Radboud University Nijmegen

Security Information and Event Management

Master Thesis

on the methodology, implementation challenges,
security issues and privacy implications
concerning SIEM environments in
large retail organizations.

Sander Dorigo
August 28, 2012

Supervisor:	Prof. Dr. Herman Geuvers	Student number:	0817880
Company supervisor:	Antony Verheijen	Thesis number:	171

For Vincent.

Abstract

This is the Master Thesis of Sander Dorigo. In this thesis the implementation methodology of Security Information and Event Management (SIEM) environments is discussed. It is based on currently available literature and new research. The research was done at a Germany-based international retail organization that plans to implement such an environment to tackle ongoing and future security problems.

The methodology consists of two parts that each have three steps. There is an organizational part that focuses on defining a scope, use cases and requirements after which the choice must be made whether SIEM would be the right step for the organization. If so, the (second) technical part consists of analyzing and preparing assets and procedures for data collection, data storage and data processing using a SIEM environment, after which the decision can be made how best to continue.

The thesis will show how the methodology works for a real life situation, since the methodology was developed as an answer to the questions of the aforementioned retail organization. The origin of the steps and decisions will be shown too.

The conclusion discusses several advantages of the methodology and some drawbacks and future points for optimization. The most important advantages are that the methodology is vendor-neutral and offers motivations for each next step even if it is a motivation not to continue; it does not steer towards any specific solution. The methodology however does not provide in choosing a specific SIEM vendor.

Contents

Introduction

This is the Master Thesis of Sander Dorigo. This thesis reports on the research that has been done on the subject of Security Information and Event Management (or SIEM).

Security Information and Event Management

SIEM combines Security Information Management (SIM) and Security Event Management (SEM). The first focuses on analysis and reporting of log data and long-term storage while the second focuses on real-time monitoring and notifications. SIEM combines these and includes real time analysis and correlation. SIEM is also referred to as **System** Information and Event Management, to emphasize the effect of the technology on the whole system, even though the focus is on security. Some researchers would rather speak of 'SIE**O**M', adding the **O** for "opportunity" since the reports and alerts of a SIEM environment provide opportunities to improve on the security of the system.

While the market for SIEM has been growing for some time, a structured methodology is still making little headway [19]. Most SIEM companies are merely pushing their own security solutions. It seems they keep their methodology a company secret, using it to stay ahead of the competition. This can be inferred from their publications. Instructional blog posts are very high-level or vague [24], white-papers contain little solid information [14] and technical details are scarce [27].

Slowly such a methodology is being developed. It started with a simple list of ideas [32] and how-to's, though it is becoming a full methodology [7]. A lot of work has yet to be done and this thesis hopes to add to this effort.

Thesis

The implementation methodology found in this thesis is partly based on existing (commercial) implementations, that lack a unifying concept and a neutral point of view. This methodology does not include a choice of vendor and is split into two parts. In the first part the organizational motivation and preparations are covered, followed by part of the technical implementation. Additionally the possible application of the methodology is demonstrated. In this thesis SIEM is observed from the perspective of a large retail organization that is planning to implement a SIEM environment as part of their IT security landscape. The methodology is based on the research at that organization. The original assignment and the background of the company can be found in the first chapter.

In short: The organization where this research was conducted started looking into SIEM to complement their efforts for better IT security. They indicated a lack of general overview and information; this lead to the observation that "we do not know if we have a security problem, which in itself is a problem". One of the suggested solutions was a SIEM environment, for which the research in this thesis was done. A methodology was developed that the company used to answer the question whether or not a SIEM environment could help them in this effort.

Structure

The first chapter details the assignment and the organization where the research was conducted. Then both the organizational and technical halves of the SIEM methodology will be explained. For the technical methodology information from the individual SIEM vendors (company websites, weblogs and personal conversations [3, 25, 26]) is used. For neutral information information from SIEM experts and communities [9] is being used and for more background on the discussed subjects, the academic community was searched.

Initially the same sources were used for the organizational methodology of SIEM environment implementations. However, since the research was done at a retail company, it was possible to fine tune the suggested methods to match the retail industry's need better.

This thesis starts with the organizational methodology because it follows a path that stops at the point where technical considerations needed for a SIEM environment come into play. It is important to note that the thesis is skipping the part where other methodologies advise to look for a specific vendor. There is much information on this subject available already, for example by commercial research company Gartner. Such a choice is largely dependent on the outcome of this methodology.

The structure of this thesis is illustrated in figure 1 on page xi. The top right corner shows chapter 1 (Assignment): The company with its structure and challenges.

On the whole left of the figure is the methodology. The three steps of each two parts are complemented with the results from the research. This is also the case for both decisions at the end of chapter 2 on page 5 and chapter 3 on page 25 respectively.

The four challenges for this company and their solutions are discussed in chapter 4 on page 39 coupled with the conclusion on page 43.

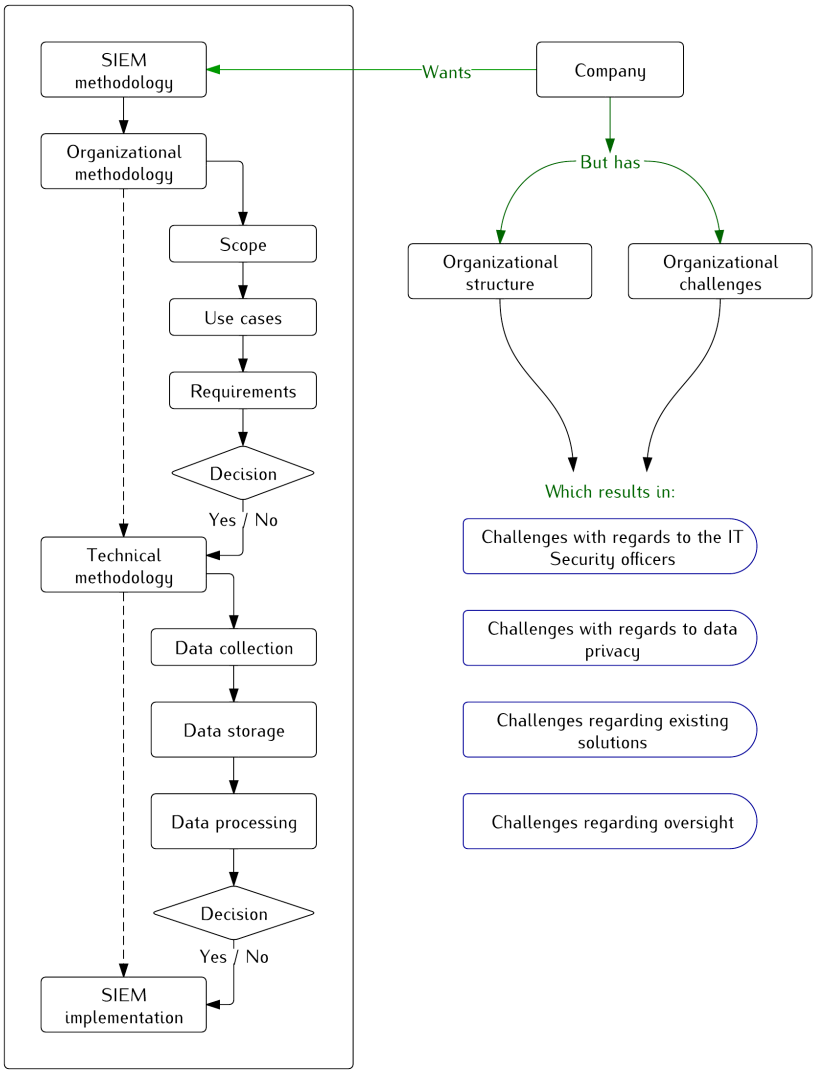


Figure 1: The overview of this thesis and the SIEM methodology.

CHAPTER 1

Assignment

In the first days of September 2011 Sander Dorigo started working at MS under Antony Verheijen[1] as a graduate research student. The original assignment started with AV's interest in SIEM as a tool to improve IT security. The question whether or not SIEM was useful depended on a few factors which he was set to research. These factors led to the methodology explained in this thesis. The original research question they defined is as follows:

> "Which information does a large retail company need, in terms of IT security, from the relevant computer systems in order to quickly and correctly respond to IT security incidents and problems?"

In the period up to March 2012 he has researched SIEM as a possible answer to this question. He has researched its maturity and has written a recommendation for its appliance within MS. This thesis discusses the research that lead to that recommendation. The recommendation is an appendix to this thesis.

At MS, the emphasis on IT security grows with AV's efforts to bring it on the agenda. More security measures are implemented daily. However, there is a source of information that now lies unused: log files. Information on possible security breaches is not extracted from these files, nor is information on possible events relayed to AV. One of the possible mechanisms to use this resource is SIEM. The analysis and consolidation such a system can offer for various critical systems would show possible problems, report on (undiscovered) incidents and show opportunities for improvement. This can greatly improve the IT security status and provide AV with a new source of information concerning said status.

1.1 Organizational structure

The research company follows a structure that may be familiar from other organizations and their supporting IT companies. It is an independent company called MS in this thesis. MS has a CEO, a board and staff departments. It is only set up

[1] Henceforth referenced to as 'AV' for clarity.

to support the main retail organization MG. They are tightly connected with each other. For example, the CEO of MS is the CIO of MG. MS was the company where the research was conducted.

MS consists of three business units. These business units are responsible for the engineering, management and operations of all IT efforts both within MS itself and within MG. All three have different departments and sections throughout. This structure can be seen in figure 1.1.

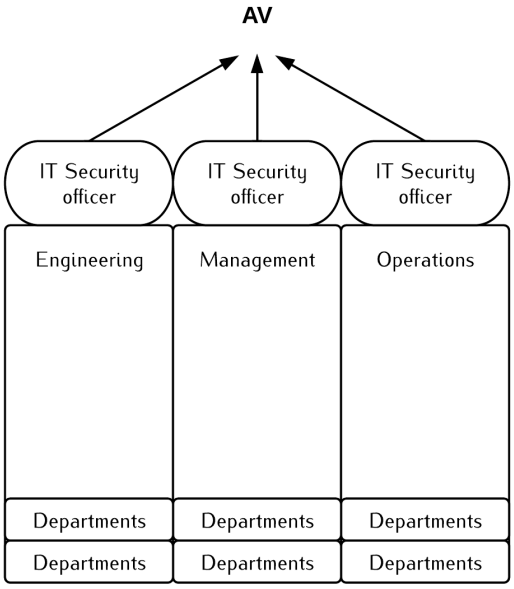


Figure 1.1: The three business units of MS.

Each of these business units has a security officer. They are responsible for all IT security efforts within their unit. AV is their manager where IT security is concerned, but they are not a part of AV's own department. Departments within these business units also have security officers. They answer to their business unit's security officer. This is the hierarchy found at MS.

The holding itself (MG) consists of several retail brands, some other support companies and a main office. The main office is AV's responsibility. The retail brands and the other support companies work in the same way: they have their own IT security hierarchy with one head IT security officer who answers to AV. This can be seen on the left of figure 1.2 on the next page.

Since MG is operating globally, each country has an IT security officer working for MS who is responsible for all the IT security in that particular country. These officers answer to a regional officer, who in turn answers to AV. This structure can also be seen in figure 1.2 on the right.

IT security officers can be found in three places: the business units of MS, the countries around the world and the brands held by MG. They answer (in)directly to AV, creating a pyramid-like organizational structure. Some roles are shared by one and the same person (for example, the regional IT security officer for Europe might

also be the IT security officer for one of the countries in that region). Others are part time functions.

1.2 Organizational challenges

The structure shown in the previous section works well enough although it presents AV with a few challenges.

First, separate incidents disappear since they occur at a level well below AV's overview. A compromised server or a large breach of privacy would of course be escalated, however not on account of AV being part of the solution-chain. AV is essentially a high-level policy maker: his IT security policies are implemented in local policies that are more specifically written for the task at hand. If something goes wrong, those local policies take effect, with all the procedures they might have. Even though AV wrote the high-level procedure, he is not involved in the local implementations. So, as incidents travel upwards in the chain, they tend to be summarized, dismissed and resolved and therefore forgotten: AV does not know something has happened. In itself, this is not a bad thing. If all the small things ended up on his desk, his efficiency would sharply decline. Plus, managers will not call AV every time an incident has occurred or is resolved. Only special incidents make it to AV's desk. The result is that unless the entire company is being hacked AV will not hear about it.

Incidents also disappear as a result of people working on different levels. Sometimes this is caused by the security officers themselves who just happen to overlook incidents although the company structure is a cause too. A lot of services are run from a few large data centers. Technically two people should inform AV of an inci-

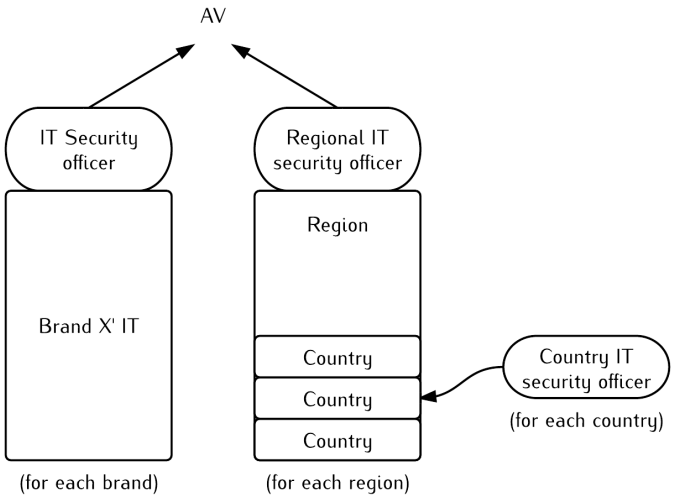


Figure 1.2: The global overview of MS' and MG's IT security officers.

dent: the security officer at the end-user side and the security officer close to the actual hardware. When both of them overlook the attack AV is in the dark as well. Even if attacks get noticed they might be mitigated by an network administrator who's report does not reach AV or his other security officers. The overview AV eventually receives is to vague.

Second, incident trends disappear. These trends are the observation that certain attack vectors are more popular than others and that certain types of incidents happen regularly. The cause is roughly the same problem happening on several levels.

Both of these problems can be handled with incident management (ICM) which another solution AV is looking into. The goal of ICM is to restore normal operations as quickly as possible after an incident [1]. AV lacks an overview of IT security incidents. Better management and handling of such incidents will also improve IT security. ICM lacks correlation, a feature SIEM does have. Problem management addresses known errors which may cause multiple incidents, but IT security incidents are not always reported by hand or even semi-automated: some incidents are created by applying deduction and correlation to multiple events which on their own, do not constitute an (IT security) incident.

The next two chapters will discuss the methodology as it was developed and applied at this company. It will be shown how the two problems that are described in the previous paragraphs can be solved using SIEM.

CHAPTER 2

Organizational methodology

The decision to deploy a SIEM environment starts with the definition of a scope and a focus. Then use cases and requirements must be defined. In this chapter all of these steps will be introduced, coupled with the results from the research at MS.

The previous chapter discussed the challenges for MS. The first thing to realize is that those challenges may not be fixable using SIEM. AV might hope to introduce SIEM within MS though other solutions might be better fitted to resolve these challenges.

2.1 Define a scope

Before the start of a SIEM environment installation, it is very important to set a scope and a focus. The scope is the driver behind SIEM and can be related to compliance, security or operations [5]. It can be a combination of all three and should encompass the entire company. If there is a compliance scope needed for one part of the company and a security scope needed for another, the work for a SIEM environment should take both into account.

