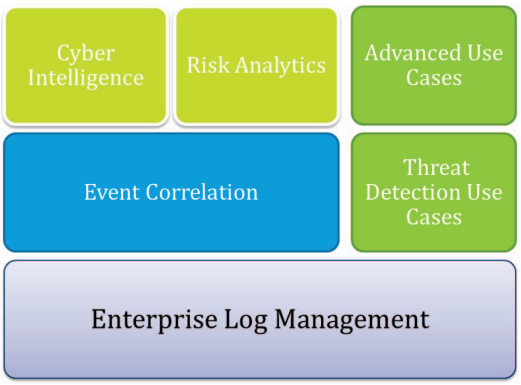

Fig. 1. SIEM tools building block.

the system. It is vital that the system should create a base profiling under normal event conditions (see Fig. 1).

The approach is layered and stage wise [5], aims at providing a structured view of how different SIEM tools focus on its implementation and working model. Listed below some of the most popular and top rated SIEM tools.

A. IBM QRadar

IBM QRadar can utilize as a single, all-in-one solution appliance and virtual appliance option to collect & process logs, net flow, full packet data, and DPI. QRadar includes incident forensics support, improved query support, flow data, threat intelligence, data storage appliances, and asset data. Authentic events can be analyzed using existing correlation rules providing an overview of logs & events with a vulnerability that delivers behavioral analysis capabilities for log events [6]. IBM Security provides an additional component, QRadar Risk Manager, for network and firewall configuration monitoring. IBM QRadar has limited capability to perform advanced use cases deployment and analytics with limited customization.

B. HP ArcSight

HP ArcSight comes with a universal log management solution that can recognize who is on the network data's they are viewing, an action they are performing with the data and how it affects the security. It has a tight integration with Big Data Analytics platform like Hadoop with a scalable architecture to support different sized organization. Even though it is scalable it has a steep learning curve for Analyst and operation making it complex for deployment and configuration. Required skilled labors to manage the solutions. ArcSight [7] has a product called Logger with maintains the storage for the huge amount of data log and an Audit App for an automated continuous control monitoring for both mobile and virtual environment.

C. Splunk

Splunk products for Enterprise and Cloud supported by Splunk Search Processing Language [8], assist searching, alerting, real-time correlation, and visualization. Can be installed as software, in a public or private cloud, or as Software as a Services (SaaS) and License is based on Volume Indexed. Splunk provides flexible analytic dashboard which improves log visualization capability. Splunk's strong visualization and behavioral predictive and statistical analytics help to detect numerous threat intelligence feed from commercial and open sources.

D. LogRhythm

LogRhythm [9] SIEM product provides a distributing log information of distributed networks also, an overall visibility of network traffic. It coalesces event management, log management, Machine Analytics with Host and Network Forensic along with File Integration Monitoring in a Security Intelligence platform. Network forensic capabilities such as deep packet inspection (DPI) and flow monitoring, provides capabilities such as file and host activity monitoring. Artificial Intelligence Engine is used for risk-based profiling and behavioral analytics. The major drawback is that no support for active directory integration for role-based access control.

III. SIEM Product Capability

Security information and event management (SIEM) systems collect and correlated security logs from numerous sources within an organization. SIEM products processes data to normalize its format performs analysis and generates alerts on detection of anomalous behavior. Machine learning [21] is fueling intelligence into analytics, changing the way we solve the problems. SIEM integrated with machine learning technologies enables organizations to aggregate and interpret data to analyze cyber-attacks before they perform damage and facilitate incident forensic and alert [10]. The different capability of SIEM products are:

Real-time Security Monitoring: A centralized storage and log correlation allow real-time analysis of an organization. Providing alerts about the live activity or attacks to take defensive measurements.

Threat Intelligence: Provides a comprehensive information about peculiar threats. Profiling and refining knowledge about potential attacks that can jeopardize an organization. Helps to understand the risks of the most common external threats, like zero-day vulnerability, advanced persistent threats & exploits.

Behavior Profiling: Learning the user activity and pattern of usage of resource in an organization. Behavior profiling builds profiles of normal activity for various event categories, such as network flows, user activity and server access. The system will assist alerts on any deviation from normal behavior.

Data & User Monitoring: Monitor the user authentication and authorization. Initially, the user authentication is done and after that, it will check for the authorized files he can access in the database. Any access or modification of the file which is not supposed to do will result in abnormal activity and creates an alert. Privileged users & sensitive data access monitoring is a requirement for compliance reporting [14].

