Author(s): Forsberg, Joonas; Frantti, Tapio

Title: Technical Performance Metrics of a Security Operations Center

Year: 2023

Version: Published version

Please cite the original version:

Forsberg, J., & Frantti, T. (2023). Technical Performance Metrics of a Security Operations Center. Computers and Security, 135, Article 103529. https://doi.org/10.1016/j.cose.2023.103529

Computers & Security 135 (2023) 103529

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Technical performance metrics of a security operations center

Joonas Forsberg, Tapio Frantti *

University of Jyväskylä, Seminaarinkatu 15, FI-40014, Jyväskylä, Finland

ARTICLE INFO

Keywords:
Security operations center
Soc
Csoc
Cyber security operations center
Metric
Measurement
Technical performance
Cyber defense
Performance indicator

ABSTRACT

This research introduces a novel framework for creating metrics intended for security operations centers (SOCs). The framework is developed using the design science research methodology and has been validated by generating four novel metrics to assess the technical performance of a SOC. Additionally, the study examines the existing landscape of metrics for SOCs and concludes that a majority of the metrics discussed in the literature primarily focus on operational aspects rather than technical performance. The absence of adequate technical performance metrics makes it challenging to accurately evaluate the tangible impact of a SOC on overall cyber defense capabilities. The research also highlights the insufficiency of current methods in constructing metrics and frameworks tailored for measuring SOCs' technical performance. The resulting framework offers SOCs means to create high-quality metrics for performance evaluation. Furthermore, the metrics the framework was validated with offer SOCs an opportunity to enhance their ability to quantify their threat detection capabilities.

1. Introduction

Cyber threats have evolved dramatically over the last few years. They have become more sophisticated and complex, and as a result, they have a greater impact on operational activities. The use of general-purpose malware has declined recently and more advanced threats, such as supply chain compromises, extortion activities, misinformation campaigns, and business e-mail compromises, are continuously increasing (European Union Agency for Cybersecurity, 2021). The increased sophistication and the associated risks require organizations to improve their defensive capabilities for combatting the evolving threat landscape because traditional malware and network defenses are not enough to protect the organization from modern-day cyber threats.

A strategy for organizations to increase their cyber defense capabilities is to acquire security operations center (SOC) capabilities. On a general level, the SOC operations can be produced in-house, outsourced to a managed service provider, or operated in a hybrid model in which the internal SOC is complemented by an external SOC provider. According to Nathans (2014), a SOC is typically responsible for detecting security incidents and initiating related incident response activities. Nathans also determined that depending on the size and the needs of the organization, the SOC can consist of a single person or a larger team working in 24/7 shifts that together form a coherent collection of different competencies to prevent, detect, and resolve cyber threats. Furthermore, Vielberth et al. (2020) mentioned there are several ways to operate and design SOCs, and that the outcome is influenced by various factors, such as regulations, company strategy, and expertise.

To quantify the operational efficiency and capabilities of a SOC, performance should be measured using a set of commonly agreed metrics. A literature survey performed by Vielberth et al. (2020) concluded that the currently established metrics are insufficient for measuring the performance of the SOC. This viewpoint is supported by Kokulu et al. (2019), Agyepong et al. (2020) and Sundaramurthy et al. (2015).

Multiple publications on SOCs focus primarily on quantitative volume and time-based operational and vulnerability-related metrics to measure the performance of SOCs (Nathans, 2014; Ahlm, 2021; Kokulu et al., 2019). The operational metrics can help managers measure and optimize human resource utilization in analysis. Other quantitative metrics, such as vulnerability-related data, can provide an overview of the overall exposure to known threats (Nathans, 2014). In practice, they are inefficient in measuring the capabilities of SOCs, because they cannot measure the effectiveness of the detection capabilities and other protective controls. Measuring the false-positive rate of the monitoring rules is a common solution to combat this issue (Nathans, 2014; Ahlm, 2021). However, considering the false-positive rate alone can be misleading. The outcome of such metrics can be manipulated either subconsciously or consciously by the people responsible for the creation of the detection capabilities. Several studies have been performed on general security metrics (Böhme, 2010; Pendleton et al., 2016; Salmi, 2018), and there are also commonly referenced industry standards, such