It might be that the company is too large to start implementing SIEM everywhere at once. If that is the case, the *focus* should be limited. The focus defines an area where SIEM is applied: a certain subset of the entire company. This focus can be as narrow as needed as long as the primary process of the organization is present. Some examples of such a focus are:

- Focus on a location, such as a store, a region or an entire country.
- Focus on the chain of a specific product, such as dairy products.
- Focus on a certain channel, such as web based or phone based customers.

The scope and the focus can be chosen separately although they are connected.

2.1.1 Audit and compliance scope

SIEM is an extension of log management and is relevant for IT compliance when it comes to log collection, retention and review. IT (security) audits check if all rules

and policies are being followed. Whenever compliance is the main driver behind a SIEM installation it will be used to make sure the necessary records are being logged and stored and are ready for retrieval upon inspection, as mandated by such rules and policies. An example of a policy that makes SIEM revolve around audits is the PCI DSS[1]. This is a security standard for companies that handle credit card information. Companies that wish to adhere to this standard are required to log all sorts of events [23, p. 55] related to credit card access or customer data. A SIEM environment makes it easy to do this. It also makes it easy for auditors to check if all records are being saved.

Saving those log files does not necessarily make a company more secure although auditors may request access to these log files to check on other rules in the PCI DSS. This is called resource access monitoring and is a very valid usage of SIEM and other log management tools.

If applied as a compliance tool SIEM monitors specific databases and servers, tracks incoming and outgoing data and logs user access to sensitive information. Most SIEM vendors equip their software with default reports that can quickly show compliance to standards such as the PCI DSS or ISO 27002[2].

2.1.2 Security

When security is the main driver the focus for SIEM is external threat monitoring and security application monitoring [17]. Firewall, web server and IDS[3] log files are monitored for events that may be indicators of a digital attack. Devices are specifically observed to see if their behavior is suspicious and may constitute a compromise. Correlation is very important since an attack can spread rapidly through the network. Logs are collected from many sources, if possible from inside the network as well, beyond firewalls. This is to ensure that malicious users do not have a free reign once they breach a firewall.

For security monitoring, extensive log analysis can show a certain heart beat. The number of attacks for example is never zero, nor is the number of suspicious transactions over the network. A SIEM environment is able to pick up this beat. From it, attack vectors can be determined. Slow patterns for example, are attack vectors where the attack is not being executed constantly. Rather once a day, once a week or even less often. Doing so averts the attention of most intrusion detection systems. Another important pattern is the anti-pattern: the sudden absence of attacks. If an intrusion detection systems scans and reports on fifty attacks every day however suddenly ceases to do so, it might be compromised itself rather than "all the hackers are having a day off".

2.1.3 Operations

For operations, the focus is on resource management, hardware limits and possible errors and warnings (bugs) [5].

[1] Payment Card Industry Data Security Standard.

[2] The ISO 27002 standard provides best practices for IT security. Some of those focus on log management.

[3] Intrusion Detection Systems; comparable with firewalls also capable of identifying heuristics and patterns.

Resource management and hardware congestion use cases track the network and hardware utilization of applications such as ERP[4] software. Servers, routers and switches have different metrics that indicate their (used) capacity. Keeping track of that shows when and where the network or the servers have too much tasks or too little bandwidth available.

Log files from servers and routers can also help developers optimize software. In the case of errors and warnings those log files are monitored for crashes, errors, redirect loops and other events. If these are combined with the log files collected for resource management, developers and administrators can determine how applications perform on busy networks.

It should be noted that client-side log collection can be very difficult. If an application has a client-server design pattern, there might be thousands of clients connecting to one server. It is my own experience that it is easier to set up developer-clients rather than configuring all these normal clients to forward their log files.

The choice for a scope and a focus define what SIEM will be about.

Initially, the scope for MS was security. As mentioned in the previous chapter the challenges for MS are IT security related and the initiative for SIEM came from the head of IT security. Once the scope was set, a focus was chosen: one (1) experimental web shop and the associated smart phone application. Customers of MG can use the smart phone application to save shopping lists and shop online. Their baskets will be ready to collect at the nearest super market. This should save them time: a customer could drive by the store and collect his or her basket, without needing to shop for all items themselves.

There are a couple of reasons this focus was selected:

1. Web shops are very visible. Contrary to email systems or login portals, every effort is made to draw more visitors to web shops. This will more likely attract malicious users as well.
2. With this visibility comes a quick win for malicious users who manage to successfully attack the web shop. Damage control is very difficult even when the attacked website is merely a static page with no sensitive data on it.
3. Web shops have a deep reach into MS' systems. A customer can change his own personal records and change MS' inventory records by buying all while not being physically present. When a hacker compromises customer passwords the MG is liable for possible losses.
4. Other systems, such as MS' web mail, have more complex security. RSA tokens and VPN connections are often mandatory, making it more difficult for a hacker to breach. Such security measures cannot be demanded from our web customers.

With this focus, the entire scope of SIEM at MS changed as well because of the experimental status of the chosen products (both the webshop and the application). Operations-based use cases can help with the development of these products.

The scope now includes operations, instead of being only about security. That means that there will be use cases focusing on the performance of the products as well. The researchers can find out if there are bugs. If so, they can determine

[4] Enterprise Resource Planning.

where these bugs occur and how they affect the performance of related devices and software. If SIEM is up and running in this scope and focus (using the rest of the methodology in this document) it can extend towards other departments, according to the "Start small, think big"-principle mentioned in section 5.1 on page 45.

It is important to note that security solutions already existed within this scope and focus. Since the web shop was an experimental product network administrators and developers already kept a close eye on server logs and errors coming from applications. Outside of the web shops scope other administrators also kept an eye out on possible security problems. The problem for AV was not the lack of such security measures, it was the visibility of their results. Now that there is a scope and a focus to try to solve such problems the use cases can be defined.

Originally the methodology was developed with just a scope and not a focus. At MS however the focus was chosen before the scope was clear. The methodology now reflects the necessity for a focus.

2.2 Define use cases

The next step is defining use cases. This is important to get a manageable environment for a new SIEM environment. Use cases will limit the number of business units, servers and processes involved in the first efforts towards an installed SIEM environment. In software engineering (using UML) use cases are defined as a list of interactions between an actor or multiple actors and a system. Use cases represent a goal or a task.

It is no different with SIEM. Use cases define goals for the SIEM environment. The SIEM environment is the actor and all the devices are the system.

Business rules are a method of defining what content a report should have. For each use case several reports can be written that show the various aspects of the use case, and business rules are a good method to get something to base a custom report on. Defining those business rules starts at the use cases, by defining terms and facts.

> A business rule is a statement that defines or constrains some aspect of the business. It is intended to assert business structure or to control or influence the behavior of the business. The business rules that concern the project are atomic – that is, they cannot be broken down further.
>
> For our purposes, this may be viewed from two perspectives. From the business ('Zachman row-2') perspective, it pertains to any of the constraints that apply to the behavior of people in the enterprise, from restrictions on smoking to procedures for filling out a purchase order. From the information system ('Zachman row-3') perspective, it pertains to the facts that are recorded as data and constraints on changes to the values of those facts. That is, the concern is what data may or may not be recorded in the information system. [28]

In order to write such statements it is necessary to define business terms and facts. These are tied together using constraints and later on derivations.

Definitions of business terms

The most basic element of a business rule is the language used to express it. The very definition of a term is itself a business rule which

> describes how people think and talk about things. Thus, defining a term is establishing a category of business rule. Terms have traditionally been documented in glossaries or as entities in an entity/relationship model.
>
> **Facts relating terms to each other**
>
> The nature or operating structure of an organization can be described in terms of the facts which relate terms to each other. To say that a customer can place an order is a business rule. Facts can be documented as natural language sentences or as relationships, attributes, and generalization structures in a graphical model. [28, p. 6]

When possible, the example use cases and the use cases defined at MS (detailed in 2.2.1 and 2.2.2 further ahead) contain such terms and facts. Business terms are normally summarized in a glossary that should be shared with all use cases.

2.2.1 Example use cases

In this section some example use cases are mentioned. One for a compliance-based scope, one for a security-based scope and one for an operations-based scope (in that order). They should give a general idea of the questions that SIEM can answer in different scopes. It is important to note that most scope/focus combinations require multiple use cases to cover all the angles.

Privileged-user monitoring and segregation of duties issues

This is a compliance use case. The segregation of duties (and powers) is an important compliance rule that prevents abuse and fraud. It means that certain sensitive tasks must always be done by two separate persons or be confirmed by a another person.

It is possible to match the actions of privileged users to changes in resources, access to resources and other interactions with the system. Segregation of duties is achieved by carefully defining each action and subsequent privilege needed. This combination of actions and privileges is analyzed (usually by hand) for combinations with risks. For example: employees with the ability to give discount on an item *and* the ability to finish the sale of that same item could give themselves or accomplices huge discounts.

This compliance use case helps to prevent such user actions: certain combinations are not allowed: the SIEM environment can raise an alert if it encounters them. The advantage of SIEM is that the previously mentioned slow patterns can be detected as well, where malicious employees wait a day or a week between actions to try and circumvent the security.

Audit and compliance use cases are based on existing rules and policies such as the PCI-DDS[5]. It can be a valid use case only to focus on such policies. Since SIEM environments are tailored for such policies business rules only hinder the use case. It is only useful to define terms and facts when the use case is specifically tailored to possible compliance issues[6].

[5] Mentioned in 2.1.1 on page 5

[6] For example segregation of duty within custom made SAP applications.

Cross-system authentication tracking

This is a security use case [6]. For this use case the SIEM environment must track authentication requests and privileges across networks and devices. The SIEM environment must keep track of (failed) authentication attempts and their identifying information such as session variables and user names on login-devices such as a domain controller. The SIEM environment must check the logs of devices using that authorization to allow access to files and records to ensure they match. Mismatches can be a sign of malicious access attempts, especially when the device itself does not realize it has been compromised.

The original goal of SIEM was to reduce the number of false positives raised by intrusion detection systems[4]. Security related use cases are usually closely related to that goal and are further away from business terms and facts. This example use case only mentions users and privileges. Both can be captured in business rules but are usually inferred. The rule "*Only authenticated customers can buy products*" implies a login system of some sorts (but not necessarily!) and the rule "*The store must verify the authenticity of each sale*" implies that the store must double check that no sales are fraudulent.

Cross-system errors and warnings tracking over complex networks

This is an operations use case. It deals with the various types of errors and warnings applications can generate, possibly when the network they operate on is unreliable or busy. The latter is a situation not often found in development environments. Testing for such problems is certainly possible though a real-life network will always display more diverse problems. A use case that tracks such errors correlates these errors with upgrades, patches and other notable events.

In order for such a use case to work the SIEM environment must correlate information from a few sources:

1. The state of network devices (bandwidth, CPU usage);
2. The errors and warnings of software applications and network transmissions;
3. Version numbers and patches.

Using that information errors are noted and compared: what was the situation when the error occurred? Its risk and priority can be determined by looking at the number of errors logged and the capacity or lack thereof of the network.

The business terms and facts mentioned on page 9 are only useful for operations use cases when the terms and facts can be related to the errors and warnings that are logged, preferably one-on-one. That is difficult because such terms and facts are very high-level (a customer, a sale) while errors and warnings are generally very low-level. To solve this discrepancy it is important to log high level events as well.

Summarizing: The scope and the focus have defined a certain cross section of the companies IT landscape. The defined use cases apply in that section. Use cases need a goal and must be feature complete. That means that no matter how small the cross section is the devices needed by the use cases should *all* come from that section. It cannot be so that one device completely outside of this cross section has to be added to the SIEM environment in order for the use case to work. As a

consequence no matter how wide the scope or the focus becomes, the use case does not need to change. All that needs to be added are the new devices found in the wider cross section.

2.2.2 MS use cases

The original goal of applying SIEM at MS (or applying other techniques) was to provide AV with an overview of the security status of the entire company. Given the focus and the scope however, it was now possible to be more specific.

Together with AV and his colleagues the researcher defined the following use cases on which MS should focus. They were also used as examples when discussing SIEM with stakeholders and technical administrators.

The previous section mentioned business rules. Where possible relevant business terms and facts have been defined for the use case.

Track customer authentication

Customer authentication occurs every time a user logs in or makes a specific personal request of the web shop. They can be difficult to track. Mobile networks often share IP addresses between clients which means that there is a pool of IP addresses where requests come from. Any client can potentially authenticate from any IP address.

The goal of tracking (failed) authentication attempts is to make sure that every request is made purposefully by clients who are who they claim to be. Mismatches and failed attempts are traced to find possible weaknesses in the system. This kind of tracking requires requires correlation. Firewalls, routers and application servers all receive such authentication requests and deal with them accordingly. By allowing the SIEM environment access to all these log files weak spots can be found before they are exploited.

The reason such weak spots exist are for a part due to user friendliness. Login attempts can be done in rapid succession and passwords have little limits. Login tokens are saved for a long time. Logging in and out is easy.

The goal is to prevent user data theft and catch other abuse.

With this use case the application business rules depends on the rights of the user in the web shop and the application. In this use case the user is a customer and it is fairly straight forward what a user is capable of. Furthermore, since the web shop and application were already in development it is not useful to define business rules. They are just another way of writing down the intrinsic rules of the systems involved. Doing so is only convenient when it is certain that SIEM will be applied. The newly written business rules can then be extended and kept coherent with the widening scope and focus of the new SIEM environment.

MS had not reached a decision for SIEM at this point. The methodology was not finished either. Customer authentication tracking was therefore limited to what applications and servers were capable of reporting and nothing more. High level reporting is set for a later stage in the development process.

When this methodology is applied to new systems it is advised to use business rules. This will save time further ahead because with the definition of business terms and facts, it will be easier to define the accompanying facts and constraints. [28, p. 6]

Analyze errors and warnings

The web shop and the application are in an experimental stage. Log files are being kept by all relevant servers and services. Firewalls block and keep track of possibly malicious URLs. This is a preventive method set up to make sure that even vulnerable systems are not easily compromised. The servers on which the web shop is running keeps track of errors in the programming and errors in the communication with the main inventory systems. The smartphone application keeps track of errors as well, but due to privacy concerns these are not sent to MS automatically [16].

Each of these kinds of log formats differ from each other. Using a SIEM environment, they can be collected, normalized and correlated. The primary goal is to find out how the SIEM environment will handle such errors. Developers and administrators can use these reports to track errors and fix them in the hopes that more errors will be fixed than there would be without such tracking.

With the experimental stage of both the web shop and the application it was difficult to define solid business terms and facts for MS. of course rules may pertain to errors and their preferred absence but that bears no relation to the actual 'business'-part of business rules.

Inventory and stock

In a perfect world, every sell will see the inventory diminish with exactly *that* amount of items. Sadly, theft and losses prevent this perfection. Even in a digital store this is something to be on the lookout for. As an experiment into business rules[7] this research suggested a use case that was inspired by such rules to see if SIEM environments may be able to track such things.

In this use case the changes to the inventory are compared to receipts generated by the web shop. These should match at all times. External changes to the inventory such as due to breakage should be accounted for as well.

The goal is to find out if there are bugs in the processes of buying and returning items as well as where the greatest losses are.

