UNIVERSITY OF PIRAEUS
DIGITAL SYSTEMS DEPARTMENT

M.SC. IN TECHNO-ECONOMIC MANAGEMENT
& SECURITY OF DIGITAL SYSTEMS

Master Thesis

SECURITY INFORMATION AND EVENT MANAGEMENT SYSTEMS:
Benefits & Inefficiencies

Katsaris Dimitrios

Piraeus, January, 2014

SUPERVISOR

Prof. Sokratis K. Katsikas
University Of Piraeus

COMMITTEE OF INQUIRY

Prof. Sokratis K. Katsikas
University Of Piraeus

Associate Prof. Konstantinos Lambrinoudakis
University Of Piraeus

Assistant Prof. Christos Xenakis
University Of Piraeus

Abstract

In this Master's thesis, the new trend in computer and information security industry called Security Information and Event Management systems will be covered. The evolution, advantages and weaknesses of these systems will be described, as well as a home-based implementation with open source tools will be proposed and implemented.

Acknowledgements

I would first like to thank my professors for the wonderful experience I had at University of Piraeus. They taught me how to overcome the challenging and exciting problems of the real world and show me how to use and utilize new technologies. They guided me and encouraged me when I felt lost in the world of security and they learnt me to focus on detail.

Moreover, I would also like to thank my parents and my friends for their support during my studies. They were beside me, every single time I needed them and they were willing to help me and calm me when I felt stressed. Without them I could not find the motivation and the energy to fulfill that dream of mine.

Table Of Contents

Introduction

We are living in a fast paced world, eating fast food, driving fast cars, seeking fast services and demanding fast solutions. The world is moving faster than any before around us; but when it comes to security, is that speed desirable? In the information era we are living, we are surrounded and exposed to extensive information. It is pretty sure, though, that each and every one of us would definitely trade some of that aforementioned speed for security's sake. "*Security, is not just a technical concern; it is not even a product. It itself is a process*", as *Schneier* states [1].

Information security plays a significant role for companies and organizations and has become the basis for business survival as much as any other issue. As nearly every device in an IT infrastructure gives security alerts, the amount of data in need of processing and archiving grows at a blistering pace. Therefore, special attention should be given when it comes to identifying security threats, mitigating security risks and assuring that all feasible precautions have been taken to handle efficiently and effectively any security incident that may arise.

SIEM (Security Information and Event Management) systems is the new trend being used in IT-SEC (IT Security) and INFO-SEC (Information Security) industry for those reasons. More and more companies and organizations tend to implement such systems, as they provide a sufficient solution to efficiently analyze and report data, respond effectively to inside and outside threats, and follow compliance regulations. However, in such a dynamic environment, where data are generated rapidly, continuously and dynamically those solutions may be proved inadequate to cope with huge volumes of data. Big data revolution has become a reality, thrusting information security into the business limelight and opening new doors for both exciting business potential and increased security risks.

The concept of SIEM systems emerged in the late 90's, around 1998, and has been evolving rapidly ever since. Due to the large amount of security log information created in modern systems and the variety of their respective formats, log handling is getting increasingly difficult to manage without the assistance of a specially designed platform. Even though resembling events occur in an organization's life cycle and relevant data about an enterprise's security is produced in multiple locations, most of the times none of them are given the attention they deserve.

Regardless of how effective or hardened an individual device might be, if monitored as a single unit with no event-log correlation taking place, minimum concerns will be raised in case of a potential security bypass. Thus, the actual impact of such an incident inside a corporate network or in regard to other devices or systems is difficult to be assessed or predicted. On the other hand, when monitored as a whole, with cross device correlation, each device will signal an alert as it is attacked raising awareness and threat indications at each point allowing for additional defenses to be brought into play, and incident response proportional to the total threat. Having collected all that signals and data in one place and being able to look at them from a single point of view makes it easier to identify unconformities and inconsistencies. SIEM systems contribute to that direction, offering a central repository where all data and events are stored and maintained.

In this thesis we will shed light on this new trend, called SIEM system. More specifically, in the first chapter an introduction to SIEM systems will be provided, describing how these systems evolved to reach their recent form, explaining their purpose of existence and specifying their benefits and advantages in modern business environments. In the second chapter the preliminary actions prior to implementing such a system will be specified, while in the third one, the basics of SIEM systems related to their capabilities, operational requirements, and architecture will be discussed. In the next two chapters, 4 and 5, a reference to the available solutions in SIEM marketplace will be

provided and their inefficiencies will be pointed out, while, in the last but one chapter an open source, light-weight SIEM solution for a SOHO environment will be implemented and tested. Finally, in the very last chapter, conclusions will be presented and future work will be proposed.

Chapter 1 – Introduction to SIEM systems

1.1 Why to use a SIEM

A SIEM system, as its half name states (Security Information Management), must provide all the appropriate means to efficiently manage information security. The objective of information security management is to ensure confidentiality, integrity, and availability of the information within an organization; having such a system implemented and configured properly, will result in the assurance that the organization's sensitive data is not disclosed to unauthorized individuals, processes or devices, will guarantee that information is sound, unimpaired and in perfect condition and that information objects and other system resources are accessible when needed and without undue delay. Moreover, the appropriateness and correctness of the control mechanisms and of the processes of data collection, retention and reporting would be more easily verified and validated.

According to the other half of SIEM's name (Security Event Management), such a system should assist in identifying and investigating anomalous events, through the analysis of large volume of data. Assuming the proper implementation of a SIEM system, an enterprise would benefit in many ways and multiple directions; from false positive and false negative reduction to normal activity omission. Among others, a unified framework will be provided, consolidating multiple log lines into single "threads" of activity and disparate logs from different systems into single events. The significance of those actions can be easily assessed, if we consider the majority of IT and security devices, as well as the variety of vendors that could be found inside a company. All these systems, devices and technologies are hard to integrate on a single solution, as they do not conform to any common format. On the contrary, each of them uses a distinct log format, serving many purposes such as analysis and troubleshooting, that adds unnecessary complexity in terms of trend analysis and makes it inevitably difficult to produce a single security detection tool for the best part of them.

Implementing a fine-tuned SIEM system can also help reducing the number of security events on any given day to a manageable, actionable list, and to automate analysis such that real attacks and intruders can be discerned. As commonly known, a defense in depth strategy utilizes multiple devices such as Firewalls, IDS/IPS, OS Logs, etc. that can easily generate millions of logs. No matter how skilful a security engineer is and how effectively an individual device operates, if logs are not monitored and correlated, each device can be bypassed individually. On the contrary, when systems are monitored as a whole, with cross device correlation, each device may signal an alert raising suspicions, offering on each device the time to prepare itself and allowing for additional defenses to take place.

However, apart from the operational advantages, after implementing a SIEM solution, an organization will also reap the benefit of regulatory compliance. Such an implementation can be used to ensure legitimate use of company's resources, validity of company's procedures, and assure company's service quality, as during an audit or investigation, an enterprise will have the information needed to demonstrate compliance or due diligence. Moreover, it can be used to make sure that the necessary records are being logged and stored and are ready for retrieval upon inspection, as mandated by such rules and policies. Log analysis, used in this manner, is not applied only to detect emerging threats and trends, but also improve and fine-tune overall security. This is of great importance, especially for businesses and enterprises that, due to regulatory compliance commitment, require collection and archiving of event logs.

Last but not least, Business Continuity optimization and Return On Investment maximization can also be achieved through such an implementation. "*A properly implemented SIEM Solution provides security analysts with a toolset that can greatly enhance their effectiveness. A more effective security team has a greater likelihood of intercepting and addressing security events in their early stages before they can significantly impact the enterprise. This effectiveness can help reduce the overall information risk profile of the enterprise*"[2]. At the same time the implementation of a single SIEM system will help the enterprise to save money and time, as the purchase and maintenance costs associated with various monitoring and analysis systems such as SEM (security event management), SIM (security information management), log management and analysis systems etc. can be lessened by having a single SIEM tool.

Demand for SIEM systems has flourished over the last decade. According to Gartner's publication [3], which took place in May 2011, only during 2010 the SIEM market grew from $858 million to $987 million, achieving a growth rate of 15%. That growing demand was a result of a stronger focus on security-driven use cases, driven by a combination of compliance and security needs, not only from new but also from existing customers. SIEM market and vendors following that tension started expanding SIEM systems and enriching them with capabilities and functionalities from adjacent areas, such as integrity monitoring, vulnerability assessment, security configuration assessment, etc. and started promoting them as "platforms" that can provide security, operations and application analytics. Despite the hard times in their early days of existence, today's SIEM systems are very competitive and mature.

Another research, which was conducted in March 2008 by Aberdeen Group [4], revealed that SIEM users exhibit superior capabilities in security Governance, Risk Management and Compliance, while these users, compared to the Industry Average, rated their performance:

- ✓ 15% higher at prioritizing security and compliance-related investments based on defined business objectives and acceptable levels of risk
- ✓ 11% higher at speed of decision-making regarding security governance, risk management and compliance
- ✓ 18% higher at optimizing business processes related to security governance, risk management and compliance.

From all the above we can argue, that the implementation of a good and efficient SIEM tool, can be proved multi-beneficiary for an organization. The increasing growth rate of SIEM market as well as the optimized results in terms of performance and governance can verify that. "*Such a system can provide the analytics and knowledge of a good security engineer, while acting automatically and repeatedly against a mountain of events from a range of devices. In that way and with the help of a SIEM tool an engineer can handle even 100,000 events per day.*"[5]

1.2 SEM + SIM = SIEM

SIEM systems are comprised of the formerly disparate, yet complementary, product categories of SIM and SEM systems. However, since the first few years of the 21st century, these two products have converged into a single solution set under the name of Security Information & Event Management

Although the terminology can be misleading, as a reference to a project of protecting a company's information in case of any computer security breach, a SIM system is typically a log management solution that aims to provide a common framework for managing security logs with the use of data collection techniques. Primitive SIM systems focused on the historical analysis of log file information for forensic investigation and reporting purposes. As they were complex to deploy and to use, they were only applied to the largest organizations with the most mature and accurate security components. The most recent SIM products are software-based agents running on different (security) devices such as IDSs, firewalls, workstations etc. and they are used to automate the collection, storage and archival of logs files from such components into a central repository. These files, after successfully collected into the centralized server, can be subsequently translated into simplified and/or correlated formats for trend analysis. Log message analysis can reveal suspicious actions, expose malicious behaviors and identify significant threats, while automating the above processes can result in reducing errors and complexity. Automated tools and procedures decrease the burden on human to process the significant data collected by different sources and the time to derive information from multiple systems.

First SEM systems, on the other hand, focused on real-time or near real-time monitoring, correlation and analysis of security events. The benefit of real time log analysis is that the security personnel can stay up-to-date with attacker's actions, and by doing so they can respond fast enough and prevent him from doing further damage. Thus, although SIM and SEM systems may have often looked at the same event, the latter ones did that in real time. In order to be considered successful, a SEM solution must be able to identify multi-type events, correlate different data sets and monitor for unusual behaviors rather than dealing with individual events. Towards providing better correlation and reporting capabilities many SEMs may attach contextual and/or identity information in event logs (such as owner, location, user info related to accounts referenced in the event like first/last name, manager's name, etc.), while the vast majority of SEMs can also integrate with external remediation, ticketing, and workflow tools to assist with the process of incident resolution. [6]

In the following lines we are going to describe the functionality and the characteristics of these 2 components, on which the modern SIEM systems are based.

SEM

A Security Event Management system, in an abstract level of description, is nothing more than a software that aims to notify, in real time, information security and information technology personnel about unusual occurrences, after analyzing and correlating input logs

and alerts from different systems and devices. Having the full picture of an enterprise infrastructure comprised of systems, devices and applications and possessing a correlation/behavioral engine that assists in monitoring and understanding of multi-type security events, a SEM system helps improve signal to noise ratio, and separate real threats from false alarms. A behavioral engine is usually applied therefore, and as it is obvious, behavioral engine's intelligence is of vital importance to SEM success. It is due to that feature that a SEM is often characterized as the brain of the SIEM solution.

