Tampere University

Juho Frigård

SECURITY INFORMATION AND EVENT MANAGEMENT SYSTEMS MONITORING AUTOMATION SYSTEMS

Faculty of Engineering and Natural Sciences
Master’s Thesis
October 2019

ABSTRACT

Juho Frigård: Security Information and Event Management Systems Monitoring Automation Systems
Master's Thesis, 75 pages
Tampere University
Automation Engineering, Master of Science
October 2019

This thesis studies how suitable Security Information and Event Management systems (SIEM systems) are for monitoring automation system log data. Motivation for this study has been the growing number of cybersecurity threats faced by industrial automation systems and the disruptive effects cyber-attacks can have on industries, vital infrastructure and the everyday life of people. The research material for this study was gathered from various literary sources as well as automation engineering lectures, studies and personal work experience in the field of cybersecurity.

This thesis provides information for adding resilience for cybersecurity threats faced by industrial automation systems. Growing cybersecurity threats and legislations such as the EU NIS Directive emphasise the need for better situational awareness and management for cybersecurity related events. The findings of this thesis help improve the design of automation system log management and SIEM systems, both of which improve the systems capabilities for counteracting cybersecurity threats.

The major finding of this study is that by default automation systems and SIEM systems are not highly compatible. Automation systems are complicated and highly tuned environments with special requirements for which SIEM systems are not originally designed. Integrating SIEM systems in a meaningful way with automation system would require major changes in both.

The biggest issue is the log data itself. In automation systems, device logs stay in the devices, whereas SIEM needs that log data to be brought to a centralised location. Furthermore, the log data will most likely not be adequate for analysis as is, and it would require enriching. Implementing modifications to correct these issues would be major transformational process to an automation system.

From the automation engineer perspective the process and its reliability and availability are the most important aspects in relation to the functionality of the system. Therefore, their focus is on operational technology security. On the other hand, SIEM system focuses on information technology security and it will be an additional service that comes in use in rare special circumstances. For this reason, selling SIEM for automation systems is difficult.

Furthermore, in automation systems, device logs are only inspected after an error situation, whereas SIEM aims to detect issues before they have a disruptive impact on the system. This mitigates damages and makes any countermeasures faster. However, systems such as SIEM would require monitoring which adds workload and the people monitoring SIEM would require expertise in both automation engineering and cybersecurity. Furthermore, monitoring the events in themselves is not enough, as the SIEM system has to be monitored as well in order to make sure it is working in the desired way.

Automation systems are unique environments and adding SIEM requires extensive customisation. Smart event detection systems use machine learning and AI to detect anomalies in a system's activity rather than identifying markers of known malicious activity. This way they require less hands-on customisation and are more compatible with industrial control systems of various types and sizes. The future of cybersecurity management in automation systems is in implementing smart detection systems and automated responses for better compatibility with operational technology environments and faster countermeasures.

Keywords: automation, information security, cybersecurity, SIEM, event management, log

TIIVISTELMÄ

Juho Frigård: SIEM-järjestelmien käyttö automaatiojärjestelmien valvonnassa.
Diplomityö, 75 sivua
Tampereen yliopisto
Automaatiotekniikka, DI
Lokakuu 2019

Tässä diplomityössä tutkitaan, kuinka turvallisuusinformaation ja –tapahtumien hallintajärjestelmät (SIEM, Security Information and Event Management) soveltuvat valvomaan automaatiojärjestelmien lokidataa. Teollisuusautomaatiojärjestelmiin kohdistuvien kyberturvallisuusuhkien kasvu sekä näiden aiheuttamat seuraukset teollisuudelle, infrastruktuurille ja ihmisten joka-päiväiseen elämään ovat olleet tämän työn keskeisinä motivaattoreina. Tutkielman aineisto on koottu kirjallisista lähteistä, automaatiotekniikan luennoilta, opinnoista ja työkokemuksesta kyberturvallisuusalalla.

Tutkimus tarjoaa tietoa kyberturvallisuusuhkien ennaltaehkäisemiseen teollisuusautomaatiojärjestelmissä. Kasvavat kyberturvauhat ja EU:n NIS direktiivi lisäävät tarvetta tarkan tilannekuvan luomiselle ja kybertapahtumien valvonnalle yhteiskunnan tärkeiden palvelujen tuottajille. Tutkimustulokset tarjoavat apua automaatiojärjestelmien lokienhallinnan ja SIEM-järjestelmien suunnittelussa, joilla molemmilla tarjotaan suojaa kyberturvauhkia vastaan.

Keskeisin tutkimushavainto on automaatiojärjestelmien ja SIEM-järjestelmien yhteensopimattomuus. Automaatiojärjestelmät ovat monimutkaisia, pitkäikäisiä ja tarkoin säädettyjä toimintaympäristöjä, joiden erityisvaatimuksille SIEM-järjestelmiä ei ole alun perin suunniteltu. SIEM-järjestelmän integroiminen mielekkäästi automaatiojärjestelmään vaatisi suuria muutoksia molempiin järjestelmiin.

Yksi suurimmista ongelmista on lokidatan siirtäminen laitteelta toiselle. Automaatiojärjestelmissä laitteiden lokit jäävät laitteiden muisteihin, kun SIEM-järjestelmä tarvitsisi niiden tuontia keskitettyyn sijaintiin. Laitteiden tuottama lokidata ei myöskään todennäköisesti sovellu sellaisenaan, jotta SIEM-järjestelmillä pystyisi tekemään tilannekuvatarkoituksiin riittävää analyysia. Lokidata pitäisi rikastaa esimerkiksi vähintään laitteen sijaintitiedolla. Tällaisten muutosten tekeminen vaatii suuria muutoksia automaatiojärjestelmiin.

Automaatioinsinöörille automaatiojärjestelmässä tärkeintä on itse prosessi sekä tämän luotettavuus ja saatavuus. Tästä näkökulmasta SIEM-järjestelmä olisi lisäpalvelu, josta on hyötyä vain harvinaisissa erikoistilanteissa. Tästä syystä SIEM-järjestelmien myynti automaatiojärjestelmiin on vaikeaa. Automaatiojärjestelmissä lokeja tarkastellaan vain, kun jokin laite vaikuttaa toimivan väärin, mutta SIEM-järjestelmillä on tarkoitus havaita ongelmat ennen suuria vaikutuksia kohdejärjestelmälle. Etukäteistoimenpiteillä voidaan vähentää seurauksien vaikutuksia ja lyhentää vastatoimien vasteaikaa. SIEM-järjestelmät kuitenkin tarvitsevat valvontaa, mikä lisää työtaakka, ja valvontaa suorittavilta henkilöiltä vaaditaan asiantuntijuutta sekä automaatiotekniikasta että kyberturvallisuudesta. Lisäksi pelkkä tapahtumien valvonta ei riitä vaan koko SIEM-järjestelmää täytyy valvoa ja kehittää jatkuvasti, jotta se toimisi halutulla tavalla.

Automaatiojärjestelmät ovat ainutlaatuisia ympäristöjä ja SIEM järjestelmän lisääminen vaatii laajaa räätälöintiä. Älykkäät tapahtumien havaitsemisjärjestelmät käyttävät koneoppimista ja tekoälyä tunnistamaan poikkeavia toimintoja järjestelmässä sen sijaan, että ne yrittäisivät tunnistaa tunnettuja haitallisia ohjelmia tai hyökkäyksiä. Tällä tavalla ne vaativat vähemmän käytännön mukauttamista ja ovat paremmin yhteensopivia erityyppisten ja erikokoisten automaatiojärjestelmien kanssa. Automaatiojärjestelmien kyberturvallisuuden hallinnan tulevaisuus onkin älykkäiden havainnointijärjestelmien ja automatisoitujen vastatoimien teknologioissa. Niillä saadaan aikaan parempi yhteensopivuus teollisten automaatiojärjestelmien kanssa ja niiden avulla voidaan toteutettua nopeampia vastatoimenpiteitä.

Avainsanat: automaatio, tietoturvallisuus, kyberturvallisuus, SIEM, tapahtumien hallinta, loki

PREFACE

This thesis was done for *Tampere University* – the university formerly known as Tampere University of Technology - and *Insta DefSec Oy*. I would like to thank my instructors *Mikko Salmenperä* and *Hannu Koivisto* from Tampere University. I would also like to thank my instructor *Santeri Taskinen* from Insta DefSec Oy for providing assistance in the process of writing this thesis.

Furthermore, I would like to express my gratitude towards *Vesa Keinänen*, *Yrjö Kinnunen* and *Teemu Salmivesi* for providing assistance in the different stages of making this thesis.

Tampere, 23 October 2019

Juho Frigård

CONTENTS

List of Figures

LIST OF SYMBOLS AND ABBREVIATIONS

AI	Artificial Intelligence
CIA	Confidentiality, Integrity and Availability
CPS	Cyber Physical Systems
CSV	Comma-Separated Values
DCS	Distributed Control Systems
HMI	Human Machine Interface
IACS	Industrial Automation and Control System
ICS	Industrial Control System
ICSaaS	Cybersecurity as a Service
ICT	Information and Communications Technology
IoT	Internet of Things
IIoT	Industrial Internet of Things
IP	Internet Protocol
ISO	International Organisation for Standardisation
MES	Manufacturing Execution System
MOM	Manufacturing Operations Management
MSSP	Managed Security Service Providers
NMS	Network Management System
PLC	Programmable Logic Controller
SAS	Finnish Society of Automation (fin. Suomen Automaatioseura ry)
SCADA	Supervisory Control and Data Acquisition
SIEM	Security Information and Event Management
SOA	Service-Oriented Architecture
SOC	Security Operations Centre
TAU	Tampere University
TCP/IP	Transmission Control Protocol/Internet Protocol
TUT	Tampere University of Technology
VPN	Virtual Private Network

1. INTRODUCTION

Automation systems are everywhere. Their functions range from controlling large industrial processes to autonomous vehicles and all the way to managing electronic devices in people's homes. They are also a key part in maintaining vital infrastructure on a global scale. Countries, businesses and individual people rely on these systems to work. One growing threat in the field of automation and technology in general is *cybersecurity*.

In 2019, Kaspersky Lab published their study on the threat landscape for industrial automation systems. Major finding in the study was that in 2018 as much as 47,2 **%** of Industrial Control Systems (ICS) contained malicious objects. In the previous year this number was 44,0 %. Malicious objects were categorised into multiple different categories most prevalent of which were Trojans, malicious scripts, worms, web miners and malicious link files. (Kaspersky Lab, 2019)

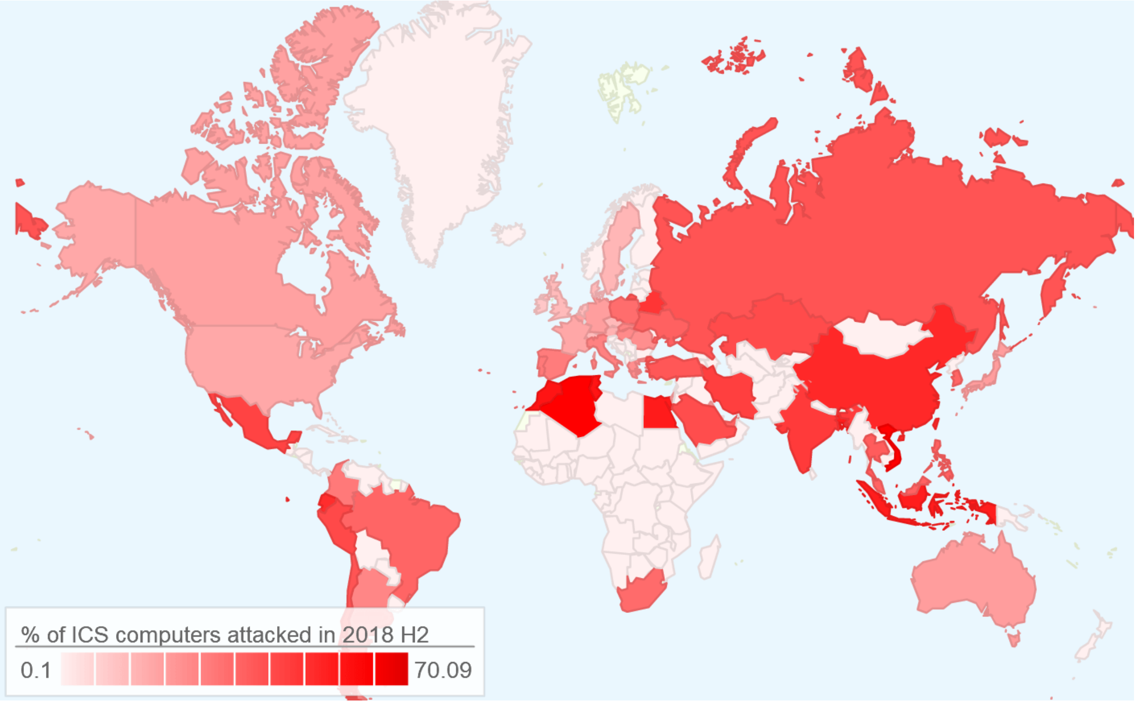


Figure 1. *Kaspersky Lab's ICS CERT research results showing the percentage of ICS computers infected with malicious software. (Kaspersky Lab, 2019)*

Figure 1 is Kaspersky Lab's ICS CERT research results concerning the amount of ICS computers being affected by malicious software worldwide. Countries with the highest percentages were Vietnam, Algeria, Tunisia, Morocco and Egypt. In Vietnam, as much as 70,0 % of ICS computers were infected. (Kaspersky Lab, 2019)

This level of infection can have major disruptive effects on not only business but also individual people and society in general as automation systems with ICSs control vital infrastructure. Furthermore, the number of infected ICS computers in industrial systems is growing every year. (Kaspersky Lab, 2019)

The most widely used malware found in the Kaspersky Lab's study was *Trojans*. These are pieces of malicious software, which are distributed inside other programs or email attachments. They deceive the user into trusting the software and once the program is used or the email attachment is opened, the Trojan's malicious code will be executed.

Important aspect in handling cybersecurity threats is situational awareness. This involves monitoring events in the system. Security Information and Event Management (SIEM) system is a tool specially designed for detecting, monitoring and categorising events based on logged data. These SIEM systems allow people to follow the state of a system they are monitoring and through that react to possible threats or other anomalies.

SIEM relies on centralised log management and constant monitoring. Both of these make it challenging to integrate SIEM systems to automation systems. Challenges also include resources, automation engineers' interests and added workload that SIEM will inevitably require during the entire automation systems lifecycle as cybersecurity is not a goal but a continuous process.

Industrial and automation systems can also have legislative obligations concerning cybersecurity. In August 2016, the EU *NIS Directive*, that is the first EU-wide directive on cybersecurity, entered into force. The directive aims to improve overall level of cybersecurity in the EU. Improvements are achieved by ensuring that member states enhance their preparedness for cybersecurity threats and co-operate with other states. In addition, the directive aims to establish a culture of security across sectors that are vital for our economy and society and rely heavily on Information and Communications Technologies (ICT), such as energy, transport, water, banking, financial market infrastructures, healthcare and digital infrastructure. (EU: Cybersecurity & Digital Privacy Policy (Unit H.2), 2019)

The NIS Directive obliges business, that the member state has identified as providers of essential services, to take appropriate countermeasures and notify the proper authorities of any serious cybersecurity incidents. SIEM is a tool for monitoring such incidents and an effective solution for monitoring cybersecurity in large systems. As some of these vital service providers or infrastructures rely on automation systems to function, SIEM is an alternative for providing monitoring, analysis and auditing for these environments.

This thesis studies the usability of SIEM systems in monitoring automation systems. SIEM is widely used in monitoring various information technology systems, however, automation systems impose their own requirements that SIEM systems will have to fulfil.

2. THESIS STRUCTURE AND METHODOLOGY

This thesis is a literary study. Information is gathered from various literary sources, automation engineering lectures, studies and work experience. The automation systems this thesis concerns, are large process automation system used in industry and infrastructure. Physical safety related automation systems are not covered in this thesis.

The purpose of this thesis is to study how well SIEM systems would perform in monitoring automation systems and how SIEM systems should operate when monitoring these automation systems. Therefore, the topic of this thesis is *Security Information and Event Management Systems Monitoring Automation Systems*.

Security information is all the information relating to the systems resilience against undesired effects from external or internal sources. External sources could be malware or physical threats such as vandalism, whereas internal sources can be malfunctioning components or operator errors.

Security Events are observable occurrences that can be categorised relating to the security of the system. Events can be initiated by components, process or human operators and they do not have to be malicious in nature. Example of a security event would be login to a service.

The *management* part relates to handling security information and events. Major part in handling this information is the processing of the input data. Through analysis, the actual information is gathered from data, as raw log data does not qualify as information on its own. Having tools that have capabilities for automated processing and analysis of data will automate this process.

Automation systems are used to control physical systems automatically. This thesis concerns process automation systems that used to control industrial processes and safety related systems and smaller automation systems are omitted.

All process automation systems require *monitoring*. Monitoring is done by collecting data from the process as well as the devices controlling the process and using this data in forming a situational image of the system. The purpose of monitoring is to ensure the system is functioning within the desired parameters.

2.1 Sources

Sources for this thesis are mainly electronic sources. The topic is related to automation, information technology and cybersecurity. The information concerning these topics changes fast and electronic sources seemed provide the most up to date information. There is also a wide range of older sources concerning automation systems and industrial control systems that offer accurate and still relevant information. In addition to electronic and other literary source, the faculty of Automation Engineering in Tampere University provided information about automation systems and their log handling. Information was also attained from automation engineering studies and work experience from the field of cybersecurity.

