Bachelor's thesis

Information & Communications Technology

2021

Konsta Kiiveri

# AUTOMATION IN CYBER SECURITY

BACHELOR'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Information & Communications Technology

2021 | 31 pages

Konsta Kiiveri

# AUTOMATION IN CYBER SECURITY

With the contemporary society's increasing reliance on IT infrastructure, the importance of cyber security is also increasing. The tools used by attackers often utilize automation to find and exploit vulnerabilities in an organization's systems, and attacks have also become more sophisticated and more time-consuming to prevent, thus automation is required in cyber security as well.

The purpose of this thesis was to provide an overview of automation methods used in cyber security, and evaluate how automation can improve security. Deploying automation is challenging because of the lack of competent professionals and the incompabitility of multi-vendor software and hardware. Security Content Automation Protocol is presented as a solution to some of the challenges.

The research was carried out as a literature review. Books, scientific journals, standards, and articles were used as references in writing this thesis.

The conclusion of this work was that automation has several use cases in security, and automation brings multiple benefits to an organization in detecting, analyzing, and preventing security threats. Automation reduces the amount of required routine tasks for cyber security personnel, so working hours can be used more efficiently to improve security. With the use of automation and orchestration accurate and effective methods can be created to deal with security threats.

KEYWORDS:

automation, cyber security, technology

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Tieto- ja viestintätekniikka

2021 | 31 sivua

Konsta Kiiveri

# TIETOTURVAN AUTOMATISOINTI

Kyberturvallisuus on tärkeämpää kuin koskaan, koska yhteiskunnan perustoiminnot ovat riippuvaisia IT-infrastruktuurin toimivuudesta. Hyökkääjien käyttämät työkalut usein hyödyntävät automaatiota haavoittuvuuksien hyväksikäytössä. Lisäksi hyökkäykset kehittyvät jatkuvasti ja niiden estäminen on hankalampaa, joten automaation hyödyntämistä tarvitaan myös kyberturvallisuudessa.

Tämän opinnäytetyön tarkoituksena oli tutkia automaatiomenetelmiä yleisesti ja arvioida, miten automaatiolla voidaan parantaa tietoturvan tasoa. Automaation käyttöönottoa hankaloittaa osaavan henkilöstön puute ja eri valmistajien tuottamien ohjelmistojen sekä laitteiden yhteensopivuus. Näihin haasteisiin on esitetty ratkaisuksi Security Content Automation Protocol.

Tutkimus tehtiin kirjallisuuskatsauksena. Kirjoja, tieteellisiä julkaisuja, standardeja ja artikkeleita käytettiin lähdemateriaalina opinnäytetyön laatimiseen.

Johtopäätöksinä saatiin, että automaatioteknologialla on useita käyttötarkoituksia tietoturvassa ja automaation käytössä on useita etuja organisaatioille tietoturvauhkien havaitsemisessa, analysoinnissa ja estämisessä. Automaatiolla voidaan vähentää tietoturvahenkilöstöltä vaadittujen rutiinitehtävien määrää, jolloin työaikaa voidaan käyttää tietoturvaa kehittäviin toimintoihin. Tietoturvan orkestroinnilla ja automatisoiduilla vastauksilla tietoturvauhkiin voidaan luoda tarkkoja ja tehokkaita menetelmiä uhkien torjumiseen.

ASIASANAT:

automaatio, tietoturva, teknologia

# CONTENTS

# FIGURES

# LIST OF ABBREVIATIONS

| Abbreviation | Explanation of abbreviation (Source) |
|---|---|
| 5G | Fifth-generation of cellular networks |
| AI | Artificial Intelligence |
| API | Application Programming Interface |
| ARF | Asset Reporting Format |
| CCE | Common Configuration Enumeration |
| CCSS | Common Configuration Scoring System |
| CPE | Common Platform Enumeration |
| CPU | Central Processing Unit |
| CVE | Common Vulnerabilities and Exposures |
| CVSS | Common Vulnerability Scoring System |
| EDR | Endpoint Detection and Response |
| IDS | Intrusion Detection System |
| IoT | Internet of Things |
| IPS | Intrustion Prevention System |
| $(ISC)^2$ | International Information Systems Security Certifications Consortium |
| IT | Information Technology |
| MD5 | Message Digest 5 |
| ML | Machine Learning |
| MSSP | Managed Security Service Provider |
| NGFW | Next-generation Firewall |
| NIST | National Institute of Standards and Technology |
| OCIL | Open Checklist Interactive Language |
| OVAL | Open Vulnerability and Assessment Language |
| PE | Portable Executable |
| SCAP | Security Content Automation Protocol |
| SIEM | Security Information and Event Management |

| | |
|---|---|
| SOAR | Security Orchestration, Automation, and Response |
| SOC | Security Operations Center |
| SSH | Secure Shell |
| SWID | Software Identification |
| TIP | Threat Intelligence Platform |
| TMSAD | Trust Model for Security Automation Data |
| UEBA | User and Entity Behavior Analytics |
| USD | United States Dollar |
| VPN | Virtual Private Network |
| XCCDF | Extensible Configuration Checklist Description Format |
| XML | Extensible Markup Language |
| YAML | YAML Ain’t Markup Language |

# 1 INTRODUCTION

Cyber security has been an important subject as long as the Internet has existed, and the importance has only increased in recent years due to a growing reliance on IT infrastructure due to telecommuting, online entertainment, and online business to name a few key areas. With faster internet connections, more users, and more devices, the amount of data moving in networks has seen massive growth. Disruptions in the flow of data can be very costly for businesses and society.

Security threats and attacks are only becoming more sophisticated and complex, and with attackers utilizing automation, organizations face a continuous threat. Keeping applications, operating systems, and devices securely patched, and identifying, analyzing and responding to threats is time-consuming to security experts, and automating these processes is a way to make systems more resilient and secure against many threats.

Automating complex security tasks is a rather new field, with plenty of room for growth. Security automation ranges from using basic scripts to more complex processes like integrating machine learning and artificial intelligence in security software.

This thesis discusses the methods and benefits of security automation, and explores its current and future prospects. The second chapter defines the need for security automation. The third chapter explains the functionalities of a security operations center, and it goes into detail of the various security tools available in the industry, and examines the level of automation the tools include. The fourth chapter explains methods of automation that are used, additionally the use of artificial intelligence and machine learning is also discussed. The fifth chapter presents challenges of security automation, and potential solutions to the challenges are listed. The sixth chapter concludes the findings and results of the literature research of this thesis.