Deployment of a SEM system can be a complex, demanding and time consuming procedure, especially for larger companies and organizations. Significant investment in money and time may be also required, at least in its installation phase. The most common challenges that SEM users face are [7,8]:

- A high rate of false-positive security alerts,
- Ineffective prioritization of security events,
- Limited contextual information on security events identified by the SEM.
- Manual configuration of sophisticated rules in order to govern event correlation and reporting.

In order to deploy an efficient SEM system, special care should be taken, at least, over the following characteristics:

- correlation mechanism

  An intelligent behavioral engine should be used to effectively identify security events in response to a correlation between log items, log items and event data, and different event data itself; and prioritize identified security events accordingly. The more intelligent the behavioral engine is, the less likely an analyst will be overloaded with false positives, and the more likely the SEM project will succeed.

  Correlation mechanism should be able to perform normalization in logs and events, so as to produce a common formatted output. The term "normalization" is mostly used from vendors to refer to the Regular-Expressions usage, in order to populate structured, relational data from unstructured information such as logs data. Normalized data and events can then be used to identify, aggregate and correlate repeated events from a single device, or same events repeated from multiple devices.

- notification engine

  After having identified malicious behaviors the appropriate incident handling team must be informed. The challenge at this point is the adoption of a notification mechanism that is not only user-friendly but also efficient, accurate and time-effective. Most SEM solutions are based on e-mail alerting, however such an approach could be disastrous for large-environment companies, as they could easily suffer from overloaded inboxes. Another solution proposed is the use of multi color alert depicting dashboards. A

graphical – visualized solution, would allow appropriate personnel to perform "peak-and-trough" analysis and identify variances that would be an indication of something abnormal is happening.

- scalability

  At the SEM layer, scalability comes down to two concerns. The first one is related to alerts database. Large number of such databases should exist in multiple servers, providing a redundant and a more robust solution. Based on such an architecture scheme, separate alert-specific databases may be implemented with their own data retention times and permissions settings. The second one is related to the correlation engine. Especially in large enterprises, where billions of logs are generated and must be handled efficiently, it is important that the SEM implementation can support multiple correlation engines that can pass each other alerts for evaluation, contributing to full scalability.

SIM

A Security Information Management system is often referred to as the dump portion of the SIEM. According to [9] "*Security Information Management (SIM), is an industry term related to information security referring to the collection of data (typically log files) into a central repository for trend analysis... SIM refers to just the part of information security which consists of discovery of 'bad behavior' by using data collection techniques...*" So, as a typical Log Management Solution, SIM must be responsible for log collection and log retention, especially, for future reference. Despite the fact that log collection is among the primary features of the SEM component, a SIM system must also have the ability to report on the collected logs and alert if needed.

Considering collection and retention functions, in [10] it is stated that a log management infrastructure can be divided into 3 layers.

The first layer, comprises of the hosts providing their logs to the second layer. This operation usually occurs in a real-time or a near-real-time and can be performed in two different ways. Log data can be either forwarded to the second layer from host-based services that run on layer-1 host or provided after second layer devices requested.

In the second layer belong the log servers that receive log data from the first layer hosts. These servers are also known as collectors or aggregators and they are responsible for log storage and analysis. Due to multiple log formats different methods of converting logs to a single standard format need to be implemented, with the syslog format being the most commonly used.

The third layer consists of monitoring apps and consoles that are mainly used for reviewing the results of log data analysis. Report generation, management dashboards and log baselines are met found this tier.

Complementary to the above mentioned functionalities, additional ones, such as forensic analysis, may be supported by a SIM system. However, in order to use the logs

collected for forensic purposes, it must be ensured they will not be dropped even if slight changes are performed on them. At the same time, alerting capabilities is a nice-to-have feature, rather than a mandatory one, as typically clever alerting would occur at the Security Event Management (SEM).

The table below depicts the main differences among SIEM's components, SEM and SIM.

SEM	SIM
Real time analysis	Historical Analysis
Correlation and aggregation.	Log collection and indexing.
Improved reporting and alerting functionality	Basic reporting and alerting functionality
Scalable	Not Scalable
Normalization & interpretation	Forensics

Fig 1.1 SEM SIM Comparison

Chapter 2. Preliminary Actions

In the previous chapter we examined the heterogeneous product categories of SIM and SEM, as the basic components of recent SIEM systems. From this chapter on we will focus on SIEM systems themselves. Before starting implementing such a system, it is of great importance to define the scope and the focus of that project, as various solutions exist in SIEM marketplace [Appendix A], each of them with its own characteristics and capabilities. The scope is usually related to the objective and the purpose of such an action, while the focus defines the size and the basis of the system that will be applied. Although the scope and the focus can be chosen separately, they together define what the SIEM system will be about.

In [11] compliance, security and operational objectives are identified as the most common reasons for installing a SIEM system.

When compliance is a priority, a SIEM system can be used to extend the capabilities provided from log management systems. In such a scenario, a SIEM installation should make sure all necessary records are being logged and stored and are ready for retrieval upon request, as mandated by company's rules and policies. Although collecting and storing log files doesn't necessarily make a company more secure, a SIEM system can be used to demonstrate compliance to standards such as ISO 27002, SoX, HIPAA, PCI DSS.

When it comes to security, a SIEM can operate as an external threat monitoring and a security application monitoring solution. Log file analysis can reveal suspicious behaviors and identify possible compromises. In order to prevent an attack to spread rapidly inside a network and limit its impact on adjacent systems, correlation techniques are used by SIEM systems. In that way not only sophisticated attack vectors can be determined but also anti-patterns can be discovered. An anti-pattern could have been the sudden absence of attacks. If, for example, an IDS reports on a significant number of attacks during the day and suddenly stops doing so, it might have been compromised itself.

Finally, when the primary reason for installing a SIEM system is performance-related, such an installation can be used to achieve continuous optimization. Each device and/or vendor usually has different metrics that indicate its capacity/load. Collecting and correlating log files from different sources can help determine interconnections and interactions among systems and lead to better understanding of how resources are being used inside a network. Based on these metrics, operators and developers can optimize software quality, maximize every-day procedure's efficiency and prevent system crashes and errors.

The focus, on the other hand, defines how narrow or extended the SIEM project will be. For example the focus may define the location where SIEM system will be applied (branch, region, whole country) or the population that will be responsible for (customers, company's personnel, both). Having defined both the focus and the scope of a SIEM system the next step is to specify the use cases.

The better defined the goals and the environment where the SIEM system will be installed, the more accurate and more effective the project will be. Among the most common ways to represent the goals of a project is through use cases. Use cases can help getting a manageable environment by limiting the number of business units, devices and processes involved in a project. As obvious the use-cases can be used to serve a compliance-based, a security-based or an operations-based scope. Some general examples of the questions that SIEM can answer in those different scopes as cited in [11] are:

Issues related to segregation of duties.

Segregation of duties is a fundamental security principle that ensures a single person does not have the ability to control two or more phases of a transaction or operation. It also prevents abuse and fraud and is achieved by carefully defining each action and subsequent privilege needed to complete a transaction or operation. A compliance-based use case related to segregation of duties would help defining and protecting against malicious actions and users who would take advantage of their privileges serving their own interests. A properly configured SIEM system with the knowledge of such a compliance use case, could easily raise an alert if it encountered a transaction like that, even if the transaction was not executed regularly, or from the same user.

Cross-system errors and warnings tracking over complex networks

A SIEM solution implemented to deal with the above issue should be able to track and correlate various types of errors and warnings generated from different sources inside a complex network topology. Based on that correlation a SIEM system should be able to define risks and priorities related to the errors and warnings that are logged, and prioritize and notify accordingly for the proper and timely resolution of those problems.

Cross-system authentication tracking

In an environment where cross-system authentication tracking is a concern a SIEM system should be able to track (failed) authentication attempts and their identifying information (session variables, usernames, login devices) through the log files preserved from various devices inside a network, aiming to reduce the number of false positives raised from IDSs. "*Security related use cases are usually closely related to that goal and are further away from business terms and facts.*"[11]

As a final step, before starting to implement a SIEM solution, the system requirements should be defined. This is a very important step for the general success of the whole project, as the requirements describe what the system needs in order to efficiently and effectively handle and operate under the use cases specified from previous stages. Requirements can be quite technical such as requirements for data collection and retention, or more generic such as those related to reporting and event management. It is essential though to make sure requirements are a little bit of both: "*technical enough to be usable and high-level enough to be flexible.*"[11] In the next chapter we are going to analyze these requirements as well as the capabilities the SIEM should possess in order to successfully meet them.

Chapter3. SIEM Systems Fundamentals

3.1 Capabilities & Requirements

"*Efficient information security management requires a security event management approach with enhanced real-time capabilities, adaptation, and generalization to predict possible attacks and to support human actions*". [12] A SIEM system is a dedicated solution towards, designed to collect and analyze log files and search for patterns which may indicate incidents or persistent problems. Despite the vast number of alternatives available in SIEM industry today, there have been identified some specific features, common to most SIEM solutions. These features are related to the requirements and capabilities that a SIEM system should meet and offer respectively.

Data Collection

First and foremost a SIEM system should be capable of collecting data. In order to do so, it should be accessible from all the systems inside an infrastructure and should support different collection techniques. There might be network devices, operating systems and security devices that may share similar logging and alerting functions, but usually there will be significant variations among them. Additionally, some devices may be able to connect directly to the SIEM system using a standard protocol while others may use a vendor-proprietary protocol or an API (application programming interface). There may be also end systems that they communicate in real time with the SIEM system while others may periodically forward system created files. A SIEM system should support all the above mentioned categories.

There are 2 major requirements concerning data collection, data collection continuity and data collection relevance. The first one addresses the need for all the log files to arrive untouched, without any of them go missing between steps, as they traverse to their final destination, while the second one assures that limitations put in place by scope, focus or use cases should not affect devices' usability and/or performance. A well defined use case, scope and focus of the project will prevent all these units (network devices, operating systems, etc.) from logging unnecessary data and the SIEM system from processing them.

Data Aggregation

From the time SIEM system collects the log files from its various sources, it starts combining that data into a single data store. Aggregation reduces the volume of event data by consolidating duplicate event records, while providing significant performance benefits by minimizing the potential overhead occurred from handling duplicate event records and by optimizing the search time required from a query to return a result. However, despite aggregation may seem straightforward, it presents a number of challenges and

considerations, such as the size of an enterprise, the amount of data being collected and the architecture of the IT infrastructure.

Data Normalization

If the aggregated data won't be normalized first, the identification or the categorization of a single event may require extensive processing and analysis, and may end up being impractical. "*Normalization is the process of resolving different representations of the same types of data into a similar format in a common database*".[1] The process of normalization extracts common information and expresses it in a consistent format, which allows for a direct comparison of different events. Normalized events can then be used to dampen repeat events from a single device, or multiple devices repeating the same event. Frequently, before normalizing data, copies of raw logs are stored in their native format to ensure that a full record of the logs is maintained. This information can be proved valuable for investigation purposes and compliance demonstration.

After being normalized and aggregated, data should be placed in a storage component. As it is obvious, new requirements should be defined for retention purposes. These requirements should contain the demands and the restrictions needed to keep the log files (consolidated). In general, there are 3 main restrictions concerning data retention, as cited in [11].

First of all, data retention techniques and retention time are affected by data privacy laws and legal frameworks. There are both national and international privacy laws and legal frameworks which define the rule-set that companies should abide by and ensure that these companies do not illegally store information on their customers or employees, or retain that information longer than allowed. As a result, that legislation applies to log files too.

But even if legislation was loose enough and allowed more storage or longer retention time, another constraint should have been taken into consideration; that of the cost. The costs related to servers, storage capacity, energy and maintenance would make it unprofitable and impossible to store every snippet of log file.

Last but not least data retention requirements are about relevance as well. Not all records need to be kept indefinitely and no records other than required should also be kept. There might be times that even within the same use case the retention period may be different for certain records, as some events trigger more often than others. Therefore, no clear answer exists to what is the optimal retention time, as the retention period should be long enough to analyze past incidents.

Event Correlation

Any SIEM system would be of no use if it only collected logs for a specific type of event or from just a single device, since it would defeat the purpose of data and event's correlation. Output's efficacy relies on having a broad set of data from a wide range of devices.