When gathering information from electronic sources, it was important to identify the type of the source. In the field cybersecurity, one can find multiple different source types from scholar papers to blog texts, and even though blog texts could provide a lot of practical advice on SIEM and cybersecurity matters, the information they gave had to be checked from other sources if possible.

2.2 Process

The process of making this thesis started from the inception of the topic. The topic concerns handling log data in automation systems, which relates to cybersecurity and the growing threat this imposes to automation systems. The research done for this thesis can offer information that helps in building systems that are more resilient to cybersecurity threats.

After deciding the topic, started the process of gathering information and refining the area of focus for the thesis. The foundation of this thesis is formed by compiling special requirements that automation systems possess, and then analysing how well would SIEM systems be able to function in accordance to these special requirements.

2.3 Structure

This thesis starts with theoretical background section in chapters 3 and 4. Chapter 3 covers SIEM systems. It describes how SIEM systems are used, who uses them and what the overall purpose of these systems is. Chapter 4 in turn, provides theoretical background on automation systems and the different industrial system levels from enterprise levels to field device levels. This provides more information on the devices that produce the actual information that is logged and monitored. Chapter 4 also covers special requirements for automation system software and log management and event handling.

In conjunction with chapter 4, chapters 5, 6 and 7 form the basis of the analysis. Chapter 5 covers log data in automation systems and chapter 6 provides information on cybersecurity in automation systems. This includes pre-existing solutions for event management in industrial control systems. Chapter 7, on the other hand, is an example of how SIEM parsers are implemented in practise. This offers additional perspective into what parser development is like and what steps must be taken to develop parsers. Parsers are a key part in analysing log data with SIEM systems.

Chapter 8 is the analysis of how well SIEM systems would function under the requirements and challenges automation systems impose on them. This chapter also provides practical solutions for integrating SIEM systems for industrial automation systems in order to achieve the best possible results from both automation system and SIEM system perspectives.

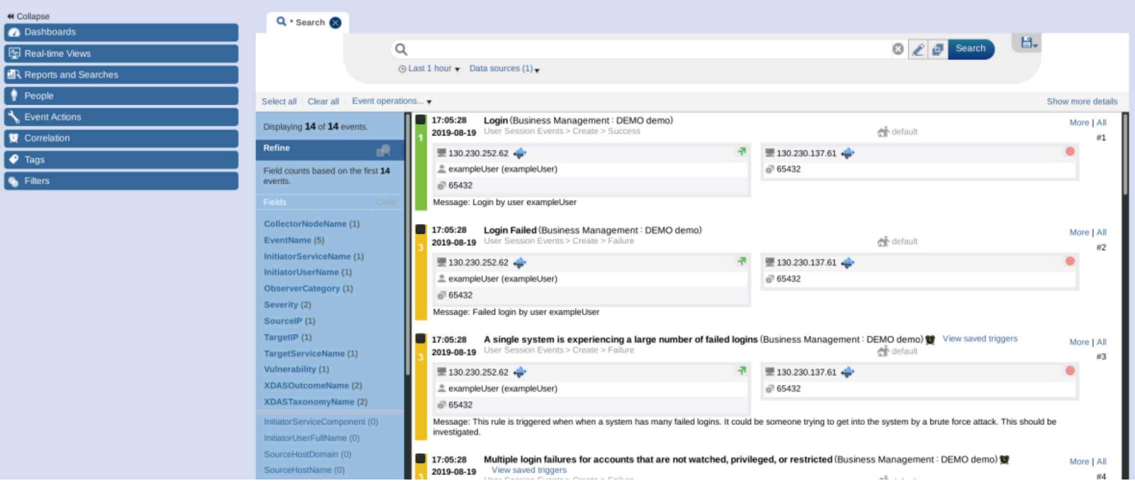
Finally, chapter 9 is the conclusion chapter of this thesis. It is a summary of the results of the analysis conducted in chapter 8. It also gives information on the future of SIEM and automation as well as future areas of research in the aforementioned topics.

3. SECURITY INFORMATION AND EVENT MANAGEMENT SYSTEMS

The term *Security Information and Event Management Systems* (SIEM) was first introduced by Mark Nicolett and Amrit Williams from Gartner Analysts in 2005. They recognised that there were so-called *Security Event Management* (SEM) and *Security Information Management* (SIM) tools on the market, the former of which meant real-time analysis of events to support incident response, and the latter was for the long-term storage of information security data and historical analysis. In the past, there were suppliers who focused on either SEM or SIM. However, there was an increasing desire from the market to have them both working in unison. Thus, in 2004 these new combinations of SEM and SIM started to appear on the market. (Williams, 2007)

The main purpose of SIEM is to detect security related activity. This is achieved by collecting data from event sources, parsing the incoming data, categorising the event and storing all the relevant data. The stored data forms a baseline for activities in the system and allows SIEM to detect and inform the operator of any possible security threats or other security related events that are worth noting. SIEM can detect security threats by using indicators in individual events to identify specific activity or correlation rules to detect activities from a collection of individual events. An example of using correlation rules is to detect when somebody is logging into a service multiple times with the wrong password. Activity such as that may be the sign of a brute force attack trying to find a password to gain an access to the system.

SIEM is an *event analysis tool*. Events in this case are forms of activity that occur in the system that is being monitored. SIEM receives or fetches data from data sources, parses the information, perform analysis and displays the relevant information to the person monitoring the system. The analysis concerns reading input data, parsing out relevant data and making observations based on rules in the SIEM system. The rules can be preprogrammed or in more sophisticated SIEM systems they can also be formed through a machine learning process. In machine learning, the system forms rules by experience rather than being programmed by a human. Figure 2 presents the user interface for Micro Focuses SIEM product *Sentinel*.



Displaying **14** of **14** events.

Field counts based on the first **14** events.

1 17:05:28 2019-08-19 **Login** (Business Management : DEMO demo)
User Session Events > Create > Success default #1
130.230.252.62 | 130.230.137.61
exampleUser (exampleUser) | 65432
65432
Message: Login by user exampleUser

3 17:05:28 2019-08-19 **Login Failed** (Business Management : DEMO demo)
User Session Events > Create > Failure default #2
130.230.252.62 | 130.230.137.61
exampleUser (exampleUser) | 65432
65432
Message: Failed login by user exampleUser

3 17:05:28 2019-08-19 **A single system is experiencing a large number of failed logins** (Business Management : DEMO demo) View saved triggers
User Session Events > Create > Failure default #3
130.230.252.62 | 130.230.137.61
exampleUser (exampleUser) | 65432
65432
Message: This rule is triggered when when a system has many failed logins. It could be someone trying to get into the system by a brute force attack. This should be investigated.

3 17:05:28 2019-08-19 **Multiple login failures for accounts that are not watched, privileged, or restricted** (Business Management : DEMO demo) View saved triggers #4

Figure 2. *Micro Focus’ SIEM product Sentinel’s user interface where different events are presented as a list.*

In figure 2, all the events are presented as a list and they are organised based on timestamps. In all of the events, there is an event name, IP addresses of both the source of the message and the target of the message. Port numbers are included for both source and target services. The events have also taxonomy definitions, which categorise the events. Categories are for instance *login, logout, password changed, malware, and database query*.

3.1 Log Data Inputs

SIEM systems take event data in various formats through various sources. Sources may include for example:

1. log files,
2. syslog messages,
3. SQL databases and
4. Windows events

Log Files are either unstructured or structured lists of log events in a file. It can be, for instance, a comma-separated values (CSV) file where individual log fields are separated by a comma or another character. Events are separated on separate lines and usually in a way which a single event occupies one line. However, it is possible that one event is split into multiple lines in which case the event has to be constructed from the different lines.

When using *syslog*, messages are sent from logging devices to a server that stores these messages. Syslog server can receive messages from multiple devices at once and they can relay all the gathered information forward for data parsing and event detection purposes. Figure 3 illustrates the function of syslog logging.

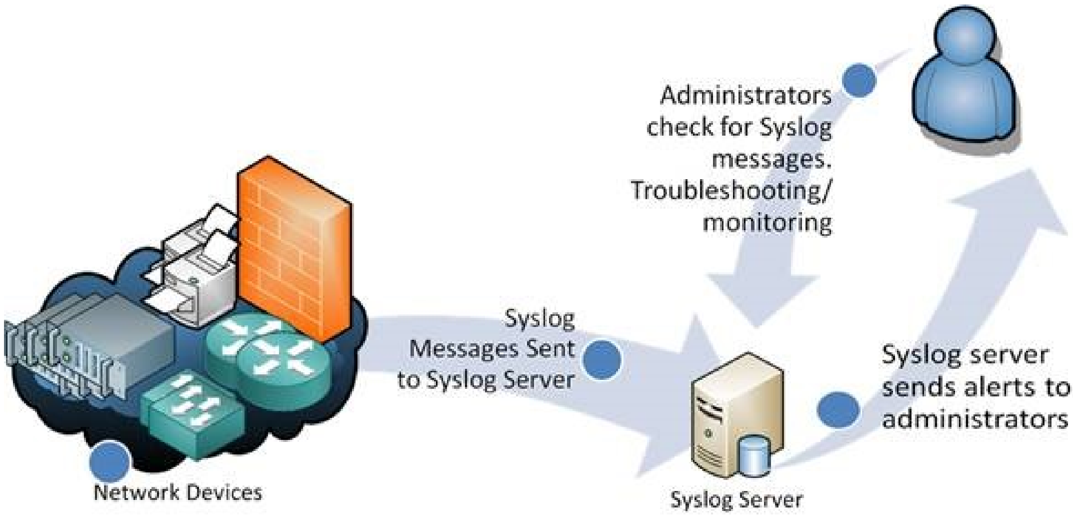


Figure 3. *Illustrative picture of how syslog message are send from the end devices to a syslog server and to administrators who check the event data. (Leskiw, 2018)*

In figure 3, the network devices represent the devices from which the syslog messages originate. They send messages to the syslog server, which stores the information and then sends it to the administrator. Between the device and the syslog server, there can be other servers or databases that store the messages before sending them to the syslog server. The administrator can be an actual person going through the data by hand or it can be an automated program that parses the information, which makes interpreting the messages easier.

In some cases, it might be useful to implement event filters into syslog servers. Event filters filter out irrelevant messages in order to automate and simplify the process of interpreting the data. This can be useful if a SIEM system is not used in conjunction with the syslog server. (Leskiw, 2018)

Syslog messages can vary widely depending on the device sending the messages. However, most syslog messages comply with a specific format and the messages are comprised of IP addresses, timestamps and the actual event information. Below is an example of a syslog message.

```
<35> 2019-01-01T09:00:00.000 130.230.252.62 applicationID[12345]:
2019-01-01T08:59:59.000 130.230.137.61 WARNING message="Failed
login by user 'admin'"
```

The messages begins with a priority number that is encased in angle brackets. The priority number is formed by multiplying the facility number by eight and then adding the severity to that result. The facility number in turn tells what kind of message it is. For instance, the example message is an authentication message, as it is a failed login warning, so it would have a facility code of four. With a severity value of three, the priority number would be in this case 35.

After the priority number, there is the timestamp of the event. This timestamp is from the device that sent the message and it describes the accurate time when the event occurred. Accurate timestamps are integral to building an image of the events in the system as they reveal information on the sequence of the events. The timestamp is followed by an IP address or a host name. This informs what device sent the syslog message. However, in some cases the IP address can change during transmission, as IP addresses are not static.

The last piece of information before the colon is the application ID. The application ID is not formally part of the header, but it is most often included. This ID tells from what application the syslog message originates. The application ID can be used to identify the format of the message and how the message is supposed to be parsed.

The actual content of the syslog message comes after the application ID and that contains the event data. The format of event data can vary and is related to the device and the event itself. It may contain the IP address of the logging device, severity information, timestamp of the event and an actual plaintext explanation of the event. In the example event, the message begins with a timestamp followed by an IP address. Then there is a value *WARNING* which is information concerning the type and severity of the message. Lastly, there is a plaintext message field that explains that this event is a failed login by the user *admin.*

Syslog has its faults. The syslog format does not concern the message content and this means that every single log source might require its own parser. Originally, syslog used UDP to transfer messages and UDP is connectionless protocol. As a result, it is possible to lose packets due to network congestion or packet loss. Now syslog messages also use TCP and TLS. The latter of which is used for sending encrypted syslog messages. Syslog messages do not have any authentication so it is difficult to verify where the messages are coming from. This in turn makes it possible to forge messages, which leaves the system open for replay attacks. (Leskiw, 2018) Replay attack means that a malicious party intercepts a message or multiple messages and uses those messages to forge events by sending them to SIEM later.

Syslog is preinstalled into many Unix-based operating systems. Windows on the other hand uses its own logging format: *Windows Event logs.* Windows event logs are formed by actions from the operator of the machine or some internal process and they are used for security management as well as debugging issues. The logs are all in the same format making automated parsing more simple and the entire analysis process more streamlined and faster. (Gough & Porter, 2007; Rouse, 2018)

In contrast to syslog messages, SIEM can also fetch event information from *databases*. In this case, data is stored into a database from which they are collected and parsed by SIEM. Whereas with syslog the messages were specifically sent to SIEM, in this case SIEM has to query them actively from an external data storage. When using databases, there is no format for how the data is stored or what kind of data there is in the first place. Similarly to syslog, this means that SIEM might need multiple parsers to parse the data. This can be aided by designing databases smartly so that they utilise the same fields regardless of where and in what format the information arrives to the database.

3.2 Event Data Parsing and Normalisation

Regardless of where the event information is coming from, the data has to be parsed before SIEM can present the data in a human readable form or perform any analysis of the events. Parsing means that all the relevant values are extracted from the raw data that comes from the logging device and are stored in the appropriate data models in SIEM. This latter part is called *normalisation*. The analysis comes from interpreting the data, identifying patterns and alerting of possible threats. This analysis requires normalisation, as without it SIEM would not be able to ably any logic into the analysis.

The format of the log message plays huge part in this as device manufactures have their own formats in use. As a result, most log formats will require their own parsers. If a SIEM system collects logs from a multitude of devices, making parser for each of them can be time consuming or even unfeasible. In addition, syslog messages may cause more issues as syslog may send a single event as multiple messages. This results in having to buffer messages until all the messages relevant to one event have been received. These single messages might not even be received in sequence as other syslog messages might be received in between them. Before parsing can begin, all messages relating to a single event must be received and identifying which messages relate to which event might be difficult.

Some SIEM products require that every single parser be programmed manually. The parser developer will write the code for the parsing logic and build the program to form a parser plugin. This plugin is then installed to the SIEM system where it can be tested and taken into use. This type of parser development requires knowledge of programming, software development and testing. Therefore, it might be outsourced to companies that offer parser development as a service.

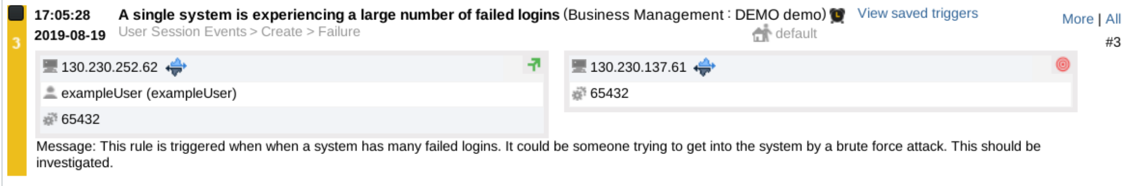
Some SIEM products offer more high-level parser development tools. With these tools, parsers are built from pre-programmed widgets on a graphical user interface. Unlike when programming parsers by hand, this might not require as much knowledge or training in software development. The information can also be extracted from the log data by the use of regular expressions that identify regular patterns in the data.

When implementing SIEM, it is vital that parsers are developed properly. Improper parsers might leave out integral information, misplace information into the wrong fields or even lose entire events. All of these on their own could cause event correlation detection to work incorrectly. This will in turn add workload to administrators who are monitoring the events. In the event that a possible threat is found, the administrators will have to find all of the required information by themselves, which is time consuming and requires a lot of resources. This also means that the administrators will have less time react to events and to decide the importance and severity of the threat. These are all important factors when it comes to resilience and resistance to security issues.

3.3 Event Correlation Detection

One of the key features of SIEM systems is *event correlation detection*. In event correlation detection, log data is analysed in order to identify relationships between multiple events by recognising patterns in the data. For example, if from a particular IP address someone is trying to login to a system with different username and password combinations a certain amount of times, it may signal a malicious attempt to access peoples' accounts or the system in general. Through event correlation, these types of activities can be recognised and the system administrators would be notified of the suspicious activity. (Zhang, 2018)

Malicious activity might not consist of only one type of activity. In situations like that, correlation rules in SIEM system have to be able to build situational awareness from a variety of events. A few failed logins do not usually mean that someone is attacking the system but if they accompanied by a port scan originating from the same IP address, you might be dealing with an attack. (Zhang, 2018)



3

17:05:28
2019-08-19

A single system is experiencing a large number of failed logins (Business Management : DEMO demo)

User Session Events > Create > Failure

default

View saved triggers

More | All

#3

130.230.252.62

exampleUser (exampleUser)

65432

130.230.137.61

65432

Message: This rule is triggered when when a system has many failed logins. It could be someone trying to get into the system by a brute force attack. This should be investigated.