# 2 THE NEED FOR AUTOMATION

In this chapter, the need for security automation is defined by listing shortcomings in the cyber security ecosystem that could be addressed with automation.

Estimations for annual economic cost of damages caused by cybercrime range from hundreds of billions to several trillions of USD, and the damages have been estimated to increase each year by 15%. This includes damage and destruction of data, stolen money, lost productivity, theft of intellectual property, theft of personal and financial data, embezzlement, fraud, post-attack disruption to the normal course of business, forensic investigation, restoration and deletion of hacked data and systems, and reputational harm. [1]

Humans are prone to making errors, whereas machines are reliable and accurate as long as the logic inserted into them is working as intended. Human errors can lead to configuration errors, information leakage, failure to patch a system on time, or overlooking an important alert in a security system. For example, the high profile Equifax data breach in 2017 was carried out by exploiting the Apache Struts CVE-2017-5638 vulnerability, leading to the disclosure of up to 143 million customer details. The vulnerability had already been identified and patched two months prior to the attack, but Equifax had failed to install the update on their Apache system in a timely manner. [2] By decreasing the amount of required human input, systems can be made more secure, because the chance of human errors will also be reduced.

There is a clear desire to further adopt security automation. The Ponemon Institute surveyed 1,859 IT and IT security professionals in Germany, France, the United Kingdom and the United States in 2018. The participants in the survey were in organizations that at the time deployed or planned to deploy security automation, and the participants were asked to rate the importance of security automation to achieve a strong security posture at the current time and in two years time. The result was that 70% of respondents rated it a very high priority currently, and 80% rated it a very high priority in two years. [3]

## 2.1 Asymmetry in attack and defence

A malicious actor only needs to find one weak point that can be exploited to gain access to a system, whereas a defender needs to make sure all points are secure and no vulnerabilities exist [4]. Needless to say, it is an impossible task. New vulnerabilities are found all the time, and once they are made public developers need to act swiftly to patch the vulnerability and release an update to a secure version, and users need to install the updates to avoid becoming a victim to attackers abusing the vulnerability in their systems.

Attackers can analyze how security tools work, and adapt their attacks to take advantage of that knowledge. As an example, traditional anti-virus programs look for signatures of known attack tools, and stop the execution of a file if a known signature of a malicious file is identified. To avoid detection, fileless malware take advantage of non-malicious and whitelisted programs already installed on the system, such as Windows PowerShell as illustrated in Figure 1. PowerShell can then be used to run malicious commands and gain further access in the network. [5]

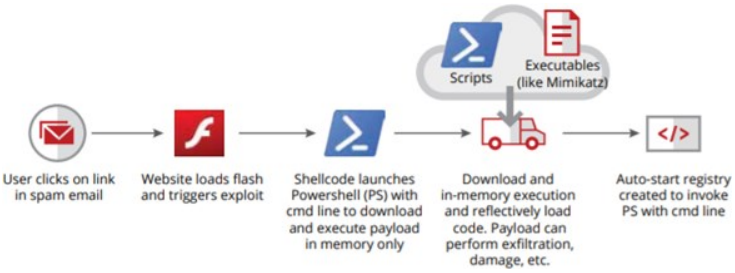


Figure 1. Example of a fileless attack kill chain [5].

Cyber security has to adapt to a changing threat landscape continuously. Adoption of new technologies, such as 5G, IoT, and cloud systems also introduce new attack surfaces which pose new threats. Advances in technology also entail more advanced and sophisticated tactics, techniques, and procedures used in attacks. Due to these facts, SOCs and security tools have to evolve to keep up with the threats organizations are facing, and automating security is a high priority operation to keep up with the increasing requirements in security. The goal of security automation is not to replace human analysts, but to empower SOCs with better capabilities to monitor, identify, and respond to threats.

## 2.2 Workforce shortage

Another important factor in favor of further expanding cyber security automation is the shortage of skilled security professionals. (ISC)² conducts a yearly study on global cyber security workforce. The study is based on survey data from individuals responsible for cyber security at workplaces. In 2020, 56% of respondents in their survey said that cyber security staff shortages are putting their organizations at risk, and the estimated workforce gap was 3.1 million. The definition of workforce gap is the amount of qualified workers needed to fulfil the needs of the industry, which is not directly the number of open job positions. [6]

There is a high demand for cyber security professionals, but the required skillset for security positions is increasing. Security analysts must operate in high-pressure situations to quickly analyze and respond to security incidents. Stressful working conditions can lead to burnout and employee churn [4]. Automation leads to higher productivity, and it can reduce the stress experienced at work which will lead to less burnout.

# 3 SECURITY OPERATIONS CENTER

In this chapter we cover what a SOC is, and what are the most important tools used in modern SOCs, and explore whether these tools include automation.

In short, a SOC is a team of security analysts, whose job is to detect, analyze, respond to, report on, and prevent cyber security incidents [7]. There are two type of SOCs, an internal SOC that exists within an organization and is in charge of that organization's security operations, and outsourced SOC-as-a-service model which is offered by managed security service providers (MSSP). Larger businesses often have internal SOCs, while mid-size businesses will often look into hybrid solutions or outsourcing their security operations with MSSPs.

The SOC is the place where all the logged events in the network are monitored, and the SOC is responsible on taking action when it is needed. The goal of a SOC is to understand the entire threat landscape of the organization and do their best at protecting the on-premises IT infrastructure, and also third-party services such as cloud services [8]. Essentially the SOC needs to stay on top of all possible threats, and figure out how to best protect against, mitigate, and prevent these threats.

To achieve that goal, continuous monitoring is conducted with tools scanning the network around the clock. The security center is notified of anomalous activity and threats immediately, so most problems can be prevented or mitigated before they become major incidents. Human analysts do not have to investigate every notification because false positives, redundant alerts and non-critical alerts are filtered out. In the case of actual threats, the SOC will determine the importance and respond to the threat accordingly by isolating endpoints around the incident if needed and carrying out an investigation, while making sure it has minimal impact on business. [8]

After responding to a threat, recovery and remediation takes place. Cleaning up the devices, restoring backups, and reconfiguring systems are some aspects of recovery and remediation. Finding the root cause is important to prevent such attacks in the future, so logs of network activity is investigated to figure out the source of the problem, and take action if needed to fix any holes found in security. [8]

Figure 3 illustrates the abstracted use of tools in a SOC, presenting how the different components interact in a security environment.