Event correlation is the function of linking multiple security events of alerts, typically within a specified time window and across multiple systems, to identify anomalous activity that would not be evident from any singular event. It is obvious that correlation should happen real-time to detect zero-day threat vectors, minimize the attack surface across the IT infrastructure and speed up detection of and reaction to security threats. As it is shown in [4], by collecting and correlating information and events from both logical and physical security infrastructures, companies improve overall management visibility and progress towards a common, enterprise-wide view of risk.

Different approaches exist over event correlation engines; each of them under different perspectives and with different requirements. The most important categories as cited in [13] are:

- Self-Learning vs. External Knowledge Correlation Engines

  "In order to be able to correlate events triggered by service and network problems, a correlation engine requires knowledge, such as information about the network structure, information about the triggers for the events, or information about service dependencies.

  Such information can either be gathered automatically, or manually from experts. Obviously, the second option requires a lot of work from experienced operators. This is economic only if the majority of the supplied external knowledge is static. On the other hand, automatic learning is difficult and may lead to incorrect information, if it is not done very carefully. A compromise is to do automatic information gathering, but leave the final decision, which information to use, to an operator."

- Synchronous vs. Asynchronous Processing Correlation Engines

  "Event correlation can be done either real-time with the incoming data, or offline with stored data." For example, in case of event correlation for log monitoring, real-time event correlation is needed; real-time correlation does not necessarily imply that the events must be forwarded immediately but that they must be processed in real-time. Offline event analysis on the other hand may be useful to find event patterns in large amounts of data. "

- Stateless vs. Stateful Correlation Engines

  "A real-time correlation engine can be stateful (i.e. it has a memory of the event history), or stateless (without any memory). Obviously, a purely stateless correlation engine is very limited, as an incoming event cannot be related with older events. A completely stateless correlation engine is limited to filtering the events according to predefined rules.

  On the other hand, stateless operations are usually very simple and fast, and may be useful to handle events at the input. A typical correlation engine is therefore usually stateful, but allows both stateless and stateful operations."

- Centralized vs. Distributed Correlation Engines

  "The fact that the events which need to be correlated are usually generated by distributed sources, suggests itself that the correlation should also be done in a distributed

fashion. The obvious advantage of distributed processing is better performance and scalability, and easy access to additional information from the source.

On the other hand, a centralized approach is better suited to find correlation between events from different sources. Additionally, a central solution is easier to manage and requires less Operating System (OS) independency. As a compromise, a possible solution is a system, where the events are pre-processed (e.g. filtered and compressed) on the sources, and then correlated centrally in a second step."

Reporting

After collecting, aggregating, normalizing and correlating data or events, a SIEM system must be able to represent the results produced from the above mentioned procedures. It is therefore obvious that, at first stage, a SIEM environment should support an automated generation of reports and at a later stage make the processes of defining, generating and exporting reports as versatile and user-friendly as possible. In that direction, although SIEM environments come with a set of pre-defined reports, non-standard reports can also be created using the SIEM application combined with some "point-and-click" add-ons. By feeding a SIEM system with the right information and by applying the suitable transformations, clear and useful reports will be produced. When it comes to reporting, these transformations usually focus on the business rules that should be satisfied and the authorization needed to access these reports and data.

The practice of business rules and the application of them within this methodology has several advantages. *"First, it helps to define which are the actual issues of such an implementation. SIEM adds no extra security to an IT landscape, it adds knowledge about that landscape."* [11] For example it can be used to assess the validity of rules configured, related to actions logged and log entries created therefore. *"Second, business rules have a rigid structure, that can help technical and non-technical people alike to define what they will use SIEM for."* [11]

Taking reporting requirements one step further, authorization concepts may become necessary in order to define who is allowed to access the aforementioned reports and data. For that purpose fine grained tools have been developed and used by SIEM systems. By using these tools the "need-to-know" privilege is applied inside an organization as there are only specific people/group of people who can access these reports and data. Furthermore, drilling down can be achieved through interactive reports; by filtering the data the report is based on, more details can be revealed allowing though, only authorized personnel to reach the original log lines the report was based on.

Alerting

Another feature that a SIEM system should possess is the ability to trigger notifications or alerts to operators or managers, when something unusual or unexpected occurs. Alerts may be based on both pre-established and custom alert triggers, while alerting mechanisms can include e-mail, SMS, or even SNMP messages. [2] Categorization of alerts, based on their criticality, can also be implemented speeding up decision making and incident handling processes. For example, in cases of high priority alerts, the most likely notification method could be an SMS or a text message to a cell phone, while in cases of lower severity

alerts, notification could occur via email or through a simple data point that would appear on a report.

Compliance

Finally, a SIEM environment apart from effectively managing security operations, should also be able to demonstrate organization's compliance to regulatory and government requirements. Regulatory and government needs include national and international rules and laws that businesses must show adherence with or industry specific rules and standards that companies must comply with. Additionally, internal policies and requirements must also be examined when compliance is a concern. As stated in [14] "*Compliance, specifically with regards to management of an organization's log retention requirements, remains the most common justification for new SIEM implementations. For industries with compliance requirements, particularly those with geographical restrictions on data movement, the location of log data should be verified to ensure it does not invalidate compliancy.*"

Today, companies and organizations in order to be able to demonstrate continuous compliance with laws and rules, recur to SIEM systems, because they find it less complex and less expensive to invest in such a solution. As cited in [15], "*While the majority of Governance, Risk and Compliance solutions tend to monitor policies and compliance processes via attestation, SIEMs complement GRC by monitoring infrastructure controls and supporting compliance orders*." This demands that companies and organizations have previously understood their applicable industry, regulatory and legal obligations for security and risk management, and that SIEM systems can meet compliance standards and provide automation in compliance monitoring. A SIEM solution for example should automatically detect violations in configuration changes, identity management, resource access and/or event log retention and provide reports on control performance. Automated notification of control violations can provide an early warning system that can minimize the impact of violations, while automated reporting can be used across the IT organization to identify opportunities where strengthening controls can reinforce defense.

3.2 Architecture

Although many variations exist in SIEM marketplace today, there have been identified some specific, yet common, parts, that a SIEM system can be broken down into, in order to provide the capabilities described in the previous chapter. The minimum parts required from a SIEM system to promptly operate, as depicted in the following diagram are: a source device, a log collector, a log parser, a correlation engine, a log storage and a monitoring mechanism.

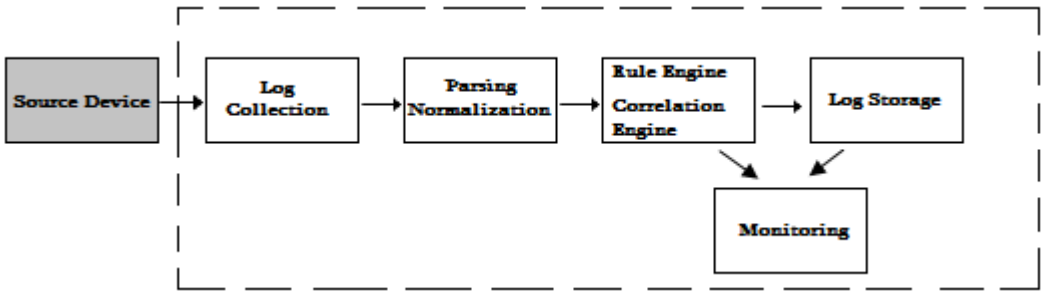


Fig 5.1 The Anatomy Of A SIEM

The very first part of a SIEM infrastructure is the source device that feeds the appropriate information into the SIEM system itself. Although many people do not consider the source device as an actual part of the SIEM, it is indeed a vital piece of the overall SIEM process. A source device can be a user's computer, a server, a network or a security device, or any other application that may be used to retrieve logs from. Without the source device and the information that these devices generate, the remaining SIEM components would be of no use.

Last thing to do before start implementing a SIEM system, is defining the sources that will be used for log retrieval. Having defined these sources, will make deployment much easier when moving down the line and configuring the remaining pieces. Once these sources are defined, it should be determined what logs must be retrieved, as not all logs are needed from all devices. Finally the approximate amount of SIEM resources that shall be devoted to the processing and storing of these logs shall be estimated.

The very next component inside a SIEM infrastructure is the log collector, which is responsible for getting all the appropriate logs from their native devices to the SIEM system itself. Log retrieval mechanisms may vary, depending on the SIEM's version being used, but in general two fundamental collection methods exist: the pull and the push method. In the following lines the advantages and the weaknesses of these methods are described.

A common method to enforce push log collection is through syslog. The advantages of that method is its ease of setup and configuration, as the only thing needed is the configuration of syslog in source and server devices. In such a configuration, there are also some disadvantages. For example using UDP syslog, means that it cannot be verified if the packets reached their destination, due to UDP's connectionless nature. Another security issue could arise if no proper access controls are implemented on receiver's side. In that way, a misconfigured system or a malicious user could flood SIEM with false information, making it harder to focus on real events.

Contrary to push method, in which the source device starts the interaction with the actual SIEM system, when pull log collection is implemented, SIEM initiates the connection with the source

device and actively retrieves the logs. The pull procedure usually runs at predefined intervals, with the logs coming into the SIEM at certain time-periods and not in real-time, as it occurs with the push log collection procedure. Such a technique could be used in cases where logs are stored in flat text files on a network share, requiring from the SIEM to have previously established a connection to that share, so as to be able to read the logs from the source devices. The key advantage of that method is that logs from fake devices cannot be transmitted as the beginning of the process is handled by the SIEM system and the identity of the source device is verified by the system itself.

Finally, mixed log collection can be used, especially in environments comprising of multi-type devices that may require both these two methods of log collection.

Having collected all the appropriate logs, from the multiple devices and applications, on a central log repository, the next thing to do is to reformat (normalize) them into a single standard output that is usable and acceptable from the SIEM system. That action is handled from the parsers. A parser may handle the act of normalization in different ways, but in the end all the logs, no matter what type of device or manufacturer, will look like the same in the SIEM. Event/Log normalization plays an integral role for the success of the SIEM project, as it can simplify the reading of these logs and also provide a common and standardized format for rule generation.

After the normalization process, the correlation/rule engine takes action. That engine is always known as the brain of the SIEM, as it is responsible for alert-triggering when specific conditions are met or identified through log files. The correlation engine, which is actually a subset of the rule engine, is responsible for matching multiple events from different sources into a single correlated event, so as to simplify incident response. In that way, same events are grouped and represented as one, helping operators to avoid diving through multiple logs and different events to identify a threat and start troubleshooting. Operators can do so by simply monitoring the console, where the actual correlated data/events are depicted.

At a more abstract layer, the rule engine is responsible for dealing with those correlated events and handling them appropriately. The rule engine uses a simple Boolean logic, but sometimes it proves to be extremely complex, especially when a large number of conditions should be checked, in order to decide whether or not an alert should be triggered. The next 2 figures describe a simple scenario, where a rule for alerting was established to prevent a possible brute-force compromise.