Here business rules are important. Every piece of programming that deals with inventories will give alerts when the inventory is off, but business rules can define regulations around such changes. They can define what is allowed and what is not. Since we are starting with business terms and facts business rules here define objects such Baskets, Products, Sales and Discounts. Facts that relate to these terms put Products in Baskets or give Discounts over Products. Such terms and facts were already defined at MS. Since this is a retail organization the knowledge required to write such rules is abundant. At MS these rules were not defined as business rules. The language and grammar are different from the business rules practice. But for the sake of this use case, these rules were enough to work with.

It is important to note that many high level business terms make it more difficult to collect the necessary log events. Log events are generally very low-level. They contain information on memory shortages, application crashes or null pointers. Most standard log events pertain to the programming language, not the application built with that programming language. It takes much custom logging to collect high level events such as the total price for a Basket, combined with the Price of a Product. This is important to keep in mind when the data collection requirements are considered.

[7] See ahead in section 2.3.3.

2.3 Define requirements

Requirements define what is needed to successfully execute the use cases. Some requirements can be quite technical, others are on a higher level. This depends on the domain experts brought in. Requirements for data collection and –retention should be set up together with network-, server- and database administrators [18]; they might become more technical. The requirements for reporting and event management are best written down with the help of the managers and team leaders who will be using those reports and alerts. Those requirements can end up being to vague. While writing these requirements it is important to make sure requirements are a little bit of both: technical enough to be usable and high-level enough to be flexible.

The previous section contains three example use cases: one for compliance, one for security and one for operations. The following section and those after this one will touch upon these examples as well as on the use cases defined for MS. Four sections were defined in the methodology where requirements should be set up for: data collection, data retention, reporting and event management. These were the four subjects that kept coming back as questions from administrators and stakeholders when SIEM was discussed.

Data collection was an issue since the agents would potentially use a lot of system resources and the network administrators did not think it would be wise to assign so many resources to the collection of log files. This was also an issue since firewalls and routers can generate many events and processing them would be time-consuming and costly.

Data retention questions were raised about the "where" of data storage. MS' current architecture is designed around shared space in data warehouses. For an experimental SIEM system which could easily use a lot of storage space dedicated servers are preferred. At the head office of MG little space was available for even more extra servers.

Managing the world wide IT for a company like MG brings about a lot of paper work. Existing compliance guidelines, rules and laws already generate many reports. The issue for many administrators with the reports generated by SIEM was two-fold. First, they simply did not *want* more reports. Second, they already had the tools to find out anything SIEM might tell them. The fact that the overview was missing for AV and others, was not an issue for them. This issue is also discussed on page 3.

Incident management was mentioned by many people since MS is already running a large incident management system. It tracks everything from technical to administrative incidents and is fed by hand as well as automatically. Initially, it was brought up by AV as an example of why he lacks an overview: the only metric it has for security issues is a check box that says: "IT-Security related incident". A list of such incidents is hardly an overview. Therefore, a SIEM environment should be able to both read from such a system and write to such a system.

2.3.1 Data collection requirements

Correlation is important for every use case; certainly for the examples in this thesis. The first requirement for data collection is that there are no gaps. Events cannot go missing between steps. An authentication attempt can go from a domain controller to a master domain controller up to an Active Directory or another server that holds

the actual user account. Each server the request hops over should be monitored. Otherwise data can be inserted or manipulated on the way.

Another requirement is relevance. With the limitations put in place by scope, focus and use case, routers, servers and other devices might log data that is not needed for the use case. Especially network nodes may have this problem. The SIEM environment must either filter diligently or ignore the records all together.

Raw log events may contain a lot of sensitive information, such as names, IP addresses or other records. Privacy and compliance laws may limit the collection of such records, require encryption or require hashing. In turn, this means collecting and processing log events will require more resources.

Most SIEM environments have an EPS[8]-based licensing model [13]. The more events processed by the SIEM environment, the more expensive the environment will become.

From a more technical perspective, the following things should be taken into consideration.

- How much bandwidth is available? Bandwidth is needed for the transportation of log files, events and reports from the originating devices to the SIEM applications[9].
- How much disk space is left on the average device for log files to be stored on? Do mission critical devices have enough wiggle room to keep log files?
- The same question for CPU and RAM capacities. If every device is running at its peak, it will be difficult to add more tasks.
- How important is the device from which log files are collected? This is important for the SIEM environment to operate properly.

Although it is not yet necessary to have the answers to these questions, they will become relevant as the methodology continues.

An important consideration to make as well is the fact that SIEM environments like HP's ArcSight come with predefined reports, depending on the scope and requirements [15]. According to a HP researcher the researcher spoke to[10] however, SIEM environments "only know what you tell them". That means that those predefined reports may be incomplete and / or inaccurate, depending on what input they get. They are not capable of anticipating missing devices or information.

Finally, if use cases require many high-level events to be collected, data collection may proof to be difficult because the information is difficult to obtain or difficult to deduce. The focus of the SIEM effort has already limited the number of devices involved. It will be necessary to search for all log entries pertaining to the defined business terms and facts, for each use case. That way it is possible to see if log collection is possible. If the terms and facts can be deduced from these log entries notes should be made so that later on, the SIEM environment can be set up accordingly.

Section 2.2.1 on page 9 holds three examples of use cases. For compliance use cases, specific servers and log files need to be monitored; for example those accessing credit card data. Security use cases benefit greatly from correlation: it

[8] Events per Second.

[9] See also page 25; the introduction of the technical methodology holds an explanation of these devices.

[10] Personal conversation on February 12th, 2012.

is important to keep an eye on firewall and IDS log files, as well as other devices operation on the edge of the network. For operations-related use cases the emphasis must be on the application servers and key network nodes.

The web shop built by MS is a single application. It runs distributed on several systems. Each of those systems can be equipped with an agent that uses standard and custom written rules to collect the log files from login procedures, errors and warnings (`stderr`) and connections with the inventory systems of MG. It is not difficult to determine from which devices log files need to be tracked.

A discussion that was going on in my time at MS was how log collection works on shared devices. Firewalls, storage devices and some servers are being used by many services across the board. When one application needs extended tracking, the server needs an agent[11], possible affecting every service running since the agent requires resources all applications share. There was no definitive answer at the time Sander Dorigo left the company. Such collective log collection efforts should be shared by all parties. This issue is further touched upon in chapter 4 on page 39 about obstacles at MS.

2.3.2 Data retention requirements

Data retention requirements contain the demands and restrictions placed on keeping (consolidated) log files. There are three main restrictions concerning data retention.

Data privacy laws

There are various privacy laws, both national and international. This legislation makes sure that companies do not unnecessarily keep private information on their customers or employees. Log files are no exception. That makes it illegal to store log files for long periods of time [8].

Costs

Even if such laws did allow more storage, it would still not be possible to store every snippet of log file. There are costs involved for servers, storage capacity, energy and maintenance, whether the log entries are compressed or not. For example, according to one of my colleagues at my research company the outer firewalls alone generate 1.5TB of log data per day. It would take less than a year to fill up all current free space within MS' storage centers with those log files.

Relevance

Data retention requirements are about relevance as well. Not all records need to be kept indefinitely. For compliance use cases, the retention period is usually set to a year: no exceptions. A combined operations use case may collect records that only need to go back a week. Even within the same use case the retention is different for certain records. Once an authentication request has been deemed clean, i.e. without suspicion, it might not be necessary to save all the log events related to the request: just the request itself. Data collection requirements and use cases do not account for such subtleties.

[11] See also section 3.1.1 on page 25 on agents.

The final answer to this question depends on the number of incidents and the response time to such incidents. Some events trigger often, others rarely. The discussion of defining a scope mentioned a heart beat[12]. This heart beat determines the ideal data retention requirements. The retention period should be long enough to analyze past incidents.

There was no definitive answer yet for data retention. That had two reasons: data privacy and EPS data[13].

Technical data privacy is an important issue. Laws and compliance rules state what MS can and cannot save. What private data is collected exactly is difficult and time-consuming to determine beforehand.

Following from the scope, use cases and data collection requirements it is not difficult to find out EPS numbers. These numbers can be combined with other important information such as the actual budget to determine how much events can be saved. Since data privacy issues may limit the number of events collected the EPS cannot be determined accurately. This problem can be mitigated by encrypting or hashing the private data.

2.3.3 Reporting requirements

One of the end results of a SIEM environment is the automated generation of (interactive) reports. SIEM environments come with a set of default reports. They are based on common compliance standards and use cases or the specific devices connected to the SIEM environment. Whether or not they provide the right insights largely depends on these devices.

Non-standard reports are created using the SIEM application, usually with some kind of "point and click"-system. Business rules are a good method to define what should go in the reports associated with the use cases and requirements so far.

Business rules

In the previous sections we have defined business terms and facts as guides for the use cases and requirements. They can help when defining custom-built reports.

SIEM environments come with a lot of predefined reports. They default to common scenarios and compliance use cases. By feeding the SIEM environment the right information, a clear and useful report will be produced. It is only when more custom work is needed that business rules are required.

SIEM environments are not tailored for retail use cases. They have no reports ready that allow you to track inventories or sale-procedures. Business rules can make it easier to do so. Table 2.1 on the facing page lists some examples of business rules on which a report can be based.

These examples are called 'facts'. Each term must be related to a log entry (or a combination thereof) so the validity of the rule can be determined. This correlation is the custom part of custom reports.

The practices of business rules and the application of them within this methodology is fairly advanced but it has several advantages. First, it helps to define what

[12] See page 6 for more information.

[13] See also section 3.1.1 on page 25.

Compliance-oriented scope	For each Change that requires Administrative Access, an Access Record must be present.
Security-oriented scope	Every Request in this scope must have an Authorization.
Operations-oriented scope	Crash reports must have a known cause.

Table 2.1: Business Rule examples in different scopes. Note that the last example is not a business rule in the traditional sense.

the actual issues are. SIEM adds no extra security to an IT landscape, it adds knowledge about that landscape. Business rules help to make explicit what extra knowledge is needed for better IT security. Second, business rules have a rigid structure, as defined in the Business Rules Manifesto[2] that can help technical and non-technical people alike to define what they will use SIEM for. It might even be the case that the custom reports defined using business rules can be generated just as well using the default options provided by the SIEM environment. If the integration of business rules in this methodology helped realizing that, nothing is lost. In order to properly do so it is necessary to get properly acquainted with business rules, which is outside of the scope of this thesis.

Authorizations

When the focus widens it may become necessary to define sharp rules as to who is allowed access to reports and data. SIEM environments have fine grained tools to define such access and are capable of monitoring that access as well[14].

Who benefits from the overview the report gives? For each scope and focus there are people in the organization that directly benefit from the overview given by the report. These people should be the only ones who can access the reports. A list of such people can be found using the following questions:

1. Devices: who manages the devices included in the scope and focus? If the reports contain information pertaining to the hardware of these devices they must have access to them.
2. Software: Application managers should know how their software is performing.
3. IT Security: if any of the above handle sensitive information or need authorized access, the status of the device or software should be known to the local IT security officer.
4. Data Privacy: If private data is being handled by any of the above, the data privacy officers should have access to such reports.

Furthermore, interactive reports allow drilling down: by clicking or filtering the data the report is based on more details can be revealed. If the reader goes down far enough SIEM applications will show the original log lines the report is based on. That means the people who end up at those possible hashed[15] lines must have access to a procedure that.

These questions are important on the grounds that servers and applications have more sets of users and administrators. The hardware is being maintained

[14] Of course *those* records need access control too, ad infinitum if must be. Who watches the watchmen?

[15] See also section 2.3.1 on page 13 on data collection.

by one group while the software is developed by another. Then there are the application managers, the normal managers, the security officers and so on and so forth. Therefore it is important to define who the target audience is; in this step it might be discovered that necessary information is not being collected. All these stakeholders have different informational needs.

Given the use cases and rules defined earlier: which groups of people need to know the results? And how detailed must it be? Here, privacy is an issue as well since each of the aforementioned groups of people has to make a solid case for access to such data.

AV's organization of his IT security officers[16] provided the ideal structure for reporting: since there is an IT security officer involved at every level they can define and receive the reports that are useful to them.

The business rules and other considerations for reports follow from the use cases. The only complex use case is "Inventory and stock"[17]. This use case was originally the use case that made me think of business rules as the base for a report:

1. In a store that is completely digitally operated inventory loss should not happen: customers cannot steal. Even with the possibility of warehouse employee theft the differences should be minimal. The business rule is: The calculated inventory must match the actual inventory.
2. A customer basket, excluding shipping costs or other costs, must be worth more than the cheapest item in stock.
3. Web shops can be very flexible in their pricing. Multiple prices can be defined for articles, so price changes are not reflected in customers baskets. To safe guard against TV's being sold for a few euro's for example, the rule is: The selling price of an item cannot be less than half the purchase price for MG. This still allows MG to sell below margins though it provides a safe guard against mistakes.

For the use case "Analyze errors and warnings" the reports are not difficult to generate either. It is a matter of organizing and displaying error type and source. In the period that the experiment was running production error logging was set to `error`: only serious problems were logged to disk. Such errors were reproduced by developers in a smaller environment with more extensive logging. This worked for most issues although sometimes the bug would be difficult to reproduce and more log information was needed which is difficult to obtain from production servers. Experiments were being ran to find out what the applications performance was while also writing *everything* to log files.

2.3.4 Incident management requirements

SIEM environments offer alerts and event reporting. When certain events or thresholds for multiple events get triggered they can alert the administrator and help them prevent security incidents or the escalation thereof. This information can also be drilled down into. The incident can be tracked down to the exact log entries that

[16] See also section 1.1 on page 1.
[17] Mentioned in section 2.2.2 on page 11.

triggered it. In turn activated triggers can prompt other triggers. Such behavior allows for the possibility that one alert will not be enough to warn everybody, while ten alerts might.

One of the issues with reporting incidents (to an external system) is privacy. The problem is that the correlation and storage of incidents might be 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) [31].

Most SIEM environments provide a one-way hashing method to hide this information or use encryption methods. That way, log files still can be correlated while the actual private data is not being disclosed. This problem has been mentioned before in this thesis. A procedure should be set up for this although there are issues there as well.

Such a procedure must be called upon when IT security officers feel confident enough they should read the private / hashed data. An example: Access to credit card records is logged as per PCI-DSS rules. Log entries contain the record number and the user accessing them. Since both can be considered private, they are hashed. In the case of data leaks, unlocking of such records can be requested.

The responsible data privacy officer must give such requests due consideration. At the same time IT security officers must not abuse this procedure to get rid of it: an effective way to abolish this procedure would be to invoke it so many times the data privacy officers will be forced to approve requests without looking into them. This is a risk especially when the procedure contains some kind of response time frame.

This must be considered before SIEM is set up. Not only is SIEM able to track spyware–, spam– and security incidents across the network, it can also say which employees browse Facebook all day instead of actually doing their job. Although that can be very useful information for managers and HR employees alike, it might be an invasion of their privacy.

Current incident management systems

Related to the incident management requirements is the observation that many companies already have event / incident management systems. Every service desk for example, utilizes such a system to keep track of tickets. The alerts generated by SIEM can usually be handled by such incident management systems. Careful configuration and testing is necessary so the SIEM environment will not overflow the existing solutions with incidents.

At the same time, SIEM environments can use these incidents to generate more complete reports. As an example: a common rule in IT Security policies defines password length and complexity. The number of 'I forgot my password'-calls to the service desk can determine the efficiency of that particular policy.

Both responding to incidents and creating incidents as part of the work surrounding SIEM can be done seperately by an incident management team or by a CSIRT[18]. A SIEM environment can be incredibly useful in this whole process although it needs human supervision to establish the correct rules and responses. This is one of the reasons why setting up a SIEM installation will initially *cost* time.