Figure 4. *Too many failed logins triggered a correlation rule in Sentinel.*

Figure 4 depicts an alarm caused by too many failed logins to a system. SIEM uses taxonomy definitions to detect failed logins and tracks the source IP and initiator username to determine whether they originate from the same source. If the number of failed logins exceed the appropriate limit set by the system administrator, SIEM system will automatically initiate an alarm event signifying potential malicious activity to the operator monitoring the SIEM system.

Event correlation detection can also be used to find root causes to incidents. For instance, in a situation where in a complex system hundreds of alarms are sounded conveying that servers and related services are no longer reachable. Event correlation tools can analyse the data in order to determine the root cause allowing the IT department to focus on implementing a solution rather than spending valuable time trying to pinpoint the cause. In an alarm situation, thousands or millions of events can be generated in just a short period. These events can range from informational to critical. While a good analyst can identify the root cause of failures, this type of knowledge is expensive to obtain. Event correlation technology was designed to automate and register interrelations between ongoing events in a more cost-effective manner. (Zhang, 2018)

Benefits of event correlation detection:

1. Real time threat visibility. Active event correlation and analysis can help IT departments to detect threats in real time. Failures, security breaches, and operational issues all affect the system.
2. Vigilance of network safety. The network can be monitored at all times. In addition, impact failures can be identified and remedied.
3. Reduces operational costs. Event correlation tools automate processes such as the analysis of large workflows to reduce the number of relevant alerts.

3.4 Security Monitoring

The effectiveness of SIEM is connected to security monitoring. SIEM performs the automated analysis of event data and presents analysed information to the operators who might be system administrators, security personnel, IT personnel or employees of security operations centres who are tasked with monitoring the state of a system. SIEM’s purpose is to parse, analyse and normalise the incoming data in such a way that requires little manual analysis but provides the user as much information as they need to form an accurate situational image of the system and perform any required countermeasures or follow-up actions to events.

SIEMs can detect events in large data sets. They can detect events that might otherwise stay hidden if the data was monitored purely by a human, as SIEM is better than humans are at detecting small anomalies in large samples of data. Furthermore, with the use of correlation rules, SIEM can detect suspicious activity by connecting a larger number of events, which is also something that is difficult for humans to do. SIEMs can be used to detect activities such as:

1. denial of service attacks,
2. brute force attacks,
3. suspicious queries,
4. port scans,
5. malware and
6. privilege escalation attempts.

From these, malware is one the most significant threat to industrial systems from the perspective of automation systems. In 2018, as much as 47,2 % of industrial control systems were infected with malware. Most common piece of malware was Trojans, which were found in 27,1 % of industrial control system computers. (Kaspersky Lab, 2019)

SIEM systems have some correlation rules built into them but end users can also create their own. Thus, correlation rules can be tailored to the specific application in order to provide analysis that is more accurate. When there is less need for manual analysis, countermeasures to error situations can be faster and more precise. This reduces damage caused to the system and the data in the system.

SIEM also provides an audit trail and forensic capabilities. SIEM can be used to trace an attack as it can collect data from multiple data sources and act as journal for all activity in the system. However, the danger is that business deploy SIEM to fulfil a legislative mandate to monitor and audit activity in a system without implementing it in a meaningful way. In these types of cases, SIEM is not configured properly and it might not serve its purpose. In the worst-case scenario, SIEM can hide important information, as the people responsible of monitoring the system rely on SIEM to provide them with all the necessary information. A poorly configured SIEM might not alert the monitors of noteworthy events and these events will stay hidden from the monitors. When implementing SIEM, it is important to focus on the purpose of the SIEM and configure it in a way that it provides a meaningful function for the whole target system.

Monitoring can be done in a *Security Operations Centre* (SOC). SOC is a centralised facility where information security professionals both monitor and implement countermeasures to threats for a client organisation. SOC and SIEM are a part of service model called *security as service*, in which information security management is outsourced to a specialist organisation. In addition, security as service offers up-to-date tools that are instantly available to use and the client organisations funds are directed towards actual security measures rather than buying software, acquiring licenses and training staff. Organisations that produce security as service are known as *Managed Security Service Providers* (MSSP). (Brook, 2018)

3.5 Benefits for Business

Excluding the development stage, handling event data can be divided roughly into three categories: automated processing of data, manual – human performed – analysis of the data and required countermeasures. In order to make this process more efficient, the automated processing should take the majority of the data analysis away from the human administrator or the person monitoring event data. This is the very purpose for which SIEM was designed. (Vesamäki, 2016)

Detecting anomalies and filtering relevant data from a large stream of events is a time consuming job that requires resources and affects businesses' resilience to information security threats. Detecting patterns and performing correlation detection would be very difficult if not impossible with large amounts of event data. SIEM takes away a lot of the workload from the administrators by automating this process.

SIEM improves reaction times to events, cuts down production shortages and redirects resources to understanding better the root causes of incidents and repairing them. SIEM also brings visibility to the organisation's operations and state. Even by preventing one cybersecurity breach, SIEM can make the entire investment financially profitable, as the costs of recuperation go down and the company might gain an advantage in the market in relation to its competitors. (Vesamäki, 2016). However, SIEM and SOC are expensive to deploy and maintain, so normally the value of these systems and services rely on continuous operations and long-time benefits.

4. AUTOMATION SYSTEMS

Automation systems can be single logic devices or factory sized systems intended to control the entire manufacturing process of the factory. In this thesis, the term automation system specifically refers to a process automation system. Process automation is used to improve the efficiency of the manufacturing process and improve the quality of the end result, which in turn improves the company's position in the market place. Automation is also used commonly to do dangerous or monotonous tasks.

4.1 Different Device Levels

Industrial automation systems can be divided into different levels based on the functionality of the devices. Common levels are control room level, cross-connection level and field level. Figure 5 illustrates these levels.

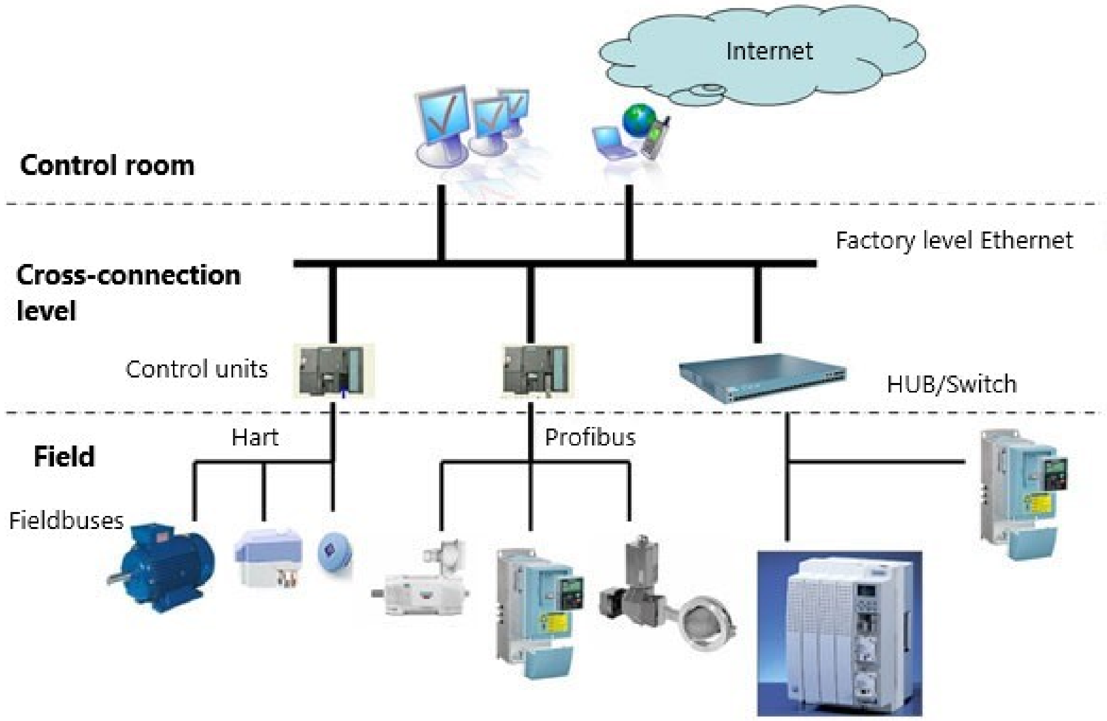


Figure 5. *Common different levels in automation systems. (Asp, et al., 2019)*

In figure 5, measurement data flows from the bottom level (field level) to the top level (control room level). Control input data in turn flows from top to bottom. However, in modern industrial automation systems, based on Cyber-Physical Systems (CPS) and Service-Oriented Architectures (SOA) and the direct communication and administration of field devices, the separation of these levels can be less clear (Bergweiler, 2016).

Cyber-Physical Systems are the amalgamations of computational and physical processes, where embedded computers and networks monitor and control physical processes (Lee, 2008). In Service-Oriented Architectures, the service and the product have a loose coupling. This means the service is less dependent on the product produced and it relies on offering a certain type of service through which the desired result is achieved. (He, 2003)

4.1.1 Field Level

Field devices are electronic devices that are located at the field level, the lowest level in the hierarchical level model for automation. This level is closest to the actual process. Sensors, actuators and general field devices are located here. Field devices are associated with sensors that detect the data of the measuring points and send the control data to the actuators. Field devices continuously supply measured data for process control and receive control data for the actuators. (Bergweiler, 2016)

Automation systems include vast numbers of field devices. A single system can have thousands of measuring devices also, known as measuring points, each of which, for instance, could send one measurement every second. This results in large amounts of data that have to be transferred via the communication network.

4.1.2 Cross-Connection Level

The cross-connection level is comprised of various automation controllers and connecting fieldbuses. This is also the level in which factory level Ethernet would reside as is shown in figure 5. Controllers' function is to gather measurement data and give new control inputs to field devices. (Mraz, 2017)

One very widely used controller type in modern factories are Programmable Logic Controllers (PLC). PLCs are units that are made from various components such as central processing units, both analogue and digital inputs and outputs and communication modules. Advantage of using PLCs is that they allow automation engineers program control functions or strategies that drive the automation process. (Mraz, 2017)

4.1.3 Control Room Level

The control room level, sometimes referred as the supervisory level, is the level furthest from the actual process. This is where automatic devices and monitoring systems aid the control and adjustment functions, with the help of Human Machine Interfaces (HMI), distribution control systems and supervisory control and data acquisition (SCADA) devices.

These are used to monitor process parameters and measurements, set control inputs for the process, store measurement data and set machine start and shutdowns. (Mraz, 2017)

Control room level devices might have access to the Internet. As a result, control room level is the only level with a connection to the outside of the system. This creates a vulnerability to the whole system as it might act as an entry point for cybersecurity threats. To counter act this issue, devices that have access to the Internet should always be separated from any devices that control the process or the system overall by the use of a *security zone* (Crevatin, et al., 2005). For high security requirements, separating the control room network from the Internet is done with a *demilitarized zone* (DMZ), where access servers and other servers that are to be accessible from the external network are placed. The DMZ is isolated from both the office network and the external network by firewalls. These firewalls are configured to allow access from the Internet only to selected servers. To help the DMZ to detect anomalies it can be connected to an intrusion detection system that monitors network traffic and uses correlation rules to detect malicious events. (Crevatin, et al., 2005)

4.1.4 ISA-95 System Level Classification

ISA-95 (ANSI/ISA-95) is an international standard for developing an automated interface between enterprise and control systems. The standard is used by manufacturers in automated batch, continuous and repetitive processes and its main purposes is to establish a common terminology and understanding between suppliers and manufacturers. It helps to provide a consistent model for operations and makes the functionality of applications more clear. (Castilla, et al., 2011; Åkerman, 2016)

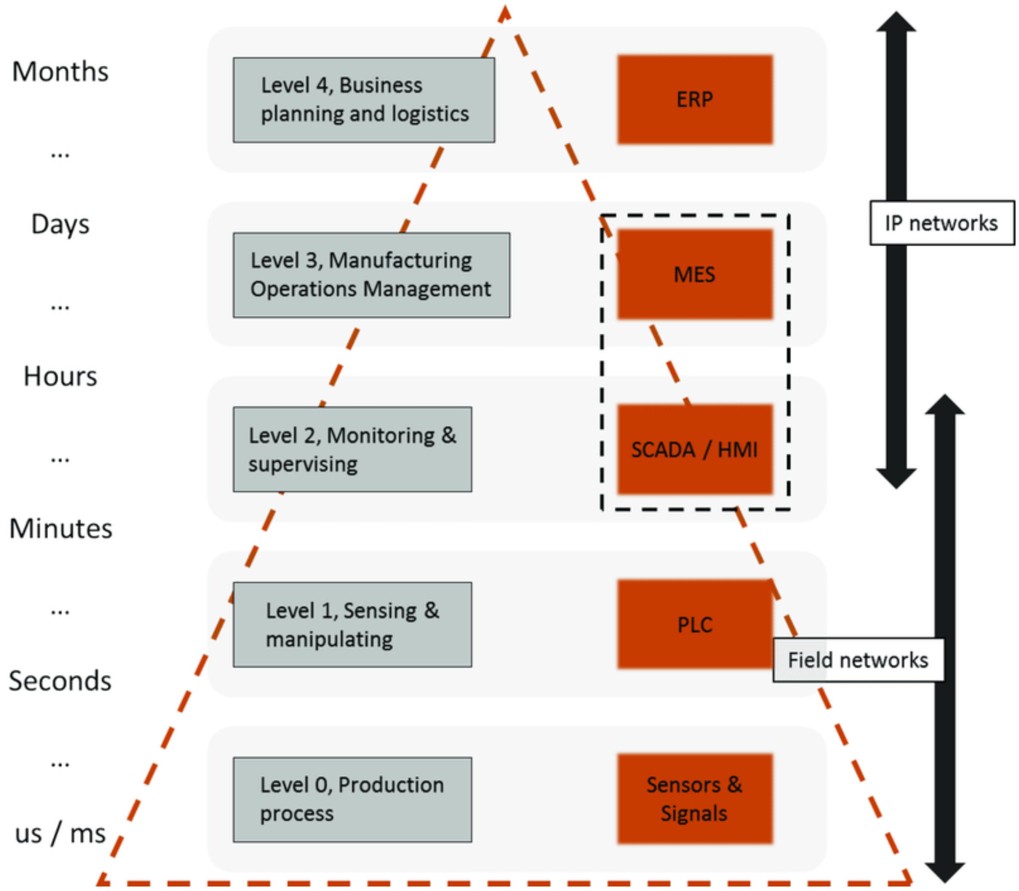

Months
...
Days
...
Hours
...
Minutes
...
Seconds
...
us / ms
Level 4, Business planning and logistics
ERP
Level 3, Manufacturing Operations Management
MES
Level 2, Monitoring & supervising
SCADA / HMI
Level 1, Sensing & manipulating
PLC
Level 0, Production process
Sensors & Signals
IP networks
Field networks

Figure 6. *ISA-95 standards multilevel hierarchy for industrial enterprises from level 0 to level 4 (Åkerman, 2016).*

Figure 6 shows the different hierarchical levels of the ISA-95 standard. ISA-95 divides the industrial enterprises to five levels. The first level, commonly referred as level zero is very similar to the field level described in chapter 4.1.1. It contains field devices and the communication in this level comprises of measurement data, control input signals and alarms. The timeframe devices in this level are concerned with is measured in microseconds and milliseconds. Levels 1 and 2 are also analogous to the cross-connection and control room levels in chapters 4.1.2 and 4.1.3. Figure 6 shows, however, that level one's timeframe is measured in seconds and level two's timeframe is measured in minutes. (Castilla, et al., 2011; Zhang, 2018)

ISA-95 introduces levels 3 and 4. Level 3 is the Manufacturing Operations Management level and level 4 is the Business planning and logistic level. Level 3 is responsible for the supervision of the manufacturing process on the whole from the field level to the control room level. Level 4, on the other hand, is also known as the Enterprise level that oversees the functioning of the enterprise and comprises of superior management policies of the enterprise. It is responsible for the development and operations of the enterprise including all of its plants and respective production lines. Levels 0, 1, 2 and 3 create the *industrial zone* and level 4 is called the *enterprise zone*. (Castilla, et al., 2011; Didier, et al., 2019)

Figure 6 also illustrates the communication technologies on the different levels. Levels from 0 to 2 use field networks to communicate and levels 3 to 4 use IP networks. IP networks refer to communications through Ethernet or Internet.

4.2 Manufacturing Execution Systems and Operations Management

Manufacturing Execution System (MES) is an integrated system of a collection of process control related functions. This can mean a collection of separate pieces of software that collect data from the process and control the production, or it could be single piece of software that handles all the necessary data collection and device control. Typical task for MES is automated data collection from the process, tracing manufacturing history of the product, handing work instructions and other documentation and managing maintenance and quality control information. Furthermore, MES acts a real time link between the different levels in automation processes. It transfers data from the process to the control room level and production orders and situation queries from the upper levels to the field device levels. (Hästbacka & Kuikka, 2017)

Manufacturing Operations Management (MOM) on the other hand is a process that is used to observe and evaluate the entire manufacturing process in order to optimise its efficiency. MES and MOM are similar in many ways. However, MOM is more focused on the observation of the operational process where as MES is more concerned with individual systems and software. (Hästbacka & Kuikka, 2017)

4.3 System Reliability

One of the most important aspects of automation systems is reliability. Industrial automation systems can operate for years without shutdowns and conditions on the field device level are often challenging. Dust, heat and chemicals impose problems for electronic devices, execution of the process is depended on functionality of the controllers and data transfer and every shutdown is expensive. Shutting down an industrial process can cost up to 100 000 dollars an hour (Armstrong, 2016).