Time	Event Number	Source	Destination	Event
10:10:01 CST	1035	192.168.1.200	10.10.10.25	Failed login to server
10:10:02 CST	1036	192.168.1.90	10.10.10.21	Successful login to server
10:10:03 CST	1037	192.168.1.200	10.10.10.25	Failed login to server
10:10:04 CST	1038	192.168.1.91	10.10.10.35	Failed login to server
10:10:05 CST	1039	192.168.1.10	10.10.10.2	Successful login to server
10:10:06 CST	1040	192.168.1.10	10.10.10.3	Successful login to server
10:10:07 CST	1041	192.168.1.200	10.10.10.25	Failed login to server
10:10:08 CST	1042	10.10.10.54	192.168.1.201	Failed login to server
10:10:09 CST	1043	10.10.10.34	192.168.1.10	Failed login to server
10:10:10 CST	1045	192.168.1.200	10.10.10.25	Successful login to server

Fig 5.2 Normalized Login Event Information

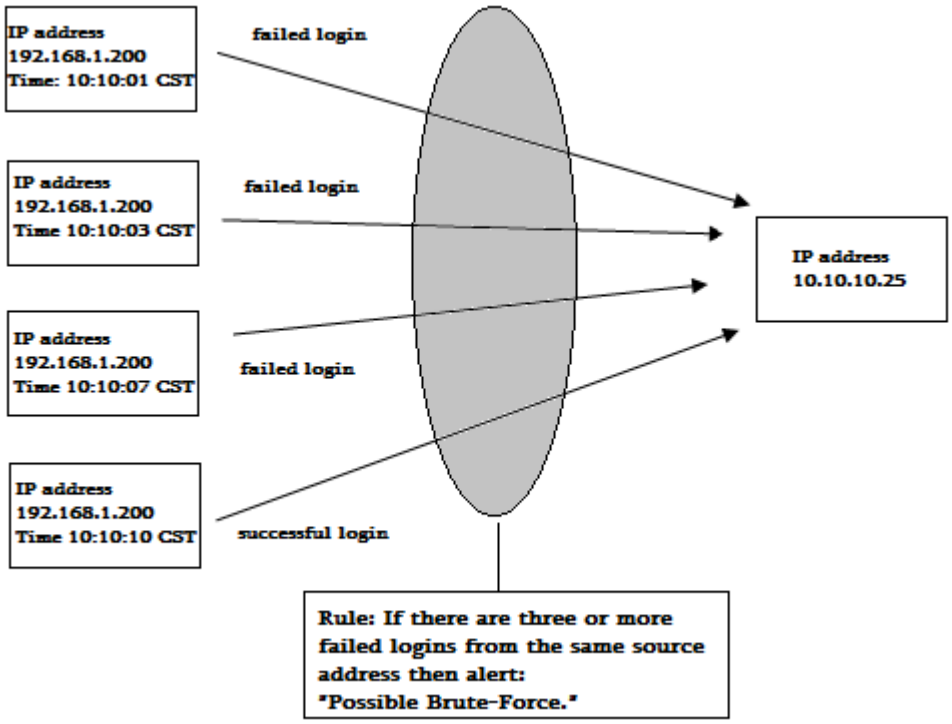


Fig 5.3 Rule Definition & Alert Triggering Example

Another element that should be present in SIEM infrastructure is a log repository. SIEM systems usually have to cope with large volumes of data. Thus adequate solutions shall be provided, when it comes to log storage.

The most common among the 3 alternatives, that SIEM systems provide to store their logs, are within a database. This alternative allows for easy interaction and retrieval of the stored data, as well as for optimized performance when it comes to accessing that data. However, a few issues may arise depending on how the SIEM implements its database. For example if SIEM is not running as an appliance but it is a hardware component, administration of the database will be required and a qualified DB administrator will be needed.

Another alternative, is the storage of the information in flat text files. Despite the fact that these files use human readable format with the help of some delimiter characters, they are not used very often, due to scalability and performance (reading-writing speed) issues especially within large environments. Among the advantages of that method are the ease for external applications to access this data and the flexibility for analysts to search through these files (grep – reg.ex).

Finally, a binary formatted file can be used. These files are exclusively used by the SIEM systems, as it is only them who know how to read and write to these highly proprietary files.

Last but not least, a SIEM infrastructure must be in possession of a monitoring tool. Such a tool can be either web-based or application-based and can be used for managing and visualizing the information emerged from the previous mentioned components. That component, as depicted in figure_5.1, interacts with both the log storage and correlation engine, in order to provide a unique and complete view of the SIEM environment, and allow system engineers or incident handlers to process the data stored efficiently and effectively. SIEM's management and monitoring console can be also beneficial, when it comes to developing the content and the rules that will be used to pull out the information from the events being processed.

As it can be seen, a SIEM is composed of many parts, each doing a separate job. Although, each of these systems can run independently of the others, if any of them is not running in unison with the others, the effectiveness of the SIEM system is under serious doubt.

3.3 Operational Considerations & Best Practices

Having defined the basic capabilities of a SIEM system, the general requirements it should meet prior to implementation and its generic anatomy, we can move on proposing some SIEM practices, that when applied, they can help enterprises gain more assured value from SIEM solutions. These practices require that most of the dealings have been previously settled between company's and vendor's side and can be proved extremely useful when having to choose between two or more solutions.

Activation and deployment

In order to successfully deploy a SIEM system the identification and activation of device's auditing functions are among the top priorities. Once respective device targets, event log capture, alerting and reporting requirements have been identified and documented, the process of expanding the scope of SIEM system can be simplified. At the same time, the best means to achieve SIEM's deployment success is through a "step-by-step" approach, starting with the most security and business critical devices that should be monitored or those that address a specific set of critical controls and moving on with less significant and relatively minor operational controls. "*The better the documentation, change management processes and initial preparation of the environment prior to SIEM deployment, the easier the implementation and ongoing maintenance of the SIEM will be.*" [15, 16]

As cited in [15], SIEM deployment, scheduling and cost management, should consider:

- Procurement and overall deployment costs
- Delivery mechanisms such as software, hardware or virtual appliance
- Necessary operating equipment (hardware, servers, storage)
- Installation and the scope of coverage (site, multiple sites, multiple divisions)
- Personnel resource requirements (responsibilities, training and support)
- Monitoring frequency (real-time or historic), by devices class or criticality of IT function

Integrity verification

Another field that SIEM solutions may become useful, is that of assuring and maintaining infrastructure's integrity. Given that corporate operating requirements evolve and vendors update their systems to address known issues, configuration changes are inevitable. Unauthorized, accidental or malicious configuration changes are among the major reasons of performance degradation, data corruption and other security issues.

Many companies and organizations therefore rely on well defined change management processes that provide the policies and procedures for provisioning,

documentation change review and maintenance. By setting and enforcing configuration and maintenance standards service reliability can be ensured, as standardization eases provisioning and management, while at the same time reduces operational faults and improves fault triage.

As described in [15,16], SIEM systems can help in that direction, as they can provide a variety of monitoring, alerting and reporting mechanisms to support maintenance of infrastructure integrity. Either directly or through data integration after collaborating with other IT management systems, they can be used to ensure that several procedures, such as identification of unauthorized changes or non-compliant systems and verification of approved changes, are properly handled. Patch management or vulnerability management solutions, for example, can feed SIEMs with operational details such as configuration changes, patches and vulnerabilities updates, supporting security operations and incident response. Additionally, SIEMs with virtualization intelligence can help monitoring virtualized environments, bridge physical and virtual environments and enable means to reduce potential security and compliance issues, through information cross-correlation from all layers in the network stack. While, at the very end, they can provide integrity checking on operational resilience measures. Such measures are typically related to disaster recovery and backup functions and SIEMs can be used to verify that replication processes are properly followed, while highlighting unscheduled and atypical processes requests.

Enhance Perimeter Defense

Two other issues related-to SIEM operations that should be considered are the monitoring and reporting capabilities on activities associated with perimeter controls. [16,17] Boundary devices such as firewalls, routers and other means of network-based access controls remain vital to defend against unauthorized access to network resources and mitigate risks and threats. As obvious plenty of boundaries can be met inside an organization; boundaries between users and systems, remote and internal users, business partners and extranets, wireless access points and corporate network.

SIEMs can be used to consolidate the monitoring of network flow information, provided from different perimeter devices. Such information may contain details about source and destination addresses, ports and amount of data and can be proved vital for incident handling purposes. Additionally, SIEMs can cross-correlate network flow information with other operational data to identify suspicious behavior and potential security threats. For example they can aggregate alerting from firewalls and IDS/IPS, conduct event consolidation on similar alerts, which can consume huge resources, and facilitate incident management.

Enhance Core Defense

Beyond perimeter defense, a SIEM solution should provide all the appropriate means to enhance defense on host and application basis, protecting against service unavailability and fraud and data privacy issues. [16,17]

Among the most popular tools for end-point security are anti-virus and anti-malware software, with the majority of information security compliance frameworks specify

employing such tools. Best practices define that these tools should be tiered from end point to perimeter and that enterprises should routinely monitor for critical anti-virus/malware issues. SIEMs provide many ways for malware and virus monitoring and incident response. One practice, among them, is through correlating event log data from anti-malware/anti-virus management systems, after having previously configured all endpoints to focus on detection-only, rather than remediation on like issues. SIEMs apart from correlation, can also facilitate processes to quarantine and remediate.

When it comes to application security, SIEM systems can perform application platform, resource, database activity and web application defenses monitoring. In terms of application platform defense, SIEMs can be used to maintain platform's operational state by monitoring for newly installed and running applications, unusual application process or processes that consume high amounts of resources and other techniques. In cases where database defense measures are considered, "*SIEMs have the means to obtain, offload and utilize database audit logs, which are in the form of database tables, to complement overall security monitoring.*"[16] SIEMs may also support additional database security measures such as database firewall technologies and database security applications that employ threat protection and auditing mechanisms, vulnerability scanning etc. Finally, SIEMs can be used in conjunction with conventional devices (firewalls, application firewalls) to aggregate and cross-correlate vulnerability and attack events to support incident response and strengthen web application defenses.

Limit False Positives

Last but not least, SIEM systems can be used to tag and tune out random noise and false positives – a detection of an attack that is actually benign. [16,17] In some cases, IDSs due to broad IDS rules or statistical profiling, may improperly identify legitimate actions as attacks and burden security staff with false alarms. For example an IDS may report a known Windows attack against a Linux system or a known attack against a patched system. SIEM systems can alleviate this condition and respective administrative burden via event correlation and exception management, as it can identify false positive conditions which are comprised of: (i) attacks against invalid systems, (ii) attacks against systems that are patched and no longer vulnerable, or (iii) attacks that are non-threats such as scheduled vulnerability scans.

Common techniques to refine IDS rules, test SIEM implementation and assess incident management processes are through offending traffic generation that will cause alert triggering. "*Organizations can generate IDS alerts by either using traffic replay and test mechanisms within the IDS or by employing IDS test tools such as Tomahawk and Metasploit*." [16,17] At the very end, for other scenarios, such as scheduled scans or penetration tests, that may equally identified as attacks, SIEM systems are well positioned to handle exception management through ad-hoc configuration.

Chapter 4. Existing SIEM Solutions

Numerous solutions with different characteristics exist in SIEM marketplace today. A list of the most notable among them, as well as Gartner's "Magic Quadrant for Security Information & Event Management Systems" for the year of 2013 can be found in Appendix A. In the magic quadrant SIEM solutions are split into 4 categories (leaders – challengers – visionaries and niche players).

In SIEM Leaders quadrant they belong vendors such as ArcSight, Splunk and LogRhythm, that *"provide products that are a good functional match to general market requirements,"*(see 3.1) *"have been the most successful in building an installed base and revenue stream within the SIEM market, and have a relatively high viability rating."*[18] Apart from providing a technology that matches current customer requirements, leaders can also show evidence of superior vision and execution for anticipated requirements. *"Leaders typically have relatively high market share and/or strong revenue growth, and have demonstrated positive customer feedback for effective SIEM capabilities and related service and support"*.[18]

The Challengers quadrant is composed of vendors such as LogLogic, RSA and Novell that have a large revenue stream, at least a medium-size SIEM customer base and products that meet a large part of the general market requirements. A great number of Challenger vendors use their SIEM solutions as an extension of related security and operations technologies. *"Companies in this quadrant typically have strong execution capabilities, as evidenced by financial resources, a significant sales and brand presence garnered from the company as a whole, or other factors, but they have not demonstrated as rich a capability or track record for their SIEM technologies as vendors in the Leaders quadrant have"*. [18]

Vendors characterized as visionaries *"provide products that are a good functional match to general SIEM market requirements, but have a lower Ability to Execute rating than the Leaders. This lower rating is typically because of a smaller presence in the SIEM market than the Leaders, as measured by installed base or revenue size or growth, or by smaller overall company size or general viability"*.[18] Among the vendors contained in that category are Sensage and Trustware.

Finally, the Niche Players quadrant is composed primarily of *"smaller vendors that are regional in focus, or provide SIEM technology that is a good match to a specific SIEM use case, a subset of SIEM market requirements."* The most famous among them is the Alienvault. *"Niche Players focus on a particular segment of the client base or a more limited product set. Their ability to outperform or innovate may be affected by this narrow focus. Vendors in this quadrant may have a small or declining installed base, or be limited, according to Gartner's criteria, by a number of factors. These factors may include limited investments or capabilities, a geographically limited footprint, or other inhibitors to providing a broader set of capabilities to enterprises now and during the 12-month planning horizon."*[18]

Chapter 5. Drawbacks & Insufficiencies

SIEM technology systems are often labeled as a blooming business in our days. From what is mentioned above, we can argue that, as an incident handling tool, a SIEM can be highly effective at expanding visibility to network threats and increasing decision makers' ability to properly identify and handle large numbers of security events. However, SIEM tools by no means are a silver bullet for identifying all attacks. As any other security tool does, they lack the essential ingredient of human intelligence; and tools are only as effective as the people using them. In the following lines, the main inabilities and insufficiencies of SIEM systems will be described.