[18]Computer Security Incident Response Team.

Since MS is working on incident management from a different perspective, this subject has not been researched extensively. However, it should be noted that there should be more meta data attached to an event than "this is a security related incident". With the level of information a SIEM environment should have about single incidents it should be able to submit more detailed reports, possibly already assigned to the responsible team or department. Such meta data should include the application, the server and type of error. That information should not just be in the free text field, it should also be in separate meta-data fields.

2.4 Choosing for SIEM

Before we continue with the technical considerations and methodology of SIEM, the question is whether or not it is worth the effort and time to continue with a SIEM implementation. SIEM itself has enough advantages. The reporting, the log management, the overview and all of its features add to a company's IT security status.

Before a SIEM environment was being considered at MS, the IT security officers, network administrators and developers already reported and solved IT security related incidents. The question is now: will the SIEM pick up the same IT security events and report them as such? If not: what is the difference and why is it there? This requires the effort from all involved personnel to keep searching for and reporting IT security related incidents. With that information the efficiency of the SIEM environment can be determined. At this point the actual SIEM environment has not been set up. It is possible to get a (partial) answer to this question already. Collect log files using the newly defined data collection requirements and analyses those for possible security incidents. Even without a SIEM environment, an improvement may be seen already.

If the SIEM environment reports more than what used to be reported manually, the question remains if these incidents are relevant or if they can be ignored. Over time, less and less incidents should be ignored by administrators. The configuration of the SIEM environment should be adapted instead. It is important to note that it should not be a goal to try to get the same number of incidents (per day, per week) as before the SIEM environment was introduced. It might just be that a lot of incidents were overlooked before and by artificially limiting the SIEM environment it may be the same in the future.

For MS thirteen distinct advantages can be listed, that they would enjoy if they would start implementing SIEM. However, these advantages were only written down after the following questions were answered. See also the outline in figure 2.1 on page 22.

1. Is the scope defined in such a way that in order to improve security, compliance or operations, SIEM is necessary? This does not imply that this definition should work *towards* a SIEM environment. It is important to find out if the required effort to improve the situation within the scope can be done easier without SIEM. If that is so, SIEM should not be set up. For example, small changes in firewall management or resource management may be enough to improve IT security within the chosen scope and focus.

2. Is SIEM necessary to answer the questions asked in the use cases? Maybe existing tools and software can answer those questions already. Alternatively,

the information may be already extractable from devices, log files and other nodes. If so, a SIEM environment is not necessary.

3. Log management can take up a lot of storage resources. Is it worth the effort in saving time, money and improvement within the scope and use cases to allocate these resources? In order to properly generate trend information and follow slow patterns and anti-patterns, it might be necessary to collect log files over long periods of time. The question is: is it worthwhile to do so? Keep in mind that some mandatory compliance rules make the answer to this question "yes", regardless of costs while privacy regulations make the answer "no", regardless of needs.

4. Are the system resources available that the SIEM environment will need? Log correlation and report generation potentially use a lot of system resources. Dedicated SIEM servers cost money, rack space and generate a lot of heat. These costs must be accounted for as well.

5. Is it possible to collect all log files? Given certain use cases, very specific or targeted log files might be needed. It might be impossible to collect log files from certain devices. There might not be an agent available for the device and the agentless collection method might be unavailable as well. Such log collecting agents require system resources to encrypt and transport their log files[19]. If critical devices cannot be used, the use cases might fail.

6. Are people actually willing to read and analyze the reports generated by the SIEM environment? It takes effort to tailor the reports in such a way that they are useful to the target audience; even prefab reports provided by the SIEM environment will not necessarily provide the right information. This is one of the reasons (see also page 19 on 'Current event management systems') why setting up SIEM can take a long time: generating good reports is an art. Will it be worthwhile?

7. Will the SIEM environment be compatible with an existing incident management system (if any)? If the SIEM environment generates incidents and saves them to an existing incident management system it is important to try and limit the number of incidents it saves there and it is equally important to make sure that these incidents are properly assigned and processed. Otherwise, all the SIEM environment will contribute is a lot of ignored incidents.

2.4.1 Choosing SIEM at MS

For MS the answers are as follows:

1. Do the scope and focus call for SIEM?

The pressing issue this research has observed at MS was a lack of overview from the perspective of AV. This lack of overview called for his initial interest in SIEM. Although other efforts *could* collect more information on possible attacks, incidents and events than is being collected currently, SIEM is necessary to get a complete birds eye view.

[19] When this was being researched Q1 nor HP could provide me with meaningful numbers as to the impact of such agents. Estimates ranged from "very little" to "10, maybe 20%" of device resources.

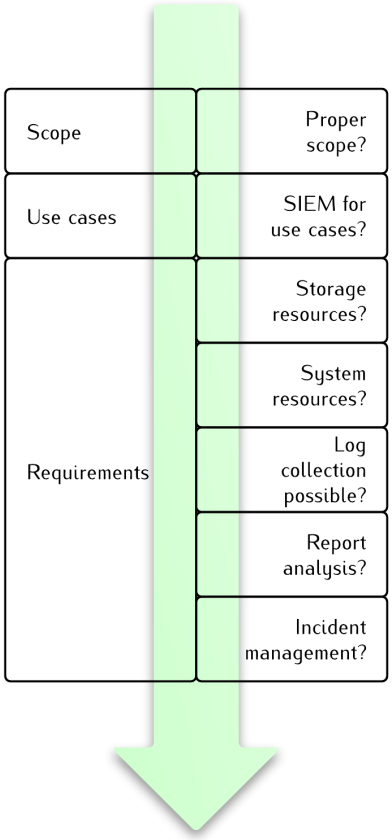


Figure 2.1: Continue with SIEM?

2. **Is SIEM needed for these use cases?**
   Technically the answer to this question is no. However that is always the case: data can be collected, correlated and poured into a report by any collection of tools, applications or people. However, that would not be fast, accurate or complete. Improving on such tooling simply means re-inventing the wheel.

3. **Are the necessary storage resources worth the effort?**
   Yes. Especially trends cannot be followed and resolved with the short term memory that current solutions have. Current scripts and tools designed to keep an eye on possible security problems simply fire off emails or create incidents. They do not keep track of trends or patterns.

4. **Are the necessary system resources available?**
   Within MS the necessary resources can be allocated for a SIEM environment.

5. **Can MS collect from all devices?**
   Network administrators have always warned AV that it would be difficult to collect log files from ALL firewalls, since they generate op to 1.5TB of log files each day. The applications and servers running the web shops generate far

less log files and it is possible to collect and save those. The collection of such firewall logs depends on the SIEM environment which given the resources can easily hold that many log files. The storage of log files is AV's problem, not theirs.

6. **Will the reports be read?**
   Yes. If anyone, AV himself would very much like to read such reports. His IT security officers[20] have also voiced their interest.

7. **What about incident management?**
   There is no definitive answer to this question yet. When the research left MS nothing was known yet about how incident management would work with a new SIEM solution, since incident management itself was being rewritten as well.

As a final remark, MS should focus on the first two use cases[21] before continuing with the third one[22]. They increase in complexity and especially inventory tracking requires much high level information that is not available at the time of writing. It might be wise to allow developers to implement logging for such information before continuing.

This concludes the organizational part of this methodology. The next chapter continues with the technical methodology and implementation. A final thought: It is important to temper the demand for a successful SIEM implementation. At MS, some administrators demanded SIEM should work as planned and solve everything. Not only is SIEM not the perfect solution (none are), it is also a sure way to shoot down SIEM efforts.

IT security can be a difficult field. Most of the times IT security efforts do not add anything except more costs and more procedures. It is the *lack* of security breaches that determines IT security success. Luckily SIEM can help showing how great the difference is that all other IT security efforts make.

2.5 Conclusion

Following this part of the methodology should give the reader a lot of knowledge on whether or not the implementation of SIEM might actually be worthwhile. It skipped over the choice of vendor. That is a very specific quest that is not in the scope of this thesis.

The decision for SIEM is not based on a list of prerequisites. Rather, by gathering knowledge about the IT landscape, the IT security status and the actual goals of IT security managers and stakeholders, the answer follows from this information: given all this, our scope, our requirements and everything: can SIEM *help*? If the answer is yes, SIEM efforts should continue. If not: well, it is then up to the reader to decide what the next step should be.

With the work done on the organizational part of this methodology it was concluded that it would be useful for MS to continue with their SIEM efforts. With a

[20] See also section 1.1 on page 1.
[21] Customer authentication and error-tracking
[22] Inventory and stock

proper incident reporting chain using the existing IT security officer structure and advanced reporting already incidents are better saved and reported on, going up the chain of command without disappearing. There are several reasons for this continuation:

1. None of the existing solutions give AV the overview he requires. The overview people have is limited to their own area of expertise. Developers track bugs, network administrators block malicious users and security officers keep an eye on possible security issues. The hierarchy of MS prevents a larger overview.
2. A SIEM solution can consolidate security efforts from different perspectives.
3. The reports can be useful for stakeholders and administrators alike.

The same can be said about the trends mentioned in section 1.2 on page 3. SIEM environments are capable of reporting and analyzing such trends given proper data retention policies. The better these are defined, the better this particular challenge can be met. Mentioned in section 2.3.2 on page 16 are the challenges with data retention. They are further discussed in chapter 4 on page 39.

Using SIEM as a base for reporting incidents and trends will not only give AV the overview he requires. It will also help other security officers and administrators. MS is currently working on a pilot program for SIEM. The final recommendation the researcher wrote before he left is attached to this thesis as appendix I and details the reasons MS used to continue with SIEM. The organizational methodology is partly based on that work and the research done for that recommendation.

CHAPTER 3

Technical methodology

Assuming the organizational part of this methodology renders a positive answer as it did for MS it is time to understand how SIEM works. So far SIEM was a black box generating reports out of our log files using some unknown magic method. This chapter takes a closer look at the approach most environments have in order to get the promised "reduced costs and improved visibility[1]". White papers and documentation from several SIEM vendors were studied. There were several instances where specific knowledge of how a SIEM environment works was needed. In those cases the technologies used by OSSIM[2] was analyzed. OSSIM was chosen since it is an open source SIEM solution with an excellent reputation [19] and various advanced features. Their technical documentation and the nature of their product allowed for more details to be included here.

Figure 1 on page xi shows the three steps of the technical methodology. Each step may make it harder or even impracticable to implement SIEM. By following these technical steps as well the final decision to apply SIEM (or not) is justified from all angles.

3.1 Technical overview

3.1.1 Agents

At the top of figure 3.1 on page 36 some of the sources for log entries are visible: routers, servers, firewalls, applications, etc. These can be both hard– and software devices. For the sake of clarity, these can be *other* SIEM applications as well. AlienVault for example uses "Sensors"[3] which keep an eye out on a select part of the network and provides a central SIEM core with correlated information from multiple devices. Other SIEM applications can be a source for events as well. Just as Sensors collect information from multiple agents (see further ahead), other SIEM applications can provide correlated and normalized event information to a main

[1] As found on `www.alienvault.com`.

[2] OSSIM is the open-source SIEM environment on which AlienVault is based. See also `http://communities.alienvault.com/community`.

[3] See `http://www.alienvault.com/solutions/architecture/sensor/`.

SIEM application which then summarizes information from those other applications as well.

A source can be provided with an agent. An agent is a piece of programming, an extension or a plugin provided by the SIEM vendor which is capable of transferring and converting the log entries from the target system to a SIEM application such as a Collector (see below). The key features of such agents are their ability to pre-filter log entries based on for example their severity and ability to normalize log entries so they can be more easily compared by the Logger. Agents send their log entries over dedicated secure connections.

The alternative is agentless. In agentless solutions the device is capable of sending the log entries to a Collector themselves, thus mitigating the need for an agent to be installed. Syslog for example is implemented in many devices and when those entries are being sent to a central Collector the net effect is the same except certain security related advantages.

When agents are not available it might be possible to set up a device in such a way that the Collector can retrieve log entries over the network. However, this solution is not preferred. Agents are available for thousands of devices and often set up quickly for new or unknown ones since the logging of errors and warnings is a fairly low-level operation: usually the logs end up somewhere in plain text and reading them is trivial.

Log collection generally uses EPS[4] as an indication of how many events are generated (by a device) every second. There is an average EPS, a total peak EPS and an average peak EPS. The total peak EPS is the number of events per second when a large attack is happening such as a DDoS. The average peak EPS defines how many events per second an average attack generates. In order to get these numbers, the PE (Peak Events) and NE (Normal Events) per second need to be defined. These two numbers make up the EPS variables. These numbers are important when setting up data collection since the Collectors, Loggers and SIEM applications used may be limited in the number of events per second they can handle. This limitation can be technical (the server cannot handle it) or artificial (license based). The details on how to define the EPS are outside of this thesis and can be found for example in [3].

3.1.2 Collectors

Collectors are an intermediate between the actual SIEM application and the hundreds or thousands of agents spread around the network. The pre-mentioned Sensor is technically such a collector. Collectors are capable of some correlation although they mainly do normalization work so SIEM application(s) further ahead receive more structured log entries. If possible this is combined with source information provided by the agent. Once classified the entries can be send to the Logger and / or the SIEM environment itself.

Collectors can run stand-alone or combined with the applications below in a single instance. Again, this split in functionality is provided so even very large networks can be properly monitored.

[4] Events per Second.

3.1.3 Loggers

Loggers are the first step towards a full SIEM environment for the reason that they provide log management. With a logger entries are saved, secured and signed for future analysis. It supports rudimentary search and alerting capabilities and it depends on the hardware, not the software, how much data can be saved. With just a Logger forensic research into log entries is possible.

They differ from SIEM applications in a few key ways. First, they are more focused on the storage of log files rather than the analysis of log files. Security standards and regulations such as the PCI DSS [23], the Sarbanes-Oxley act [30] and HIPAA [29] (the latter are US laws) and many others require organizations to collect and analyze log files from many sources for which Loggers are used. By reason of these requirements Loggers are capable of processing more events than SIEM applications are. All they need is storage; the search capabilities they offer are not that extended.

3.1.4 SIEM applications

A possible (second) target for the before mentioned Collection is a SIEM application. It can handle less events per second although it is capable of doing more with those events. It is usually the end-point for all mentioned applications and thus the core of the whole system. The SIEM application is the core of a SIEM environment. It handles things like:

- Risk assessment
- Event correlation
- Vulnerability Scanning
- Data Mining
- Real-Time Monitoring [20]

A SIEM application generates the reports and shows the overviews. Although it can handle less events than a Logger, it is important to realize that a SIEM application can rely on other devices to do event and flow processing. In the end, this application is where the scope, use cases and business rules are set up.

With this overview it is easier to find out if SIEM is indeed feasible.

3.2 Data collection

SIEM uses several techniques to collect log events from connected devices. Each of these methods has their advantages.

Data collection is the largest challenge for all use cases. Potentially log collection can occur from each device in the company. Every router, every firewall and every server would route their log files to Collectors, Loggers and SIEM applications which in turn would collect all these log files, generate their own log files which would also require collecting etcetera. This is an unworkable situation.

Starting at the scope however, we can draw some useful conclusions.

3.2.1 Collection using agents

Agents are processes running on a device. These processes can collect log events from multiple sources such as multiple applications and normalize events so they can be compared more easily. They are sent over a secure line to the Collector. This collecting and sending may cost a lot of resources. Setting up a secure connection and normalizing events costs CPU power. Sending those events over the network will cost bandwidth.