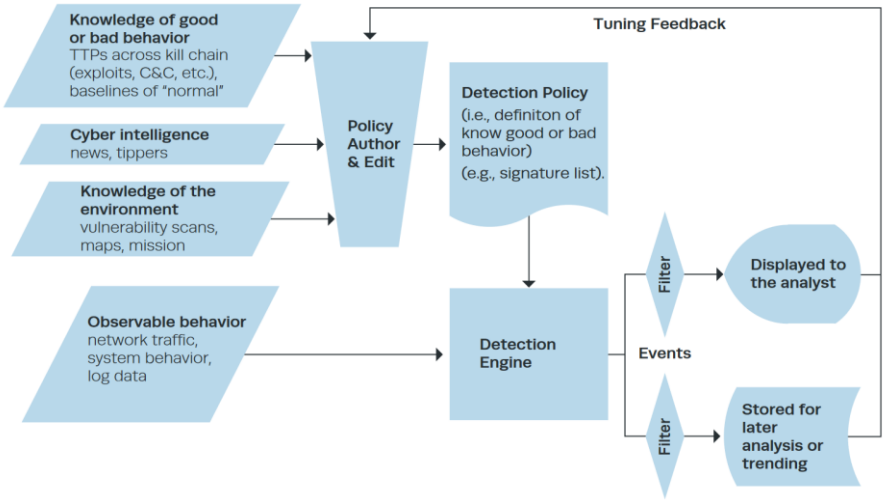


Figure 2. Interaction of SOC components [7].

## 3.1 Tools in use

A SOC relies on a variety of security tools to achieve its goals in keeping an organization's information systems secure. Many of these tools are designed for specific tasks, with the main purpose of generating, collecting, analyzing, storing and presenting the data.

The amount of data generated by security tools in modern complex networks is so large, that it is not feasible to monitor everything through individual consoles and dashboards.In a well-designed SOC environment these tools are centrally managed and the information is presented to human analysts in a cohesive format so important incidents do not go unnoticed and the analysts are not overwhelmed by a large amount of unimportant data.

### 3.1.1 SIEM

The security information and event management (SIEM) system is the most integral part of a SOC. System logs and events from various security components are collected and aggregated in the SIEM, which in turn uses its correlation and analysing capabilities to detect anomalities and threats to alert the SOC staff of potential security incidents. SIEM enables an effective use for all the data collected to provide real-time monitoring, quick detection, and response to attacks. [9]

Machine learning is used to further improve the capabilities of SIEMs. Next-generation SIEM solutions integrate machine learning techniques to better analyze the massive amount of data the systems are handling, and these ML models are trained to find both known and unknown threats, and they also involve behavioral analytics [10]. User and entity behavior analytics (UEBA) monitors and models normal behaviour of entities and users to create a baseline. The baseline includes where and when the user logs in to systems, which files and servers they usually access, and which devices they use. When anomalies and suspicious activity is detected, the SOC is alerted. UEBA can help defend against insider threats as well as cases where an employee's credentials are stolen and used maliciously. [11]

### 3.1.2 IDS and IPS

Intrusion detection systems (IDS) and intrusion prevention systems (IPS) both monitor network traffic and compare the contents of packets to a database of known threats. The difference between them is that IDS is a passive system that only works in detection and monitoring and does not take action on its own, while IPS is a control system which can drop packets if their content is found malicious. [12]

There are two categories of IDS and IPS systems depending on their detection technique: signature-based detection and statistical anomaly-based detection. Signature-based detection is vulnerable to zero-day attacks, because they are not yet in a database of known threats. Anomaly-based detection creates a network baseline and when there is deviation, the administrator is notified. Anomaly-based detection is stronger against new types of attacks, but has a higher rate of false positives. Modern IDS and IPS systems use a combination of the two techniques. [13]

IPSs are a good example of automation in security. These systems only require the threat databases to be kept up to date, and they handle the identification and response to threats and attacks on their own. IPSs have replaced IDSs to some extent due to the automatic response features IPSs have.

### 3.1.3 NGFW

Traditional port-based firewalls are ineffective in protecting corporate networks, because a port-based approach is limited to inspecting the TCP or UPD header of a packet to determine the application protocol and then either allowing or blocking the traffic through a port. To address this problem, IDS/IPS, URL filtering, and other security solutions were implemented, leading to more convoluted and harder to configure systems. [14]

Next-generation firewalls (NGFW) were developed to “restore firewalls as the cornerstone of enterprise network security” [14]. NGFWs are defined by their user identity-awareness and application-awareness, and they integrate functionalities of IPSs and traditional firewalls. The primary features of traditional firewalls are packet filtering, network- and port-address translation, stateful inspection and VPN support. NGFW was one of the first technologies to take advantage of a zero trust architecture. Zero trust works on the basis of “never trust, always verify”, as opposed to always trusting network traffic which originates from inside the network, which leaves the network vulnerable for lateral movements of attacks. [14] IPS functionalities are described earlier in 3.1.2, and NGFW is able to deliver the same functionalities with deep packet inspection.

NGFWs incorporate workflow automation, policy automation, and security automation. Workflow automation includes APIs so the firewall can be programmed by other tools and scripts the user might use. NGFWs are also able to use the APIs of other devices to make consistent policy changes when necessary. Policy automation means that the firewall is able to adapt to changes in the environment, and it can take in threat intelligence from third-party sources and act on that intelligence automatically. Security automation enables the firewall to block new threats when information of them is delivered to the firewall by other security tools. [14]

### 3.1.4 EDR

Endpoint detection and response (EDR) is an endpoint security solution that combines real-time monitoring and collection of endpoint data with automated response and analysis capabilities. A high degree of automation is integrated to enable security teams to quickly identify and respond to threats. [15]

Endpoints refer to end devices, such as servers, desktops, laptops, IoT devices, and mobile devices, which can all act as entry points for attackers. The main features of EDR solutions are monitoring and collecting activity data, analyzing this data to identify threat patterns, automatically responding to identified threats and notifying the SOC, and forensic tools to research identified threats or hunt for threats that may still be undetected on an endpoint. [15]

EDR solutions were developed to fill gaps in security that other tools do not take care of. The amount and diversity of end devices accessing networks has grown massively, and if left unaddressed they pose a large threat to security. EDR offers a unified solution to these threats, as opposed to a large amount of overlapping security tools for different types of devices. With automatic identification and responding capabilities they are an effective way to deal with the growing amount of endpoints and attacks targeting them.