* Corresponding author.
E-mail addresses: joonas.forsberg@iki.fi (J. Forsberg), tapio.k.frantti@jyu.fi (T. Frantti).

https://doi.org/10.1016/j.cose.2023.103529
Received 2 June 2023; Received in revised form 28 August 2023; Accepted 3 October 2023
Available online 6 October 2023

as ISO/IEC 27004:2016 (2016) and NIST SP 800-55 (Chew et al., 2008), that organizations can use to measure the effectiveness of their information security.

An argument could be made that an industry standard framework for measuring SOCs does not exist at the moment. The lack of commonly accepted methods to measure the technical performance of a SOC is particularly prominent when the SOC is being outsourced to a third-party vendor. This often leads to a situation where the tendering process produces a suboptimal result, either by selecting the vendor with the lowest total service cost or the most convincing sales material. In Finland, a trend has been observed, where vendors are chosen on a proof of concept phase, in which an attack simulation is performed in an environment temporarily monitored by the vendor and afterward their threat detection capabilities are evaluated. The trend is essentially a manifestation of the difficulty in evaluating the technical capabilities of different vendors. Although it can provide useful insights when comparing vendors, the results are not truly comparable between vendors because the number of resources allocated to the proof of concept can significantly impact the final results and thus skew the vendor selection.

The *objective of this study was to determine the metrics and other performance indicators relevant for measuring the technical performance of a SOC*. Without properly constructed metrics, it is difficult to evaluate how substantial impact the SOC has on the overall cyber defense capabilities. A technical performance metric within this paper is defined as *a qualitative or quantitative indicator derived from one or more measurements resulting from the activities performed by the SOC, which describes how well the SOC can utilize technologies to prevent, identify, detect, and respond to cyber threats affecting the organization.* Compared to an operational performance metric, a technical performance metric attempts to quantify the defensive capabilities, rather than assessing the effectiveness of processes and people associated with the SOC. Understanding both the operational and technical efficiency of the SOC is required to successfully protect an organization from cyber threats and measure the effectiveness of the SOC.

The objective leads to the following *research questions*: (1) *What frameworks are available to measure the performance of a SOC?* (2) *What are the commonly mentioned key metrics used to measure a SOC?* (3) *Can the common metrics be used to measure the technical performance?* (4) *How can the metrics be improved to enhance the reporting capabilities of technical performance?* The expected outcome of the research is a novel framework that could contribute to an industry-standard way to measure the technical performance of a SOC and apply to most SOCs.

2. Research methodology

Design science research methodology is used as the primary research methodology in this research. The research follows the principles defined by Peffers et al. (2007), in which a design science artifact is generated as an outcome of an iterative process with the following activities: (1) *Identify the problem and motivation*, (2) *Define the objectives of a solution*, (3) *Design and development*, (4) *Demonstration*, (5) *Evaluation* and (6) *Communication*.

The first activity is covered in sections 1, 3 and 4 of this research, the second in section 5, the third in section 6, the fourth in section 7 and the fifth in sections 8 and 9. This research as a whole covers the sixth activity. The second activity is supported by a brief literature review of both academic publications and commercial sources to establish sufficient theoretical background on the subject, and to answer the research questions (1) and (2). A keyword search is used as the search method for the discovery of literature related to SOCs and metrics. The search query for SOC-related literature is *'(security OR cyber security) (operations OR operation) (center OR centre)'* and the metric-related search query is *'(metric OR measurement OR "performance indicator")'*. The literature review was performed between May and September in the year 2022 and updated in August 2023. JYKDOK,[1] IEEE Xplore[2] and Google Scholar[3] were used as the search engines.