Systems reliability can be improved in many ways. Some improvements come from the hardware devices and their software and some come from control software. The system can have for instance duplicate devices or fieldbuses so that when the primary device stops working all operation can be transferred to the secondary device. Systems could use self-diagnostic tools to detect possible problems and find root causes for them. Also, when designing the system, paying attention to having proper alarm systems that indicate when problems occur and designing the entire system in a way that problems in one part of the factory influence the other parts as little as possible.

4.4 Distributed Automation Systems

Process automation systems are typically Distributed Control Systems (DCS). DCS typically consists of process control stations, monitoring stations, fieldbuses, programmable logic controllers (PLC) and data management stations. These data management stations might not exist in small systems. (ABB, 2007)

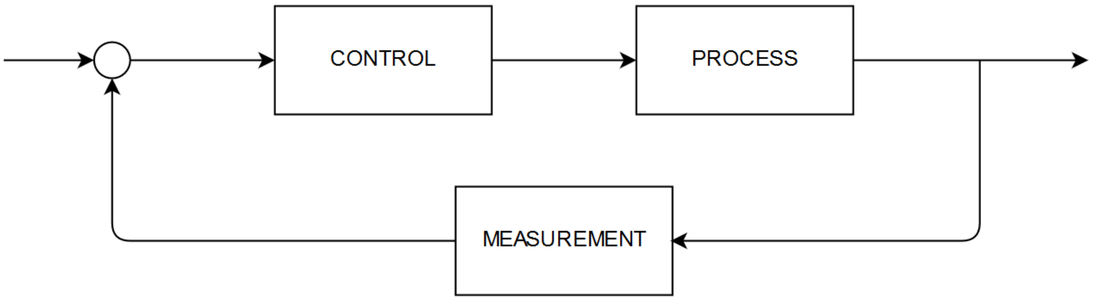


Figure 7. *Block diagram of a simplified automation system.*

Figure 7 shows a simplified block diagram of a general automation system. Automation systems collect measurement data from the process. This is shown in the *process* and *measurement* blocks in figure 7. Based on the measurements, the system calculates the required control inputs for controllers. Controllers are located in the *control* block in figure 7. This is called a feedback loop. The measurements can also be displayed to system monitors in the control room from where human operators can change inputs if needed. Overall, the purpose of the measurements is to help create a good understanding of the state of the process so that it can be improved. The quality of the data is an important factor in achieving this. (ABB, 2007)

Process control’s real time database is distributed by taking process control stations or process control systems close to the process. In distributed control systems, process control stations are capable of handling measurement data, calculating control inputs and issuing new input for controllers. By doing this, the data does not have to be sent to a central computer and back for the calculations. Furthermore, automation systems can be built in accordance to the process layout. Taking process I/O units close to the process helps in saving resources when it comes to cables as the distance between the processes and processing units is short. (ABB, 2007)

4.5 Requirements and Challenges of Automation Systems

Process automation system is a challenging environment for both hardware and software. It imposes special requirements for all devices, connectors and the logic that drives the system and even short downtimes can be expensive. Therefore, implementing new features into an automation system is a long and costly process where every single detail is looked at with high precision. For SIEM to be included in an automation system, it must be able to comply with the special conditions and requirements.

One major challenge for implementing SIEM in an automation system is that to an automation engineer information security is an additional service. The process itself is the main purpose and main function of the entire system. Additional services require additional resources and investments, and if they do not relate directly to the process, justifying them might prove difficult, as *efficiency* and *reliability* of the process are the main aspects that engineers are concerned with in an automation system. All additional services and functionality should be justified through these aspects.

4.5.1 Design Phase

When adding functionality to an automation system, the planning stage can be expensive compared to the cost of the products added to the system. In automation systems, replacing single wires can be costly even though the wires themselves might be inexpensive. Planning involves multiple steps to ensure that the added functionality or devices do not disturb the process, and that the modifications would yield the best possible results.

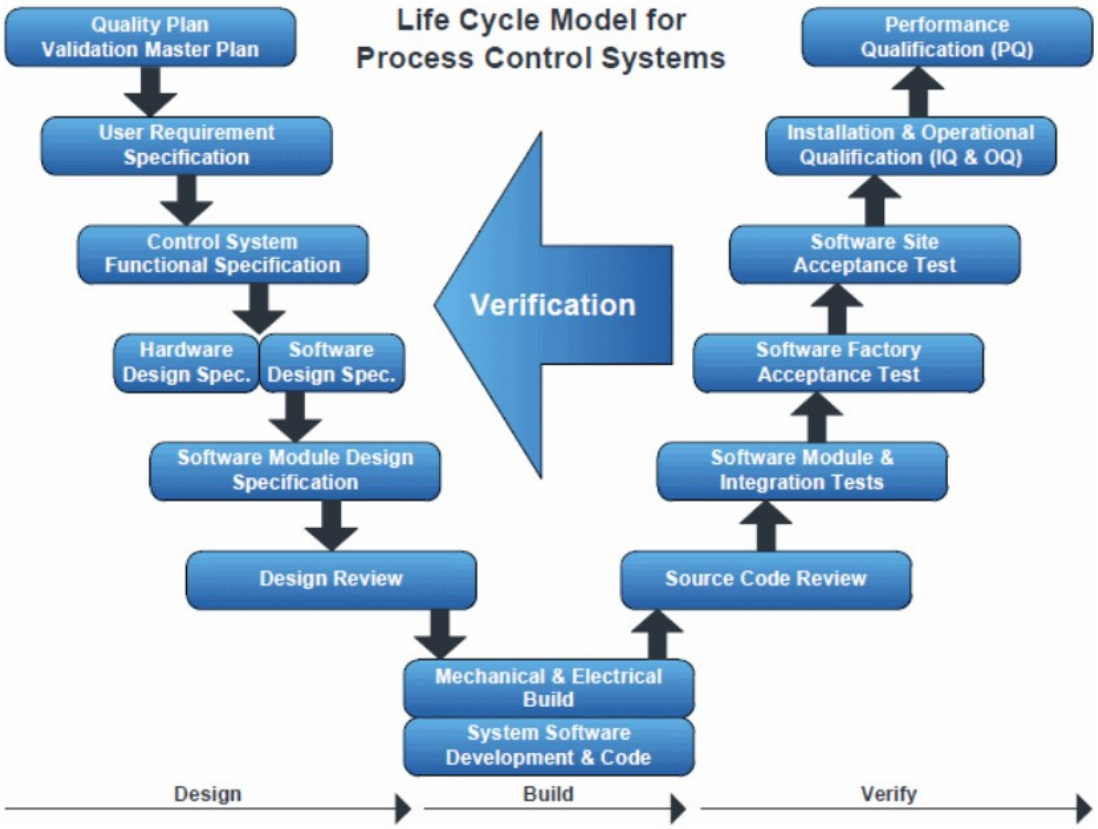

Life Cycle Model for Process Control Systems
Quality Plan Validation Master Plan
User Requirement Specification
Control System Functional Specification
Hardware Design Spec.
Software Design Spec.
Software Module Design Specification
Design Review
Mechanical & Electrical Build
System Software Development & Code
Source Code Review
Software Module & Integration Tests
Software Factory Acceptance Test
Software Site Acceptance Test
Installation & Operational Qualification (IQ & OQ)
Performance Qualification (PQ)
Verification
Design
Build
Verify

Figure 8. *Life cycle model for process control system. (ServicesParis, 2010)*

Figure 8 shows a model of a process automation system's development and modification life cycle. The left cascade in the model describes the design phase, the middle part is the building phase and the right cascade is the verification stage. The planning stage is comprised of multiple phases from quality planning to design review.

Every modification to the system requires an extensive inquiry to the effects of the modification. Changing one component in one place might influence other components elsewhere and understanding this is vital. Mistakes are usually cheapest to fix in the planning phase, however, mistakes that that are caused by inadequate planning and are fixed later on in the life cycle of the modification process are often much more expensive. Configuration changes in the field device level cause a medium risk while changes in the cross-connection level cause a higher risk, as that is the place where a lot of the control software and logic is housed (Goss, 2010).

Furthermore, when choosing new components to be added, deciding between a multitude of vendors and different specifications of the component is not a trivial process. The field of automation is full of product manufactures, which offer their own versions of components, software or protocols. Manufacturers might also supply different lifecycle services such as maintenance and log handling.

4.5.2 Performance and Reliability Requirements

One of the more important factors in automation is the *performance* of the system. The end product or purpose of an automation system, be that an actual manufactured product or an end result of a function that the system provides, is dependent on the performance of the system. Issues in performance can have a negative economic, quality and safety implications, so performance to an automation engineer is perhaps the most important factor in an automation system.

Automation systems contain controllers that require certain performance capabilities in order to control devices with the precision required. During a regulatory task, automation system must measure output of the system or part of a system, calculate the feedback loop value from the measurement and the input of the device and give a new control input to the system on a constant cycle. The feedback loop can be seen in figure 7.

Performance capabilities are an important factor in automation system *reliability*. Reliability describes the probability of the system to function as desired and disruptions in performance will have an effect in system reliability. As reliability is one of the important factors from the perspective of an automation engineer, this puts more emphasis on the system's performance requirements.

4.5.3 Software Requirements

Software in automation system has to operate within certain real-time requirements. Real-time requirements can be categorised into two categories: *soft* and *hard real-time* requirements. Soft real-time means that the response from a piece of software or system is not highly time critical, whereas hard real-time means that the response has to come within a certain time frame. The internet for instance is a soft real-time environment. When a user requests a web page, the time it takes for all the packets of the page to arrive can vary significantly depending on internet speeds, server performance, congestion and the route the packets take. From the users standpoint all this means is that the time they have to wait for the page to load will vary, but disregarding user experience, the service is still successful.

The timing of functions in an automation system are often precise. How precise they are, depends on the application and its scheduling policies. *Fair scheduling policy* means less strict timing is required and *priority-scheduling policy* means that application requires timing that is more precise. Measurement, control inputs and computational algorithms precision on the other hand is very precise.

Other important factor in software for automation systems is usability. In a control room, the amount of functionality that is available to the user is vast and control room software can contain multiple measurement or system variable displays, warning and error information, video feeds and other data that the user of the software must monitor. User interfaces are designed individually to every system and the users of the system. Its main function is to assist the users in their tasks.

Quality in automation system software can be obtained by adhering to systematic software development processes and software quality factors that are crucial in any professional software development project. One important factor related to quality is dependability and it can divided into following categories that create the acronym RAMS:

1. reliability,
2. availability,
3. maintainability and
4. safety.

Reliability can be defined as the probability that a system will perform a specified function within prescribed limits, under certain environmental conditions, for a specified time. The limits under which reliability is judged can be quantified by defining constraints for acceptable performance. *Availability*, much like in the information security acronym CIA, means that the required resources are available when they are needed. Problems in availability can be disruptive to the entire system. *Maintainability* can be defined as the probability that a failed item can be restored to an operational condition within a given period of time. Maintainability is an important factor in a system’s recuperation and it can have significant effects in reducing automation system downtime. Lastly, *safety* describes the probability of a failure. (Stapelberg, 2009)

Defects that have negative impacts on automation system software dependability may be divided into software faults, errors, failures, accidents, incidents and hazards. Software faults are caused by malfunctioning code, which includes code that is incorrect or missing. This is a fault where the software is broken and therefore will never function correctly. An error is either a human action that produces a software fault or a manifested fault in the software, which is caused by the software coming across an unpredictable scenario and does not know how to operate. A failure, on the other hand, is a scenario where a system or component cannot operate within the specified limits. Failures can be classified into two categories: systematic and random. Systematic failures are failures caused by a specific set of environmental conditions or factors inherent to the system. In software all failures can be classified as systematic as random failures are caused by hard to predict variables caused by variations in materials, manufacturing process or environmental stress. The rate of random failures can be predicted. However, predicting individual events is nearly impossible. (Hästbacka, 2017)

Accidents refer to the possible damage to the system, components or users caused by a failure. Whereas, incidents are unplanned events that do not result in damage to the system or its users, but it is still an event that had potential for damage. Hazards are set of conditions or scenarios that might lead into an accident. All failures are the results of hazards but not all hazards cause accidents. (Hästbacka, 2017)

4.6 Lifecycle and Maintenance

Automation systems have long lifecycles. The lifespan of these systems overall can be measured in decades. Devices and software added to the system also increase the amount recourses that need to be directed towards maintenance throughout its entire lifecycle. New technologies bring new requirements for maintenance, and all hardware and software in automation systems is generally minimised. Correct maintenance insures the system produces the highest possible results while minimising risks (Goss, 2010).

Table 1. *Typical lifespans of devices in the different device levels of automation systems.*

Device level	**Typical lifespan (years)**
Field device	10-15
Cross-connection	15-20
Control room	4-6

Table 1 displays the typical lifespans of devices in the different device levels. On the field device level, device lifespan is typically 10-15 years, on the cross-connection level it is 15-20 years while on the control room level typical lifespan is only 4-6 years (Goss, 2010).

The control room level brings the most challenges. When the system becomes more complex, the integration and interaction of the levels also becomes more complex. This is especially the case when using Custom Off-The-Shelf (COTS, alt. Commercial Off-The-Shelf) products, which are software products that are not bespoke to every individual application, but are customised to some degree for every customer. In other words, instead of making the software for every environment separately, they are merely adapted to the environments. As COTS products are being implemented more and more, automation engineers, system administrators and maintenance engineers are confronted with new releases that might cause compatibility issues. Furthermore, as time goes by software develops even faster. In order to keep systems up-to-date they have to easily upgradable, which automation systems traditionally are not, because upgrades on automation systems might require system downtime, which in turn, is expensive. (Goss, 2010)

Having professional automation engineers working at an automation system might be challenging. Finding people with the appropriate level of expertise, knowledge and experience in all of the systems areas and the managing a system is often delegated to a system administrator. The administrator is responsible for the operation and maintenance of the system. Furthermore, automation technology is improving constantly making it challenging for administrators to keep the system's hardware and software revision levels up to date for both administrative and technological standpoint. Staying in budget and in the allowed schedule are also challenges in updating automation systems. (Dedzins & James, 2015)

In addition to being challenging, having around-the-clock monitoring of an automation system is also cost prohibitive. Resources for monitoring an automation system are often scarce so they have to be invested wisely. Automation systems have maintenance and diagnostic tools to display information about its operational status and system alarms to indicate of any error situations. However, the information is provided after the incident so it does not act well as a precaution. (Dedzins & James, 2015)

4.7 Correlation Rules in Automation Systems

Correlation rules form a key functionality in SIEM systems. Just like in systems that monitor IT security, automation system SIEM will combine security related events with correlation rules. These rules can backtrack historical event data or analyse incoming log data in real-time to detect abnormal behaviour. Correlation rules in automation systems can vary from IT security correlation rules. (Bajramovic, et al., 2016)

Correlation rules can be used in automation systems to detect for instance:

1. increase in network traffic,
2. changes in the statistical distribution of message types,
3. improper maintenance tasks,
4. suspect usage of employee ID or access card,
5. suspect logging events in the system or
6. activity outside normal working hours.

Increase in network traffic and changes in the statistical distribution of message types can indicate of an attacker gaining access to the communication network and manipulating messages. (Bajramovic, et al., 2016)

Warnings from improper maintenance task might occur for example when maintenance actions are performed in the wrong order. Maintenance engineer could be patching the system while cabinet lock-monitoring system seems to be in place and the proper maintenance sequence requires an unlocking event of the cabinet to occur prior to any maintenance events. Correlation rules can also be used to monitor that a locking event happens after maintenance procedures in order to make sure no areas remain unlocked. (Bajramovic, et al., 2016)

Suspect usage of employee ID or access card can happen when an employee’s ID card is used to access physical areas, systems or resources in the monitored plant after they have left the plant. Any events related to a specific ID card or the employee logging into a system or device in the plant after the leaving event has occurred is a possible indication of malicious activity, improper functionality in the system or operator error. Related to this is detecting activity from employees after normal working hours or detecting that employees have not exited from the plant in certain amount of time. (Bajramovic, et al., 2016)

5. LOG DATA IN AUTOMATION SYSTEMS

Automation systems generate a lot of log and other measurement data. Behind a single switch there can be hundreds of measurement points each producing measurements on one-second intervals. This means that there is a great deal of data to be transferred through communication buses. The measurement data is integral to the function of the process' control system, whereas log data might not as necessary.

5.1 Transferring Log Data

In automation systems, device logs are not monitored. The control room receives information about measurements of the system but it usually does not get any log data. These logs are only looked at in an error situation. Specifically, when a device is not working in the desired way. As a result, transferring logs would require unnecessary bandwidth on the communication network of the system. Incorporating this with the performance requirements of the process, would mean that some headroom would have to be reserved on the network for transferring data that is useful in rare instances. Transferring log data cannot interfere with the performance of the system.

For this reason, centralised log gathering is not common in automation system. Log data stays with the devices on the field device level and in the case of a malfunctioning device, an engineer will physically go the device and access the log data. Furthermore, this engineer is most often an employee of the device manufacturer or device supplier. Manufacturers and suppliers offer this as a service as they have intricate knowledge of the device. A single automation system contains devices from a multitude of suppliers so it is difficult for system administrators to have extensive knowledge of each device and their log formats and policies.

5.2 Log formats

All device manufacturers have their own log formats. They all will probably contain similar information, such as timestamps, error codes and error descriptions, but the format and the interpretation of the error codes, descriptions or other information would still be manufacturer-specific.