Application Monitoring: Weaknesses in an application such as bugs or vulnerability are exploited using targeted attacks.

TABLE I
SIEM PRODUCT FEATURE [11–13].

Subject	IBM QRadar	HP ArcSight	Splunk	Log Rhythm
Logging format	Event extended format	Common event format (CEF)	Binary convert ASCII format	Log4j Format
Portfolio	IBM QRadar SIEM, IBM QRadar Log Manager, IBM QRadar Risk Manager, IBM QRadar QFlow, IBM QRadar VFlow	HP ArcSight Logger, HP ArcSight Identity view, HP ArcSight Connector, HP ArcSight ESM, HP ArcSight Audit App	Splunk Indexes, Splunk Search Heads, Splunk App for Enterprise Security	LogRhythm Log Manager, LogRhythm Event Manager, LogRhythm Network monitor
Underlying DB	Proprietary based on Ariel data store and Ariel Query Language (AQL)	Oracle till 2012, then a combination of MySQL, PSQL etc.	DB connect and Hadoop/ NoSQL stores and	SQL server database for query and reporting purposes
Unique feature	Network events, application, logs, network activity and user context	Network Monitor, Login, Logoff, File Access, DB Query	Extensive log collection capabilities with flexible dashboard-ing	Independent Host Forensic and File Integrity
Scalability	Highly scalable but limited multi-tenancy	Highly scalable support multi-tier and multi-tenancy	Minimal scalability	Minimal scalability
Configuration	Simple	Complex	Difficult	Minimal
Correlation engine	Custom Rule Engine (CRE)	Correlation Optimized Retention and Retrieval (CORR) Engine	Log Correlation Engine (LCE)	AI Engine
Organization	Small, medium & large scale	Medium & large scale	Small sized	Small and mid-size
Example	The University of Chicago, ECS Tuning	MacAulay-Brown, ITC Global Security Ltd.	Adobe, BOSCH, Duke University	Nasa, EY, Fujitsu

Ability to parse activity streams from applications allows application layer monitoring [14].

Analytics: Discovery, interpretation, and communication of meaningful patterns in data security analytic composed of dashboard views, reports, and query functions. Carries investigation of user's activity & resource access to identify a threat, breach or the misuse of privilege.

TABLE II
PRODUCT RATING CAPABILITIES [11], [13], [16], [17].

Capability	IBM QRadar	HP ArcSight	Splunk	Log Rhythm
Real-time security monitoring	4	4.1	3.7	4
Threat intelligence	4	4	3.5	3.5
Behavior profiling	4.5	4	3.5	3.5
Data & end user monitoring	3.8	4.2	3.7	4.1
Application monitoring	4.3	4.5	4.3	4.1
Analytics	3.7	3.8	4.2	3.8
Log management & reporting	4	4	4	3.8

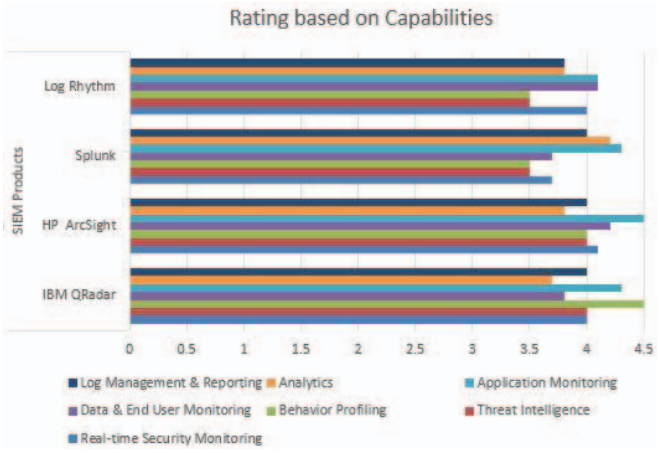

Fig. 2. Graph based on service rating on capabilities.

Log Management and Reporting: Collection of log files from different sources such as server logs, system logs, event logs, firewall etc. These files will be in the large size of 10–100 TB of data. Managing, storing and analysis these files for reporting an alert is a tremendous task for the SIEM tool.

The security team should identify products that support requirements of the organization to match internal project and support capabilities. The principle requirement of security analysts, to identify the critical capability of SIEM products. Each capability is graded in terms of its relative performance as well as for specific product use cases [17].