The research is expected to result in a design science artifact, a framework that can be utilized to create metrics to measure the technical performance of SOCs. The design science artifact is validated following the guidelines defined by Hevner et al. (2004), by following the experimental evaluation method, which consists of controlled experiments and simulations. The artifact is evaluated by utilizing it to create metrics to measure the technical performance of a SOC. The resulting metrics are evaluated with simulated data to ensure they behave as expected and produce meaningful insights for SOCs. The metrics are further evaluated in a controlled environment to ensure the measurements required for the metrics can be collected in a fully operational SOC.

3. Security operations center

To tackle the first phase of the design science research, *problem identification*, we first introduce, how technical performance metrics target environments, *i.e.*, SOCs, are described in the literature. One method to describe a SOC is through a People, Processes, and Technologies (PPT) framework (Knerler et al., 2022; Vielberth et al., 2020). Vielberth et al. (2020) summarized that the people block describes the people associated with the SOC and their required competencies, the process block describes how the people interact and how security incidents are handled, and the technology block describes the tools used to do the work. They also argued that the PPT framework could be expanded to include governance and compliance, enabling organizations to utilize the SOC as a function to ensure compliance with various standards, such as ISO/IEC 27001, GDPR, or PCI-DSS.

Vielberth et al. (2020) stated, that the work performed at a SOC is heavily driven by processes, as the work is structured around the prevention, detection, and response of security incidents. The Computer Security Incident Handling Guide (NIST SP 800-61, Cichonski et al. (2012)) describes the security incident management process, which consists of four phases "preparation," "detection and analysis," "containment, eradication, and recovery," and "post-incident activity," as depicted in Fig. 1. The purpose of the preparation phase is to ensure the SOC has the necessary visibility and potential to detect and respond to security incidents. The preparation stage also includes activities aiming to prevent security incidents altogether, such as malware prevention or user awareness training. The detection and analysis phase consists of responding to security incidents by analyzing, documenting, and prioritizing them and finally notifying the necessary people. The third phase is about containing the security incident by limiting the potential damage the incident may cause, followed by evidence gathering and identifying the attackers' host. Once the impact of the incident has been limited and the source of the attack has been identified, the threat can be eradicated, and recovery actions can be started (Cichonski et al., 2012). The final stage, post-incident activity, does not contain activities for which the SOC would typically be responsible. Other processes in the SOC can, for example, include data collection and creation, validation, and tuning monitoring rules (Knerler et al., 2022).

A SOC can be further split into multiple collections of tightly interlinked functional areas, which according to Knerler et al. (2022), are the following:

1. Incident triage, analysis, and response
2. Cyber threat intelligence, hunting, and analytics
3. Expanded SOC operations
4. Vulnerability management

[1] https://jyu.finna.fi/.
[2] https://ieeexplore.ieee.org/.
[3] https://scholar.google.com/.

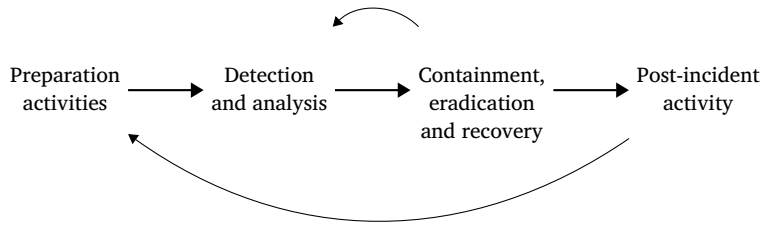

Fig. 1. Incident response life-cycle as depicted by Cichonski et al. (2012, p. 21).

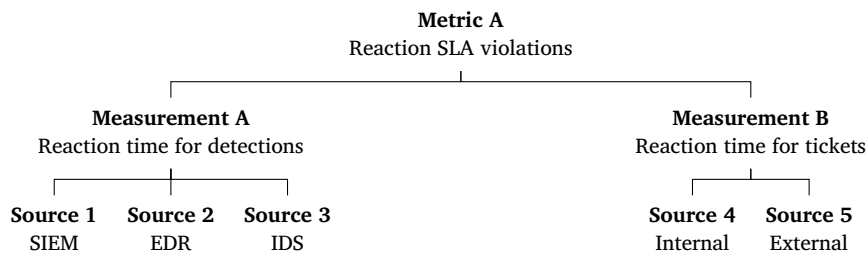

Fig. 2. Structure of a metric.