Timestamps in particular have an important role in the log data. In order to meet real-time requirements of ICS systems, timestamps have to comply with a predefined accuracy on all different components. Universal clocks are for ensuring the timestamps in events sent from one device are accurate in relation to timestamps from all other devices. Universal clocks are not, however, mandatory in ICSs and individual components usually have their own clocks. If devices use their own clocks, proper protocols have to be implemented for reliable time-synchronisation between logging devices. If logging devices have time differences in their clocks, determining the sequence of events is difficult. Furthermore, the accuracy of the clocks plays an important role, as the sequence of events that occur within a certain minimum timeframe cannot be determined. (Bajramovic, et al., 2016)

Timestamps have an important role in gathering log data. Events might not be recorded in the correct sequence. Events that occur later might be recorded before events that occurred earlier, but the timestamps will inform of the proper order. Time-synchronisation of device clocks is best implemented during the design phase of the automation system. If automation system logs are only analysed when problems occur and they are analysed from one device at a time, timestamps do not play such an important role. When analysing multiple logs from multiple devices the sequence of events is important in understanding where the problem originated and how the error situation proceeded. (Bajramovic, et al., 2016)

5.3 Handling Automation Systems Log Data

The importance of communication in automation systems has grown in the last two decades. This is the result of transferring from using traditional fieldbus solutions to using TCP/IP protocol. TCP/IP has brought more complicated and thus more vulnerable communication solutions to automation systems. As a result of automation systems starting to use TCP/IP solutions, a new sub-system has been brought to automation systems – communication networks. However, as the measurement data is integral to the function of the system where as monitoring communication components and infrastructure are not seen as vital as monitoring the process itself, the development of these aspects is usually not at the same level. (Ahonen, et al., 2019)

Monitoring an automation system is not done only for controlling the process. Through monitoring, you can anticipate the service needs helping maintenance, make repairing issues caused by failures easier or collect data for environmental reports for example. In addition, monitoring is used to save resources and time, help quality control and ensure security. Monitoring helps improve situational awareness of the system, which in turn aids engineers operating the system, maintenance measures, security and quality control. (Ahonen, 2017) (Ahonen, et al., 2019)

On the other hand, automation engineers try to limit the amount of data to be followed in the control rooms, as the more data there is to be followed the more difficult it will be to form a good understanding of the state of the system. This is one of the reasons why monitoring communication networks might be sidelined in the way of process monitoring. The additional value that monitoring communication networks would bring have not been able to demonstrate enough to justify their importance to automation engineers. It has remained as additional service that is only used in special circumstances. It has been largely sold as a tool against information and specifically cybersecurity threats and its benefits for production overall have not been emphasised enough. (Ahonen, et al., 2019)

6. CYBERSECURITY IN AUTOMATION

Cybersecurity can be divided into two categories: physical cybersecurity and non-physical cybersecurity. The former is the protection of information and assets related to information from physical threats such as breakages, thefts, vandalism, fire, floods and other natural disasters. The latter, on the other hand, is concerned with protecting systems, networks and programs from threats emanating from the digital world. These threats are not just attacks but also include non-malicious threats.

Cybersecurity can also be divided into three distinct levels: *strategic***,** *operative and technical*. These levels refer more to the execution of cybersecurity countermeasures. At the *strategic* level, a company will create a strategy that defines all events and actions that might have a negative effect on the operations of the company or system. From these actions, all risks relating to cybersecurity can be identified and resources for preventing these events can be allocated. An important part of this level is defining acceptable residual risk that describes the risk that remains after all preventive measures have been implemented. Sometimes a particular risk cannot be eliminated entirely, and in these situations it will have to be reduced to an acceptable level. (Kääriäinen, 2019)

The *operative* level directs the strategy towards real actions, and the *technical* level implements the technical actions according to the strategy. Managing cybersecurity in a company would be done by maintaining the cybersecurity strategy that will set in motion all actions on the operative and technical levels. This requires close teamwork from the management and technical departments in a company. (Kääriäinen, 2019)

6.1 Cybersecurity Objectives

Cybersecurity has three key elements that form the information security acronym *CIA: Confidentiality, Integrity* and *Availability*. Confidentiality means that only the authorised parties, individuals or systems have access to the information. Integrity refers to the assurance that data remains intact throughout its lifecycle and that no one, without proper permission, has tampered with it in any way. Availability means that the required data, systems, networks or programs are available whenever they are needed. This triangle applies well for IT security. However, with automation systems the more important factor is operational technology (OT) security. In OT security, the process and the control of the process are the most important aspect to secure rather than for instance information ending up in the wrong hands. This is a large distinction between automation systems and other conventional systems involving IT. In automation systems the acronym is more accurate in the form *AIC* (Availability, Integrity and Confidentiality), which emphasises the importance availability over confidentiality.

In automation systems specifically, availability concerns all IT elements such as control systems, safety systems, operator workstations, engineering workstations, manufacturing execution systems and the communication systems between these elements and to the outside world (Crevatin, et al., 2005). Problems in availability can interrupt the entire automation process causing system downtime and corruption of the manufactured product, both of which are costly and can take a lot of time to repair and resume to normal operations. Furthermore, it can cause safety issues if certain safety features are off-line.

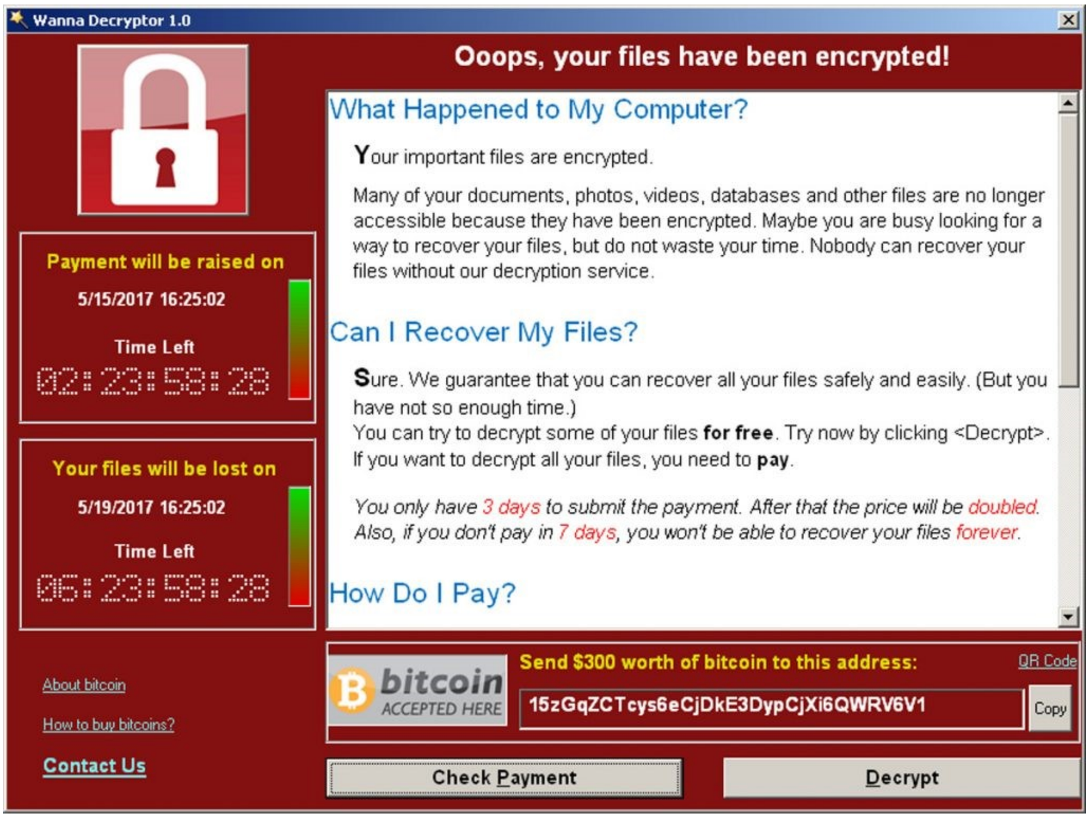
6.2 Cybersecurity Threats

Automation systems might encounter a multitude of cybersecurity threats. Most threats that regular systems face concern also automation systems. However, automation systems tend to have long lifespans. Devices, programs and operating systems are renewed seldom so systems might include old and unsupported software that is vulnerable to attacks. These attacks include, for instance: denial of service, eavesdropping, man-in-the-middle and various types of malware (Crevatin, et al., 2005).

Denial of service relates to availability, in that, during an attack the target system is prohibited from performing its normal functions causing financial or physical harm to the system and possibly its operators. During eavesdropping, an attacker would follow and capture communications between programs, devices, systems or people. It might not involve any manipulation of the data and it could be just a method of collecting sensitive data or perform reconnaissance for future attacks. Unlike in eavesdropping, during a man-in-the-middle attack, the attacker will take an active role in the communication between systems or people. As the name suggests, the attacker would place themselves in the middle of the line of communication so that none of the other participants would notice. When one participant sends data to another participant, the data would instead end up to the attacker, who, in turn, would relay the data modified or unmodified to the correct recipient. The attacker can change or collect data and the other participants would think they are talking to one-another directly.

Malware is a malicious piece of software. Malwares can be categorised into multiple different types such as Trojans, viruses, worms, ransomwares and rootkits. Malware's purposes include stopping system operation, gathering information or harnessing devices, or parts thereof, for other purposes such as bot networks. Malwares know as ransomware lock a user's computer and encrypt their data so the user cannot access it. The ransomware then informs the user that they get their data back if they pay a certain amount of money, usually in the form of Bitcoin or other cyber currency because of their untraceability.

In 2017, a ransomware known as WannaCrypt or alternatively WannaCry infected tens of thousands of computers in over 150 countries. The ransomware used a vulnerability found in older Windows operating systems, such as Windows XP. These older operating systems were (and still are) used in businesses, industrial systems and other professional environments, but not too much in peoples' personal use. In Europe, many hospitals came infected crippling healthcare systems for several hours. The British National Health Service had to turn away all non-emergency cases as 16 of its computer networks were shut down. Legacy vulnerabilities, which are vulnerabilities that affect older devices, are a problem in automation systems, as computers, software and operating systems are not updated to newer ones very often. (Wagner, 2017)



Wanna Decryptor 1.0

Ooops, your files have been encrypted!

What Happened to My Computer?

Your important files are encrypted.

Many of your documents, photos, videos, databases and other files are no longer accessible because they have been encrypted. Maybe you are busy looking for a way to recover your files, but do not waste your time. Nobody can recover your files without our decryption service.

Can I Recover My Files?

Sure. We guarantee that you can recover all your files safely and easily. (But you have not so enough time.)
You can try to decrypt some of your files **for free**. Try now by clicking <Decrypt>.
If you want to decrypt all your files, you need to **pay**.

You only have 3 days to submit the payment. After that the price will be doubled.
Also, if you don't pay in 7 days, you won't be able to recover your files forever.

How Do I Pay?

Payment will be raised on
5/15/2017 16:25:02
Time Left
02:23:58:28

Your files will be lost on
5/19/2017 16:25:02
Time Left
06:23:58:28

Send $300 worth of bitcoin to this address:
15zGqZCTcys6eCjDkE3DypCjXi6QWRV6V1

Figure 9. *WannaCrypt's user interface informing a user that the data on their computer has been encrypted. (Wagner, 2017)*

Figure 9 shows the user interface for a computer infected by WannaCrypt. It is designed in simple way that even people with little technical knowledge could get the necessary information and pay the ransom.

One well-known malware created for automation systems is Stuxnet. It was originally aimed towards Iran’s nuclear facilities that were trying to enrich uranium. Stuxnet would specifically target certain type of Siemens programmable logic controllers in the automation systems and managed to break multiple centrifuges in Iran’s Natanz uranium enrichment facility by burning out the centrifuges used in the process. In addition to Siemens’ devices, it also contained code that affects frequency inverters made by Finnish company Vacon. Vacon denies selling frequency converters to Iran. (McAfee, 2019; Drivers&Controls, 2010)

Stuxnet is commonly believed to have been created, although both parties denied it, by the U.S. National Security Agency and Israeli intelligence. After its creation, it has been modified by different groups to target power plants, water treatment facilities and gas lines. Stuxnet was distributed via a USB memory stick that an employee must have inserted into one of the computers in the facility to contaminate the system. Subsequently Stuxnet has been spread to other industrial systems by accident. (McAfee, 2019; Rosenberg, 2017)

There was a time when automation systems relied on security through obscurity. It involves thinking that the system, devices and protocols used were too obscure for attackers to exploit. However, this has not been the case for a long time now, as Crevatin et al. (2005) wrote about the insufficiency of this mind-set already back in 2005.

6.3 Current State of Cybersecurity Management

Systems with Information and automation technologies are developing and spreading further into different areas of society. Many automation control systems, including the critical infrastructure systems, were designed and built a few decades ago. These systems are developing into structures that are more complex and the threats they face are becoming more diverse. In the past they were designed for physical threats, however, currently, the amount of threats emanating from the digital world increases. Furthermore, as the dependencies of these systems grow, they become more difficult to manage and the risks increase. (Ahonen, 2017; Gold, 2009)

Risk are managed by:

1. Improving detection and reaction capabilities to prevent risks.
2. Choosing industrial components and systems from suppliers that are proven to be reliable.
3. Designing and sharing cybersecurity strategies with different levels of the organisation or system from the strategic level to the operational level or from the enterprise level to the industrial level according to ISA-95.
4. Identifying critical systems in the entire organisation from enterprise level to industrial level, all remote access solutions and cloud solutions in order to protect them from threats.
5. Managing services and supply chains from investment decisions all the way to the end of the systems lifecycle.
6. Taking into consideration international customer requirements, recommendations and legislation.
7. Implementing tested and proven technical implementations to prevent risks.

Many of these items require extensive team effort from the different levels of an organisation. Cybersecurity solutions cannot be designed and implemented separately for different levels as these levels have dependencies to each other that have to be managed through unified cybersecurity solutions. (Ahonen, 2017)

The implementing cybersecurity proceeds in different stages. The work begins with risk assessments to determine the cybersecurity risks of the automation system. This is part of the strategic level of cybersecurity. If the company, to which the automation system belongs, has a cybersecurity strategy, it will come into play at this stage. (Kääriäinen, 2019)

6.3.1 Implementing Cybersecurity

From the perspective of automation systems, a substantial source of cybersecurity threats is the communication links to the outside of the system. Communication technologies in automation systems have partially moved towards TCP/IP solutions and having multiple systems integrated together has widened the potential for cybersecurity threats. Therefore, networks in automation systems are segmented in a way that prevents communications from one level of the system to another if it is not necessary. (Kääriäinen, 2019)

In many cases, industrial networks are separated from the internet. These networks are only accessible on-site and this is done in order to mitigate the risk of the system being infected by malware. On the ISA-95 standard, levels 3 and 4 in figure 6 are separated by a *demilitarized zone* (DMZ). In this case, levels 0, 1, 2 and 3 have no access to the internet. Devices outside the DMZ, on level 4, may have access to the internet. These devices can have communications to devices on the lower levels through a secure node in the network. This node acts as an interface for communications over the DMZ. (Kääriäinen, 2019; Ahonen, 2017)

DMZ reduces the risk of infection but does not remove it completely. Malware can be for instance brought inside the DMZ through infected devices such as external memory devices. Improperly configured firewalls in the DMZ or any routers in the system can also let in undesired software. Routers should block all unnecessary traffic, however, in the worst-case scenario they are just left in default settings without doing any configuration. (Kääriäinen, 2019)

Automation systems require communications with enterprise level (level 4) applications to exchange manufacturing and resource data. However, direct access to the automation system might not be required. An exception to this having remote access for managing the system. (Didier, et al., 2019)

Remote access is done with a *Virtual Private Network* (VPN) connection. When using VPN, all internet traffic is routed through a specific VPN server and from the target service's point of view, all this traffic originates from the VPN server regardless of its true origin. If the VPN server is in the on-site network of the automation system, it can used to access the devices in the network remotely. VPN is widely used outside industrial solutions but industrial-grade VPN equipment should be used when accessing industrial systems. Furthermore, industrial-grade VPN is recommended instead of cloud-based remote connectivity services (Kääriäinen, 2019).

Modern industrial automation and control systems (IACS) are commonly required to have network infrastructures that are:

1. scalable,
2. reliable,
3. safe,
4. secure and
5. future-ready.

The term future ready refers to the support of future technologies such as *Internet of Things* (IoT). (Didier, et al., 2019)

Internet of Things is comprised of computing devices that communicate over the internet without the need of human operators. Therefore, IoT devices are sometimes referred to as smart devices. An example of an IoT device is an oven that is connected to the internet. This type of oven could be controlled remotely and it could transfer data about the temperature of the oven and any food items being cooked. IoT devices such as this create cybersecurity threats if they are not configured properly. No access control or the usage of default passwords may allow malicious parties to access and control the devices. When IoT devices are industrial devices the term *Industrial Internet of Things* (IIoT) is often used.