### 3.1.5 TIP

Threat intelligence is evidence-based information or knowledge of the context, mechanisms, indicators, implications, and actionable advice of existing or emerging threats [16]. Threat intelligence feeds are data streams that share gathered threat intelligence, allowing organizations to use shared knowledge in their security. Often these feeds are free and they use a collection of open source intelligence, but there are also paid feeds which combine open and closed sources of intelligence. A threat intelligence feed can also come from internally collected and analyzed data within an organization.

Threat intelligence platforms (TIP) combine multiple threat intelligence feeds to put relevant intelligence into use, augmenting SIEMs, endpoints, firewalls, and other security systems with up-to-date threat intelligence. [17]

The use of machine learning to aid in producing accurate cyber threat intelligence from various data sources has been researched in recent years. The identified issue with the current model is too much irrelevant threat intelligence being included in threat intelligence feeds, with a distinct lack of strategic threat intelligence, such as attack patterns and techniques that represent the behavior of an attacker or an exploit. [18]

# 4 AUTOMATION METHODS

In the early days of security automation, the methods mainly consisted of Unix scripts that would automatically collect logs or configure devices. Modern automation tools are centrally controlled automation platforms which can orchestrate the network security of organizations. However, all forms of security automation have their limits on how much they can do on their own. Human input is required in decisionmaking and intervention no matter how sophisticated an automation system is, but automation is an effective way to reduce the workload of SOCs, so that security professionals can put their efforts into more productive tasks and futher improving security, rather than carrying out routine work that can be done by machines.

Automation methods can be split into two main categories: robotic automation and cognitive automation. Robotic automation includes the use of scripts and software to automate routine and repetitive tasks. Cognitive automation uses more advanced techniques, artificial intelligence (AI) and machine learning (ML), to learn about network and security baselines to detect anomalies, respond to and analyze threats. [19]

## 4.1 Scripts

In the early 2000s, system administrators could create basic automation of security tasks with scheduled Unix scripts. They were easy to create, implement and maintain because creating shell scripts only requires basic knowledge of Unix commands. Examples of such scripts would be file integrity checks with MD5 checksums, detecting incoming port scans, and log parsing [20]. A more modern approach are scripts which can read information from devices or make configuration changes to devices or network security products with the help of APIs. These tasks can be scheduled or triggered by events in the network [21].

There is a limit to how much can be accomplished with scripts in creating security policies and maintaining a secure network. More advanced software is required for automation in large enterprises because there is such a large number of protocols, networking devices, endpoint devices, and applications in use.

## 4.2 Central control automation tools

A more modern approach to automation are security automation platforms, which run separately on servers and manage different security tools and devices in a network. Some of these automation tools require software agents to be installed on the controlled devices to be able to send and process information to and from the central automation tool [21].

One example of a security orchestration platform is Ansible. Automation tasks are written in the form of Ansible Playbooks using YAML as the programming language. Playbooks describe a series of actions to take, and instructions to devices are sent via SSH by default [22]. Ansible is a modular platform that can be used to control multiple security tools such as firewalls, intruder detection & prevention systems, security information and event management, and privileged access management.

## 4.3 SOAR platforms

Security Orchestration, Automation and Response platforms take orchestration and automation one step further. It is defined as technologies that enable an organization to collect inputs from a variety of sources, and it covers capabilities of vulnerability management, security operations automation and incident response. SOAR combines the use of human and machine power to analyze and respond to security incidents. [23]

In their early stages the tools were quite limited in scope and only offered minor time savings on select tasks for SOCs. Since then, these platforms have only become more effective with the ability to orchestrate and automate larger and more significant operations with more security tools being integrated to SOAR platforms. [24]

Processes in SOAR platforms can be handled using playbooks and runbooks. Playbooks and runbooks direct a standardized approach to incident response, allowing repeatable, enforceable, and effective incident response workflows. They help with following regulations, reporting procedures, automation, and orchestration. The difference between the two is that playbooks offer more of a step-by-step approach for one type of incident, while runbooks have conditional steps depending on the type of incident. The playbook and runbook approach ensures similar incidents are handled in a similar

fashion, which simplifies the tasks and also helps new staff learn and handle tasks quickly. [25]

A proper implementation of SOAR takes advantage of the entire security stack of an organization, orchestrating multi-vendor tools to streamline processes in security. The technology has reached a level where the users do not have to be experts on automation, instead the platform can be integrated into an existing framework with the use of pre-built automation playbooks, guided investigation workflows, and automated alert prioritization. The objective with orchestration and automation is to have a quicker response to security threats. Identifying and reacting to a threat earlier will limit the amount of damage and the severity of a potential data breach. [26]

The vision of SOAR technology is to create a single dashboard which includes all the metrics an analyst would need to assess the security situation, and take action if needed. The technology is not quite there yet, but it is something to keep an eye out for because progress is being made towards this goal.

### 4.4 Artificial Intelligence and Machine Learning

Automation is the act of programming a machine to perform a repetitive task without human involvement, but higher usability in advanced levels of security require machines to perform dynamic tasks, and also learn and analyze data, and that is when AI and ML come into the picture. In recent years academic and institutional research has been focused on leveraging AI and ML in security. With more advancements in technology, AI and ML could become very powerful instruments in the field of cyber security, further reducing human workload while also making systems more resilients to many threats.