5. SOC tools, architecture, and engineering
6. Situational awareness, communications, and training
7. Leadership and management

Larger SOCs may have elements from all of the functional areas above. However, a single operations center does not necessarily need to cover all of the areas to be a functional part of the overall cyber defense capabilities (Knerler et al., 2022).

4. Metrics and measurements

The primary purpose of a metric is to measure how well a business process, product, or resource performs so that a business decision can be made Savola (2007). Typically, a metric consists of one or more discrete point-in-time measurements, from which metrics are then derived. Savola (2013) continues that to produce quality metrics for security, the metrics must conform to four fundamental quality criteria: correctness, measurability, meaningfulness, and usability.

4.1. Constructing metrics

Metrics can be structured in multiple ways. A metric can simply be a direct relation to a measurement, such as the number of security incidents, or it can be a composite metric constructed by multiple measurements collected from multiple sources. An example of a composite metric is depicted in Fig. 2, in which the metric is constructed from multiple measurements. Measurement A collects the detections generated by security tooling, such as Security Information Event Management (SIEM), Endpoint Detection and Response (EDR), or Intrusion Detection System (IDS). The detections generated by the tools are analyzed by security analysts. Measurement B consists of reports from end-users and other third parties about possible security incidents, such as suspected phishing or security policy violations. Similar to measurement A, the tickets in measurement B are investigated by the analysts. The metric configuration selects the measurements, for which the reaction time by the analysts exceeds the pre-determined value. The reaction time could for example be defined in a Service Level Agreement (SLA) between a SOC and other stakeholders.

There are several ways to construct or select metrics that measure the outcome of an activity. A study by Doran (1981) describes a method that utilizes the S.M.A.R.T. method, which is an acronym for specific, measurable, assignable, realistic, and time-related, to select a suitable metric. Brotby and Hinson (2013) introduces the PRAGMATIC method that consists of nine meta metrics (predictive, relevant, actionable, genuine, meaningful, accurate, timely, independent, and cheap), which are essentially scoring criteria for the metric itself.

The S.M.A.R.T. and PRAGMATIC methods are more about selecting metrics and which principles should be embraced when a metric is constructed. There are also other methods to construct metrics. For example, the Annex A of the ISO/IEC 27004:2016 (2016) standard summarizes a measurement information model contained in the ISO/IEC 15939 standard and describes how specific attributes related to an entity can be converted into an information product that can be used for conducting business decisions.

4.2. Problems and pitfalls of metrics

Metrics can provide valuable insights into the performance of an organization. However, when incorrectly constructed, they can become counterproductive. The metric can be incorrect due to a problem with the raw measurements, a programming error in the algorithm used to derive the metric, or someone is either maliciously or non-maliciously misinterpreting the metric (Brotby and Hinson, 2013). Hauser and Katz (1998) argues that it might also be possible that certain metrics, which are hard to influence by the activities of team members, might lead to a situation where short-term decisions are favored over long-term decisions. A metric may also be precisely wrong, meaning that something is measured with high accuracy, but the metric does not improve the business process it is supposed to improve (Hauser and Katz, 1998).

We can also approach the problems with metrics from the quality perspective. An attribute closely related to correctness is unbiasedness and objectivity, meaning the interpretation of the metric should not be influenced by personal beliefs or biases (Savola, 2007). Similarly, reproducibility is closely related to measurability. If a result from a metric cannot be reliably reproduced, the metric may produce an incorrect value.

There may also be a problem with the quantity of the metrics. If the metrics are incomplete, it is impossible to understand the overall situation. However, having too many metrics can also be a source of incorrect decisions (Brotby and Hinson, 2013).