Among the most important issues that a SIEM has to solve is finding and using the right balance in data collection, storage and analysis. In the question of "how much and what data should a SIEM collect?" no clear answer exists. Due to organizations' physical constraints (storage, hardware, personnel) and finite resources (operating costs) the volume of data to be collected and managed is limited. In cases like these, data should be handled with the "less is best" principle in mind. But then the question changes to "how few is enough?". To overcome the challenge of data overload, the relativity among the data being processed and its applicability with SIEM's initial scope and focus should be fulfilled .

Another major problem in the SIEM space is the difficulty in consistently analyzing event data. Every vendor, and indeed in many cases, different products from the same vendor, uses a different proprietary event data format and delivery method; even in cases where a "standard" is used for some part of the chain, like Syslog, no adequate documentation exists. As the standards don't typically contain enough guidance to assist developers in how to generate events, administrators in how to gather them correctly and reliably, and operators in how to analyze them effectively, additional overhead and complexity is introduced in the various actions required to properly collect all that data and handle them within the SIEM environment.

Additionally, as already described in previous chapters, SIEMs have to undertake the difficult task of resolving false positives, which are among the most costly headache for enterprises. In order to lower that rate and operate effectively, SIEM systems require pre-deployment and integration with several security devices. The data related with these devices as well as the correlation between them are of equal significance to SIEM's optimum effectiveness, not only during identification phase but also during incident handling phase. A tuning procedure can also occur post deployment to weed out false positive and meaningless data. This is why, fine-tuning SIEM and striking a balance between simplicity and effectiveness of controls and alerts should be handled as an on-going, rather than as an one-way, process. The following diagram depicts the steps needed to move from complete log ignorance to near-real-time security monitoring. "*The trend here is from being ignorant, to being slowly reactive, to being quickly reactive, to eventually being proactive and aware of what is going on across the IT environment.*" [19]

Fig 6.1 Maturity Curve of Log Management [19]

Further to the above, the ever increasing pace of security threats and complexity of regulatory environments are two additional limiting factors for SIEM systems. The number of threats faced by well known firms and organizations are constantly increasing, while the strategies and methods being used are becoming more and more sophisticated, demanding continual awareness and on-going assessment and evaluation of the controls in use. At the same time the impact of new regulations and compliance requirements may render existing SIEM solutions less competitive or obsolete and in need of reevaluation, updating or even replacement.

One more issue, concerning SIEM Infrastructure, is that it presents a valuable target to attackers, as the information contained in a SIEM system can become a magnet for any of them. The data collected by a SIEM reveals a great deal about the weaknesses and vulnerabilities of an enterprise and can be used to determine the best method of attacking a network. At the same time, since it is known that SIEMs are usually the central repositories of notification around attackers' activities upon their victims, the attackers will make efforts to ensure that their activities will not register in SIEM systems wherever possible. Therefore, SIEMs must be protected just as any other high-value system, if not more – Limiting their exposure and accessibility from the rest of the infrastructure to the absolute minimum and minimizing the learning surface from the rest of the systems inside the network, should be among the first-line treatments of security personnel, since the more an attacker can deduce of a SIEM deployment, the easier it becomes for him to evade it.

Having in mind the fact that a SIEM acts as a central repository of all security data inside an organization, an ethical issue rises. Correlation and retention of incidents can be considered an intrusion into the privacy of the people ,who's data is being collected; especially now with the state of IP addresses as personally identifiable information (US) or personal data (Europe). As stated in [11], raw log events may contain a lot of sensitive information, such as names, IP addresses or other records and despite the fact that most SIEM environments provide a one-way hashing method or encryption techniques to hide this information and protect against disclosure of the actual private data, log files can still be correlated. In that way, "*SIEM achieves its primary goal of tracking security*

incidents across the network, but it can also reveal which employees browsed social networking sites. Although that can be very useful information for manages and HR employees alike, it can be regarded an invasion of their privacy". [11]

Another important consideration to make as well is that SIEM environments are shipped with predefined reports, depending on the scope and requirements. That means that those predefined reports may be incomplete and / or inaccurate, depending on what input they get, rendering SIEM systems incapable of anticipating missing devices or information. Requirements concerning event prioritization and reporting capabilities depend on the unique characteristics of an organization and the environment in which it operates. Thus, in order to be most effective, a SIEM must take into account the unique business processes that exist inside an organization. For example monetary transactions might be the focus on a financial and retail enterprises, while sensitive information on national security matters might be the highest concern to military environments.

Finally, as many security managers and IT professionals state, there is a distinct lack of actionable security provided, and threat intelligence implemented on these systems, leaving the network in a vulnerable state. To overcome these impediments a SIEM environment needs an attack detection system to supplement itself and provide an efficient security monitoring solution. In order to achieve this, great customization is required to perform suitability, which is proving increasingly difficult due to the lack of security experts available to fine tune it.

Chapter 6. SIEM Implementation

As already described in chapter 4, SIEM market today is a very crowded and complex place. Not only there are many solutions out there, but since one of the major components of a SIEM is third party device/application integration, there are major differences in product support between the various solutions.

In this chapter, an open source SIEM implementation for a SOHO (Small Office Home Office) network is proposed, similar to the one described in [20]. Our proposal is based on Splunk and OSSEC (), 2 open source tools that when combined, they can provide a basic, free and worthy SIEM solution.

OSSEC [21] is an Open Source Host-based Intrusion Detection System. It runs on most operating systems, including Linux, OpenBSD, FreeBSD, Mac OS X, Solaris, and Windows and provides the following core capabilities:

- Log Monitoring and Collection
- File Integrity Checking
- Windows Registry Integrity Checking
- Active Response

OSSEC works on a server-client model. The server must be a Unix machine, while the clients, which OSSEC calls them "agents" can be of any operating system type. For Windows machines, however, an agent installation is the only option.

Splunk [22] searches, indexes and analyzes machine-generated data including logs, config files, messages and alerts, from applications, systems, and other IT infrastructure devices. It also correlates real-time data in a searchable repository from which it can generate graphs, reports, alerts, dashboards, and visualizations. Among its key features are its scalability and efficiency in processing "big data".

The first step in using Splunk is feeding it with data. Once Splunk gets some data, it immediately indexes it, so that it's available for searching. Splunk can handle almost any type of data and from any kind of sources. [23] Considering the latter, data can be either on the same machine as the Splunk indexer (local data), or it can be on another machine (remote data). Remote data can be, respectively, collected into Splunk, either by using network feeds or by installing Splunk forwarders on the machines where the data originates.

In our example, however, an alternative solution is chosen for implementation. Splunk located in the same machine with OSSEC server, will be used to visualize the data collected from OSSEC agents. The most basic steps through that implementation are described in the following pages.

Once network interfaces on our machines are properly configured, we can move on with SIEM implementation. An architecture overview is depicted in the scheme below, in order to provide the readers with a better understanding of our SOHO network. As it can be seen, our "lab" consists of 2 machines with 3 O/S: 1 Windows 7 and 2 Ubuntu 12.04. The first Ubuntu machine will be used to host OSSEC-Splunk server, Windows-based machine will be used as a web-server (Wamp platform) and the other Ubuntu-based machine (VM) will be used to emulate a desktop for everyday

operations. Ubuntu-based VM and Windows-based machine will have OSSEC agents installed. A diagram of our SOHO network is provided below.

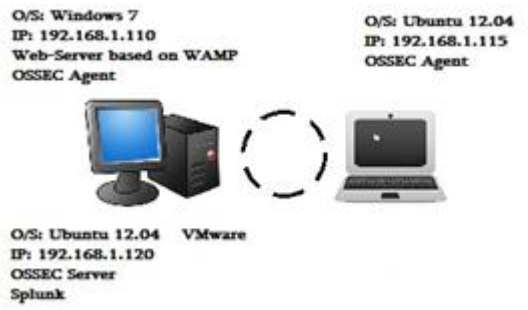


Fig 5.1 Network Diagram

@ 192.168.1.120

In this machine OSSEC server and Splunk will be installed. After downloading ossec-hids-2.7.1.tar.gz and splunk-6.0-182037-Linux-i686.gz from OSSEC's and Splunk's official sites [21,22], we can start the installation process.

As a next step, we should integrate OSSEC with Splunk. Splunkbase [24] will be used therefore, which is a repository of Splunk applications, that among other it also contains an OSSEC application. OSSEC for Splunk [25] contains parsing logic, saved searches, canned reports, and dashboards and as long as OSSEC is installed in the default path, it will automatically configure Splunk to pull in the OSSEC logs and alerts. Having downloaded and installed the above application, we should restart it, in order to start using it.

After restarting Splunk's, we can see a new tab appearing in the left upper corner "App: Splunk for OSSEC". If we expand that tab, choose Splunk for OSSEC and then data summary we can see all the events recorded from the predefined OSSEC agents and server. Monitoring optimization and event's evaluation is of great importance in order to narrow the width of potential irrelevant events and focus on these that really matter and more information on how to achieve that can be found in [23].

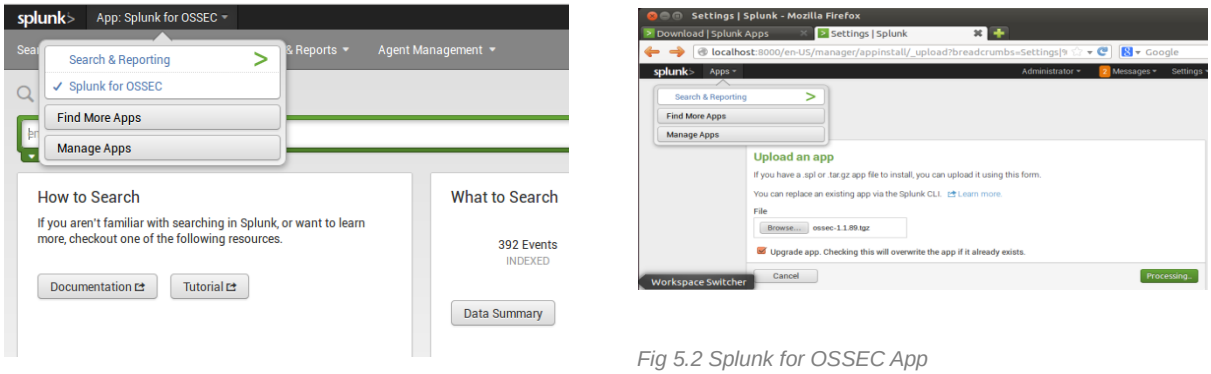


Fig 5.2 Splunk for OSSEC App

@ 192.168.1.110

In this machine OSSEC agent will be installed. As it is a Windows machine, the appropriate file should be downloaded. ossec-agent-win32-2.7.1. The windows installation is pretty

straightforward, since it requires only the OSSEC Server IP and the Authentication Key to function properly. That key is obtained from OSSEC Server through the key extraction process.

As with server's configuration file, agent's configuration file can also be viewed or modified. This can happen by choosing View Config, as shown in the next figure.