It is important to consider these costs. If the log entries are sent through over the network, they will no longer take up space on the guest device. According to a researcher at HP, agents can cost up to 10% of the resources of a running system. These costs come from the following activities:

- Collecting log events, possibly from multiple sources.
- Normalizing events for easier correlation later on.
- Filtering out events, as determined by the security policy.
- Transferring these events, over a secure connection.

These steps may have to be repeated for multiple log sources on one device such as a server running multiple applications and depending on the total EPS, the agent might have to save many log events locally before sending them out in burst. Agents can handle virtual machines without a problem although they may have to be installed outside of the virtual machine(s) in order for the agent to work smoothly.

The greatest advantage of agents is their capability to quickly adapt to changed rules. If the SIEM environment is set up to focus on something else (a changed scope or different requirements), agents will pick up new instructions automatically.

For MS: for the scope and focus mentioned earlier the web shop and the application could easily be monitored using agents. Experimental systems are by no means legacy systems: there are no old or unused software products being used any more. If the application(s) run in the cloud or on virtual machines, one agent could monitor several servers. Although MS has limited resources for such agents, specially on shared devices it is feasible to run them on the most critical systems.

The exact device(s) that need monitoring first can be determined by the requirements set forth for data collection.

3.2.2 Agentless collection

Agentless collection involves all methods in which the device sends the log events to a Collector or where the Collector retrieves all events from a network share, – drive or another (protected) source. Instead of an agent the collector will take on most of the normalization and categorization tasks. This also means policy rules will be applied by the Collector and not by not the agent.

This approach should be used only when there are no agents available for the specific device (which is unlikely, unless it is an old or mission critical device) or when the device is running on or close to its maximum capacity. Most `syslog` daemons are capable of sending their events to another device and if that fails it might be possible to open a network drive or a share. The latter can be very unsafe. The share should be protected using authentication.

The largest advantage of using agentless collection is that the original device can keep running, without changes to its system. A disadvantage is the possibility that the log files have to be transferred without encryption or compression and that the host system must save the log files, at least until the SIEM environment has collected them.

Agentless collection is less feasible for MS. Since IT resources should be used for the primary process first, it might not be possible to allocate device resources for the transport and encryption of log files. This is not different from log collection using agents.

Such log collection however is less efficient and less secure. Syslog for example sends its messages in clear-text. It would need another layer on top of that to encrypt the messages. Furthermore, since filtering is impossible (an agent could decide not to send events based on a policy) *all* messages will be encrypted and send through. Then the question not only becomes "can we set up syslog in such a way?", it also becomes "can we spare the bandwidth and device resources?".

3.3 Data storage

In order to describe data storage we look at the technologies powering large databases such as KDD[5], distributed data and data warehousing.

KDD is a way to go from low level data to high level information in the form of reports, models and approximations[10]. It describes the process of searching large volumes of data with the goal of finding patterns[10]; patterns which can be considered as knowledge about the data. This way of searching and processing is developed out of data mining and shares many of its methodology and terminology. For SIEM environments the application of new knowledge from this field is very simple: when normalization has been applied patterns in the event logs are more easy to find. SIEM uses these technologies to connect events to each other and generate reports. This is a large advantage a SIEM environment can have over its competitors: the ability to give (rule based) information based on all that data. The Logger is responsible for the storage and signing[6] of raw event logs. They provide interoperability with existing SAN/NAS solutions in order to provide better scalability.

The main technical consideration for data storage is hard drive space. SIEM environments allow for both vertical and horizontal load distribution: it is possible to set up multiple agent / Collector / SIEM / Logger sets (vertical), as well as setups where multiple Collectors report to one SIEM application (horizontal). The preparation from section 2.3.1 on page 13 should be used here.

With the space and resources allocated, data storage and data processing are a matter of budgets. Effectively, the question should be forwarded to the SIEM vendor who would be delivering the SIEM platform: how much resources do they need for such-and-such amounts of data?

[5] Knowledge Discovery in Databases.

[6] Used in digital forensics.

Thanks to the EPS calculations, requirements and reporting structure such questions can be asked in a very specific manner. During the research at MS there was no direct answer for this question although MS is confident about getting an accurate answer in the future.

3.4 Data processing

Data processing in SIEM environments only deals with log events. The sources can be agents, collectors and SIEM applications. In order to explain how events are processed, the following sections will walk through the process; figure 3.2 on page 37 shows how this process is organized. The section numbers match the numbers in the figure. Data processing is the core of SIEM. Here the advantage of SIEM over other processing mechanisms can be seen.

3.4.1 Receiving and categorization

Events that are received by the SIEM environment are previously collected from log files by agents or are collected agentless[7]. By standardizing those events the SIEM application can deal with the log data in an equal manner. Storage and processing are simpler and more coherent. [22]

SIEM agents are capable of processing a variety of different types of logs. They can handle `syslog`, the standard logging solution for Unix and Linux systems, as well as `SNMP`[8] for example. The preferred way to collect log files from Windows machines is NtSyslog or Snare which work the same way as `syslog` does. Not all protocols supported by SIEM agents are encrypted by default. Agents and SIEM applications connect and communicate using encrypted lines (these can be provided by stunnel, ssh, openVPN or any other remote encryption method). For improved security, agents should be used to transfer log events.

Agents send events to SIEM servers. Most SIEM servers differentiate between a number of types of events. As an example, AlienVault distinguishes the following events and categorizes events as such:

- "Normal" events: these events are the ones that come from all the agents installed across the network.
- OS events: These events usually come from p0f[9] detectors. p0f is an OS fingerprinting tool. p0f identifies the system on machines that make a connection, machines that connections are being made to and if possible, machines that connect through the current system. These events constantly keep track of the OS's on the network. Changes indicate possible attacks.
- MAC events: These events usually come from `arpwatch`. `arpwatch` keeps track of the MAC address that is associated with an IP address. The MAC change events are very useful to detect ARP poisoning[10] and other layer 2 attacks.

[7] Technically, agentless collection still works using agents. Log collection using `syslog` sends its log entries to a SIEM server's `syslog` instance, where the entries are processed by an agent running locally. Since no agents are installed on the system itself however, it is called agentless

[8] Simple Network Management Protocol

[9] Passive Operative system Fingerprinting

[10] Address Resolution Protocol poisoning or —spoofing is a technique whereby false ARP messages are being sent over the network.

- Service events: These events keep the SIEM database updated with the open ports in the deployment network machines. This makes it possible to carry out cross correlation and store an inventory of all the machines in the network. [21, 12, 33]

Once the events are received and categorized, certain policies can be applied.

3.4.2 Applying policies

There are two types of policies. First, the general server policy defines what should happen to all events the server receives. The events may be correlated, cross correlated or reprioritized, forwarded or stored depending on the server role. If there is one central SIEM server, these server policies are not applied. More specifically, all tasks that exist for events (normalization, risk determination, etc) will be applied to all incoming events.

Besides these server policies it is also possible to set up specific policies for specific events. Events coming from specific agents or specific devices can be set to be correlated (or not), stored (or not), etc. The usage of such policies allows administrators to fine tune the flow of events, especially over multiple SIEM servers.

There are two advantages of setting up such policies in a proper way. First, they allow for filtering of events. Some agents may generate many events which need not to be stored while it *is* necessary to correlate the event with past events. Policies also define the priorities that events have, for risk determination later on.

The second advantage has to do with multi-tier SIEM installations. When there are many more SIEM servers in the network, it might be useful to set up server policies that allow a strict group of servers to do the correlation, forwarding the events to other servers that handle the storage.

Server policies automate the process of keeping track of what is important. Log sources will not be overlooked but diligently checked by the SIEM environment. Policies can be changed on all SIEM servers which makes maintenance easier. The events that should be normalized by the server end up at the next step.

3.4.3 Normalization

There are two properties that are normalized: the priority and the reliability.

First the priority is normalized. The risk assessment uses this priority and it must be equalized. Policies are used to define the priority. If the event is not mentioned in a policy, the default priority is used (a database is checked). Events cannot determine their own priority.

The reliability defines how reliable it is that the attack will be successful, assuming the event signifies an attack. The reliability is defined using policies and a general knowledge base provided by the SIEM vendor.

If the event has been normalized already, it is not normalized again. This is important because events can be received from other SIEM servers as well, and that would mean events are being processed two or multiple times.

The priority and reliability are also influenced by the cross correlation done in the next step.

3.4.4 Correlation

Correlation is very important in SIEM. It greatly reduces the number of false positives. It is used to reduce the (possible huge) number of events down to a limited number of alarms and events using various methods of correlation.

Cross correlation

Cross correlation is applied to events that have a destination. It checks if there are known vulnerabilities for that destination. If so, cross correlation can redefine the priority of an event. This process is usually executed as part of the normalization of an event. The aforementioned OS and MAC events cannot be cross correlated because they have no destinations, only a source.

An example: An event is received from an IDS signifying attack vector *x* on IP address *z*. If the SIEM environment has knowledge of that IP having a vulnerability and the event contains traces of that same attack the reliability of the event is increased to the maximum level.

There are different characteristics saved for each destination address. The operating system for example is saved when an OS event arrives[11]. SIEM vendors provide known vulnerabilities for OS's and software installations.

Custom correlations can be defined as well so different alarms can be raised. This is important when working with use cases and data collection. As an example, motion alerts raised by security camera's can be intercepted using custom correlation.

Inventory correlation

Inventory correlation is a collective noun for several types of correlation: OS –, port –, protocol –, service – and version correlation.

The incoming event may signify an attack. The priority of the event can be further defined by trying to determine if the destination of the event is vulnerable to the attack in the event. The more matches are found (for OS, port, etc) the higher the priority. In general the service/version correlation is the most important one. If the version of the destination matches with the events attack vector the priority of the event will be increased significantly.

Logical correlation

Logical correlation allows administrators to raise alarms without knowing the specific type of attack. It allows the combination of events such as "If *a* and *b*, but not *c*, and *a* stays connected to *d*, raise an Alarm" [21]. Such correlations are called directives, and they tell the SIEM server what to do.

In OSSIM, these directives consist of rules. There is an opening rule which enables the other rules. An example would be "a connection from any server on any port to any server on ports 25, 80, 135 or 137[12]". If this rule is matched, other rules from the directive can be applied.

One such a connection probably does not constitute an attack. The follow-up rule is: "more than 300 connections over one of these ports". If matched the reliability for this event (which indicates the likelihood of an attack) is increased. Then other

[11] OS events cannot be cross correlated but when OS event `A` tells the SIEM server the operating system for destination `X` we can use that information when "normal" event `B` has destination `X`.

[12] Ports commonly used by worms

rules from the directive can be run by the SIEM server as well. They may increase the reliability (or decrease it).

The rules inside directives generally increase in complexity as they are matched. The opening rule is applied to each event that is received by the SIEM server while third– of fourth level rules rarely get applied.

Logical correlation is costly to set up because writing good directives is very time consuming and requires a lot of domain knowledge.

These three types of correlation correlate between events and destination vulnerabilities, between events and destination characteristics and between between events from different sources respectively. Each type of correlation can change the priority and reliability of events and greatly reduces the number of false positives for each event. Once the event has been correlated, the risk can be determined.

3.4.5 Risk determination

In SIEM the risk of an event is determined by three factors, two of which have already been determined: the priority and the reliability. The third one is the value of the asset: the server holding the customer records is a valuable asset.

Every event has two risks. There is a risk for the source (measured as the probability the machine is compromised), and a risk for the destination (measured as the potential risk due to attacks launched against the machine).

> These two variables exist because they characterize different situations: the level of attack indicates the probability that an attack has been launched, an attack that may or may not be successful; while the level of compromise provides direct evidence that there has been an attack and that it has been successful.
>
> The importance of both variables depends on the situation of the machine. Mainly due to the exposed state of perimeter networks, which are exposed to an enormous number of attacks, most of them automated, unfortunately a high level-of-attack value is a "normal" situation. However, any sign of compromise or movement that might lead us to think there is an attacker residing in these networks should be immediately pointed out and reviewed.
>
> On the other hand, there are cases in which a machine generates anomalies within the network due to the nature of its function (such as a security scanner, a service with random passive ports, bad-configured DNS servers, and so on) and these will normally have a high C and a low A. We must bear in mind that although a machine may have a very high C value, it doesn't mean that it's always compromised, this machine just needs to be fine tuned. [21]

The risk is calculated using the priority, reliability and asset value (for both the source and the destination asset). The highest of these two values is used as the final risk factor. For OSSIM, this is a number between one and ten.

When the risk is higher than a certain threshold, an alarm is raised. This alarm can be related back to all the events involved, from risk calculation back to the policies. Figure 3.2 on page 37 shows that this alarm is treated as another event

and can be sent to the next SIEM environment. This is true for multi-tiered SIEM environments.

3.4.6 Storage

If the server policy allows it, all events received by the SIEM environment are stored in a database. There are usually two types of databases. One is for general storage and/or subsequent analysis. The other one is a backlog for events that are still being correlated. SIEM environments are capable of forwarding the general storage to any type of compatible storage system. Data warehouses or NAS's can then be used as external storage. See also the previous section on data storage.

In a multi-tiered environment[13] server policies determine which of the above steps should or should not executed.

At the moment of writing there were no specific demands for data processing coming from MS. The effort for MS lies in data collection and storage, not processing. So, for MS data processing requirements did not raise any objections.

3.5 Result

This chapter shows the technical overview of a SIEM environment and discussed data collection, data storage and data processing. Each SIEM vendor has a different approach to them although the essence remains the same: log files are collected, analyzed and stored given certain rules and procedures. Some of those are defined by the SIEM vendor. Others are defined by server policies, use cases and requirements.

Each of these three steps came with different questions. For data collection it is important to find out which device logs are needed to answer the question(s) in the use case and the subsequent reports. The business rules written for the organizational methodology are good sources for this. Data collection also requires technical knowledge to get to the result: the collection methods that will be used to collect log files each device.

Data storage mainly deals with databases. There are no specific considerations to make regarding SIEM: the implementation is handled by the storage-environment itself. It is only when there are a lot of events and a lot of devices that are being monitored that it may be necessary to step away from the default SIEM storage implementation and move to data warehousing or other storage solutions.

Data processing is influenced by the size of the use cases. Complex use cases and complex reports require much correlation from many devices, possibly using custom correlation and risk determination. The server – and event policies may need to be adjusted to make sure that this is possible, and may even require a multi-tiered set up. This can be determined by comparing the SIEM capabilities (per server) with the demands as defined in the organizational methodology.

[13] Such as discussed in Data Storage in section 3.3 on page 29.

If the SIEM environment is set up there are two products: a dashboard / management console and a set of reports. The dashboard provides a general overview. It mentions the current security status (based on alarms and events), it allows the manager to change policies and settings and it provides tools for analysis of security events and raw log files. This dashboard is usually web based, running on the SIEM server that has been set up to be such an end point.

The reports are usually available in the dashboard as well, but can be set up to be mailed to specific persons or a group of people. Reports have a date range, a range of assets to be included and custom alarms and events that signify something noteworthy. They can be build to custom specifications, based for example on the business rules set up in the organizational part of the methodology. There is no one-on-one correlation between the business rules and the creation of a new report. That could be a subject of further research.

For MS there are still issues surrounding data collection. It is a difficult topic mainly as a result of the shared devices, such as routers, firewalls and (virtual) servers. Data storage and processing do not have these problems. A central Collector or SIEM application only needs resource (storage and system) to do its work. It does not matter if the SIEM environment is set up distributively. Storage and processing requirements are limited by budgets and physical space and not by technology.