4.3. Security operations center metrics

Vielberth et al. (2020) concluded that the general level of governance and compliance-related aspects of SOC-related research are

immature. There is significant research on security metrics, but the metrics are mostly operational and non-technical. However, they identified some technical metrics, such as false-positive rate, mean time to detect, threat actor attribution, and defensive efficiency. Nonetheless, they failed to show meaningful metrics in several areas, such as automation & orchestration, threat hunting, and detection engineering & validation.

Nathans (2014) discussed SOC-related qualitative and quantitative metrics utilized in managing a SOC. He also discussed the importance of vulnerability-related information in detecting potential security incidents by SOCs. The metrics related to the vulnerabilities are categorized as management-related metrics. He presented qualitative metrics, such as the top 10 vulnerable endpoints, and quantitative metrics, including the number of vulnerable endpoints, the number of vulnerabilities per severity, the number of unknown assets, and the time it took to apply a patch that fixed the vulnerability. However, the proposed metrics were not sufficient to be used to demonstrate the technical capabilities of the SOC.

Agyepong et al. (2020) presented a framework in which a SOC was split into multiple functions. Each function was measured separately to determine the actual performance of it. These functions included monitoring and detection, analysis, response and reporting, intelligence, baseline and vulnerability, and policies and signature management. The framework proposed that each function should monitor its performance in quantitative and qualitative metrics. The framework did not provide concrete measurement mechanisms upon which organizations can implement the metrics defined in the framework.

Keltanen (2019) used results from a customer survey to measure the performance of an outsourced SOC. He ranked the metrics based on the PRAGMATIC method. The resulting score will help evaluate different metrics between one another, making it possible to determine which metric is considered to be the most important (Keltanen, 2019). The study did not present concrete metrics that SOCs could use to measure their performance, instead focused on how the metrics could be constructed.

Kokulu et al. (2019) presented a qualitative study on the issues observed by SOC practitioners. The study was based on interviewing eighteen persons working in a SOC. The interviewed analysts mentioned that the metrics selected for measuring the performance were used to measure completely irrelevant things and to demonstrate a false improvement to the upper management rather than the actual performance of the SOC.

Onwubiko (2015) presented a framework consisting of metrics used to evaluate SOCs' performance and determine the return on investment. The framework does not contain a concrete set of metrics, but instead, it provides the top five examples that should be considered, which are the number of incidents, the performance of the cyber operations, the top ten cyber attacks, a summary of policy violations and a summary of privileged user misuse detections.

Alahmadi et al. (2022) present a comprehensive qualitative study of SOC analysts' perspectives on security alarms. The authors state that security alarm validation is a tedious task that can cause alarm burnout and eventually desensitization. They continue that vendors and researchers must be able to make distinctions between types of false-positive alarms. They also argue that researchers have used false-positive as a metric for evaluating system performance when proposing security tools, seeking few false-positives for optimal performance.

In addition to scientific literature, there is a plethora of commercial material that discusses the metrics to measure SOCs. Zimmerman and Crowley (2019) presented seven high-level groups for the metrics. The groups are the health of the data feed, coverage of the monitoring, vulnerability-related metrics, monitoring rules, analysts' performance, incident handling, risk priorities, and general hygiene.

Gartner published a guide (Ahlm, 2021) discussing the industry's best practices for building and operating a modern SOC, among which a set of fundamental metrics were defined. The metrics mainly focused on quantitative metrics and were grouped into four categories: incident volume, incident detection, incident response, and incident impact. The metrics in these categories, for most parts, are not relevant to the technical performance of the SOC.

SANS Institute (Crowley and Pescatore, 2019) surveyed 355 organizations about common and best practices for SOCs. There were questions about which metrics are used in the SOC or measure its performance. The results of the survey conclude that quantitative metrics such as the number of incidents handled and time from detection to containment to eradication are the most common. The three least commonly deployed metrics were "Losses accrued vs. losses prevented," "Monetary cost per incident," and "Avoidability of the incident," all of which are relatively difficult to implement. Out of the metrics mentioned, there were a few metrics that are relevant for measuring the technical performance of the SOC, which are "Threat actor attribution," "Thoroughness of eradication," and "Thoroughness and accuracy of enterprise sweeping."