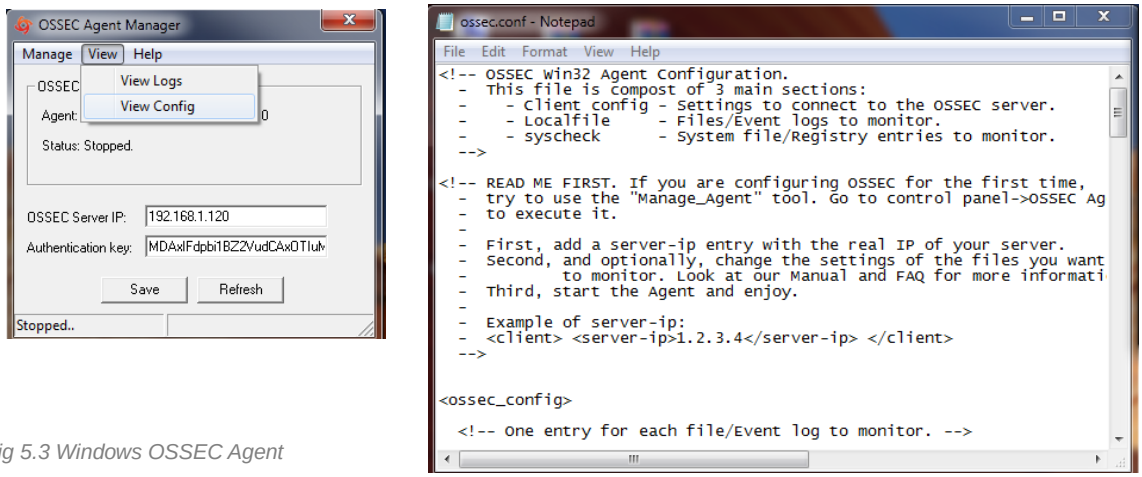


Fig 5.3 Windows OSSEC Agent

@ 192.168.1.115

In our last host, the agent version of OSSEC will be installed with the help of ossec-hids-2.7.1.tar.gz. The same steps, as with OSSEC server installation, are followed except for choosing *agent* as the kind of installation and typing OSSEC server's IP when asked. By the time the installation is complete, the authentication key should be imported from the OSSEC server. This is done with the use of manage_agent tool..

Having our OSSEC Server, Splunk and OSSEC agents configured as described above, we are ending up in a fully operational SIEM system implementation. The next big challenge is OSSEC's ruleset creation and optimization in order to fit the different needs that may arise. Although no standard procedure and specific guidelines exist towards, as the scope and the framework of each implementation may differ depending on the occasion, in [26] some useful suggestions are provided. Some testing and results from our implementation are however available in Appendix C.

Chapter 7. Conclusions & Future Work

At this point, it is worth mentioning the reasons that led us to choose that specific implementation. First of all, Splunk is a cutting-edge, sophisticated tool to handle the threats, networks and hosts are exposed to today, which according to [18], is placed in Leader's category. That places it among the most successful vendors in building an installed base and revenue stream within the SIEM market, reflecting its good functional match to general market requirements and its proven evidence of superior vision and execution for anticipated requirements. Moreover, this solution is totally cost-free and this is of great benefit especially to SOHO environments, where budget constraints exist. Additionally, it is lightweight with low hardware requirements that can be run on home-based PCs and laptops. More information about minimum and recommended hardware requirements can be found in [27]. Among its strengths, is also its integration and compatibility with other well-known tools such as Cisco Security Suite, Snort and Nagios. While finally, is the usage of a host intrusion (HIDS) instead of a network intrusion detection system (NIDS), saving this project from inheriting network intrusion detection systems' impediments. As already described in [26], "*several very popular evasion techniques exist to bypass, or sidestep, the watchful eyes of a NIDS solution*." For example fragmentation attacks, session splicing, denial-of-service (DoS) attacks or even encrypted communications can be used to bypass a NIDS, rendering it useless. A HIDS, on the other side, is more efficient against these techniques offering better protection. That's because a HIDS is able to detect events on a server or a workstation and generate alerts similar to an NIDS, but is also able to inspect the full communications stream. In that way, "*NIDS evasion techniques, such as fragmentation attacks or session splicing, do not apply because the HIDS is able to inspect the fully recombined session as it is presented to the operating system. At the same time encrypted communications can be monitoring and HIDS signatures will still be able to match against common attacks and not be blinded by encryption, as HIDS inspection can look at the traffic before it is encrypted*." [26]

For future development this thesis would propose, the integration of Splunk with additional tools such as Snort and Cisco Security Suite. Having implemented more tools, the inter-correlation of the events could be studied and the efficiency of the SIEM could be further tested. Another interesting challenge would definitely be the mining of collected events to suggest new rules and to provide some metrics on how these rules optimize the event aggregation-correlation processes. Additionally, a comparative analysis between Splunk and other open-source SIEM solutions could be carried out, while as a final step, that solution could be tested in more complex environments and against more sophisticated techniques, similar to the ones referred above.

Appendix A – SIEM Vendors

ArcSight (HP)

LogLogic (formerly Exaprotect)

LogRhythm SIEM 2.0 Security Intelligence

NitroSecurity (McAfee)

Novell (NetIQ)

Qradar (Q1 Labs)

RSA (EMC)

SenSage

Splunk (Splunk)

SSIM & SEP (Symantec)

Tier-3 (Huntsman)

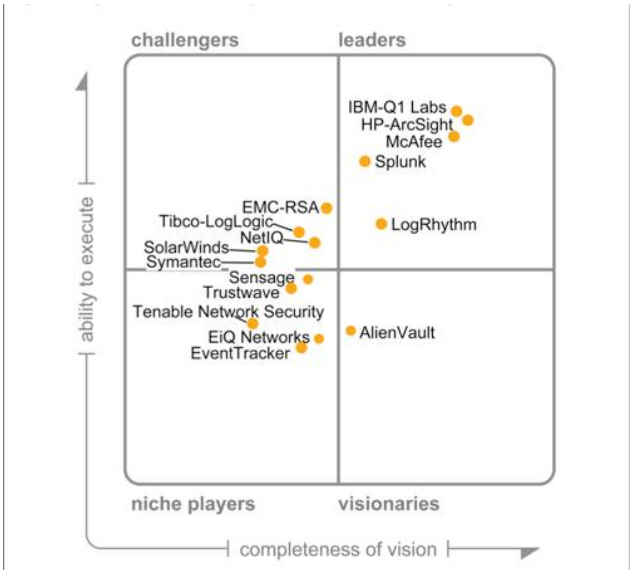


Fig A.1 Magic Quadrant For SIEM [18]

Appendix B – What - where to look for

A checklist for reviewing critical logs in case of a security breach, as well as some common guidelines in event log monitoring are provided below, according to [28].

Potential Security Log Sources

- Server and workstation operating system logs
- Application logs (e.g., web server, database server)
- Security tool logs (e.g., anti-virus, change detection, intrusion detection/prevention system)
- Outbound proxy logs and end-user application logs
- Remember to consider other, non-log sources for security events.

Typical Log Locations

- Linux OS and core applications: /var/log
- Windows OS and core applications: Windows Event Log (Security, System, Application)
- Network devices: usually logged via Syslog; some use proprietary locations and formats.

What to Look for on Linux

Successful user login	"Accepted password", "Accepted publickey", "session opened"
Failed user login	"authentication failure", "failed password"
User log-off	"session closed"
User account change or deletion	"password changed", "new user", "delete user"
Sudo actions	"sudo: … COMMAND=…" "FAILED su"
Service failure	"failed" or "failure"

What to Look for on Windows

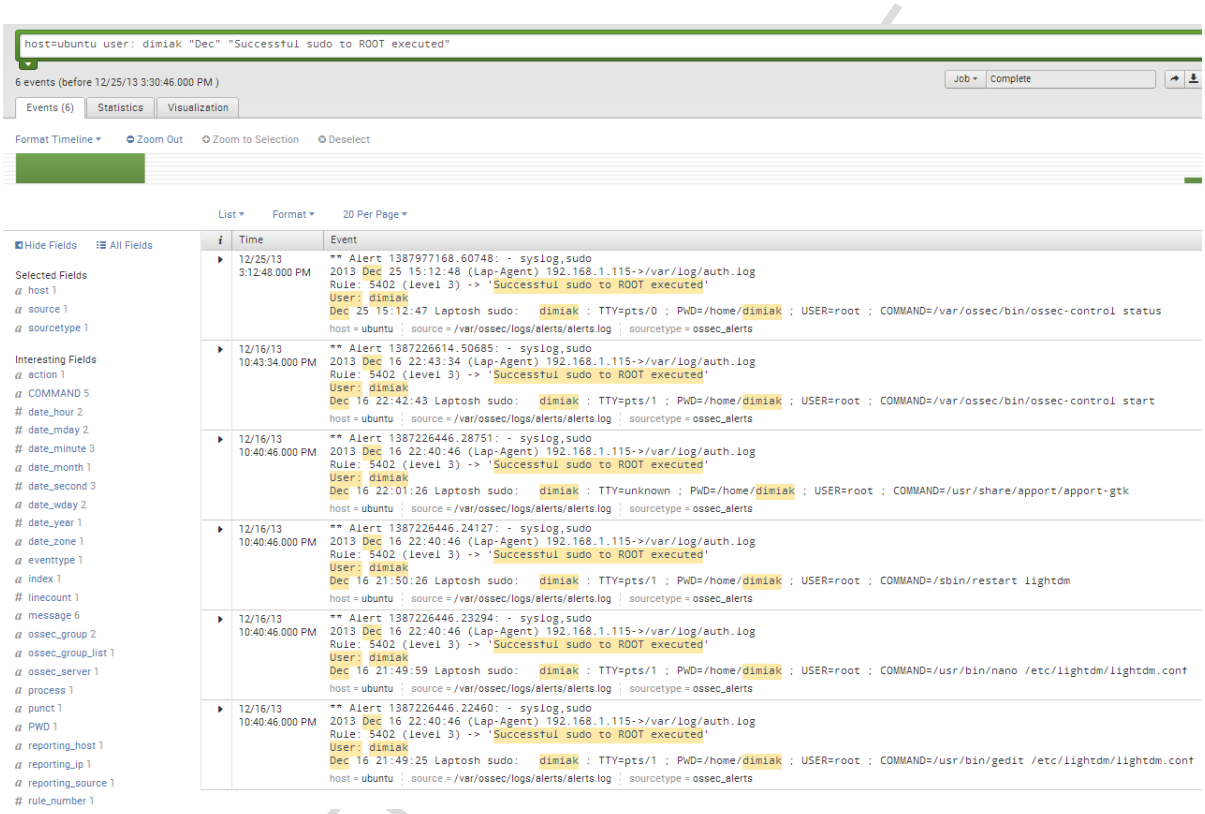
Event IDs are listed below for Windows 2000/XP. For Vista/7 security event ID = 4096 + Windows 2000/XP event ID. Most of the events below are in the Security log; many are only logged on the domain controller.

User logon/logoff events	Successful logon 528, 540; failed logon 529-537, 539; logoff 538, 551, etc

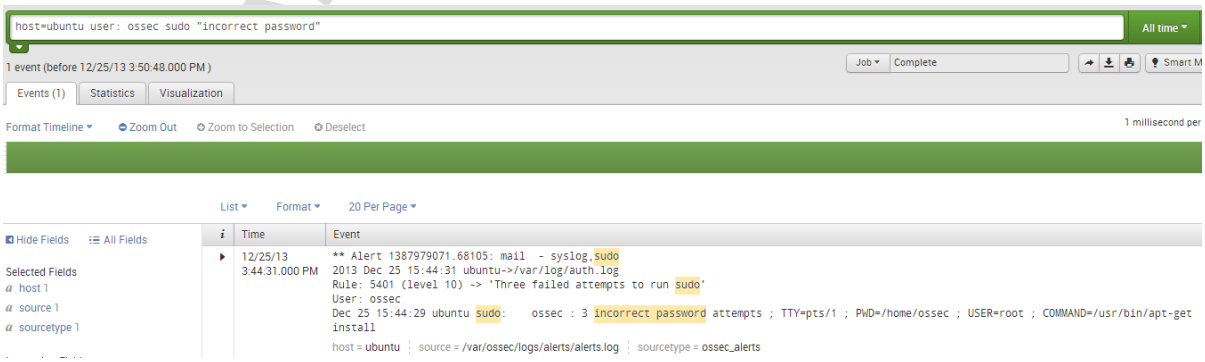
User account changes	Created 624; enabled 626; changed 642; disabled 629; deleted 630
Password changes	To self: 628; to others: 627
Service started or stopped	7035, 7036, etc.
Object access denied (if auditing enabled)	560, 567, etc

Appendix C – Use Cases

Suppose that we need to detect all the successfully executed **sudo to ROOT commands** from the user **dimiak**, during month **December** in order to investigate a potential resource misuse. In the following scheme, the query we should use and its output are depicted.



Respectively, for unsuccessfully executed sudo command for user ossec we should run the following.