More than the organizational part of this methodology is the technical methodology based on research and less on feedback from MS. However, combined with the organizational questions the final step can be made. The choice for a SIEM solution (if any) and the best approach towards such a solution.

In the next chapter the obstacles for SIEM found at MS will be discussed. Similar obstacles may exist elsewhere too.

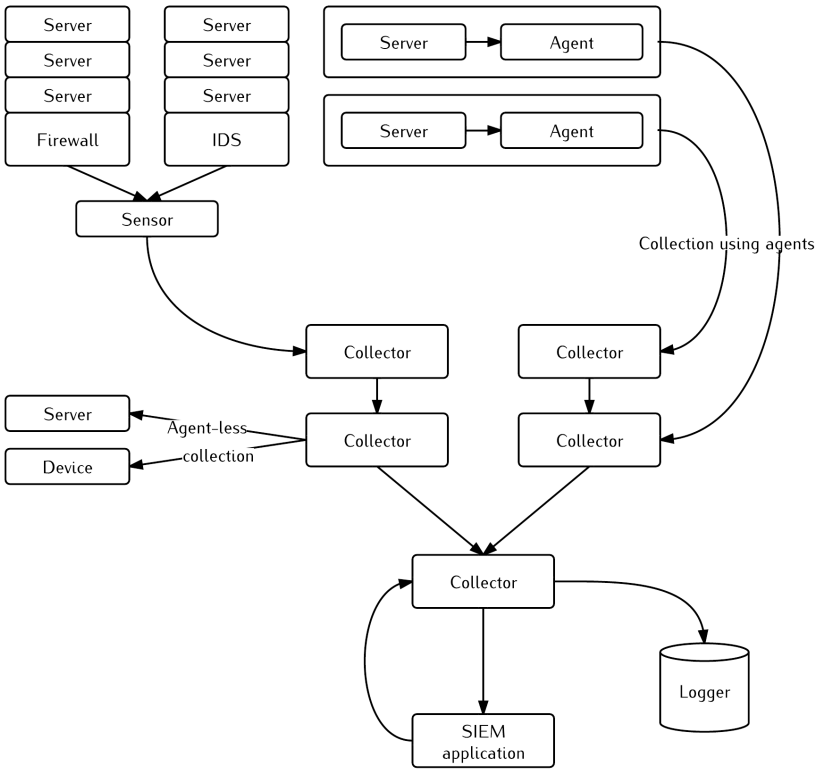


Figure 3.1: An overview of a full SIEM environment

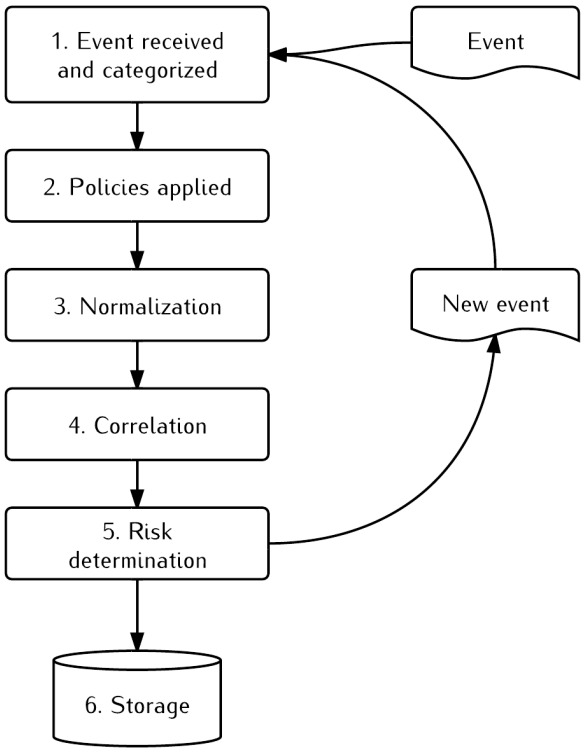


Figure 3.2: SIEM data processing. Figure 3.1 on the facing page reflects that new events may be sent back to the SIEM application through the Collector, not directly.

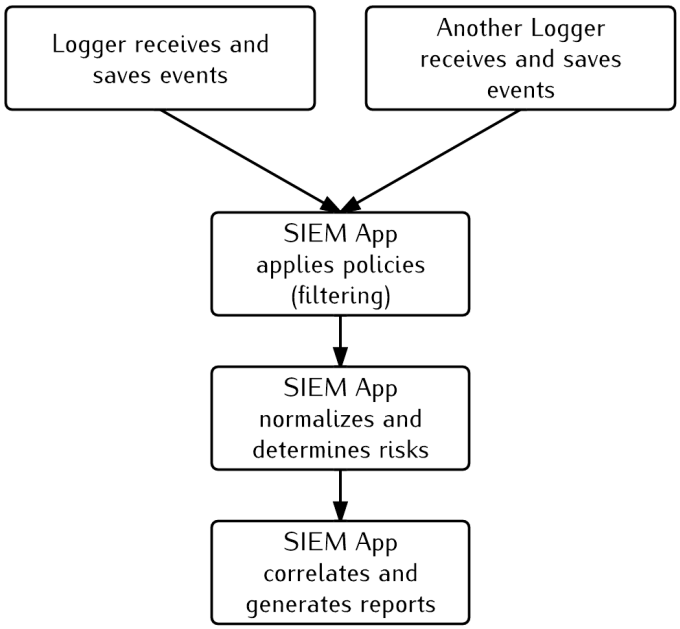


Figure 3.3: An example of multi-tiered SIEM processing.

CHAPTER 4

Obstacles

The previous chapters worked through the methodology and discussed several challenges, possible solutions and other issues. Although SIEM can address the issues AV has (as mentioned in "organizational challenges" on page 3), there are still some other obstacles that need to be addressed regardless of SIEM's actual future implementation.

With all the security officers in the three business units of MS (as explained on page 1) it is very tempting to simply pick one of them and use their department as an experimental site for SIEM.

The problem is that no matter how small that site is there will always be tens or hundreds of different systems down the line. There are many resources being used in each of these departments. An iPhone application for example is a very solid thing to grasp for AV. A focus could be that application. However while the application is communicating with MG it touches many systems: network systems, storage servers, other applications, inventory stocks and their applications, tools, programmers and debuggers. All these items are somehow connected to the one iPhone application. No matter how narrow the focus is MS will always hit many systems in many layers, managed and checked by people who may not even want SIEM.

Privacy is another barrier. The collection, storage and analysis of log files brings a lot of privacy concerns to the table, especially with today's privacy laws. The issue is comparable to the ever increasing number of CCTV cameras hanging in streets and cities. They might prevent crime[1] although they also invade people's privacy. Likewise, these SIEM correlated and analyzed log files may help prevent digital attacks or compliance issues, they also invade the privacy of the entities in the log files. If those are devices, there is no harm. What does a router know of privacy? When tracing people the issue is more pressing. Proxy server log files for example, may tell where a digital virus entered the company's network, they will also tell which employees are spending most of their time on the internet. Another example is the tracking of customer data and access thereof. Then there are names in *two* places. The original record and the access record.

[1] Although that is debatable [11].

This means there are two sets of records a company needs to protect. With the aforementioned privacy laws, this is not a simple task. Especially not in Germany, where this is an even more pressing concern. If the policies regarding that data are not solid, the whole project will fail.

Another remaining issue is that the people who are currently responsible for databases, servers, routers and other devices are already drilled in security procedures and technologies. They already scan log files, secure files and connections. The first advantage of a SIEM environment is for AV. From the perspective of these technical administrators all security measures are taken care of and AV is coming in with a SIEM environment which will simply reinvent the wheel. On top of that this wheel needs up to 10% of their resources to run agents and forward log files. It does not matter how this problem is solved. At MS it was observed how people felt checked and mistrusted after SIEM was mentioned. These stakeholders feel they do a good job in preventing security issues and problems. The problem is that it is very difficult to convince them that SIEM can work for them as well.

Another problem can be summarized as statistics. Large networks such as the one MS has, get attacked all the time. There is not a moment where there are no hackers probing for access or running port scans. Sometimes malicious users actually manage to breach a test server or a honey pot. These sort of problems have always existed within MS. Bringing in SIEM will not change that.

Over time the network administrators have gotten used to a steady flow of a hundred malicious access attempts a day. They know the heartbeat of their network. They will only respond when the attacks increase five or ten fold.

AV and the other security officers do not know this heart beat. They generate reports using SIEM that tell them that MS gets attacked 24/7. The administrators now face the task of educating these security officers so they know that deflecting fifty hackers a day is not an exceptional thing to do. That may take much of their time. Their workload will increase as they have to fine tune reports until reports match the IT security situation more properly.

Two problems come from heart beat / statistics problem. First, valuable time is taken from network administrators who have to keep explaining the anatomy of their network and devices. Second, with all these statistics there is the risk of micro-management trying to get attacks down and security up.

4.1 Possible solutions

This chapter started with four problems. The last one is a bit of a combination between two problems:

1. The fact that any combination of scope and focus will *still* affect a lot devices as they need to be included in the SIEM effort.
2. Privacy issues with log collection.
3. The problem that SIEM may seem to overrule local IT security efforts.
4. the problem that SIEM allows IT security officers to keep very strict tabs on administrators, which might lead to overzealous (micro) management.

Luckily these problems can be prevented.

1. Implement SIEM in *new* devices and software. That way SIEM will gradually become integrated in the network. New routers and must be equipped with an agent, software must be able to keep log files. That way the impact of SIEM for a new scope and focus will become less and less severe since more and more devices already have an agent anyway.
2. Often mentioned is the technical solution all SIEM environments hold with which they are capable of hashing (one way) or encrypting privacy sensitive information using regular expressions to match the private data. The solution is not just simply turning this feature on. The solution is to design a procedure that allows IT Security officers to get the decrypted information from the log files, overseen by Data Privacy officers. An artificial limit on the number of requests each [day / week / month] will keep abuse down[2].
3. The custom solutions administrators have developed themselves should stay. Their domain knowledge and experience will be better than anything SIEM can offer. If there is a need to see these custom solutions integrated into the SIEM environment, make their output compatible with the SIEM environment. An agent installed on (several) server(s) may be best helped with low level log entries although high level summaries can be very useful as well.
4. It is unavoidable that there will be administrators who feel the new SIEM environment is spying on them. This is a political issue although it is relevant none the less. Reports should not mismatch the actual environment. If they do the business rules defined for the reports do not match that environment. This is mentioned in section 2.4 on page 20 as well. The best solution is to give the administrators access to the reports and the rules that make them. They have read access to the raw log files any way and are able to tell whether the reports match the actual situation.

[2] See also section 2.3.4 on page 18.

CHAPTER 5

Conclusion

In this thesis the implementation methodology for SIEM is being discussed, using existing research and work complimented with research at MS. Both have been worked into a two-part methodology. The previous chapters discussed the inner working of SIEM technologies, the consequences and the results when applied properly. It is important to note that just like most IT security technologies, applying SIEM will not make a network more secure. All the logging in the world cannot stop malicious users if nobody acts on alarms. The goal is to make sure that nothing suspicious is happening on a network. But it is impossible to prove that malicious events *may have* occurred when no IT security measures would have been implemented.

Simply setting up SIEM servers and installing agents will only add overhead to networks and devices.

Data collection is an important factor of this overhead. It is difficult to balance between overzealously logging everything and sitting in the dark, not knowing what is going on. Storing log files is costly as well: the private data in log files must be protected. Processing all those log files will cost time and energy. The processing of log files is the easy part.

Implementing SIEM can be an expensive process, but it does not have to be. If a small scope and a small focus are chosen (after which all the steps are followed) the costs will be low (both financial and resource-wise). Section 5.1 on page 45 mentions several observations which can be used to evaluate the SIEM implementation / environment so far.

Figure 5.1 on page 47 shows the structure of this thesis again. On the left the methodology can be seen again that was followed in the previous chapters. The organizational part answered questions surrounding the need for SIEM. The technical part answered questions surrounding the possibility of SIEM.

The figure shows a Yes/No choice for both parts. It will not always be that easy to answer the question of SIEM. At MS the four obstacles observed (shown on the right) during the application and development of the methodology show a complex situation. The important thing is that these issues are now visible and can be addressed.

Even if the results in other situations are similar, with more questions raised than answered, the methodology has been applied successfully. By going over these questions and answering them, one works towards a situation where the decision for SIEM can be positive.

The previous chapter discussed the four obstacles to SIEM that were observed at MS.

1. The implementation of SIEM in a complex IT network can be very difficult, unless it is made a policy to implement SIEM in new devices and servers. Then SIEM slowly becomes a part of the IT landscape.
2. Storing log files means storing (even more) private information. SIEM environments have the technology to encrypt or hash that information. Set up a process with artificial limits (to prevent abuse) to make that information available again.[1]
3. SIEM makes custom solutions and existing security measures absolute, unless they are incorporated into the SIEM environment.
4. Administrators and managers may feel spied upon. If they are given access to the reports generated by SIEM they can benefit from them and improve the accuracy of the reports.

When this methodology was first developed, during the researchers time at MS, it was assumed that SIEM would *always* have a useful place within the IT (security) landscape. That is not so, and the methodology reflects this.

The main reason for this is domain knowledge. Although SIEM is superior at correlating large quantities of information coming from for example firewalls, IDS's and server parks there are a lot of smaller sections within MS where dedicated administrators are far better at gauging the security status of their devices. They control less devices (which makes it manageable for them), have constructed dedicated security programs (which do exactly the kind of automated work SIEM does) while their knowledge and experience give them an edge over SIEM. Installing SIEM agents on such servers would undo all that work for just a small gain in IT security. This applies to the central phone servers for example, and some of the network storage servers. SIEM should be set up to receive alarms from these area's and nothing else.

There are places where SIEM would have a useful position. The maintenance of international data communication lines is for the most part out sourced to phone companies and other communication specialists. The lines are checked every now and then using manual pings back and forth. The results are not saved and when something goes wrong a cryptic email is sent to the responsible manager. A SIEM environment cannot force these checks to be executed but it can store, process and correlate the results of such pings and instead of sending an email, it can raise an alert. Not just when the line is down but also when the transmission speeds are slower than they should be, possibly indicating future failure.

[1] If a SIEM environment hashes the private data it will be impossible to retrieve it. Usually though, these systems have an encrypted database with the original data and their hashes.

SIEM unlocks information. Managers, administrators and developers alike can learn very much from the reports and information a SIEM environment can visualize. Especially in a large (and international) company like MS there is a lot of IT and a lot of data going back and forth and SIEM can show the heart beat of the system. The largest advantage is that any administrator can get an overview.

The knowledge about that network and all the information thereon is very local. There are islands of knowledge all over the company. There are departments for data storage, for networking and for applications and services. This division becomes even more fine grained when one takes a closer look: international data networks, redundant backup networks (within one data center), server network, NAS networks, etc. Each of these fields have their own experts with specific domain knowledge. SIEM does not only unlock information, it can also consolidate these islands. When SIEM is integrated in a great enough part of the IT network, a new kind of overview gets created. Not one where one department or one service is being analyzed, rather one where these islands disappear.

SIEM adds overhead yet creates knowledge. The amount of overhead can be limited by being smart with requirements, use cases and data processing. The amount of knowledge can be maximized by collecting the right information and by writing / generating smart reports. With this methodology, both can be accomplished making SIEM an essential part of any IT network.