AI and ML have not been integrated to a full degree yet, but early stages of utilization has already taken place in SOAR platforms. The ML models are trained on prior incidents and historical data, which they take into accordance when analyzing new alerts, so patterns can be detected and predictions can be made based on the patterns. There are many use cases for AI and ML in SOAR platforms, and the top SOAR products were studied in the article *AI/ML in Security Orchestration, Automation and Response: Future Research Directions*. [27]

A common use for ML models in SOAR platforms is prioritization of alerts. Alerts that are deemed critical are placed in the front of an analyst's case queue, and alerts that are not

critical, such as likely false positives, are given low priority. The algorithms also provide a list of past incidents similar to the one at hand, so the human analyst can use precedent cases in their decisionmaking in their current investigation. Threat forecast, prediction, behavioral analytics, and anomaly detection are important uses for ML, because analyzing large datasets and monitoring datastreams can be done efficiently with the help of ML. This accelerates threat detection and incident response time. [27]

ML can be trained for malware detection and anti-virus defence. Whereas traditional anti-virus programs rely on signatures of known malicious files, artificial intelligence can detect whether new suspicious samples are malicious by training the ML model on analyzing a PE file's features, which include byte sequences, opcodes, API and system calls, network activity, file operations, CPU registers, PE file characteristics, and strings. [28] However, commercial anti-virus programs are unlikely to be replaced by programs heavily utilizing machine learning any time soon, because extracting PE features and executing ML algorithms is computationally very demanding, and the required computing power is out of range for consumer computers.

### 4.4.1 Implementation of AI-assisted security tools

In 2019 Capgemini released a report with the following chart (Figure 1) with potential use cases for AI in cyber security across information technology, operational technology and internet of things. Use cases were ranked based on the complexity of implementation and ability to save time for security specialists. 850 executives from IT Information Security, Cyber security and IT Operations were surveyed on their organization's deployment of the AI-assisted cyber security tools, and the results showed that almost a half of organizations in the survey had deployed AI-assisted tools to some degree. [29]

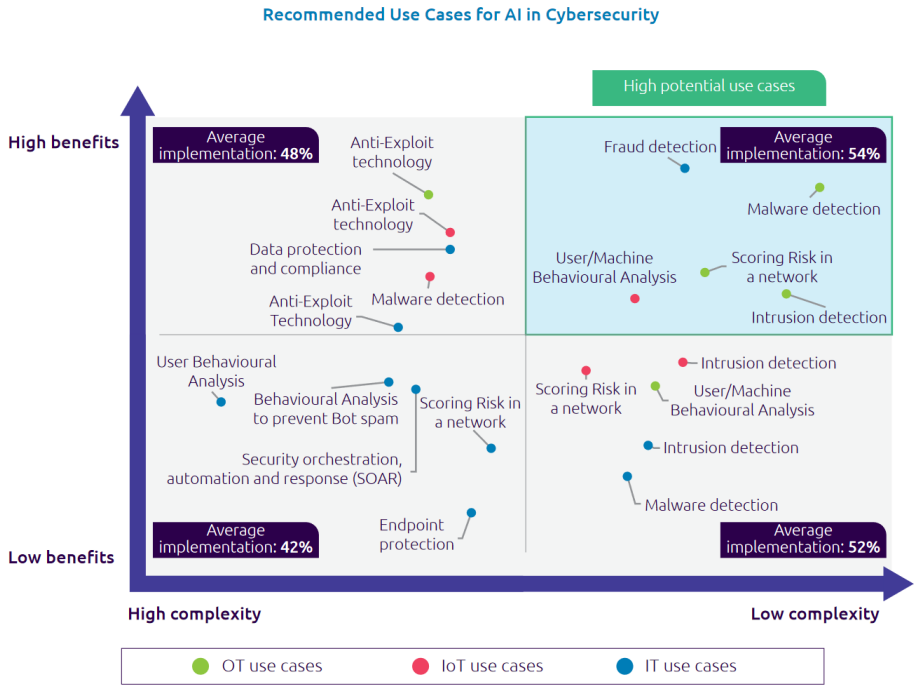


Figure 3. AI in Cyber security executive survey [29].

### 4.4.2 Market research of AI in cyber security

According to a report published by Zion Market Research [30], AI in the global cyber security market was valued at 7.1 billion USD in 2018 and is forecasted to reach 30.9 billion USD by 2025. This promises massive growth in the industry, showing that AI does not simply have limited niche uses in cyber security, but rather has a growing role in the future of security. Industry leaders operating in the AI-assisted cyber security market include IBM, Intel, Symantec, and Cisco among many other IT and security organizations. [30]

# 5 CHALLENGES

Despite the many advantages of automation in security, organizations are still not taking full advantage of this technology to further improve their defensive capabilities. Attackers have an easier time adopting automation in their attacks, because the attacker can afford to fail in an attack, only to try again with different parameters or try a different attack technique, but businesses cannot afford to fail in their defence.

Companies have to decide how, when, and where to implement security automation to improve productivity, reduce costs, and strengthen their current security. It is not feasible to jump straight from mostly manual processes in security into highly automated ones. To get the most out of automation, many existing security routines and infrastructure will have to be restructured, and knowing how to achieve that requires expertise. A lack of knowledgeable and skilled personnel is one of the biggest challenges in deploying security automation. [3]

Enterprise networks employ multi-vendor devices and security software, and interoperability between all of the systems is hard to achieve, which creates a large challenge in creating an effective security automation architecture. The integration of legacy systems presents another problem due to their limited or entirely lacking support of automation. The complexity of security configuration management due to the large amount of heterogeneous systems makes it difficult to create an overarching automation infrastructure.

## 5.1 SCAP

Some of the aforementioned challenges can be solved with standards created for security automation. With standardized specifications and requirements aimed for security automation, designing automation features in security software, and planning and deploying a highly automated security infrastructure requires less expertise and the problems with interoperability can also be solved.

Security content automation protocol (SCAP) was created to standardize the format and terminology used by security software products to communicate information about software identification, software flaws, and security configurations to machines and

humans. The National Institute of Standards and Technology has defined the technical specifications of SCAP in NIST SP 800-126. [31]

SCAP incorporates multiple standards to achieve its goals in compability between security tools by different vendors and to establish a shared content repository between tools. The standards and specifications in SCAP have five categories: languages, reporting formats, identification schemes, measurement and scoring systems, and integrity. [31] To build an understanding of the standards included in SCAP, the following sections describe the specifications and their uses are explained mostly through NIST's documentation and definitions.