Logsign (2020) groups recommended metrics into two categories, metrics for security operations and metrics related to business requirements. The number of security incidents was raised as the most important metric for security operations. It also depicted metrics, such as the number of alerts per analyst, the number of alerts closed by automation, the number of false-positive alerts, and the average time to detect a security incident.

A blog post by Simos and Dellinger (2019) presented some of the key metrics used at the Microsoft SOC. These include time to acknowledge, time to remediate, escalation between tiers, and the number of incidents remediated grouped per response type (manual or automated). The metrics are mostly about measuring the response capabilities of the SOC and are highly relevant to the technical performance of the SOC.

In addition to SOC-specific frameworks and methodologies, several publications describe ways of measuring information security on a general level. The Performance Measurement Guide for Information Security (NIST SP 800-55) describes how the organization can create, select and implement metrics for monitoring the state of the security program on an overall level (Chew et al., 2008). The ISO/IEC 27004:2016 standard describes guidelines that can be utilized for measuring the effectiveness of the information security management system (ISO/IEC 27004:2016, 2016).

Salmi (2018) surveyed information security metrics implemented in large Finnish corporations. He identified 28 security metrics categorized into either management, operational, or technical metrics. The study contained no metrics that directly relate to the technical performance of a SOC. However, several metrics are closely related to the typical activities of the SOC, such as the business impact of security incidents, characteristics of security incidents, and system vulnerabilities.

As a summary of SOC metrics, Table 1 describes the top thirty most commonly mentioned metrics in the literature, *i.e.*, the number asterisks depict how many articles the metric was mentioned in. A full list of the metrics observed has been published in Mendeley Data (Forsberg, 2023). In Table 1, one of the metrics was combined into a generalized term instead of having two separate rows for "time to resolution" and "time to incident closure." Another general level observation was that the terminology does not appear to be consistent in the literature. For example, the metric "mean time to detect" is used to describe at least two different behaviors, the time it takes to react and perform the analysis of the alert (Ahlm, 2021; Agyepong et al., 2020; Crowley and Pescatore, 2019) and time it takes for a SOC to become aware of the incident (Logsign, 2020; Vielberth et al., 2020). Some of the publications did not provide enough information to clearly distinguish between the two (Kokulu et al., 2019; Zimmerman and Crowley, 2019). However, Vielberth et al. (2020) provided a separate metric for the average analysis time and detection time. Thus a conclusion could be reached that the correct definition of time to detect metric would be the time between the initial activity of the adversary and the first detection caused by the activities. This viewpoint is also supported by the Computer Security Incident Handling Guide (NIST SP 800-61, Cichonski et al. (2012)).

Table 1
Top 30 metrics in the literature.

	Chew et al. (2008)	ISO/IEC 27004:2016 (2016)	Salmi (2018)	Vielberth et al. (2020)	Nathans (2014)	Keltanen (2019)	Agyepong et al. (2020)	Kokulu et al. (2019)	Onwubiko (2015)	Alahmadi et al. (2022)	Crowley and Pescatore (2019)	Ahlm (2021)	Logsign (2020)	Simos and Dellinger (2019)	Zimmerman and Crowley (2019)
Number of security incidents	*	*	*	*	*		*	*	*		*	*	*		
Mean time to reaction						*	*	*				*	*	*	*
Number of vulnerabilities	*	*	*	*	*			*							*
False-positive rate				*	*	*			*	*		*	*		
Mean time to detect				*			*	*				*	*		*
Mean time to resolution			*			*		*			*	*			*
Cost of security incidents		*		*		*					*	*			
Detections per category		*			*							*	*		
Mean time to vulnerability remedy				*	*			*						*	
Number of vulnerable devices			*	*	*										*
Percentage of employees trained	*	*	*												
Percentage of standard systems	*	*	*												
Analyst productivity				*									*		*
Coverage of vulnerability scanning		*	*												*
Downtime due to security incidents			*								*	*			
Incident avoidability					*						*				*
Incidents with business-impact				*								*	*		
Mean time to containment								*			*				*
Mean time to triage				*	*										*
Number of incidents per shift				*			*				*				
Number of monitored assets			*	*											*
Number of patched vulnerabilities	*		*												*
Number of risk per severity		*	*	*											
Resolution SLA breaches		*	*		*										
Severity of security incidents				*				*				*			
Threat actor attribution				*							*				*
Number of automated incidents													*	*	
Mean time to escalation								*							*
Quality of eradication											*				*
Reaction SLA breaches	*														*