The methodology consists of two parts: one technical and one organizational. Both have three steps, which are connected. Data collection requirements can be too broad, which results in storage problems for actual data collection. They can also be too narrow, leading to the need for non-collectible information. Whatever the case may be, storing log files has an impact on the privacy efforts of a company, as well as what is demanded of IT security. Privacy laws are strict in regards to log files (as mentioned before) and log files need security too, as they may contain much confidential information.

Processing all those log files will cost time and energy. Not only are those costly and possible scarce the SIEM environment might start to lag due to processing times. The usefulness of reports diminishes quickly when the information is outdated. Another consideration is the type of reports generated. When they are all snapshots in their own right, it might not have been necessary to collect and process many log events; a shorter lifespan might have done the job as well. Reports which show trends need more events to be able to detect such trends.

Another important issue is data presentation and the target audience of the reports. If the audience is non-technical, it would seem there is no need for a drill-down into more detailed (technical) information. However, even non-technical audiences might feel the need to dive deeper into reports, get more information and see what is going on. They may not require just an overview. This however, all depends on the amount of data actually collected.

Now should be the time to search for a vendor.

5.1 Start small, think big

This methodology has a waterfall-method feel to it. A sequential set of steps to be followed strictly. However, that is not how this method should be applied.

Start small. Choose a small scope and a small focus. Continue with the steps and get SIEM up and running and evaluate. It might be difficult to say if SIEM

works as intended. Rather there are several symptoms that indicate that SIEM is *not* working as intended:

1. If it is often necessary to sift through raw log files, the reports do not tell what needs to be told and the rules for reporting should be rewritten.
2. If other sources (network administrators) are reporting incidents before the SIEM environment does, its rules, alerts and policies are set up wrong. It means the SIEM environment disregards relevant information and does poor risk assessment. This is also the case when incidents raised by the SIEM environment are being dismissed too early, sometimes without checking.
3. Good network administrators with solid knowledge of the systems under their care will be faster than a SIEM environment when it comes to finding the root of an incident; people can correlate very well. However, the SIEM environment should encourage this, not be a competitor. Reports and data must match the administrators efforts.

If none of these symptoms are found it is time to expand. Choose a wider focus, a wider scope and start with the first steps again. Iteratively, SIEM can grow until it encompasses every device in the IT landscape.

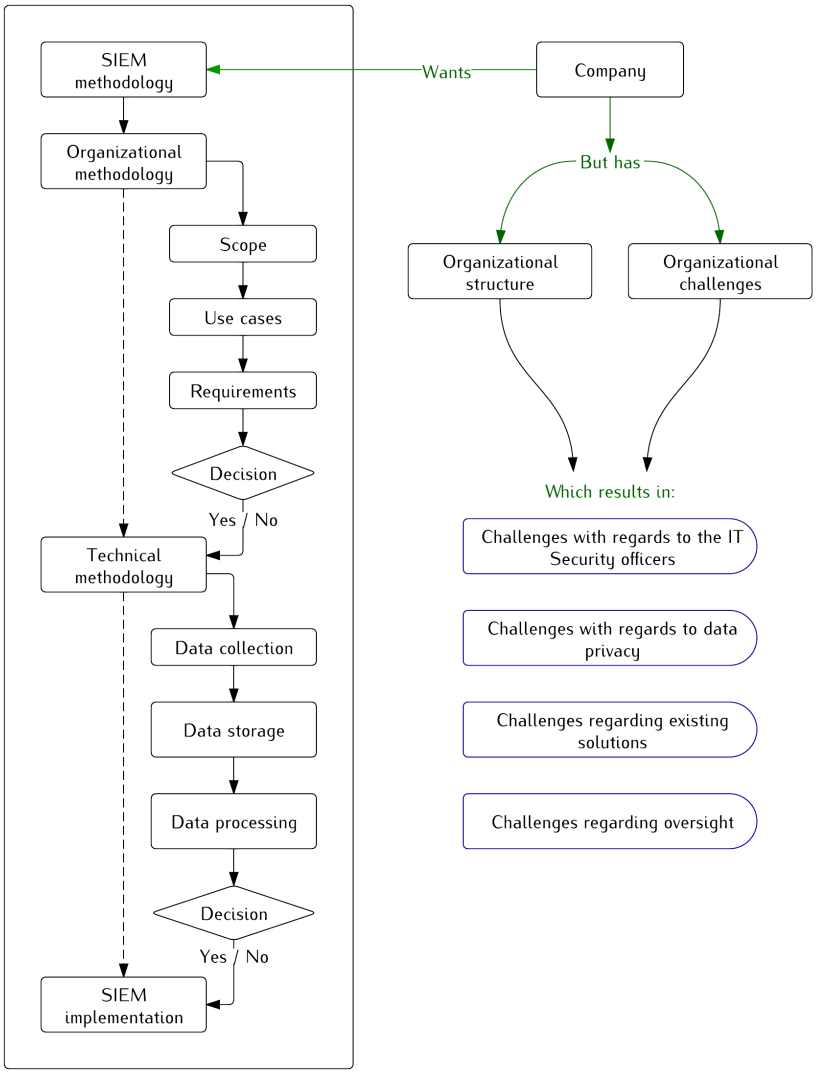


Figure 5.1: The overview of this thesis.

APPENDIX A

Final recommendation

This is the final recommendation Sander Dorigo wrote for MS*, detailing his ideas for SIEM.*

Executive summary

This document contains my advice and recommendations for MS in regards to the appliance of SIEM[1].

The first chapter shows how the market for SIEM software is now mature. Then I argue why e-commerce and markets are the best place to start with a SIEM solution. There are three important arguments for running SIEM software.

1. Firewalls and other intrusion blocking systems do not see patterns, nor do they report on events that do not brake their rules. Thus, attackers disappear "under the radar".

2. Events created by SIEM can provide extended information on possible incidents. This is useful for incident management.

3. Managers and auditors can use SIEM to provide a bottom-up or top-down analysis of the current security status.

These and more arguments are discussed in chapter 4.

In chapter 5 I discuss what MS should to to set up a SIEM environment and how to make sure that SIEM is used properly.

My final conclusion is that SIEM should be used within MS to get a better picture of the current security situation. As argued, a SIEM environment in itself will not make MS safer. Instead, it will provide valuable information to help making the systems and software of MS more secure.

[1] Security Information and Event Management

Introduction

In the first days of September 2011 I started working at MS under Antony Verheijen as a graduate research student. In the period up to March 2012 I have researched SIEM. I have researched its maturity and I have written my recommendation for its appliance within MS. In this report I will give a short overview of the recommendations and findings I have. My thesis will contain the same recommendations, backed up by literature and interviews.

Motive

The problem for this thesis is the following. In MS, the emphasis on IT security grows with Antony's efforts to bring it on the agenda. More security measures are implemented daily. However, there is a source of information which now lies unused: log files. Information on possible security breaches is not extracted from these files, nor is information on possible events relayed to Antony. One of the possibilities to use this resource is SIEM. The analysis and consolidation such a system can offer for various critical systems would bring out possible problems, report on (undiscovered) incidents and show opportunities for improvement. This can greatly improve the IT security status and provide Antony with a new source of information.

Research question

Following from this analysis, the research question as defined by Antony consisted of three parts.

1. Is the market for SIEM environments mature enough to support MS?
2. If so, can SIEM support MS in its efforts towards better IT security?
3. If so, where and how should MS apply SIEM?

In this recommendation I will answer all three questions.

Document structure

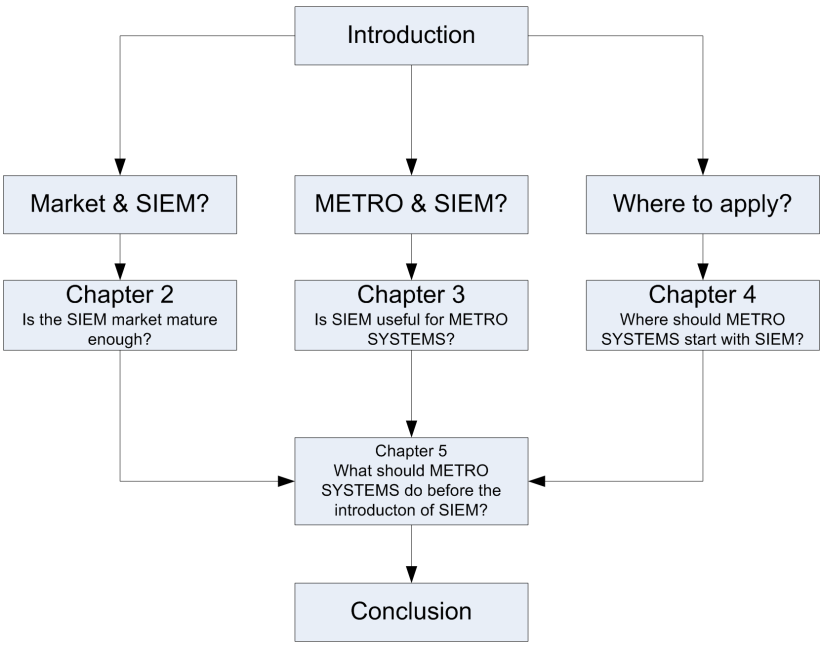
The document is divided in three parts, each answering a part of the question. All three parts come together in chapter A ("What should MS do before SIEM is introduced?") on page 56. Then, the conclusion will follow naturally.

Is the SIEM market mature enough?

A definition of SIEM

SIEM stands for Security Information and Event management. Some add Oppurtunity, making it "SIEOM". A SIEM environment or -solution is a dedicated system designed to collect and alayse log files and search for patterns which may indicate incidents or persistent problems.

SIEM environments need CPU capacity to do this anaylsis as "live" as possible. They also need storage capacity to store their own log files and reports. SIEM environments also the capacity to retrieve log files from systems such as `syslog`



which removes the need for those systems to store their own log files. A SIEM environment can be hooked to virtually every type of hard- or software there is.

The result of such an analysis is a report. Reports can be static or dynamic. Dynamic reports offer a live "overview".

SIEM is not a firewall, nor is it an intrusion detection system. Rather, SIEM reports on conclusions made by such systems. A SIEM environment is capable of analysing low-level technology such as routers, but is also able to analyse applications and business flows.

The maturity of underlying technologies

SIEm is the culmination of several research fields. In order to show their maturity, I've researched several of these underlying technologies. These are:

1. Statistics;
2. Data mining;
3. Data warehousing;
4. Distributed data;
5. Machine learning;
6. Intelligent systems.

This list is sorted on complexity and age. Especially the last item in the list is an important one, for it works on methods to make a machine learn just like people learn: by changing their behaviour based on input. AI these fields benefit greatly from the huge growth in available processing power, allowing the way for more

complex algorythms. Currently, researchers are working with bigger data sets such as collected by Facebook.

Data warehousing and data mining research results in better insights into the possibilities of such large systems. SIEM companies use this knowledge to make their SIEM environments capable of being applied to the IT landscape of an entire company,

Other improvements come from the advantages in artificial intelligence. Computers are capable of finding patterns themselves, relying less on hints and pre-defined search patterns.

The maturity of SIEM environments

According to Gartner, SIEM technologies are now reaching a stable peak on their "hype cycle" which shows that SIEM is being implemented by a lot of companies. This is beneficial for MS. Antony has given me information about the previously run SIEM trial.

All mayor SIEM companies are converging towards the same solution:

- A central log collector which does preliminary analysis and storage.
- A system capable of polling that collector in order to generate reports. Often this system is combined with the collector.
- Software designed to intercept, parse and send log files to the log collector. These 'agents' usually run on the actual servers.
- 'Agent-less' log collecting. In these implementations the collector reaches out to shared drives or network resources to get log files. Often devices are set up to deliver their log files using `syslog` or other technologies.

The maturity of SIEM companies

Reports from Gartner, Forrester and The Register show that the companies themselves are stable and mature. As I am not a business researcher myself, I have relied on these reports. From a technical standpoint I can say SIEM is stable and mature.

Is SIEM useful for MS?

At first glance, a new SIEM environment within MS sounds like a sensible idea. Log management, analysis and better incident and event response can help any company.

The company is big and has a lot of different departments. In such a complex environment SIEM technologies can support Antony with an "overall picture" of the IT security status. There are many similar systems being used which already mimic SIEM features such as reporting and alerting. In this section, I will explain why I feel that a SIEM environment can further help the company, where it should start and what it should focus on.

When I started at MS, I was at loss as where to begin. With the support of Antony I've set up interviews with colleagues all over the company. I've asked them about log management, security incidents (and handling them), procedures and

guidelines regarding IT security. I was also curious about their own motivations for solutions which mimic SIEM environments.

I've made a lot of observations in this way, and I will share the relevant ones below. They are within the scope of single departments and administrative areas.

1. Log management and analysis is done for a limited set of scenario's. This means it takes more time to catch up with new threats or new compliance / policy demands.

2. Some administrators have set up extensive logging targeted specifically at changes made to their systems. The argument for this is compliance and privacy. Their own (administrative) changes are not checked this extensively. This means that fraud and data theft are *still* possible.

3. Some administrators are capable of editing and/or removing the log entries which are generated by their own actions. This creates a 'segregation of duty' conflict.

4. Logging is only enabled for debug purposes. Related, logging is only enabled on development servers. This means that an application, when released to the customer, is no longer collecting valuable information.

5. Logging is only enabled for errors and warnings. This means that a malicious user is 'off the radar', as soon as he tricks the system into giving him access.

6. It is difficult to change or update log settings without releasing a new version of the application. That means that it is difficult for administrators to respond to new threats or policy changes.

For the first three findings, a SIEM environment can provide administrators and managers with reports on the IT security status. It can easily be adapted to the IT security and compliance policies. This is one of the mayor advantages of a SIEM environment. There is no need to search or scan for specific scenario's. A SIEM environment is capable of doing that based on the IT security policies which are in place already. The reports are based on these policies. Vice versa, the policies can be adapted based on such reports.

The issues numbered 4, 5 and 6 above are not instantly solved when installing or using a SIEM solution. According to Dr. Bär and others MS generates too much log content to properly save it. Firewalls and intrusion detection systems (or IDS') can also be managed centrally and have their log files correlated. Dr. Bär also took note of the speed of reporting. The more complex the analysis we demand from a SIEM solution, the more difficult it is to provide it 'real time'.

Both of these objections are based on a different idea of wat SIEM is. In subsection A ("A definition of SIEM") on page 50 I have defined SIEM as a non-intrusive way of making use of opportunities when they arise.

SIEM as a tool for incidents and problems

The issues I've listed in the previous section do not show the power of a SIEM environment. I consider the first three issues as originating from 'alternatives to SIEM', and the last three the result of not having proper log management.

In a discussion with Kerstin Luck, she mentioned that SIEM should be able to help in two occasions: detecting incidents and reporting problems. Detecting incidents is a matter of quickly scanning and analysing log files. Firewalls and IDS' are very good at this. SIEM environments are as well, but the question is if we should set up SIEM to do this. Reporting problems based on incidents (as defined by ITIL) or based upon log observation is a typical SIEM activity. Good SIEM analysis can bring out unknown problems by analysing application and server activity.

In a discussion about a security breach, Dr. Bär mentioned that "*if the attack went slowly [instead of over flooding us], we might never have noticed*". *This* is where a SIEM environment can support the company.

Looking for such patterns and learning from other devices, the SIEM environment *adds* to MS' security.

SIEM as a tool to report on business rules

A logged event can come from a simple firewall or a complex system. For a SIEM environment, there is no difference. Any log entry can be analyzed as long as the rules are known.