Each product is evaluated [15] in terms of its services and performance capability it delivers, on a scale of five-point where Poor = 1, Fair = 2, Good = 3, Excellent = 4, Outstanding = 5 [Table II]. Fig. 2 shows a graphical representation of different capabilities based on the different products from the data on Table I.

IV. CORRELATION PROCESS AND ENGINES

SIEM (Software Information and Event Management) systems are becoming a Critical infrastructure, providing the means for processing and analyses distributed data's and

TABLE III
TYPES OF CORRELATION [20].

Characteristics	Similarity-based	Knowledge-based	Statistical-based
Functionality	Based on similarities between previous and current alerts	Based on defined set of rules and known attacks, signatures	Based on similarities between previous and current alerts
Integrate sensors alerts	Y	Y	N
Initial profiling	Y	Y	N
False alerts detection	Y	Y	Guessing
New attacks	Y	N	Y
Rate of error	Average	Low	High

events, for analyzing security overview of an organization. The main component of SIEM tool architecture is correlation engine, which is used to normalize, filter, reduce and aggregate events from a set of miscellaneous set of inputs. From a cybersecurity point of view, correlator plays a vital role in SIEM architectures, providing the inner security information from existing event sources. Correlator performance is important to process large amounts of data inputs. Hence, we concentrate on popular open source rule-based correlation engines like Simple Event Correlator (SEC), Esper, Nodebrain and Drools [18].

A. Types of Correlation

Types of Security-specific correlation can be loosely categorized into rule-based and statistical (or algorithmic).

Similarity-based Correlation: Compare the similarity of an alert with a cluster of alerts or any two alerts [19]. The algorithms aim to cluster similar alerts in time. An important advantage of these algorithms requires no need for a precise definition of attack types.

Knowledge-based Correlation: Needs some pre-existing knowledge of the attack- the rule, to detect only the precise terms. Hence, the algorithm should be updated with all new types of attack knowledge.

Statistical correlation: Does not requires pre-existing knowledge of attacks but instead depends on the knowledge of normal activities profiled.

A characteristic comparison of different types of correlation is described in Table III. Where Y represents 'Yes' and N represents 'No'.

B. Event Correlation Engine

Profiling of several event correlation tools, aiming at their architectures, operational behavior, configuration and management, modules and rule format. The engine selection of correlator depends on representative components for the rule-based correlator category [18].

A comparison of different correlation event engine is described in Table IV.

V. CONCLUSION

SIEM systems have become an integral part of corporate defenses to detect, respond, and forensic of incidents. Here we have profiled SIEM tools and correlation engines. Enterprise can deploy SIEMs with a wide range of options by constant monitoring as per the Security Analytics needs and fine tune the data flows the systems in takes. A full sized SIEM system can cost hundreds of thousands of dollars, and while that will give top line capabilities. Some vendors offer a lightweight version that gives basic log management and reporting capabilities without the advanced analytic capabilities. Organizations should use capabilities analysis as a major concern regarding a product before making a final decision. It is also recommended that organizations or security team should consider capabilities criteria as the most important criteria for making a proper decision.

Acknowledgements: At the outset, I express my heartful gratitude to K Sethuraman Srinivas and Kamalanathan and other faculties for their valuable guidance, timely suggestions and help in the completion of this paper. I express my heartfelt gratitude to all the staffs of the department of Amrita Cybersecurity Systems & Networks, Amrita School of Engineering. I also thank the Evaluation panel for their valuable feedback, which made it possible in completing my paper. I thank all my friends who have helped me directly or indirectly, in the completion of this paper.

REFERENCES

[1] K. Scarfone and P. Mell, "Guide to Intrusion Detection and Prevention Systems (IDPS) Recommendations of the National Institute of Standards and Technology," Nist Special Publication, pp. 800–894, vol. 127, 2007. Retrieved from http://www.reference.com/go/http://csrc.ncsl.nist.gov/publications/nistpubs/800-94/SP800-94.pdf.

[2] Karen Scarfone (2016), "Basic SIEM analytics steps," [Online] Available: http://searchsecurity.techtarget.com/tip/Basic-SIEM-analytics-steps-to-know.

[3] en.wikipedia.org,"Security information and event management," 2016 [Online] Available: https://en.wikipedia.org/wiki/Security_information_and_event_management.

[4] Bhawna Gupta and Kiran Jyoti, "Big data analytics with hadoop to analyze targeted attacks on enterprise data," *(IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, pp. 3867–3870, 2014dfg.