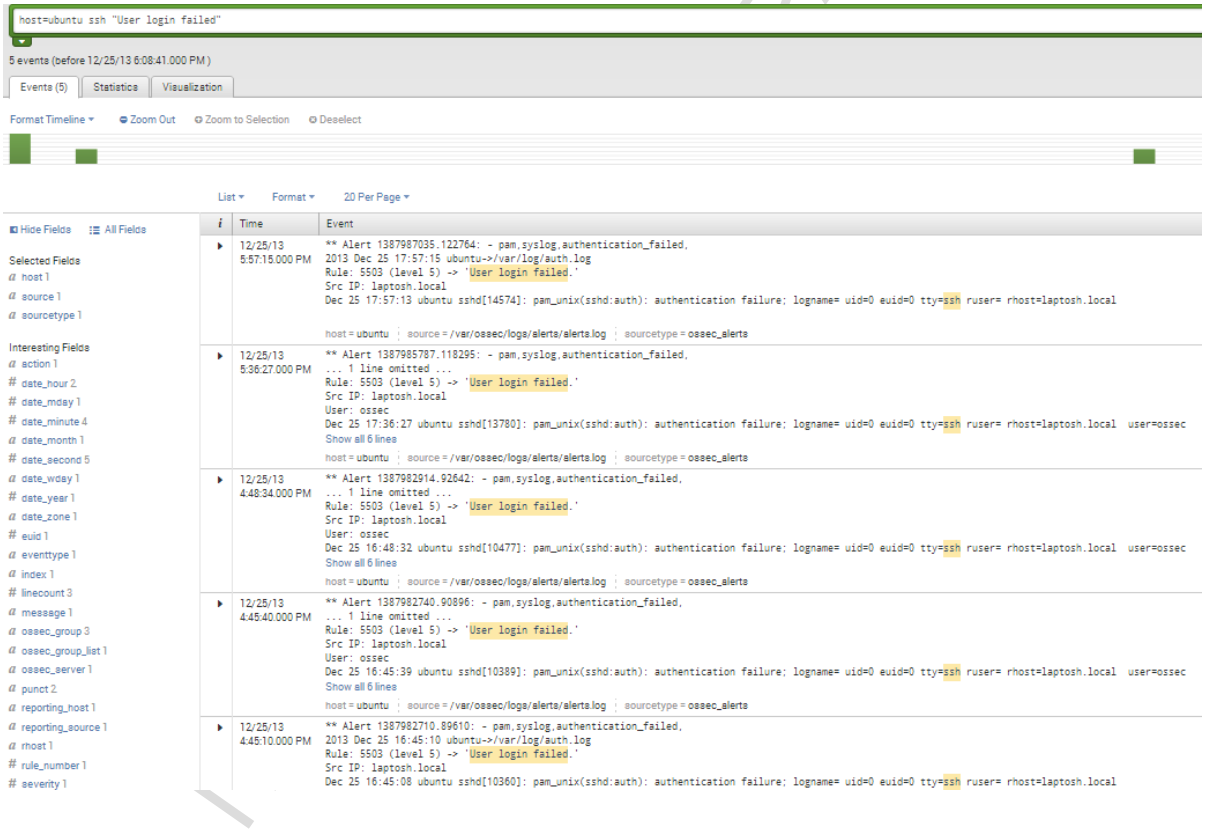


In another case we may have to deal with SSH login attempts. In order to be able to monitor SSH activity a different query shall be used. As it is known, all login attempts are logged to /var/log/auth.log and by default /var/log/auth.log is checked automatically from OSSEC.

In the picture below the failed logins, as logged in /var/log/auth.log, are depicted.

```
ossec@ubuntu:~$ grep sshd.\*Failed /var/log/auth.log
Dec 25 16:45:04 ubuntu sshd[10360]: Failed none for invalid user dimiak from 192.168.1.115 port 42908 ssh2
Dec 25 16:45:10 ubuntu sshd[10360]: Failed password for invalid user dimiak from 192.168.1.115 port 42908 ssh2
Dec 25 16:45:23 ubuntu sshd[10360]: Failed password for invalid user dimiak from 192.168.1.115 port 42908 ssh2
Dec 25 16:45:41 ubuntu sshd[10389]: Failed password for invalid user ossec from 192.168.1.115 port 42909 ssh2
Dec 25 16:48:34 ubuntu sshd[10477]: Failed password for invalid user ossec from 192.168.1.115 port 42914 ssh2
Dec 25 17:36:28 ubuntu sshd[13780]: Failed password for ossec from 192.168.1.115 port 43299 ssh2
Dec 25 17:57:16 ubuntu sshd[14574]: Failed password for invalid user dimiak from 192.168.1.115 port 43507 ssh2
Dec 25 17:57:20 ubuntu sshd[14574]: Failed password for invalid user dimiak from 192.168.1.115 port 43507 ssh2
Dec 25 17:57:22 ubuntu sshd[14574]: Failed password for invalid user dimiak from 192.168.1.115 port 43507 ssh2
```

while in the next picture, the output of "User Login Failed" string is shown.



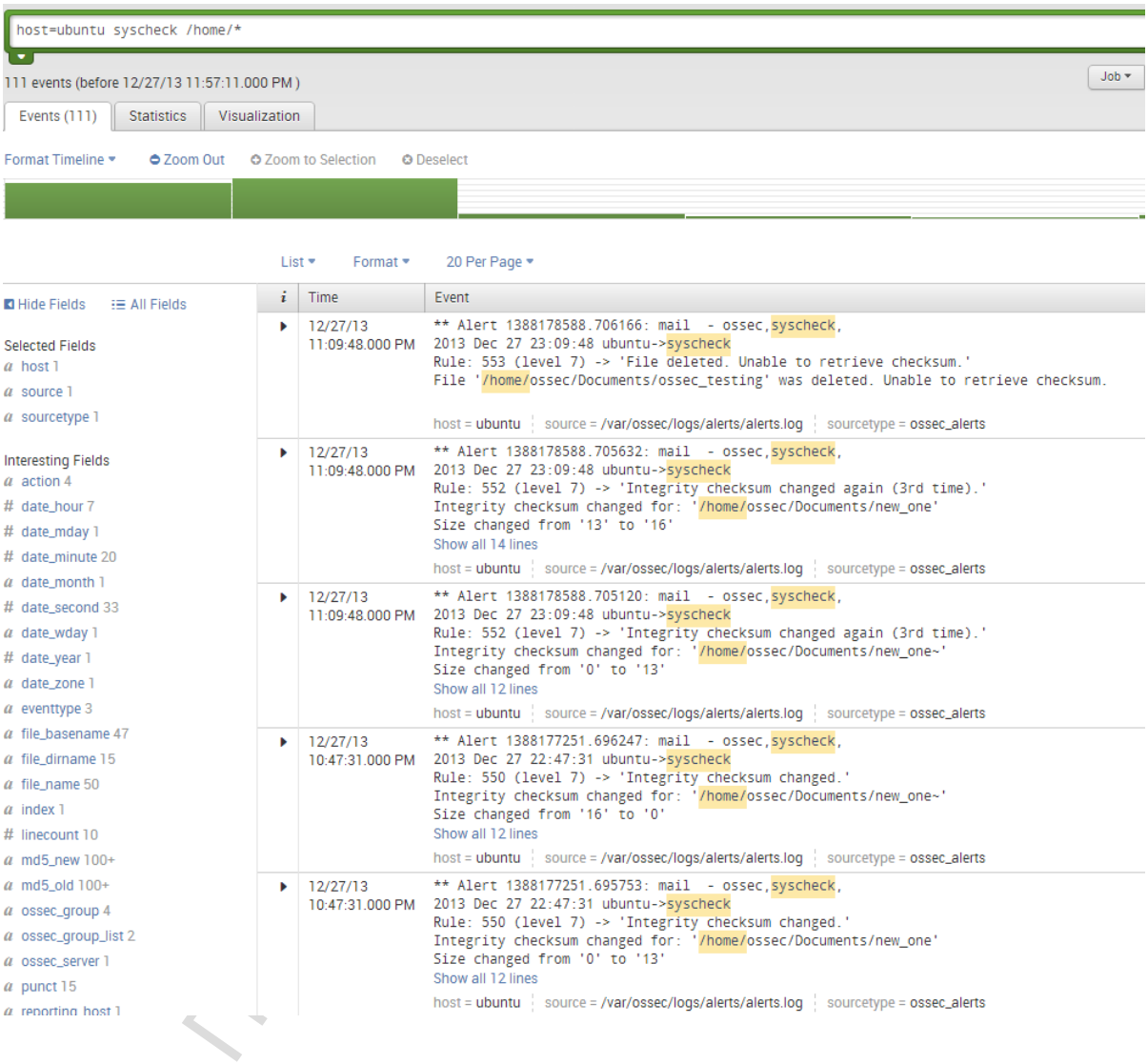
Comparing those 2 outputs, we can notice that only the user ossec is identified and his appears in failed attempt logs, while the IP of 192.168.1.115 is behind the laptosh.local description in Src IP field.

OSSEC, as already referred above, is a useful integrity monitoring tool. Thus, with some extra configuration (modify ossec.conf and local_rules.xml files) it can be used to verify the integrity of important files and directories on a critical system. In the next use case, we are requested to ensure the integrity of the Documents folder.

2 files exist in the Documents folder, as shown in the picture below (new_one and testing_ossec).

```
ossec@ubuntu:~$ ll /home/ossec/Documents/
total 12
drwxr-xr-x  2 ossec ossec 4096 Dec 27 19:25 ./
drwxr-xr-x 27 ossec ossec 4096 Dec 27 19:25 ../
-rw-rw-r--  1 ossec ossec    5 Dec 27 18:55 new_one
-rw-rw-r--  1 ossec ossec    0 Dec 27 17:50 testing_ossec
ossec@ubuntu:~$
```

According to this scenario, testing_ossec file will be deleted and new_one's content file will be changed. At the very end will be searching in Splunk's GUI to see how these events are depicted.



All these alerts can also be sent via email to a predefined contact. (Samples provided below)

OSSEC HIDS Notification.

2013 Dec 27 23:09:48

Received From: ubuntu->syscheck

Rule: 553 fired (level 7) -> "File deleted. Unable to retrieve checksum."

Portion of the log(s):

File '/home/ossec/Documents/ossec_testing' was deleted. Unable to retrieve checksum.

--END OF NOTIFICATION

OSSEC HIDS Notification.

2013 Dec 27 23:09:48

Received From: ubuntu->syscheck

Rule: 552 fired (level 7) -> "Integrity checksum changed again (3rd time)."

Portion of the log(s):

Integrity checksum changed for: '/home/ossec/Documents/new_one'

Size changed from '13' to '16'

What changed:

1c1

< At last ! :O

> This is a test!

Old md5sum was: 'f98d81c057efcdb73643aea251fd41a9'

New md5sum is : 'a28bca1b906f539ba70ca3a0b1f2e773'

Old sha1sum was: '82507fa1db877f2eeb46696b3740979f8cdb347d'

New sha1sum is : 'cc4bc53ee478380f385721b45247107338a9cec3'

--END OF NOTIFICATION

OSSEC HIDS Notification.

2013 Dec 27 23:09:48

Received From: ubuntu->syscheck

Rule: 552 fired (level 7) -> "Integrity checksum changed again (3rd time)."

Portion of the log(s):

Integrity checksum changed for: '/home/ossec/Documents/new_one~'

Size changed from '0' to '13'

What changed:

0a1

> At last ! :O

Old md5sum was: 'd41d8cd98f00b204e9800998ecf8427e'

New md5sum is : 'f98d81c057efcdb73643aea251fd41a9'

Old sha1sum was: 'da39a3ee5e6b4b0d3255bfef95601890afd80709'

New sha1sum is : '82507fa1db877f2eeb46696b3740979f8cdb347d'

--END OF NOTIFICATION

In the last but one case study, we will focus on command monitoring. Suppose that we are requested to keep track of the status of our Web-server (WAMP) and our OSSEC server. The first mail notification informs us that there are new ports opened or closed in OSSEC server, compared to the previous output. Ports 8000 and 8089 are being used by Splunk.

OSSEC HIDS Notification.