The SIEM environment can ask itself "Should we flag this HTTP request as suspicious?" as easily as "Isn't this order being sent to a known fraudulent shopper?". SIEM environments are capable of making very consious decisions about the log entries it analyses. That ability should be exploited to implove the IT security.

Where should MS start with SIEM, and what should it be used for?

MS is too big to simply implement a SIEM solution for *everyone*. Instead it should focus on one department or activity and apply SIEM technologies there. From that point MS can extend towards other departments. "Start small, think big". Existing security solutions can exist within such an environment. The one does not exclude the other.

Start with e-commerce systems

Together with Antony, I've chosen e-commerce and web shops as the best candidate for a new SIEM trial. There are a couple of reasons why we made this choice.

1. Web shops are very visible. Contrary to email systems or login portals, every effort is made to draw more visitors to the web shops. This will more likely attract malicious users as well.

2. With this visibility comes a 'quick win' for malicious users who manage to successfully attack the web shop. Damage control is very difficult even when the attacked website is merely a static page with no sensitive data on it.

3. Web shops have a deep reach into our systems. A customer can change his own personal records and change our inventory records (by buying) all while not being physically present. When a hacker compromises customer passwords the MG is liable for possible losses.

4. Other systems, such as our webmail, have more complex security. RSA tokens and VPN connections are often mandatory, making it more difficult for a hacker to breach. Such security measures cannot be demanded from our web customers.

Of course, the same can be said for other systems of MS. For example, if somebody breaks into the goods management system or a SAP system. But since the internet gives the MG such a great visibility, both me and Antony concluded this was the place to start.

Use SIEM to analyze log files from security devices

It is not just about log files. Any program can scan log files. The most simple of C scripts can look for SELECT statements in HTTP requests.

SIEM environments are capable of scanning the (already combined) log files of firewalls, intrusion detection systems, applications and servers. SIEM environments can look for incidents and problems by combining errors and warnings from multiple systems and by searching for patterns. There where firewalls will only block (and drop) traffic, the SIEM environment will save a mention and use it to find a malicious user's attack pattern.

In a minimal SIEM environment not only the web server itself is analyzed by a SIEM environment. The routers, switches, firewalls and IDS's are scrutinized as well. The SIEM environment has to be set up to ignore certain warnings (for example, known false positives made by Qualys) and prioritize according to our data privacy and IT security policies.

Use SIEM to analyse e-commerce business rules

First: more and more divisions and companies in the MG and in MS are putting emphasis on IT security. Developers, administrators and network technicians all know and stress the importance of web security. They all implement the measures available to them. For example:

- Coded guards against SQL injection by sanitizing input;
- (Web Application) Firewalls;
- Intrusion detection systems;
- Disaster recovery;
- File and folder security.

From this point of view, there is a lot of work being put in IT Security already. A SIEM environment seems to clash with these solutions. But even with all these measures higher level applications can still be insecure. Take for example this news article about the Dutch retailer HEMA:

> *Department store HEMA is denying a special offer on its website – a berry clafoutis pie for €0 – was a publicity stunt to celebrate the high-street chain's 85th birthday.*

> *Hundreds - perhaps thousands - of people ordered the pie as news of the bargain spread via the microblogging service Twitter.*
>
> *A spokesman for Hema told the Telegraaf the very special offer was accidental and the company had no idea how many pies had been ordered. The 'mistake' has since been corrected.*

The above newspaper article[2] shows exactly what the weaknesses of such security systems are. No amount of firewalls will prevent a customer from ordering a pie. It is traffic that they are supposed to allow through. The fact that the pie costs nothing is not detected by such systems. Nor is the increased traffic: on the whole, the HEMA website was doing just fine.

This is where a SIEM environment *can* help. It can help with the enforcement of high level rules and regulations, written and maintained by IT Security and domain experts alike. These rules determine the inner workings of web stores and prevent such mistakes. These rules are the key to better IT Security. SIEM environments can and should check those rules.

What should MS do before SIEM is introduced?

With this focus on e-commerce and a set of general questions, the following steps should be taken *before* a specific SIEM environment is considered.

Topics to address and focus on

Segregation of duty

In MS as well as in the MG it is common to separate certain rights and duties to prevent even the hint of possible fraud. When dealing with log files, this should be the case as well. MS should work towards a situation where the person who initiated the event which created the log entry[3] cannot overwrite or change that log entry.

This is a difficult issue, because 'log files' have so many appearances. For a web server or a router log entries consist of single line entries (up to a few million an hour) which are visible to any administrator. High level systems generate less events. But these events contain much more information: user data, originating server, relevant files, etc. The issue is not if the administrator can be trusted with it, but if he should be allowed to delete such records. The answer is no, and MS should strive to take away such rights in sensitive (financial) systems.

Incident reporting and escalation

At the moment, Radim Kolar is working for MS on a new system for reporting and escalating findings made by Qualys. Kerstin Luck is working on a similar system for general IT Security events. Their procedures should be capable of handling SIEM reports, either automated (like the web findings now) or by hand.

---

[2] Original source: `http://www.dutchnews.nl/news/archives/2011/11/hema_celebrates_its_birthday_w.php`

[3] For example: the system notes that a certain setting has changed and has written this down in a log file

Current threat detection systems rely on incoming HTTP requests. They can report these problems to administrators who have a consolidating overview. A SIEM environment can provide the same thing on a higher level, where a threat detection's algorithms are not advanced enough or not capable of correlating.

Protection against common attack vectors

OWASP, an open internet security community, publishes a top 10 threat list. Their latest list includes, in my opinion, threats and alerts all firewalls and IDS's should look out for. Since this is a 'pre-SIEM'-list, focus should be on making sure these top 10 vulnerabilities do not appear on MS' sites. However, 100% security against such threats is impossible, as has been seen recently. While firewalls and other systems can detect possible leaks, a SIEM environment can report on these findings and help administrators act accordingly.

Define metrics and benchmarking

How big should the SIEM installation be? Do we need servers, complete server rooms, or is a virtual machine enough? It is a difficult question to answer beforehand. It can be done by narrowing down the exact number of systems to be monitored and the software that will be monitored. These should all be put to numbers. This can be done using a special number, the EPS (events per second). For example, a Nokia high-availability firewall is capable of handling more than 100,000 connections per second. In theory, this could create 100,000 events (per second). Most SIEM environments can handle about 15,000 events per collector.

> Common sense tells us that we should be able to handle as many events as ALL our devices could simultaneously produce as a result of a security incident. But that isn't a likely scenario, nor is it practical or necessary. Aside from the argument that no realistic scenario would involve all devices sending maximum EPS, so many events at once would create bottlenecks on the network and overload and render the SIEM collectors useless. So, it is critical to create a methodology for prioritizing event relevance during times of load so that even during a significant incident, critical event data is getting through, while ancillary events are temporarily filtered.

In the SANS report which can be found here[4], a method is layed out which can help determining the 'average peak EPS', on which the system requirements for the SIEM environment can be based.

Not all features of a SIEM environment can be benchmarked. This list mentions some of the requirements which are hard to put a number to:

- Requirements for integration with existing systems.
- The ability to process connection-specific flow data from network elements.
- The ability to learn from new events.

[4]For more information, refer to the SANS paper at **http://www.sans.org/reading_room/analysts_program/eventMgt_Feb09.pdf**.

- Filter out events from known infected devices.

This also means that it is difficult for the vendor to show it's excellence regarding these points. According to the research I found, it is best to ask existing customers about their experiences. Keep "start small, think big" in mind.

Get live demonstrations from vendors

MS should ask to get features demonstrated.

1. Can the SIEM environment handle a flexible change of rules? Can the company change alert settings? The things we worry about today, might not be our worries tomorrow.

2. Can the SIEM environment report on custom procedures? A company preparing a demonstration should receive some basic information about the 'emergency user'-procedures. How did they implement such a procedure in their demonstration? If it was an existing feature, can they actually implement a *new* procedure?

3. The SIEM environment should adapt to the company, never the other way around. What are the requirements to the systems that will be logged? Must we install log agents? What are their system requirements? [5]

4. Regarding systems that do not generate much log files: can logging agents support a system in generating more detailed logs? MS should specifically ask about Active Directory and SAP. Those were 'weak points' according to the last trial, even though we suggest focus to be on web security.

Live demonstrations show the basic features of a SIEM environment, as well as the unique selling points. After all, a demonstration is there to sell a system. MS must make sure that the complexity of its systems can be handled by the SIEM environment. The company should also seek demonstrations from multiple vendors, as well as resellers such as the Dutch company Kahuna[6].

The company should try and get all the log files used for the demonstration, if possible beforehand. Try to find irregularities or security issues by hand (or with the help from competing SIEM solutions) in these files. Check if the SIEM demonstration showed the same issues.

A more labour-intensive but rewarding process is to collect log files from within MS and allow the demonstrators to use those. Use the expertise from the company to find possible security issues in those log files. Check if the SIEM demonstration alerts you to these as well. Be careful though, because log files may contain personal or sensitive data.

Are 'loggers'[7] capable of extracting relevant information from the company's systems? What are their system requirements? Is the demonstration running on a 'live' environment, or are the log files cached? In case of cached log files, the real life analysis will be slower and possibly incomplete. According to an HP technician, a

[5] When I asked this at a webex session with HP, the only answer I got was 'they are very light'. Not very clear!

[6] I've talked with this company in particular at a recent (January 2012) convention on IT Security.

[7] Loggers are small applications that collect log files. See also "Definition of SIEM" on page 52.

SIEM environment "only knows what it knows". That means that missing data or 'gaps' will not be noticed by the system unless specifically told to look for them. The system will do so when the reporting is based on existing compliance standards. That means that there must be a list of 'involved' systems and a list of systems monitored by SIEM.

Conclusion

There are many advantages for MS when a SIEM environment is integrated.

More information from firewalls

1. Firewalls and intrusion detection systems do not see what happens under the radar. Approved requests are not saved or traced.
2. Patterns in attacks can be detected by SIEM. Hourly, daily or even 'slower' attack vectors can be reported on.
3. Anti-patterns can be detected by SIEM. Firewalls will *always* have something to report about. What if the firewall falls silent? Is that a sign of improved security, or a hacked device?
4. A SIEM environment can detect problems inside the system as well, such as suspicious behaviour coming from inside a ring of firewalls.

A source of information for incident management

5. SIEM generated events provide a solid base for any incident solving administrator.
6. When escalating events, the SIEM report provides valuable information directly from the source.
7. SIEM environments are patient. They can report on long running or indicental problems as well.

Audits and check-ups

8. By setting up SIEM according to current policies, deviations are reported automatically.
9. Bottom-up: A administrator can check the events generated by his systems or applications.
10. Top-down: A manager or auditor can see if systems are compliant.

Rules and log entries

11. SIEM can check complex and high-level rules.
12. By forcing application developers to provide the information needed to do so, they are also forced to implement those rules themselves. That alone makes the systems safer.

Infrastructure

13. With SIEM applications as an integrated part of the IT landscape, any new application or website is checked from day one.

Bibliography

[1] N. Bruton. *How to Manage the IT Helpdesk: A Guide for User Support and Call Centre Managers*. Computer Weekly Professional Series. Butterworth-Heinemann, 2002.

[2] Business Rules Group. Business Rules Manifesto. Retrieved from: `http://goo.gl/GZ1OE`, november 2003.

[3] J. Michael Butler. Benchmarking security information event management (SIEM). Retrieved from: `http://goo.gl/yMvq5`, february 2007.

[4] A. Chuvakin. The complete guide to log and event management. Available from `http://goo.gl/0lr7f`, 2010.

[5] A. Chuvakin. Practical strategies to compliance and security with SIEM. Presented at the "Vendor T" webinar, available at `http://goo.gl/kifj2`, october 2012.

[6] Dr. A. Chuvakin. Leveraging compliance for security with SIEM and log management. Available at `http://goo.gl/m3lLQ`, june 2011.

[7] E. de Bueger. White paper: Security Information & Event Management, available from: `http://goo.gl/M1b8K`. Technical report, Kahuna Group, january 2012.

[8] Bundesrepublik Deutschland. Bundesdatenschutzgesetz. *Bundesgesetz*, pages §6a, §4f III sentence 5–7, §28 IIIa, §32, §38 V, 2009.

[9] Dr. Dobb. SIEM: A Market Snapshot. Available from `http://goo.gl/Blc7u`., 2007.

[10] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, Fall:37–54, 1996.

[11] D.F. Greenberg and J.B. Roush. The effectiveness of an electronic security management system in a privately owned apartment complex. *Sociology Department, New York University*, 2008.

[12] LBNL's Network Research Group. arpwatch – Homepage of LBNL's Network Research Group. Available at `http://ee.lbl.gov`, 2012.

[13] Brad Hale. Splunk has it wrong – the flaw of volume-based licensing. Available at `http://goo.gl/wU6sK`, june 2012.

[14] Hewlett Packard. Demonstrating the ROI for SIEM. Available at http://goo.gl/lK8K8, 2012.

[15] Hewlett Packard. HP ArcSight express: powered by the CORR-engine. Available at: http://goo.gl/cZnBG, 2012.

[16] Apple Inc. Technical Q&A QA1747: Debugging Deployed iOS Apps. Available from http://goo.gl/7R5z1, 2012.

[17] M. Nicolett. How to implement SIEM technology. Technical report, Gartner Research, november 2009.

[18] M. Nicolett. How to deploy SIEM technology. Technical report, Gartner Research, december 2011.

[19] M. Nicolett and Kelly M. Kavanagh. Magic quadrant for security information and event management. Available from Gartner Research. Distributed via Q1., 2012.

[20] AlienVault Open Source Community. AlienVault Installation Guide. Available from http://goo.gl/V4LKb, 2012.

[21] AlienVault Open Source Community. OSSIM Management Server. Available at http://goo.gl/olg4F, 2012.

[22] AlienVault Open Source Community. OSSIM Sensor – technical documentation. Available at http://goo.gl/xlqOO, 2012.

[23] PCI. Payment card industry (PCI) data security standard – requirements and security assessment procedures. Available at: http://goo.gl/LzjBY, october 2010.

[24] M. Rothman. Watching the watchers: monitoring privileged users. Available at http://goo.gl/gY6jz, 2012.

[25] RSA. An integrated approach to risk, operations and incident management. Retrieved from: http://www.rsa.com/products/sms/sb/11508_h9010-iaroim-sb-0811.pdf, 2011.

[26] Securosis. Securosis blog. Available at: https://securosis.com/blog, 2012.

[27] TechTarget. Technical guide on SIM. Available at http://goo.gl/UjxIG, 2011.

[28] The Business Rules Group. Defining Business Rules – What are they really?, 2000.

[29] U.S. Government. U.S. Public Law 104 - 191 - Health Insurance Portability and Accountability Act of 1996. Public record, available through the GPO at http://goo.gl/K4Y1e., 1996.

[30] U.S. Government. U.S. Public Law 107 - 204 - an act to protect investors by improving the accuracy and reliability of corporate disclosures made pursuant to the securities laws, and for other purposes. Public record, available through the GPO at http://goo.gl/lcRSh ., 2002.

[31] A. Whitten. Are IP addresses personal? Available from: `http://goo.gl/rv1Xf`, february 2008.

[32] A. Williams. The future of SIEM – the market will begin to diverge. Retrieved from: `http://goo.gl/bnE0I`, january 2007.

[33] Michal Zalewski. p0f v3. Available at `http://goo.gl/apgzt`, 2012.