Smart factory devices can be used to streamline automation system's operations. However, manufacturers do not always secure the products enough. This can make IIoT devices vulnerable to disruptive remote exploits over the internet. (Gold, 2009)

New TCP/IP technologies, remote access capabilities and the convergence of logical and physical resources means that the divide between IT and OT security is becoming less clear. At one point, OT and IT did generally different things and they had limited collaborations. As a result of the convergence of IT and OT in automation systems, it can be difficult to determine who is responsible for protecting the system that are owned and operated by the organisation. IT has experience and budget for digital security but lacks direct oversight of the automation system. OT supervises the automation system but is not mainly responsible for defending the organisation against digital threats. The enterprise level of the organisation is concerned with IT security, which has to be protected as well and the responsibility of this is on the enterprise level. (Gold, 2009)

Focusing on OT security over IT security when implementing cybersecurity in automation systems is no longer given. As there is no clear divide between the two categories, it creates confusion. However, the spread of IT malware, such as WannaCrypt, has indicated the growing need of IT security solutions in industrial control systems. (Gold, 2009)

6.3.2 System Alarms

Automation systems signal security risks with the use of alarms. If the process fails to maintain the control input values, set in the control room level and controlled by the control logic in the cross-connection level, and measured parameters exceed security levels, the system will sound alarms that are shown in the control room. In the case of a plant disturbance, usually several process variables, such as pressure, level, flow, temperature, are affected simultaneously and the operator is confronted with several alarm messages within a short time interval, and might not be able to act correctly. Single disturbance in the process can result into large number of alarms, which creates an overload of information that is presented to the user in the control room and hide relevant data because there is too much redundant information. For this purpose, there are alarm management concepts that aim to reduce the number of alarms in situation like this. This helps the operator to detect the important information and react faster. (Christiansen, et al., 2013)

Alarms can be blogged by malicious software in the system. Malware such as Stuxnet give falsified data to the control room and therefore the system appeared to work normally without any disturbances. Stuxnet also gave false information on the commands it was running so they would not be identified as malicious commands. This means the system could be giving information that would indicate that it is not functioning as it is supposed to, but the information might never be seen in the control room. (Rosenberg, 2017)

6.4 Integration of Cybersecurity into Automation Systems Lifecycle

Cybersecurity management has to be included into the lifecycle of the automation system. This is necessary as the cybersecurity threat landscape is constantly changing and developing countermeasures for threats is a continuous process. Mitigating the increasing number of threats cannot always be done by adding new technologies to existing ones. Having contingency plans for continuous cybersecurity management helps in keeping this as a part of the lifecycle management of the overall automation system. This also helps in establishing rules and distribution of responsibilities among employees. (Ahonen, 2017)

There are different classifications for the stages of automation system lifecycles. The Finnish Society of Automation (SAS, fin. Suomen Automaatioseura ry) divides the lifecycle into seven stages:

1. definition,
2. design and planning,
3. implementation,
4. installation,
5. testing,
6. verification and
7. production.

These stages are executed in sequence and each stage builds heavily on the results of the previous stage. This model also describes the information, support and resources that are required or produced during the process. These in turn help the in developing cybersecurity. (Ahonen, 2017)

With cybersecurity management, it is important to iterate this process to build resistance for changing cybersecurity threats. This involves designing plans for the long-term development of cybersecurity solutions that describe how solutions are deployed with the help of the lifecycle model when new threats are detected. Combining cybersecurity and physical security into one larger category simply referred to as *security* will also help in involving cybersecurity in the automation system lifecycle. Cybersecurity should not looked as a separate service from regular security as they are not completely separable from each other. (Ahonen, 2017; Kääriäinen, 2019)

6.5 Existing Event Monitoring Solutions for Industrial Control Systems

Systems very similar to SIEM are being used to monitor industrial control systems already. This chapter describes three existing solutions for monitoring cybersecurity in industrial control systems.

6.5.1 Darktrace Enterprise Immune System

Darktrace Enterprise Immune System is a product developed by the company Darktrace. It offers cybersecurity protection for industrial ICSs, cloud environments and IoT environments from both IT and OT security standpoints. The product is made for providing full cybersecurity coverage for the entire organisation. (Darktrace, 2019)

The Enterprise Immune System uses Artificial Intelligence (AI) and machine learning to create patterns of normal behaviour in a system. Then it detects any activity that differs from this baseline for normal activity instead of trying to detect well-known patterns of malwares or attacks. This is the reason Darktrace gave the system the name Enterprise Immune System. It acts like an immune system, but in a technical enterprise environment. (Darktrace, 2019; Hall, 2017)

Its technology is based on Bayesian probabilistic mathematics, which focuses on accumulating knowledge rather than historical frequency, to estimate risk. Using an appliance attached to the network, it analyses the behaviour of every user, device and network element to build patterns of normal behaviour. (Hall, 2017)

The Enterprise Immune System is designed to be a passive observer in a network. Connecting devices to corporate networks is relatively free of risks. However, this is not true for industrial networks, where in some applications even minor interruptions in services can be damaging. The Enterprise Immune System runs on a server that is connected passively to the ICS network. It does not interfere with the operation of the control system. It flags anomalies for investigating but it does not influence the situation. (Darktrace, 2019)

Darktrace's Enterprise Immune System does not offer automated countermeasures to anomalies. For this purpose, the company has developed a separate autonomous response solution Antigena. It has capabilities to:

1. Stop or slow down activity related to a specific threat.
2. Quarantine people, systems or devices.
3. Mark specific pieces of content, such as emails, for further tracking and investigation.

Automated responses are designed to make countermeasures faster. With this the impact of anomalies are minimised. (Darktrace, 2019)

6.5.2 Siemens SINEC Network Management System

Siemens SINEC Network Management System (NMS) is software designed for monitoring and administrating networks and their devices. It is designed to incorporate FCAPS model made by the International Organisation for Standardisation (ISO). FCAPS acronym is comprised of the management categories *fault, configuration, accounting, performance* and *security* and it is a model that defines what network management systems are required to do. (Siemens, 2019)

Siemens describes in their technical article (Siemens, 2019) that here are different concepts for tackling the challenges of digitalisation at the network level. They include proprietary solutions, industry-specific solutions and special custom-made building blocks for systems of varying complexity. However, these concepts are united by the five management categories of FCAPS. SINEC complies with the different categories with:

1. Fault management: Quick and easy fault localisation.
2. Configuration management: Centralised configuration that saves resources.
3. Accounting management: Security by testing the network and documenting events reliably.
4. Performance management: flexibility through network optimisation, transparency through the generation of statistics, and high availability through the continuous monitoring of the network.
5. Security management: management of procedural and technical security requirements according to standard IEC 62443.

Siemens' SINEC aims to be a network management system for industrial networks of various sizes. The digitalisation aspect of Industry 4.0 has brought the need for network management systems that are scalable for different industries. SINEC achieves this by dividing its structure into two levels: control level and operation level. (Siemens, 2019)

SINEC consists of one control component that is located at the control level, and one or more operation components that are located at the operation level. The control component is used to monitor and administrate the entire network and operation components do the same for subnetworks or even singular devices. These operation components are configured through the control component. Operation components collect data from the subnetwork and parse out relevant information that is then sent to the control component. They can also choose which messages are ignored entirely. Operation components can be installed on the same computer as the control component or completely separate computers. (Siemens, 2018)

SINEC NMS supports 25 separate Operation components. Each of these components can be connected to 500 devices. However, the maximum number of devices the system supports is only 500. (Siemens, 2018)

6.5.3 Indegy Industrial Cybersecurity Suite and CIRRUS

Indegy Industrial Cybersecurity Suite is designed for protecting industrial networks from cybersecurity threats that are either malicious or non-malicious. Just like Darktrace's Enterprise Immune System, it aims to combine the cybersecurity management and handling of both IT and OT environments. (Indegy, 2019a; Indegy, 2019b)

The suite handles three aspects: Threat detection and mitigation, asset tracking as well as risk management. Threats are detected by identifying anomalies in the network. It has its own pre-made detection policies and it allows custom-made policies to be created for better compatibility with the target system. In addition to these policies, it establishes a baseline for normal activity in the system and compares current network traffic to this baseline. Siemens refers to using both policy-based detection and anomaly-based detections simultaneously as Dual Threat Detection. (Indegy, 2019a)

With asset tracking, Indegy tracks automatically controllers and devices on the network and visualises their topography to offer a better situational awareness of the system. Risk management, on the other hand, is handled by maintaining an up-to-date database of vulnerabilities that might pose a threat to the assets. Indegy does not wait for information to be delivered from the devices. It proactively queries the state of the devices including firmware versions, open ports, hardware configurations and installed patches. This is a part of maintaining situational awareness of the system. (Indegy, 2019a)

Indegy also offers **Industrial Cybersecurity as a Service** (ICSaaS). They have a product called CIRRUS that operates in Indegy's cloud. CIRRUS is a fully virtualized software that has no footprint on the system being monitored, as no hardware or software would be placed in the system's network. It uses VPN to query information from devices in the network. (Indegy, 2019a)

7. PARSER DEVELOPMENT

One of the central steps in analysing log data with SIEM systems is the parsing of the log data. SIEM systems are designed to take input data from various sources but it cannot function if the data is not normalised. Normalisation is the process of separating values from the log data and storing them in the appropriate data models in the SIEM system. For example, parsing out the username in a login event and storing it in the username data field in SIEM so that SIEM can correctly interpret this piece of log data. Alarms and correlation rules in SIEM rely on this stage as they use the systems own data fields as a basis of their detection and analysis.

Different SIEM solutions offer different means for implementing the parsing logic for log inputs. However, this chapter illustrates how the parsing logic is programmed manually. The parsers is designed for a SIEM product Sentinel and the implementation is done with JavaScript.

7.1 Design Phase

Parser development starts with planning. Depending on the customer and the application, developers of the parsers might have a full specification of the system with detailed descriptions of the log data format or they might only have example logs as their basis for planning and implementing the parser. Planning involves studying the log data and identifying the important pieces of information that are integral to forming situational awareness. The key question that should be kept in mind is what the purpose of analysing this log data is. This helps in identifying the important pieces of information that should be collected and included in the analysis.

The amount of information available to the developers is integral. Log data in itself can have a lot of data and not all of it is useful for the purpose of analysis. However, the log data can have information that is integral to analysis but it is not evident from simply viewing the data.

Below is an example log event:

```
100001; 2019-01-01 09:00:00.000; 130.230.252.62; 65432;
130.230.137.61; 65432; Login; exampleUser; Login by user exam-
pleUser.
```

Important data in this example would include event ID (100001), timestamp (2019-01-01 09:00:00.000), target system IP address (130.230.137.62), source IP address

(130.230.137.61), user that initiates the event (exampleUser) and type of the event (login). The fields seem simple and easy to understand, however, they do have some ambiguity. The Event ID could be a specific value given to the event by the device or system that produced the log or it could an archival ID that is used in the storage of the log event. Furthermore, it could be a primary key in a database or it could be an ID that indicates the type of the event.

The timestamp could be in local time or in UTC time. In addition, it is not evident from the timestamp whether the date is presented in the format: year, month and date, or in the format: year, day and month. In addition, it is important to note that the timestamp uses a dot as a decimal rather than a comma. This is important when converting the date string to a date object in the SIEM system.

It is also important to know what values a particular data field can have. For example, the message type field in the example log event is *login*. Knowing what values this particular field can have is important in defining taxonomies. Taxonomies are the definitions of the log event types. These types are used the analysis as SIEM can use taxonomies to categorise events into groups.

Usually the more information is being parsed from the log data, the more useful SIEM will be, as it will reduce the need for manual analysis of the event. The further the analysis and handling of events can be automated the better it will be at producing useful information and developing a situational understanding of the system in question.

During the planning phase, events are also categorised into different taxonomies. Taxonomy is a categorisation of the event. Taxonomies may include login, logout, password change, malware, query and system start and shutdown. SIEM will than use these taxonomies to identify patterns in the events with the use of correlation rules. The example event is categorised as a login event, which can be deduced from the event type *login* and the plaintext description of the event at the end of the log line.

7.2 Development and Testing Phases

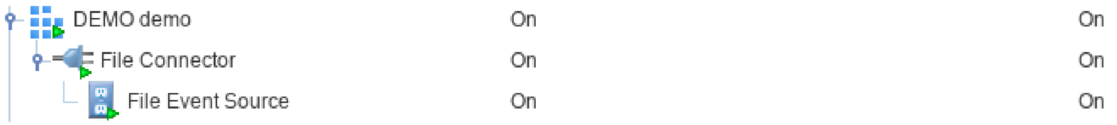
Planning phase is followed by development phase. In this phase, the parsing logic is implemented in accordance to the plan. In MicroFocus' SIEM product Sentinel, every parser is implemented by writing the code for the parsing logic with JavaScript. In appendix A, is the code for parsing the information in the example event log and stored in *RXMap* data structure. Each of the variables stored in the RXMap are then transferred to Sentinel's own fields.

After the parsing logic is implemented, either by writing the code for it or building it from already existing parsing widgets depending on the application, the parser is tested. The purpose of the testing phase, as in any development project, is to ensure the implementation is working as planned. Well-executed tests require extensive log data to ensure the parser works with all different types of events and final testing is to be done in a similar environment as the customer's environment.

Tests can also include automated tests. Unit tests can be used for testing smaller units in the parser and some SIEM products have their own automated test frameworks for testing that events are normalised correctly.

7.3 Deployment

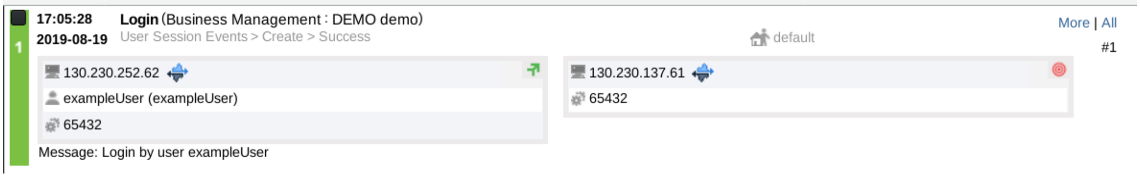
When the parser is deemed to be working correctly, it is ready to be deployed. Deploying a parser in Sentinel requires defining a collector, connector and the event source. Figure 10 shows these elements and their hierarchy for the example parser.



DEMO demo	On	On
File Connector	On	On
File Event Source	On	On

Figure 10. *Functioning parser running in Sentinel*

The event source represents the input log data and it has to be configured according to the type of the data source, which in this case is a file event source as the log data is read from a csv file. The event source is connected to a connector that, in turn, receives the data from the event source and transforms it into contextual map form for the connector. The connector will then read map, parse the data according to the parsing logic and store it in Sentinel's data fields. The connector in figure 10 is named *Demo demo*.



17:05:28 2019-08-19 **Login** (Business Management : DEMO demo)
User Session Events > Create > Success

default

More | All
#1

130.230.252.62
exampleUser (exampleUser)
65432

130.230.137.61
65432

Message: Login by user exampleUser

Figure 11. *Login event presented in Sentinel's user interface after parsing.*

Now that the event data is parsed, it is then displayed to the user in the user interface of the SIEM software. Figure 2 shows Sentinel's user interface and in figure 11 is a single login event. The event in figure 11 is a login event by user *exampleUser* from the IP address *130.230.137.61* to IP address *130.230.137.62*. In the upper left corner of the event, is the timestamp of the event and next to it, is the event's name *Login*. Below the event name is the taxonomy definition and outcome of the event. As the event describes a successful login, the taxonomy definition is *User session Event>Create* and the outcome is *Success*. At the bottom of the event is the message field that contains the written out description of the event *Login by user 'exampleUser'* that was parsed directly from the log data.

8. HOW SIEM SHOULD OPERATE IN AN AUTOMATION SYSTEM

Probably the most important factor to note is that using log data in a meaningful way by implementing log data handling would be a transformational project for an automation system. Like all projects, this would require setting goals, making risk assessments, planning resource usage and scheduling the project and its various steps. (Ahonen, et al., 2019)

As SIEM is an additional system that has to be planned, implemented, deployed, maintained and monitored, it will add costs and take additional resources throughout its lifecycle. It will add workload to various people in different stages and it will also add complexity to systems that are usually already complex by themselves.

8.1 Selling SIEM for Automation Systems

Automation engineers are interested in the process and the reliability of the automation system. Information security is an additional service to an automation system, as it does not directly affect the control process of the system. This aspect makes selling anything related to information security difficult. Information security incidents happen rarely and nobody wants to invest in a system that is needed in very special and seldom occasions. In order to sell SIEM to automation systems, its benefits will have to be justified to being beneficial to the automation process.

Automation engineers are interested in the reliability of the system. As described in chapter 4.5.3 when covering automation software requirements, reliability can be defined as *the probability that a system will perform a specified function within prescribed limits, under certain environmental conditions, for a specified time*. On a more general level, reliability could be described as the quality of being trustworthy and performing well. This is something cybersecurity measures can have an effect on, for the reason that with improvements in cybersecurity a system can detect issues faster, react with countermeasures better suited to the issue. Recovery from incidents can also be made more streamline and quicker. This reduces system downtime, additional damages and production losses.

One issue any counter measures for cybersecurity threats will not fix is the rarity of the cybersecurity threat related events in automation systems. Even though reliability could be improved with SIEM, it would only serve its purpose in exception situations, which in turn lowers the willingness to invest in such a system. Preparedness for cybersecurity threats is largely proactive measure and relies on the prospect that systems will be protected when the threat becomes a reality. When the system is working properly, cybersecurity might not produce any real and tangible value. This accompanied with the fact that SIEM will add workload to employees, and therefore consume resources during its entire lifecycle from planning to maintenance and inevitably decommissioning, will lower the desirability of incorporating SIEM systems into automation systems.

When developing and selling SIEM, suppliers of the system might focus on better integration to the automation system. Planning and deployment phases are expensive and take a long time and adding SIEM to a pre-existing automation system will always be an alteration process. If adding SIEM will require system downtime, it will add costs and it probably would not be performed until the next scheduled maintenance shutdown. Shutdowns are not performed often.