2013 Dec 27 23:42:13

Received From: ubuntu->netstat -tan |grep LISTEN |grep -v 127.0.0.1 | sort

Rule: 533 fired (level 7) -> "Listened ports status (netstat) changed (new port opened or closed)."

Portion of the log(s):

ossec: output: 'netstat -tan |grep LISTEN |grep -v 127.0.0.1 | sort':

tcp 0 0 0.0.0.0:22 0.0.0.0:* LISTEN

tcp 0 0 0.0.0.0:8000 0.0.0.0:* LISTEN

tcp 0 0 0.0.0.0:8089 0.0.0.0:* LISTEN

tcp6 0 0 ::1:631 :::* LISTEN

tcp6 0 0 :::22 :::* LISTEN

Previous output:

ossec: output: 'netstat -tan |grep LISTEN |grep -v 127.0.0.1 | sort':

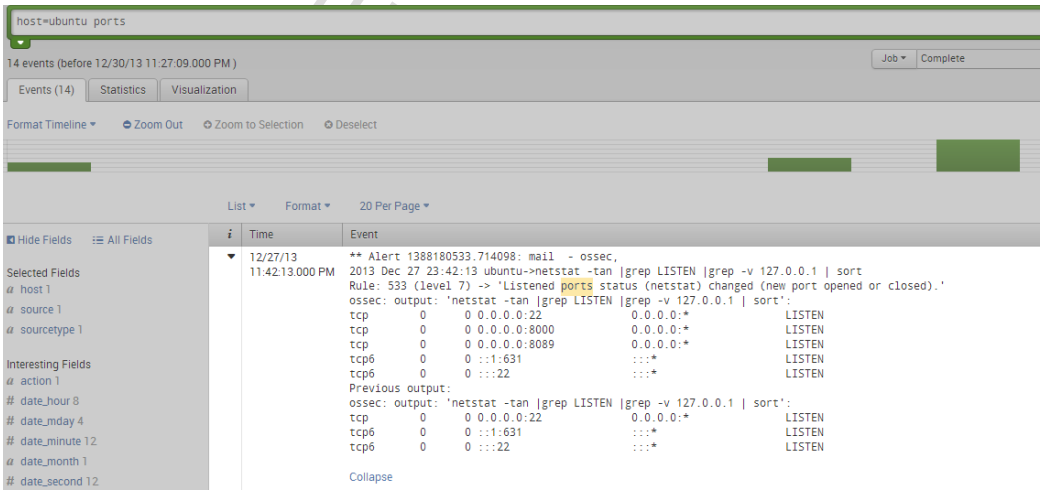
tcp 0 0 0.0.0.0:22 0.0.0.0:* LISTEN

tcp6 0 0 ::1:631 :::* LISTEN

tcp6 0 0 :::22 :::* LISTEN

--END OF NOTIFICATION

This is how it is depicted in Splunk.



while in case of change in Apache's status an alert like the following is being triggered

From: OSSEC HIDS <ossecm@ubuntu>

To: <ossec@ubuntu>

Date: Thu, 02 Jan 2014 01:27:36 +0200

Subject: OSSEC Notification - (Win-Agent) 192.168.1.110 - Alert level 7

Content-Length: 519

OSSEC HIDS Notification.

2014 Jan 02 01:27:32

Received From: (Win-Agent) 192.168.1.110->netstat -ona | findstr 0.0\80 | sort

Rule: 140126 fired (level 7) -> "Apache status changed !!!"

Portion of the log(s):

ossec: output: 'netstat -ona | findstr 0.0\80 | sort':

TCP 0.0.0.0:80 0.0.0.0:0 LISTENING 6960

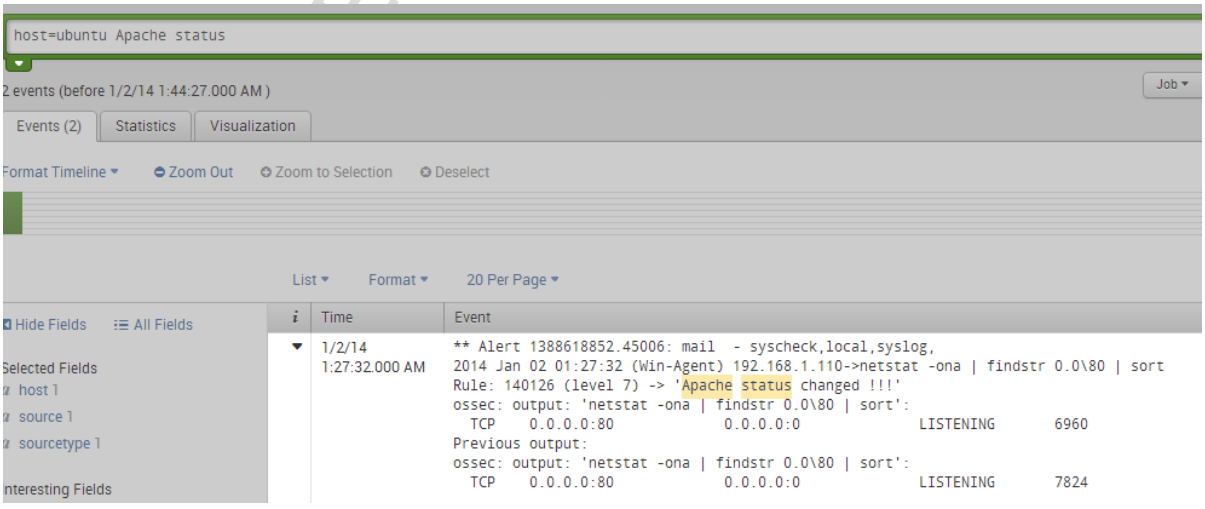
Previous output:

ossec: output: 'netstat -ona | findstr 0.0\80 | sort':

TCP 0.0.0.0:80 0.0.0.0:0 LISTENING 7824

--END OF NOTIFICATION

This is how these alerts are depicted in Splunk.



Finally, in that last case study, a notification will inform us about changes in Apache and USB devices of our Windows Machine. Considering the USB drive monitoring, once a new USB device is added an alert like the following will be triggered.

OSSEC HIDS Notification.

2014 Jan 02 01:05:21

Received From: (Win-Agent) 192.168.1.110->reg QUERY HKLM\SYSTEM\CurrentControlSet\Enum\USBSTOR

Rule: 140125 fired (level 7) -> "New USB device connected !!!"

Portion of the log(s):

ossec: output: 'reg QUERY HKLM\SYSTEM\CurrentControlSet\Enum\USBSTOR':

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_JetFlash&Prod_TS512MJF110&Rev_0.00

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_Kingston&Prod_DataTraveler_2.0&Rev_

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_SMI&Prod_USB_DISK&Rev_1100

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_ST950032&Prod_5AS&Rev_0002

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_WDC_WD80&Prod_0UE-11KVT0&Rev_01.0

Previous output:

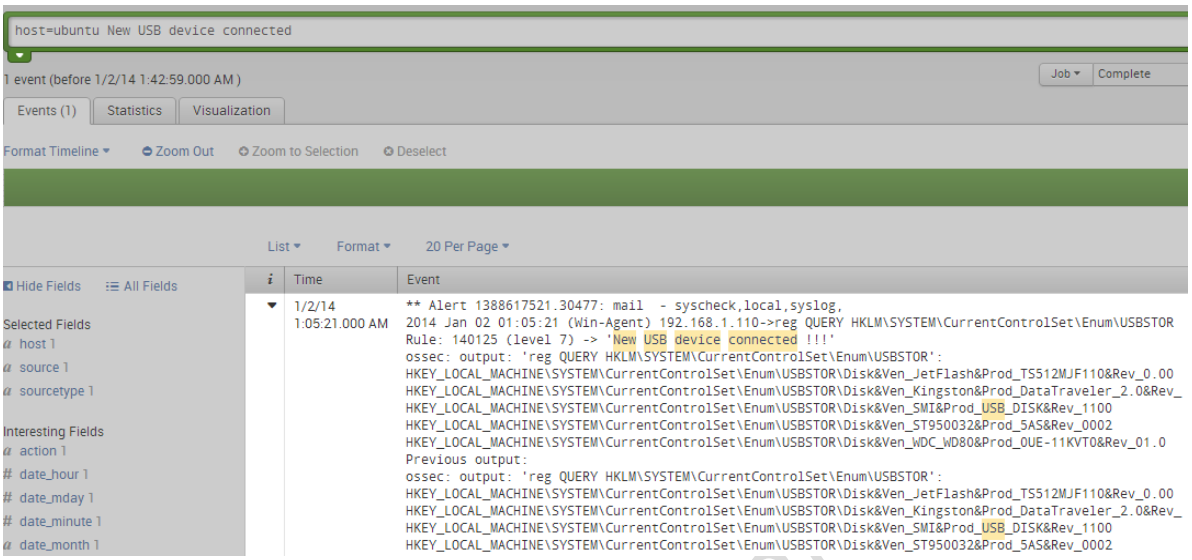
ossec: output: 'reg QUERY HKLM\SYSTEM\CurrentControlSet\Enum\USBSTOR':

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_JetFlash&Prod_TS512MJF110&Rev_0.00

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_Kingston&Prod_DataTraveler_2.0&Rev_

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_SMI&Prod_USB_DISK&Rev_1100

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Enum\USBSTOR\Disk&Ven_ST950032&Prod_5AS&Rev_0002



References

[1] Online: Bruce Schneier's Blog On Security, "https://www.schneier.com/", (Accessed August 2013)

[2] ISACA, "*Security Information And Event Management: Business Benefits And Security, Governance And Assurance Perspectives*", 2010

[3] Gartner, "*Magic Quadrant For Security Information And Event Management*", 2011

[4] Aberdeen Group, "*The Role of Security Information and Event Management (SIEM) in Security Governance, Risk Management and Compliance*", 2008

[5] SANS Institute, "A Practical Application of SIM/SEM/SIEM Automating Threat Identification", 2006

[6] Online: Wikipedia http://en.wikipedia.org/wiki/Security_information_and_event_management (Accessed August 2013)

[7] Patent Application Publication "*Method For Simulation Aided Security Event Management*", 2007

[8] Online: http://securityinformationeventmanagement.com (Accessed August 2013)

[9] Online: Wikipedia http://en.wikipedia.org/wiki/Security_information_management (Accessed August 2013)

[10] NIST SP800-92, "*Guide to Computer Security Log Management*", 2006

[11] San Dorigo, "*Security Information And Event Management - Master Thesis*", 2012

[12] Mariana Hentea, "*Intelligent System for Information Security Management: Architecture and Design Issues*", Vol. 4, 2007

[13] Andreas Müller, "*Event Correlation Engine – Master Thesis*", 2009

[14] 2012 Cloud Security Alliance, "*SecaaS Implementation Guidance Category 7: Security Information and Event Management*", 2010

[15] Mc-Graw Hill Companies, "*Security Information & Event Management (SIEM) Implementation*", 2011

[16] AccelOps Inc., "Top10 SIEM Implementer's Checklist", 2012

[17] Scott Gordon, "Putting the Top10 SIEM Best Practices to Work - Processes, Metrics, Technologies", 2010

[18] Gartner, "*Magic Quadrant For Security Information And Event Management*", 2013

[19] Dr. Anton Chuvakin, "*The Complete Guide to Log And Event Management*", 2012

[20] Online: Increased Visibility Information Security Blog http://intellavis.com/blog/?p=201 (Accessed December 2013)

[21] Online: OSSEC's official website http://www.ossec.net/ (Accessed December 2013)

[22] Online: Splunk's official website http://www.splunk.com/ (Accessed December 2013)

[23] Splunk Inc, "Splunk Enterprise 6.0 – Search Manual", 2013

[24] Online: http://apps.splunk.com/ (Accessed December 2013)

[25] Online: http://apps.splunk.com/app/300/ (Accessed December 2013)

[26] Syngress Publishing Inc, "*OSSEC HIDS Host Based Intrusion Detection*", 2008

[27] Online: Splunk's official website http://docs.splunk.com/Documentation/Splunk/6.0.1/Installation/Systemrequirements (Accessed December 2013)

[28] SANS Institute, "*Logging and Monitoring to Detect Network Intrusions and Compliance Violations in the Environment*", 2012