### 5.1.1 Languages

Extensible Configuration Checklist Description Format (XCCDF) is a language for writing security checklists, benchmarks, and reports in a structured and uniform format, with the intent to support integration with multiple configuration checking engines. The important elements supported by this specification are information interchange, automated compliance testing, and compliance scoring. [32]

Open Vulnerability and Assessment Language (OVAL) is a language that standardizes the main steps of vulnerability and configuration assessment process, which includes presenting configuration information, analyzing the system for specified machine states, and reporting the results of the assessment. This enables exchanging this information between security tools and services in a machine-readable format. [33]

Open Checklist Interactive Language (OCIL) is a language that is used for data collection from non-automated sources, such as questionnaires on security controls presented to people as a part of a security audit. The standardized format allows software to process and harvest the collected information. [34]

The main intention of the language standards and specifications is to define consistent and standardized means of formatting and collecting information to enable exchanging information between security software and devices.

### 5.1.2 Reporting formats

Asset Reporting Format (ARF) is a data model that defines the transport format of information about assets and the relationships between assets and reports. ARF standardizes how products produce and receive reports, increasing data interoperability of ARF comforming products, which enables automation processes of disparate products. [35]

Asset Identification is a format used to uniquely identify assets based on known identifiers and attributes. Asset identification improves asset management processes by allowing uniquely identifying an asset, correlation of asset data, and reporting of asset information. Asset identification was developed to support the needs of security automation, because existing specifications did not consider asset identification. [36]

### 5.1.3 Identification schemes

Common Platform Enumeration (CPE) is a method of naming systems, software, and devices in a standardized and structured format. The naming method is descriptive of the class of the product in question, giving a unique name for each product which includes metadata like software creator names, product edition, software version. Security tools can identify products by their CPE names, and detect if there are any known vulnerabilities in the product. [37]

Software Identification (SWID) tagging is a structured format for representing software identifiers and associated metadata. It has many similarities to CPE, and CPE as a standard is being retired in future revisions of SCAP in favor of using SWID tagging. [38]

Common Vulnerabilities and Exposures (CVE) provides a public database of known security vulnerabilities and exposures with standardized CVE identification numbers assigned for all known vulnerabilities. Security software can use CVE identifiers to refer to known vulnerabilities. [39]

Common Configuration Enumeration (CCE) is a list of security-related software configuration issues, and similarly to CVE, unique identifier numbers are assigned to them. The CCE entries have a description of the configuration issue, and each issue has

references to documents where the configuration issue is described in more detail to make it easier to correct the issue. [40]

The purpose of the identification schemes is to standardize the use of common terminology and identifiers between security software and security professionals, so when information of threats and security incidents is shared between security software or between organizations, there is no conflicting identification schemes used.

### 5.1.4 Measurement and scoring systems

Common Vulnerability Scoring System (CVSS) provides a measurement system to communicate the characteristics and severity of vulnerabilities, and assign a numerical score ranging from 0 to 10 based on its severity. [41]

Common Configuration Scoring System (CCSS) is inspired by CVSS, but instead of measuring the severity of vulnerabilities, CCSS measures the severity of software security configuration issues. [42]

The scoring systems help security teams in prioritizing the correction of flaws in their systems. Many vulnerabilities can exist in an organization's systems at once, but some of them may be completely inconsequential while another can be critical to security with an immediate need to be fixed. Having a score measuring the severity of the vulnerability will help the security team in deciding where to focus their resources without having prior experience of the vulnerability.

### 5.1.5 Integrity

Trust Model for Security Automation Data (TMSAD) provides a way to securely process and exchange security automation data. Information in the security automation domain is primarily exchanged using XML, so the main focus of TMSAD is the processing of XML documents. The model specifies the use of signatures, hashes, key information, and identity information in exchanging XML documents. [43]

To protect the integrity of security automation, standards must be set for the authentication methods used in the exchange of security information. If such standards were not set, automation could end up being a vulnerable part in a system.

## 5.2 SCAP version 2

SCAP version 1.0 was first published in 2009, and it is currently in its third revision, SCAP 1.3. With the technological progress in the IT industry and security automation, there are some gaps that need to be addressed, and SCAP version 2 is being developed to fill these gaps. The critical gaps in the current version of SCAP are limited coverage of endpoint types, stale security posture information, no component-level interoperability, difficult content creating and limited content availability, and limited software inventory and patch support. [44]

The design goals in SCAP v2 have the intent of addressing the issues and supporting new capabilities. The specifications and standards in SCAP v1 had focused on supporting standard endpoints like desktops, laptops, and servers. The focus is being expanded to a full endpoint type support, which includes networking devices, mobile devices, IoT, and medical devices. [44]

To address the stale security posture information in SCAP v1, new standards will be introduced to support endpoints sending notifications to security management servers when security incidents are detected. This will allow better integration with automation platforms like SOAR. Support of component level interoperability is introduced with data models with standardized protocols, to improve the exchange of data and results between security products. Additionally specifications for accessible data repositories are incorporated in SCAP v2 to enable multiple processes to access the same datasets for analytics. To address the difficulty of content creation and limited content availability, XCCDF and OVAL will be updated to support simpler security automation content creation and to improve its availability. To improve the limited software inventory and patch support, SWID tags are used to exchange information of installed software on a network. [44]

The issues in SCAP v1 and design goals for SCAP v2 were identified to build a strong foundation for advanced cyber defence capabilities and the next generation of security automation solutions. [44]

# 6 CONCLUSION

The goal of this thesis was to research and build an understanding of the use of automation in cyber security, exploring its use cases, and presenting the benefits that automation offers. Challenges regarding implementation of security automation are also discussed, and a solution to them is presented with the standardized approaches of SCAP. The research was carried out by studying books, scientific journals, technical standards, and articles related to the subject.

A basic level of understanding on security automation, the methods of security automation, and the various use cases and benefits have been presented in this thesis. Some future capabilities of automation leveraging artificial intelligence and machine learning have also been discussed. The specificiations of standards in SCAP have briefly been explained with the challenges they seek to address. This thesis can be used as a starting point to look into implementing automation in an organization’s security environment.

Cyber security automation reduces the workload of security personnel, so the amount of hours spent on routine tasks can instead be used on more productive tasks, such as assessing and improving the current security landscape of the organization. The use of SOAR platforms enables consistent responses to security threats with orchestration and automation, enabling accurate and effective responses to various threats.

With more technological advancements and research into security automation, the technology will see further adoption by organizations, leading to a more secure cyberspace. Further research subjects could be the use of AI and ML in consumer level security software.