Furthermore, in the beginning stages when all different log sources, with their different log formats, would be joint to SIEM will also take time. If the deployment process would be streamlined, it would help in selling SIEM. SIEM suppliers will have to focus more on the deployment phase in order to improve their appeal.

Automation systems are on the most part all prototypes. Every system is built for a specific purpose in order to handle a specific function. Having SIEM operate in these types of environments will require them be highly customised. This makes their development as a general purpose or so called “off the rack” systems challenging. It will also make SIEM difficult to design to be more integrable, as all target systems are different.

8.2 Effect on Performance

Gathering log data might require a new network. Communication networks in automation systems are designed for the communication between devices in order to control the process. Messages that go through the network are measurement data, alarms, control inputs or other configuration data integral to the process. As hardware is often minimised in automation systems, the network might not have the capability to transfer log data from all the devices without affecting the performance of the system.

With SIEM, all log sources require parsers to parse log data and plugins to connect the log source to SIEM. In a system with a large number of devices, this would mean that the parsers and plugins might take a considerable amount of storage. In addition, SIEM software - with any additional libraries or software - would have to be installed to the computers or designated servers running SIEM, which may not be equipped to handle additional resource usage as again hardware is minimised in automation systems.

8.3 Log Data Contents

Automation system logs contain a lot of data. When implementing solutions for handling this log data, it is important to figure out how to take advantage of the data as much as possible in order to reduce additional costs. This step will also help to understand and map out what data is missing, and what exactly needs to be done in order to repair the deficiencies in the data or the infrastructure. (Ahonen, et al., 2019)

The process of implementing log data handling in an automation system would start by finding out what kind of data is already collected. This involves going through the collected data and analysing how it could be used to control error situations that include both failures in communication infrastructure and information security events. Collecting and storing the data is not enough on its own as using log data in a meaningful way requires that you carefully research on how the contents of the data could be employed. Information about the collection of the log data and its contents can be requested from the suppliers of the devices. (Ahonen, et al., 2019)

After analysing the existing log data, you have to find out what objectives regarding error control are missed. In automation device logs, there can be many information security events that are missing, such as unauthorised or unnecessary sign-in events, software upgrades, configuration modifications or changes in process control inputs. (Ahonen, et al., 2019) When implementing SIEM, it is important to figure out what kind of events is the system trying to detect. SIEM uses correlation rules to do automatic analysis of events making it vital that the types of the events are mapped out carefully as otherwise the correlation rules will not work.

At this stage, you can go through the data and find out whether the contents of the data is enough. Is there enough data to detect the necessary events as they are happening and what even constitutes as an event. SIEM is used to analyse events, but a single log entry in automation system might not represent a single event. Events might have to be constructed from multiple different log entries that may even come from different devices. For instance, when powering on an automation system or part of the system, is opening a single valve or turning on an actuator events on their own or is the start-up of the system the event. Making all operations, which devices make, events in SIEM means that there is a vast amount of input data, but it might be necessary to help produce a meaningful image of the operation of the system and to help find root causes to problems.

Log data coming from automation devices would most likely have to be enriched. For instance, as traditionally log data would stay at the devices, it does not include information concerning the location of the device. In this case, it would be necessary to add the device IP-address and, through the IP-address, positional information from an external database telling where the device is located in the system. This would reduce the need of manual analysis and make the whole process more efficient as the handling of the event would be automated as far as possible.

Challenge that automation systems bring to the collection of log data is that a single automation system contains devices from a multitude of suppliers and device manufacturers who have their own log formats and means of extracting the data from devices. This creates the issue that collecting log data in an environment like this will not be easy. SIEM would require some form of centralised log handling as all the data must be brought to the server running SIEM in order for it to work efficiently in creating situational awareness. If there is no centralised and automated log management, the data would have to be collected using queries. This means the data would be fetched from the devices when they are needed. Centralised log management could be expensive as the number log sources grows, and using queries could generally be more cost effective, as queries can be done in a distributed manner or locally if the necessary tools exist (Ahonen, et al., 2019).

In addition to different methods of extracting logs from devices, the problem with having multiple device suppliers and manufacturers is that there are many log formats. SIEM requires that the information content of the logs are parsed and stored in the appropriate fields in the system for SIEM to be able to analyse it. SIEM will not be able to do this on its own and for this purpose, SIEM uses parsers to extract and store the required information. Parsers would almost always have to be custom made for every single log source, or log connection as they are sometimes called, as the parsing logic will be different for all log formats.

Being that there are so many different devices producing log data, there would have to equally as many parsers made for SIEM and every new device would require its own new parser. Parser production is an added expense, and it would require new resources, skills and time. There are companies that offer parser development as a service. The service might be offered in conjunction with security operations centre (SOC) service that includes monitoring and analysis of security events, alerting and support for recovery from security incidents. Outsourcing log monitoring has the added benefit that you can invest in around-the-clock monitoring without adding workload to existing employees. However, this might not be as useful if service actions are performed only during normal hours (Ahonen, et al., 2019). Errors might be detected, but there is still nobody to fix the issue.

Parsers for SIEM can be implemented on various ways. With some older SIEM products, parsers have to be programmed manually by writing the code for every operation, but SIEM systems that are more modern, use higher-level methods for making parsers. This involves assembling parsers from pre-made parsing widgets instead of writing the code for them separately. Regardless of the parser development method, every parser requires a plugin that is used to connect it to both SIEM and the log data. These plugins are often implemented with Java, which has to be installed and configured on the SIEM server and possibly the monitoring computers. Again when the number log sources grows, the resources the plugins require grows as well requiring processing power, data storage and memory as well as maintenance.

8.4 Transferring Log Data

Collecting log data is only the first step in utilising log data. SIEM requires logs to be brought to it for it to work, however, on estimate 80 % of the work and costs put into log handling will go into the effort for getting all the log data to be analysed (Ahonen, et al., 2019). Instead of collecting all the log data from devices to SIEM, SIEM could be taken to the devices. As it is customary currently, only whenever something goes wrong the logs would be inspected. This could, however, be done by taking a computer with SIEM to the device and it would read, parse and analyse the data. This would have to be done in a way that does not require a reboot of the device.

Taking SIEM to the device would, nevertheless, mean that all counter actions to events would still be reactionary rather than proactive, which is a problem for fast incident response. The point of SIEM is to be constantly analysing and monitoring logs. If the system administrators or operators have to wait for something to go wrong and then read the logs, it is negating the purpose of SIEM, as first you have to detect the failures on your own. This might also conflict with device suppliers’ interests, as it is usually their employee that comes to inspect the logs.

Having no centralised log collecting will also have the disadvantage of having small data sets for SIEM to analyse. SIEM can use multiple log sources to build a situational image of the state of the system, but this is difficult if only the logs from one device are analysed at once. This would not help in finding the root causes to incidents either.

If log data is brought to a centralised location for SIEM to analyse, SIEM could be used to gather both log and measurement data. SIEM systems are advertised primarily as cybersecurity solutions but they could be used to monitor and analyse varying types of data. For example, SIEM can analyse measurements from a particular part of the process and give alarms if the values exceed accepted limits or if some measurements are displaying abnormal amount of fluctuation. These can be programmed into a SIEM system's correlation rules.

8.5 Analysing and Understanding Log Data

One growing trend in SIEM and log handling is the use of cloud platforms. Cloud platforms could be used to store and analyse data and they offer high customisation with relatively manageable costs. Implementing cloud platforms in automation has not been without its problems as they have issues with the real-time requirements of automation systems. Handling log data for event analysis would not be subject to as strict real-time requirements. There are products, such as ServiceNow, that utilise cloud platforms for log analysis and incident response. They can create automated countermeasures and further analysis to events, which is something SIEM systems do not normally offer. On the other hand, transferring large quantities of log data that an automation system produces, via the Internet or Ethernet could be an issue. Furthermore, storage needs will grow and that can become a large expense in cloud platforms.

After storing the data and making it available for analysis tools, comes the difficult phase of analysing the data with proper understanding of the system that is being monitored. SIEM can use general or system specified correlation rules to detect the type of activity happening in the system based on alarms or other log data. However, as automation systems are complicated systems with an abundance of devices that can produce alarms, building the proper understanding and making of the activities and making correlation rules is difficult. This can be aided with machine learning and implementing smart correlation rules.

Historical data from known error situations in the automation system can be used to teach SIEM systems with machine learning capabilities. These SIEM systems will analyse the test data and build their own correlation rules based on the activity found from the test data. Machine learning can also detect patterns in system activity that human operators could not find due to the complexity of the system and the event sequence.

Machine learning in SIEM can help in reducing the need for human monitoring of SIEM and help in investigating security related alerts. An issue that reoccurs in enterprise SIEM solutions is false positive alerts. The false positive alerts can hide legitimate alerts as they might be bundled in with wrong alerts in an effort to detect correlations between events that in reality are not connected. (Canner, 2019).

Machine learning can also reduce the need of cybersecurity specialists in the monitoring stage. Cybersecurity understanding can be built into the system through machine learning just as well as correlation rules. This results in not needing as many specialised engineers for monitoring SIEM while reassuring its effective operation at all times. (Canner, 2019)

On the other hand, implementing machine learning requires a lot of data that is used in the learning process. Data like this might not exist for automation systems and integrating smart SIEM systems with machine learning into new automation systems would be difficult. Using data from other systems is problematic because of the unique nature of each automation system.

Furthermore, implementing machine learning in itself does not guarantee good security. SIEM systems with machine learning have to be configured and taught properly. Even though these systems are sometimes referred to as intelligent SIEM systems, they can just as well work improperly or produce false results. As all cybersecurity solutions, they require monitoring and maintenance. They cannot be just set to work and then forgotten.

8.6 Detecting Anomalies Instead of Attacks

One big obstacle in finding malicious objects or identifying attacks is that they are constantly changing. In order to find them, the anomaly patterns they are searched with have to be developed with a continuous rate. However, the patterns of normal activity in automation systems does not change as fast. If there are changes to the systems normal activities, they are already known beforehand. Making modifications to the detection system based on the changes is therefore easier. Any deviation from normal activity that is not known beforehand is a possible threat situation.

Moreover, cybersecurity threats are not limited to malicious activity. They can be operator errors, configuration errors or malfunctioning devices that result in compromises in cybersecurity. Monitoring systems that detect anomalies instead of known patterns of malicious activity can also identify non-malicious threats. This is what Darktrace’s Enterprise Immune Systems aims to do.

Darktrace's Enterprise Immune Systems also uses AI and machine learning to make the baseline for normal activity. This reduces the need for tailoring correlation rules to fit a specific system. The anomaly detection system uses data of the monitored system to learn by itself what normal behaviour is. Anomaly detection based on machine learning makes the anomaly detection system more applicable to both IT and OT environments, as it is not highly dependent on the target system. This in turn helps in managing the entire organisations cybersecurity on all industrial and enterprise levels.

8.7 Monitoring Events

SIEM's purpose is to take log data in, analyse it and present meaningful information to the user or operator concerning the event and the state of the operational system. For this to work there has to be someone monitoring SIEM at all times. Part of SIEM's efficiency is that error situations are noticed immediately and the appropriate countermeasures are taken to mitigate the effects of the error. Without outsourcing monitoring, this step will add workload for existing staff in considerable way and it will act as a strain on resources. Furthermore, as process automation systems run continuously, the need for monitoring would also be continuous.

Currently in automation systems, log data stays at the devices. The logs are not monitored per say and once an error situation occurs the logs are read from the device. Furthermore, this is usually done by an employee of the supplier. This is means that the response would almost always be reactionary rather than proactive. From information security standpoint, this is not a very good approach, as the damage will most likely be done once the logs are read.

Properly configured SIEM might detect problems before they happen while displaying the user only the integral parts of the event data. SIEM can use the whole log data as part of its analysis but only display parts of it to the user hiding everything that is unnecessary, sensitive or irrelevant altogether. This will reduce the need of manual analysis, as only integral parts of the data that are needed for any follow-up actions are shown.

People responsible for the monitoring would require knowledge of both industrial automation systems and cybersecurity. Cybersecurity specialists have expertise on analysing and handling cybersecurity events but, if they lack the knowledge of automation systems, forming an understanding of the state and situational awareness of the system is difficult. Automation engineers have the understanding of the automation system but might lack the expertise to analyse log data from a cybersecurity perspective. Furthermore, the design phase of the SIEM implementation process requires extensive knowledge of automation systems in order to make it effective. For instance, correlation rules for automation systems SIEM would have to be customised for the system being monitored. Automation systems in general require their own correlation rules as the systems are themselves unique.

8.7.1 Security Operations Centre (SOC)

Security Operation Centres (SOC) are centralised operational centres that monitor event data and provide incident response and incident recovery. Their service is to provide accurate situational awareness. SOC is usually operated by cybersecurity specialists, engineers and analysists who work with the incident response teams of the monitored organisation or system. (Brook, 2018)

SOC is used to monitor critical infrastructure. In these cases, they do not only analyse data from their clients, but they also analyse data from other similar fields and systems from different countries in order to create a better understanding of threats their clients might face. SOCs are most likely situated off-site and its operators can do operations remotely, which requires extensive knowledge of the client’s system.

In automation systems, this can be difficult as almost all systems are unique, possible remote operations are not extensive and SOC operators might lack knowledge of automation engineering. In addition, no operations can be done without the permission of the administrators of the client’s system. This means that having extensive around-the-clock monitoring can be cost-prohibitive in smaller less critical systems if there is nobody on-site to assist in the countermeasures.

8.7.2 Follow-up Actions to Error Situations

As stated in chapter 3.5 event data handling can be divided into three parts: automated processing of data, manual analysis of the parsed data and follow-up actions. SIEM traditionally is responsible for the first part but human interaction has been required in the two other parts. The further the process of handling events goes automatically the better, because there is less manual analysis to be done. While some are of the mind that the human operator should be removed from the chain altogether, others might be hesitant for legal and ethical reasons. With a human in the chain, countermeasures to error situations would be initiated by that person which puts emphasis on the importance of monitoring. Getting near real-time event data to the control room will not be meaningful if the response to any errors is too slow, and the more manual work the user has to perform the slower the response will be. This in turn puts emphasis on the importance of the automated data analysis and SIEM.

Algorithms are good at detecting complicated and quiet correlations in events. Smart algorithms, which might be marketed as artificial intelligence, have their use in analysing log data. Algorithms usually require large data sets to provide material from which they learn patterns but there are also algorithms being developed that can work with smaller data sets. The downside to algorithms is that they are not good at assessing risks caused by incidents to the process or the organisation. Therefore, human operators still often perform an important function in the process. (Ahonen, et al., 2019)

SIEM are good at taking data in from various sources, however, the data usually stays at SIEM. If you want to automate the process of handling events further, event data has to be exported from SIEM. If data can be exported from SIEM in compatible formats, it could be integrated to the rest of the control room software to make automatic alarms and safety measures. If it cannot be exported, it will stay completely separately at SIEM. Companies such as ServiceNow offer products that help in the automated response of security events. However, they are often aimed at business' IT management and not industrial technology management. Integration with existing control room technology would still be an issue, but they could create platforms for faster incident response.

If SIEM is deployed in a system, it should be configured to provide a meaningful function in the system it is monitoring. The danger of taking a new log handling system in use is that it is never utilised properly and it is only used to appeal to legal mandates or provide surface level assurance that the process is further secured with automated log handling. If SIEM is not working properly, it will only act as an added strain on the system and its operators. It might also hide problems rather than reveal them as the operators trust in that it would do a lot of the detection and analysis of events when in reality it might not.

8.8 Maintaining SIEM

Log handling and SIEM both require maintenance. It is not enough to implement and deploy SIEM, as it needs to be maintained in order to insure it works properly over time and especially after any modifications to the system. Maintaining SIEM should be connected with regular service operations so that its function could be monitored. For example updating devices or their software can change the format of their logs. These format changes, and especially new devices, require new parsers and plugins. New parsers require people to do the planning, implementation and testing. This might take a large amount of time and resources and it is important to weigh the costs and benefits to decide whether it even is beneficial. Does the system provide enough value to the whole automation system to justify its costs? Enriching log data should be done close to the devices that send the data to minimise the costs that would be caused by changes in the structure of the automation system. (Ahonen, et al., 2019)

SIEM will add workload in multiple areas. SIEM and log handling in general would have to be taken into consideration when making any changes to the system. When a device is added, you also have to plan and implement the means of retrieving the logs. The skills to do this might not exist in-house nor on the supplier’s side, which would mean that another party would have to be employed to implement the retrieval of the logs.

SIEM would also be a new system for operators to learn. SIEM requires users to monitor and use it and it could be done by designated SIEM specialist or it could be an added functionality for operators in the control room to monitor. In any case, employees would require training to use SIEM so it would be utilised properly.

8.9 Team Effort from SIEM Suppliers and Customers

Better implementation of SIEM would require team effort from the suppliers of SIEM, device suppliers and customers who are planning to use SIEM. Device suppliers could put effort on creating easily interpretable logs and provide information on the contents of the data. This would make their integration with SIEM easier and cheaper to the customer. It would not make it cheaper to the suppliers and manufacturers of the devices but added business opportunities and partnerships with SIEM suppliers would improve their position on the market.

From the viewpoint of automation systems, SIEM developers should aim to make products more open and lighter. Currently SIEM systems are not very open and require a considerable amount of resources with their dependencies. Many of them use Java, which is relatively heavy software to be added to multiple devices in an automation system. SIEM would have to be something that does not put a strain on the existing resources for it to appeal to automation engineers.