[5] infosecnirvana.com, "Enterprise SIEM implementation- building blocks," [Online]. Available: http://infosecnirvana.com/enterprise-siem-implementation-building-blocks/.

[6] WhitePaper, "IBM QRadar security intelligence platform," [Online] Available http://www-03.ibm.com/software/products/en/qradar.

[7] Aamir Sohail and Dr. Sandeep Josh, "IT security using arcsight SIEM," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 4, Apr. 2014, ISSN: 2277 128X, pp. 383–388. 2004.

[8] Techincal Paper, "Using splunk software as a SIEM," [Online]. Available: https://www.splunk.com/pdfs/technical-briefs/splunk-as-a-siem-tech-brief.pdf.

[9] Whitepaper, "LogRhythm - the security intelligence company," [Online]. Available: https://logrhythm.com/products/security-intelligence-platform/.

[10] Flanagan, Enda Fallon, Abir Awad, and Paul Connelly, "S. A. V. I. O. R: security analytics on asset vulnerability for information abstraction and risk analysis," in *2016 UKSim-AMSS 18th International Conference on Computer Modelling and Simulation (UKSim)*, 2016.

TABLE IV
COMPARISON OF DIFFERENT CORRELATION ENGINE [17].

Characters	Esper	SEC	Drools	Nodebrain	Prelude
Deployed platforms	Java &. NET	Perl	Java	C	Python
Usages	Used to detect inter-domain stealth port scans, analyzing the establishment of TCP connections	Reduces amount of information between log generator and servers	Monitor object in a board environment	Configure distant network element and perform actions on another, based on predefined policies.	Normalize, reduction and aggregation of events produced by different network IDS
Approach	Rules declared in SQL-like approach, labeled as Event Processing Language (EPL) and pattern matching	Text based approach, as string occurrences and regular expressions	Rules declared in expression based format and events in Java Classes	Rules uses predefined format, events based on text input	Uses modular approach, contains components like correlation engine, log analysis tool.
Memory usage	High	Less compared to Java based libraries	High	Less compared to Java based libraries	Less compared to Java based libraries
Format	No unified format	No unified format	No unified format	No unified format	Intrusion Detection Message Exchange Format (IDMEF)

[11] Infoesecnirvana.com, "Clash of the titans- arcsight Vs QRadar," [Online]. Available: http://infosecnirvana.com/clash-titans-arcsight-vs-qradar/.

[12] John P. Mello Jr., "Magic quadrant for SIEM," [Online]. Available: https://techbeacon.com/latest-gartner-magic-quadrant-siem-takeaways.

[13] Deepak Kumar, (2016), "SIEM product comparison" [Online]. Available: https://www.linkedin.com/pulse/siem-product-comparison-2016-deepak-kumar-d3pak-.

[14] Tae Kyung Kim, Hyung Jin Lim, and Jae Hoon Nah, "Analysis on fraud detection for internet service," *International Journal of Security & its Applications*, vol. 7, no. 6, p. 275, Nov. 2013.

[15] Jeff Edwards (2016), "Evaluating SIEM solutions" [Online]. Available: http://solutionsreview.com/security-information-event-management/five-questions-you-need-to-ask-yourself-when-evaluating-siem-solutions/.

[16] Kelly M. Kavanagh and Mark Nicolett, "Magic quadrant for security information and event management" Published: 10 Aug. 2016 ID: G00290113.

[17] K. Agrawal and H. Makwana, "A study on critical capabilities for security information and event management," vol. 4, no. 7, pp. 2013–2016, 2015.

[18] L. Rosa, P. Alves, T. Cruz, P. Simões, and E. Monteiro,"A comparative study of correlation engines for security event management," University of Coimbra, Portugal.

[19] Seyed Ali Mirheidari, Sajjad Arshad, and Rasool Jalili, "Alert correlation algorithms: A survey and taxonomy," in *Cyberspace safety and security*, vol. 8300 of Lecture notes in computer science, pp. 183–197, 2013.

[20] Iansresearch.com, "Anchor your security with a well-honed SIEM strategy," [Online]. Available: https://www.iansresearch.com/insights/guides/siem-campaign/anchor-your-security-with-a-well-honed-siem-strategy.

[21] V. Das, Pathak and K. T. Gireesh, "Network intrusion detection system based on machine learning algorithms," *International Journal of Computer Science Information Technology (IJCSIT)*, 2010.