# REFERENCES

[1] Morgan S. Cybercrime To Cost The World $10.5 Trillion Annually By 2025 [Internet]. [place unknown]: Cybercrime Magazine, 2020 [cited 2021 Apr 9]. Available from: https://cybersecurityventures.com/hackerpocalypse-original-cybercrime-report-2016/

[2] Brewster T. How Hackers Broke Equifax: Exploiting A Patchable Vulnerability [Internet]. [place unknown]: Forbes, 2017 [cited 2021 Apr 9]. Available from: https://www.forbes.com/sites/thomasbrewster/2017/09/14/equifax-hack-the-result-of-patched-vulnerability/?sh=de6f86c5cda4

[3] Ponemon Institute. The Challenge of Building the Right Security Automation Architecture. Michigan, USA: Ponemon Institute, 2018 [cited 2021 Apr 9]. p. 1. Available from: https://junipernetworks.lookbookhq.com/c/security-operations-ponemon-architecture?x=CsrL1v

[4] Wendt D. Driving Forces for Security Automation [Internet]. [place unknown]: IACD, 2019 [cited 2021 Apr 9]. Available from: https://www.iacdautomate.org/driving-forces-for-security-automation

[5] McAfee. What is Fileless Malware? [Internet]. [place unknown]: McAfee, 2021 [cited 2021 Apr 9]. Available from: https://www.mcafee.com/enterprise/en-us/security-awareness/ransomware/what-is-fileless-malware.html

[6] (ISC)². Cybersecurity Professionals Stand Up to a Pandemic. [place unknown]: (ISC)², 2020 [cited 2021 Apr 10]. p. 14. Available from: https://www.isc2.org/-/media/ISC2/Research/2020/Workforce-Study/ISC2ResearchDrivenWhitepaperFINAL.ashx?la=en&hash=2879EE167ACBA7100C330429C7EBC623BAF4E07B

[7] Zimmerman C. Ten Strategies of a World-Class Cybersecurity Operations Center. McLean, USA: MITRE Corporation, 2014. p. 8-9. Available from: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.662.545&rep=rep1&type=pdf

[8] McAfee. What Is a Security Operations Center (SOC)? [Internet]. [place unknown]: McAfee, 2021 [cited 2021 Apr 10]. Available from: https://www.mcafee.com/enterprise/en-us/security-awareness/operations/what-is-soc.html

[9] Logpoint. What is SIEM? A complete guide to Security Information and Event Management [Internet]. [place unknown]: Logpoint, 2021 [cited 2021 Apr 10]. Available from: https://www.logpoint.com/en/understand/what-is-siem/

[10] Securonix. What is Next-Generation SIEM? [Internet]. [place unknown]: Securonix, 2021 [cited 2021 Apr 10]. Available from: https://www.securonix.com/what-is-next-generation-siem/

[11] FireEye. What is UEBA? Definition and Benefits [Internet]. [place unknown]: FireEye, 2021 [cited 2021 Apr 10]. Available from: https://www.fireeye.com/products/helix/what-is-ueba.html

[12] Petters J. IDS vs. IPS: What is the Difference? [Internet]. [place unknown]: Varonis, 2020 [cited 2021 Apr 10]. Available from: https://www.varonis.com/blog/ids-vs-ips/

[13] Nilă C, Apostol I, Patriciu V. Machine learning approach to quick incident response. In: 2020 13th International Conference on Communications, 2020 [cited 2021 Apr 11]. p. 291-292. Available from: https://doi.org/10.1109/COMM48946.2020.9141989

[14] Miller L. Next-Generation Firewalls For Dummies. New Jersey: John Wiley & Sons; 2019 [cited 2021 Apr 11]. p. 3-5. Available from: https://incom.co.uk/wp-content/uploads/2020/10/Next-Generation-Firewalls-For-Dummies.pdf

[15] McAfee. What Is Endpoint Detection and Response (EDR)? [Internet]. [place unknown]: McAfee, 2021 [cited 2021 Apr 11]. Available from: https://www.mcafee.com/enterprise/en-us/security-awareness/endpoint/what-is-endpoint-detection-and-response.html

[16] Gartner Research. Definition: Threat Intelligence [Internet]. Stamford, USA: Gartner Research, 2013 [cited 2021 Apr 11]. Available from: https://www.gartner.com/en/documents/2487216/definition-threat-intelligence

[17] Palo Alto Networks. What is a Threat Intelligence Platform. Santa Clara, USA: Palo Alto Networks, 2021 [cited 2021 Apr 12]. Available from: https://www.paloaltonetworks.com/cyberpedia/what-is-a-threat-intelligence-platform

[18] Ghazi Y, Anwar Z, Mumtaz R, Saleem S, Tahir A. A Supervised Machine Learning Based Approachfor Automatically Extracting High-Level Threat Intelligence from Unstructured Sources. Islamabad, Pakistan: 2018 International Conference on Frontiers of Information Technology, 2018 [cited 2021 Apr 12]. p. 129-130. Available from: https://doi.org/10.1109/FIT.2018.00030

[19] Pitt L. Security Automation Challenges to Adoption: Overcoming Preliminary Obstacles [Internet]. [place unknown]: Securityweek, 2020 [cited 2021 Apr 13]. Available from: https://www.securityweek.com/security-automation-challenges-adoption-overcoming-preliminary-obstacles

[20] Newstrom H. Using Linux Scripts To Monitor Security. In: SANS Institute Information Security Reading Room; 2002 [cited 2021 Apr 6]. p. 3-5. Available from: https://www.sans.org/reading-room/whitepapers/linux/linux-scripts-monitor-security-197

[21] Nagy R, Christensen T, Horne G. Cybersecurity Automation For Dummies. New Jersey: John Wiley & Sons; 2019 [cited 2021 Apr 6]. p. 13-18. Available from: https://www.infoblox.com/wp-content/uploads/infoblox-ebook-cybersecurity-automation-for-dummies.pdf

[22] Ansible. How Ansible Works [Internet]. [place unknown]: Ansible, 2021 [cited 2021 Apr 6]. Available from: https://www.ansible.com/overview/how-ansible-works

[23] Neiva C, Lawson C, Bussa T, Sadowski G. Innovation Insight for Security Orchestration, Automation and Response. [place unknown]: Gartner Research, 2017 [cited 2021 Apr 7]. p. 3-4. Available from: https://www.gartner.com/en/documents/3834578