9. CONCLUSIONS

Security Information and Event Management systems, or SIEM systems for short, are log management tools that provide automated log analysis and event detection. They parse large amounts of log data and form events that they use as the basis for building situational awareness of the monitored system. As humans are not good at interpreting large quantities of data, SIEM systems are used to reduce the amount of information displayed to system monitors while still using as much of the available data as possible for their analysis. When a SIEM system detects anomalies in the data, either from a single event, such as port scans, or from a collection of different events by applying correlation rules, it alerts the monitor of the activity. Good automated log analysis requires little manual analysis making the first phase of event detection more precise and faster and any countermeasures will be able to be performed quicker minimising the consequences of the anomaly.

9.1 Compatibility of SIEM and Automation Systems

Automation systems are complicated systems with long lifecycles. They are comprised of a multitude of devices many of which produce logs resulting in large quantities of log data. Automation systems control manufacturing processes and vital infrastructure making their reliability vital. Furthermore, increasing number of communication in automation systems is TCP/IP based which has brought a new set of cybersecurity threats to the field of automation. There was a time when automation engineers could rely on security through obscurity, however, the technology used in automation systems are much more well-known and tools for creating damage are more readily available meaning the importance of protecting automation systems from cybersecurity threats is necessary. Event detection is one way to improve preparedness.

To an automation engineer the performance and reliability of the system are the most important aspects, and anything that does not improve these factors is difficult to sell to an automation engineer. SIEM systems are seen as additional services and cybersecurity threat related events are rare. Engineers do not want to invest in a product that becomes useful in rare exception situations and on other times acts as a strain on resources.

However, as information security in automation systems has to focus more on IT security and not only on OT security, new ways of handling the security of IT have to be implemented in automation systems. SIEM is primarily a service and a tool for IT security and it can be used as a protective tool for the growing landscape of digital threats.

One fundamental issue with applying SIEM to an automation system is that in automation systems centralised log management is not a common practice. SIEM requires log data to be readily available, but automation systems are not traditionally built for this as logs stay at the devices. They are not transferred anywhere. In order to implement SIEM, first step would be to implement a way to bring the log data from the devices to SIEM. This would be a large and expensive transformational process.

Moreover, all devices produce logs in different formats. Hence, each device would require its own custom-made parser to accommodate the log joint. The log data would most likely not suffice as is, and it would have to be enriched with for instance location information in order for it to provide meaningful information that could be used to build situational image of the entire system.

Having SIEM will also add workload and consume resources during its entire lifecycle. It requires planning, deploying and maintenance. SIEM would have to be taken into consideration during every maintenance operation to the automation system because changes in the automation system will cause changes in SIEM and log management in general. Furthermore, SIEM will have to be monitored at all times for it to a meaningful tool in event management.

Furthermore, for SIEM to be a viable way to do event management, it would require team-effort from SIEM suppliers, device suppliers and the end-customers. Devices should produce logs with more uniform formats and more quality informational content, automation systems would need to implement centralised log management in larger scale and SIEM producers would need to create lightweight, open pieces of software with good integrational capabilities with automation systems. As they are now, automation systems and SIEM systems are not very compatible.

9.2 Future of SIEM

In the future, applications for systems like SIEM will most likely increase. Future application, for instance, might be access control, network monitoring on a large scale and performance analysis. All for the effort for automating data analysis and improving situational awareness. (Vesamäki, 2016) In addition, SIEM systems could be incorporated with modern data collection devices by collecting biometric data in the field of healthcare or integrating SIEM to much larger systems outside the purely technical industries. On the other hand, one issue SIEM is facing is performance under very large input of data.

Scaling up might be a limiting factor with certain SIEM systems. A single SIEM system can handle only a certain amount of events per second. When it receives too many events in one small succession, it will have to store some events to a buffer to wait until it has successfully handled the previous events. This is a problem for the systems situational awareness, as events are not handled in real time.

SIEM might expand from monitoring systems to monitoring the security awareness of entire organisations. However, using employee filled reports about security anomalies and implementing security teams' guidelines might not be straightforward but not impossible either. This would be achievable with SIEM technology as it exists today, but implementing it would be laborious. (Vesamäki, 2016)

One very futuristic idea of SIEM is that it could be self-learning and self-acting. Vesamäki writes about this in his article for Viestintävirasto (2016). He considers that the technology for that already exists in some form. However, he also states that this is not likely happen any time soon, as that would remove the human actor from decision process between event and reaction.

Topics for future studies could be studying the applications of SIEM in monitoring organisations at large. Instead of using SIEM to monitor singular systems, it might be used for overseeing the whole of the organisation. The challenges of this might include the wide variety of events to be monitored and privacy concerns raised by having an all-encompassing monitoring system.

9.3 Event Management System Alternatives for Automation

Anomaly detection systems are an alternative for using SIEM in industrial control systems. These anomaly detection systems do not try to find specific behaviour of known malwares or attacks, but they form models of normal activity of the monitored system. This allows anomaly detection solutions to detect both malicious and non-malicious activity regardless of whether their patterns of activity are known or unknown. Furthermore, anomaly detection accompanied with **machine learning** and **AI** creates the possibility for **smart detection systems** that are not highly system-specific.

Automation systems were previously mainly concerned with OT security. However, currently the divide of OT and IT security in industrial control systems is becoming less clear and the need for cybersecurity solutions that can handle both is growing. Smart anomaly detection systems, for the reason that they are not highly system-specific, are a good fit for monitoring both IT and OT security.

Industrial Cybersecurity as a Service is a service model that offers cloud-based solutions for cybersecurity monitoring and management. Their advantages are that they need minimal hardware and software on-site while offering computing power and data storage off-site. This means that these systems have a small footprint on the system that is being monitored. Disadvantages are that all data has to be transferred through VPN. The amount of data that is transferred can be vast and using VPN can cause issues in real-time requirements.

Overall, SIEM products as they are now will most likely not be incorporated to automation systems on a large scale. Both automation systems and SIEM systems would require extensive modifications for this to happen.

REFERENCES

[1] ABB, 2007. OAMK. Available at: http://www.oamk.fi/~kurki/automaatiolabrat/TTT/24_Prosessiautomaatio.pdf (Accessed 10.07.2019)

[2] Ahonen, P., 2017. KYBER-TEO - tuloksia 2014-2016, Tampere: Utgivare.

[3] Ahonen, P., Seppälä, J., Pärssinen, J. & Vahala, P., 2019. KYBER-ENE Energia-alan kyberturvaaminen 1-2, Helsinki: VTT.

[4] Armstrong, D., 2016. Reliable Plant. Available at: https://www.reliableplant.com/Read/30641/manage-plant-shutdown (Accessed 11.07.2019)

[5] Asp, R., Tuominen, T. & Hyppönen, H., 2019. Edu.fi. Available at: http://www03.edu.fi/oppimateriaalit/kunnossapito/sahkotekniikka_a2_automaatiojarjestelma.html (Accessed 10.07.2019)

[6] Bajramovic, E., Gao, Y., Parekh, M. & Xie, X., 2016. SIEM: Policy-based Monitoring of SCADA Systems. Darmstadt, IEEE, p. 559.

[7] Bergweiler, S., 2016. Smart Factory Systems - Fostering Cloud-based Manufacturing based on Self-Monitoring Cyber-Physical Systems. International Journal on Advances in Systems and Measurements, 9 (1 and 2), p. 12.

[8] Bisson, D., 2018. Tripwire. Available at: https://www.tripwire.com/state-of-security/ics-security/ics-security-challenge-organizations/ (Accessed 14.10.2019)

[9] Brook, C., 2018. Digital Guradian. Available at: https://digitalguardian.com/blog/what-security-service-definition-secaas-benefits-examples-and-more (Accessed 31.08.2019)

[10] Canner, B., 2019. Solutions Review. Available at: https://solutionsreview.com/security-information-event-management/6-questions-about-machine-learning-in-siem-answered/ (Accessed 02.10.2019)

[11] Castilla, M. et al., 2011. Hierarchical Control of Droop-Controlled AC and DC Microgrids—A General Approach Toward Standardization. IEEE Transactions on Industrial Electronics, 58(1), pp. 158-172.

[12] Christiansen, L., Fay, A., Schleburg, M. & Thornhill, N. F., 2013. A combined analysis of plant connectivity and alarm logs to reduce the number of alerts in an automation system. Journal of Process Control, 23(6), pp. 839-851.

[13] Crevatin, M., Dzung, D., Naedele, M. & Von Hoff, T., 2005. Security for industrial communication systems. Baden, IEEE.

[14] Darktrace, 2019. Infosecurity Europe. Available at: https://www.infosecurityeurope.com/__novadocuments/588383?v=636917895698800000 (Accessed 15.10.2019)

[15] Dedzins, R. & James, M., 2015. emerson.com. Available at: https://www.emerson.com/documents/automation/article-continuous-shm-assures-pfizer-s-automation-system-performance-pharmaceutical-manufacturing-nov-2015-pss-en-67986.pdf (Accessed 20.08.2019)

[16] Didier, P., Siddeswaran, A. & Wilcox, G., 2019. Industrial Ethernet Book. Available at: https://iebmedia.com/index.php?id=12300&parentid=63&themeid=255&showdetail=true (Accessed 07.10.2019)

[17] Drivers&Controls, 2010. Drivers&Controls. Available at: https://drivesncontrols.com/news/archivestory.php/aid/3030/Stuxnet_targets_Vacon_inverters_.html (Accessed 20.08.2019)

[18] EU: Cybersecurity & Digital Privacy Policy (Unit H.2), 2019. European Commission. Available at: https://ec.europa.eu/digital-single-market/en/network-and-information-security-nis-directive (Accessed 15.09.2019)

[19] Gold, S., 2009. The SCADA challenge: securing critical infrastructure. Network Security, 2009(8), pp. 18-20.

[20] Goss, P., 2010. library.e.abb.com. Available at: https://library.e.abb.com/public/ab5861c0a6e19d85c12576c6002d0dd1/02_Article_The_new_way_to_maintain_a_modern_automation_system_in_Tissue_world_Des2009_Jan2010.pdf (Accessed 15.08.2019)

[21] Gough, M. & Porter, T., 2007. Active Security in Monitoring. In: M. McGee & A. Rebello, eds. How to Cheat at VoIP Security. Rockalnd: Syngress Publishing Inc, pp. 185-206.

[22] Hall, S., 2017. The New Stack. Available at: https://thenewstack.io/darktrace-applies-math-unsupervised-machine-learning-automate-network-security/ (Accessed 15.10.2019

[23] He, H., 2003. O'reilly's webservices.xm.com. Available at: http://www.cdrlm.com/whitepapers/other/xml_what_is_service_oriented_architecture_sep2003.pdf (Accessed 15.09.2019)

[24] Hästbacka, D., 2017. Lecture material: ASE-6030 Automaation reaaliaikajärjestelmät - lecture 10, Tampere: Tampere University of Technology.

[25] Hästbacka, D. & Kuikka, S., 2017. Lecture material: ASE-6030 Automaation reaaliaikajärjestelmät lecture - 1 Introduction, Tampere: Tampere University of Technology.

[26] Indegy, 2019a. Indegy. Available at: https://www.indegy.com/industrial-cyber-security/ (Accessed 16.10.2019)

[27] Indegy, 2019b. Hubspot. Available at: https://cdn2.hubspot.net/hubfs/2755567/The%20Indegy%20Industrial%20Cybersecurity%20eBook%202019.pdf (Accessed 16.10.2019)

[28] ISA, 2010. ISA. Available at: https://www.isa.org/isa99/ (Accessed 14.10.2019)

[29] ISA, 2016. ISA. Available at: https://www.isa.org/belgium/standards-publications/ISA99/ (Accessed 14.10.2019)

[30] Kaspersky Lab, 2019. Kaspersky Lab ICS CERT. Available at: https://ics-cert.kaspersky.com/reports/2019/03/27/threat-landscape-for-industrial-automation-systems-h2-2018/#_Toc4416091 (Accessed 15.09.2019)

[31] Kääriäinen, J., 2019. Elomatic. Available at: https://www.elomatic.com/en/elomatic/expert-articles/cyber-security-enables-more-intelligent-automation-solutions.html (Accessed 07.10.2019)

[32] Lee, E., 2008. Cyber Physical Systems: Design Challenges. Orlando, FL, USA, IEEE.

[33] Leskiw, A., 2018. Network Management Software: Understanding Syslog: Servers, Messages & Security. Available at: https://www.networkmanagementsoftware.com/what-is-syslog/ (Accessed 08.07.2019)

[34] McAfee, 2019. McAfee. Available at: https://www.mcafee.com/enterprise/en-us/security-awareness/ransomware/what-is-stuxnet.html (Accessed 07.08.2019)

[35] Mraz, S., 2017. MachineDesign. Available at: https://www.machinedesign.com/industrial-automation/differences-between-field-control-supervisory-and-enterprise-levels-automation (Accessed 05.08.2019)

[36] Rosenberg, J., 2017. Security in embedded systems. In: T. Green, ed. Rugged Embedded Systems: Computing in Harsh Environments. Cambridge, MA, United States: Elsevier, pp. 149-205.

[37] Rouse, M., 2018. Tech Taget. Available at: https://searchwindowsserver.techtarget.com/definition/Windows-event-log (Accessed 15.09.2019)

[38] ServicesParis, 2010. Paris Controls. Available at: http://www.pariscontrols.co.uk/services04Val.html (Accessed 11.08.2019)

[39] Siemens, 2018. Siemens. Available at: https://cache.industry.siemens.com/dl/files/749/109762749/att_969274/v1/BA_SINEC-NMS_76.pdf (Accessed 16.10.2019)

[40] Siemens, 2019. Siemens. Available at: https://assets.new.siemens.com/siemens/assets/api/uuid:af9c3152-8940-4562-a1c5-a966545fd79c/version:1559106812/sinec-nms-fav-iee-219-0219-en-web.pdf (Accessed 16.10.2019)

[41] Stapelberg, R. F., 2009. Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design. 1st ed. London: Springer.

[42] Wagner, A., 2017. pbs.org. Available at: https://www.pbs.org/newshour/science/everything-need-know-wannacrypt-ransomware-attack (Accessed 16.08.2019)

[43] Wald, R., Khoshgoftaar, T. M. & Zuech, R., 2015. Intrusion detection and Big Heterogeneous Data: a Survey. Journal of Big Data, 2(1), p. 42.

[44] Vesamäki, P., 2016. Viestintävirasto. Available at: https://legacy.viestintavirasto.fi/kyberturvallisuus/tietoturvanyt/2016/03/ttn201603151527.html (Accessed 02.07.2019)

[45] Williams, A., 2007. Tech Buddha. Available at: https://techbuddha.wordpress.com/2007/01/01/the-future-of-siem-%E2%80%93-the-market-will-begin-to-diverge/ (Accessed 01.07.2019)

[46] Zhang, E., 2018. Digital Guardian: What is Event Correlation? Examples, Benefits, and More. Available at: https://digitalguardian.com/blog/what-event-correlation-examples-benefits-and-more (Accessed 09.07.2019)

[47] Åkerman, M., 2016. Towards interoperable information and communication, Gothenburg: Chalmers University of Technology.

APPENDIX A: PARSING LOGIC EXAMPLE IN JAVASCRIPT

```
Collector.prototype.initialize = function () {
   this.CONFIG.params.usernameIsCaseSensitive = true;
   this.CONFIG.params.datanameIsCaseSensitive = true;
   this.CONFIG.params.hostnameIsCaseSensitive = false;

   return true;
};

Collector.prototype.cleanup = function () {
   return true;
};

Connector.prototype.sendQuery = function () {
   return true;
};

Record.prototype.preParse = function (e) {
   if (this.CONNECTION_ERROR != null || typeof this.RXMap ==
        "undefined") {
        return false;
   }
   var headerRegex = new RegExp("id;timestamp;source_uri;" +
              "source_port;" +
              "destination_uri;destination_port;type;username;" +
              "message");

   if (headerRegex.test(this.s_RXBufferString)){
        return false;
   }

   return true;
};

Record.prototype.parse = function (e) {
   var input_data = this.s_RXBufferString.safesplit(";", "\"");
   this.id = input_data[0];
   this.dateTime = DateTime.parseExact(input_data[1],
        "yyyy-MM-dd HH:mm:ss.SSS");
   this.source_uri = input_data[2];
   this.source_port = input_data[3];
   this.destination_uri = input_data[4];
   this.destination_port = input_data[5];
   this.type = input_data[6];
   this.taxonomy_key = input_data[6];
   this.username = input_data[7];
   this.message = input_data[8];

   if (this.taxonomy_key == "login" || this.taxonomy_key ==
        "logout" || this.taxonomy_key == "password_changed"){
```

```
        this.severity = 1;
   }
   else {
        this.severity = 3;
   }

   if (this.type == "login"){
        this.event_name = "Login"
   }
   else if (this.type == "logout"){
        this.event_name = "Logout"
   }
   else if (this.type == "login_failed"){
        this.event_name = "Login Failed"
   }
   else if (this.type == "password_changed"){
        this.event_name = "Password Changed"
   }

   if (false) {
        this.sendUnsupported();
        return false;
   }
   return true;
};

Record.prototype.normalize = function (e) {
   e.setObserverEventTime(this.dateTime);
   e.setTaxKey(this.taxonomy_key);
   instance.SEND_EVENT = true;
   return true;
};

Record.prototype.postParse = function (e) {
   return true;
};

Record.prototype.reply = function (e) {
   return true;
};
```