[24] Brooks C. SOAR cybersecurity: reviewing Security Orchestration, Automation and Response [Internet]. [place unknown]: AT&T, 2020 [cited 2021 Apr 7]. Available from: https://cybersecurity.att.com/blogs/security-essentials/security-orchestration-automation-and-response-soar-the-pinnacle-for-cognitive-cybersecurity

[25] Cyware. What is the Difference Between a Security Playbook and a Runbook? [Internet]. [place unknown]: Cyware, 2020 [cited 2021 Apr 9]. Available from: https://cyware.com/educational-guides/security-orchestration-automation-and-response/what-is-the-difference-between-a-security-playbook-and-a-runbook-ddc4

[26] Brooks C. SOAR cybersecurity: reviewing Security Orchestration, Automation and Response [Internet]. [place unknown]: AT&T, 2020 [cited 2021 Apr 7]. Available from: https://cybersecurity.att.com/blogs/security-essentials/security-orchestration-automation-and-response-soar-the-pinnacle-for-cognitive-cybersecurity

[27] Johnson K, Lawrence A. AI/ML in Security Orchestration, Automation and Response: Future Research Directions. In: Intelligent Automation & Soft Computing, vol. 28, 2021 [cited 2021 Apr 7]. p. 534-537. Available from: https://doi.org/10.32604/iasc.2021.016240

[28] Ucci D, Aniello L, Baldoni R. Survey of machine learning techniques for malware analysis. In: Computers & Security, vol. 81, 2019 [cited 2021 Apr 8]. p. 126-128. Available from: https://doi.org/10.1016/j.cose.2018.11.001

[29] Tolido R, van der Linden G, Delabarre L, Theisler J, Khemka Y, Thieullent A, Frank A, Buvat J, Cherian S. Reinventing Cybersecurity with Artificial Intelligence. [place unknown]: Capgemini Research Institute, 2019 [cited 2021 Apr 8]. p. 14-15. Available from: https://www.capgemini.com/wp-content/uploads/2019/07/AI-in-Cybersecurity_Report_20190711_V06.pdf

[30] Zion Market Research. Artificial Intelligence (AI) In Cyber Security Market Will Reach to USD 30.9 Billion By 2025 [Internet]. New York: Zion Market Research, 2019 [cited 2021 Apr 11]. Available from: https://www.zionmarketresearch.com/news/artificial-intelligence-in-cyber-security-market

[31] Waltermire D, Quinn S, Booth H, Scarfone K, Prisaca D. The Technical Specification for the Security Content Automation Protocol (SCAP). Gaithersburg, USA: National Institute of Standards and Technology, 2018 [cited 2021 Apr 14]. p. 1, 4. Available from: https://doi.org/10.6028/NIST.SP.800-126r3

[32] Waltermire D, Schmidt C, Scarfone K, Ziring N. Specification for the Extensible Configuration Checklist Description Format (XCCDF) Version 1.2. Gaithersburg, USA: National Institute of Standards and Technology, 2012 [cited 2021 Apr 14]. p. 5. Available from: https://csrc.nist.gov/CSRC/media/Publications/nistir/7275/rev-4/final/documents/nistir-7275r4_updated-march-2012_clean.pdf

[33] MITRE Corporation. OVAL Frequently Asked Questions [Internet]. McLean, USA: MITRE Corporation, 2015 [cited 2021 Apr 14]. Available from: https://oval.mitre.org/about/faqs.html#a1

[34] Waltermire D, Scarfone K, Casipe M. Specification for the Open Checklist Interactive Language (OCIL) Version 2.0. Gaithersburg, USA: National Institute of Standards and Technology, 2011 [cited 2021 Apr 14]. p. 1. Available from: https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7692.pdf

[35] Halbardier A, Waltermire D, Johnson M, Specification for the Asset Reporting Format 1.1. Gaithersburg, USA: National Institute of Standards and Technology, 2011 [cited 2021 Apr 14]. p. 1. Available from: https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7694.pdf

[36] Wunder J, Halbardier A, Waltermire D. Specification for Asset Identification 1.1. Gaithersburg, USA: National Institute of Standards and Technology, 2011 [cited 2021 Apr 14]. p. 1. Available from: https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7693.pdf

[37] Cichonski P, Waltermire D, Scarfone K. Common Platform Enumeration: Dictionary Specification Version 2.3. Gaithersburg, USA: National Institute of Standards and Technology, 2011 [cited 2021 Apr 14]. p. 1. Available from: https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7697.pdf

[38] National Institute of Standards and Technology. Software Identification (SWID) Tagging [Internet]. Gaithersburg, USA: National Institute of Standards and Technology, 2020 [cited 2021 Apr 14]. Available from: https://csrc.nist.gov/projects/software-identification-swid/guidelines

[39] MITRE Corporation. CVE Frequently Asked Questions [Internet]. McLean, USA: MITRE Corporation, 2015 [cited 2021 Apr 15]. Available from: https://cve.mitre.org/about/faqs.html#is_cve_another_vulnerability_database

[40] National Institute of Standards and Technology. Common Configuration Enumeration [Internet]. Gaithersburg, USA: National Institute of Standards and Technology, 2020 [cited 2021 Apr 15]. Available from: https://csrc.nist.gov/Projects/Security-Content-Automation-Protocol/Specifications/Common-Configuration-Enumeration-(CCE)

[41] National Institute of Standards and Technology. CVSS Vulnerability Metrics. [Internet]. Gaithersburg, USA: National Institute of Standards and Technology, 2021 [cited 2021 Apr 15]. Available from: https://nvd.nist.gov/vuln-metrics/cvss

[42] Scarfone K, Mell P. The Common Configuration Scoring System (CCSS): Metrics for Software Security Configuration Vulnerabilities. Gaithersburg, USA: National Institute of Standards and Technology, 2010 [cited 2021 Apr 15]. p. 1-3. Available from: https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7502.pdf

[43] Booth H, Halbardier A. Trust Model for Security Automation Data 1.0 (TMSAD). Gaithersburg, USA: National Institute of Standards and Technology, 2010 [cited 2021 Apr 15]. p. 1-2. Available from: https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7802.pdf

[44] Waltermire D, Firtzgerald-Mckay J. Transitioning to the Security Content Automation Protocol (SCAP) Version 2. Gaithersburg, USA: National Institute of Standards and Technology, 2018 [cited 2021 Apr 15]. p. 1-4. Available from: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.09102018.pdf