The most common metric was the total number of security incidents, which was mentioned in eleven of the fifteen publications included in Table 1. This is not a surprise, given the number of incidents is mentioned as a specific metric in both NIST SP 800-53 and ISO/IEC 27004. It is a metric that can easily be collected and used outside of the scope of SOC. Vulnerability-related metrics are also relatively common, with the count of vulnerabilities being mentioned in seven publications. The mean time to vulnerability remediation and the count of vulnerable devices were mentioned in four publications. Operative metrics, such as mean time to reaction, detection, resolution, containment, triage, and escalation, were also commonly mentioned. The publications also contained a few technical performance metrics. They were false-positive rate, threat actor attribution, the number of security incidents closed with automation, and the quality of eradication.

The literature review emphasizes the need for a common framework that could be used for the performance evaluation of SOCs globally. The metrics presented in Table 1 are scattered broadly, and there appears to be no common group of key metrics used to measure the technical performance of a SOC. The published literature mostly focuses on operational SOC or general security metrics. Although commercial whitepapers provide slightly better technical performance metrics, they fall short in several ways. For example, the lack of proper justification is seen throughout them. Despite the limited scientific research on the subject, several studies have reached a similar conclusion (Agyepong et al., 2020; Keltanen, 2019; Kokulu et al., 2019; Vielberth et al., 2020). The lack of standard technical performance metrics could be attributed to the lack of a sufficiently mature governance model for SOCs, as pointed out by Vielberth et al. (2020).

4.4. Other work on measuring SOCs

In addition to the literature mentioned earlier, there are several publications on the ways to measure SOCs. For example, Jacobs et al. (2013), Van Os (2016) and Schlette et al. (2021) are proposing to measure SOCs based on a capability maturity model, Schinagl et al. (2015) approaches the assessment through a set of questions to understand the perceived level of effectiveness and Rosso et al. (2022) introduces a way to measure SOCs performance by injecting simulated attacks into an operational SOC and utilize pre-existing metrics to determine the level of effectiveness.

All of these methods suffer from the same fundamental problem: they do not provide concrete metrics a SOC can be measured with. The method proposed by Rosso et al. (2022) succeeds in measuring the technical performance of SOCs, but relies on existing metrics to do so. Some studies discuss methodologies and provide concrete metrics to measure the SOC performance, such as the papers by Shah et al. (2018) and Agyepong et al. (2023), but as they focus on the operational aspects of the SOC, they do not provide ways to measure the technical performance of SOCs.

5. Solution objectives

The objective of the solution is the documentation of an approach for creating SOC-related metrics. Additionally, the solution should be used to create three to seven metrics that can either augment existing technical performance metrics or introduce completely new metrics and provide capabilities to measure the technical performance in an

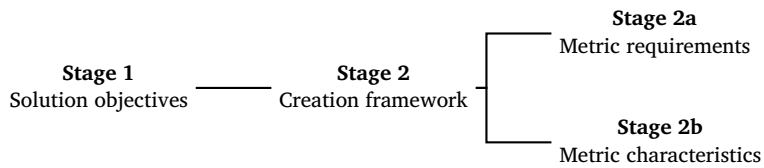

Fig. 3. The metric construction stages.

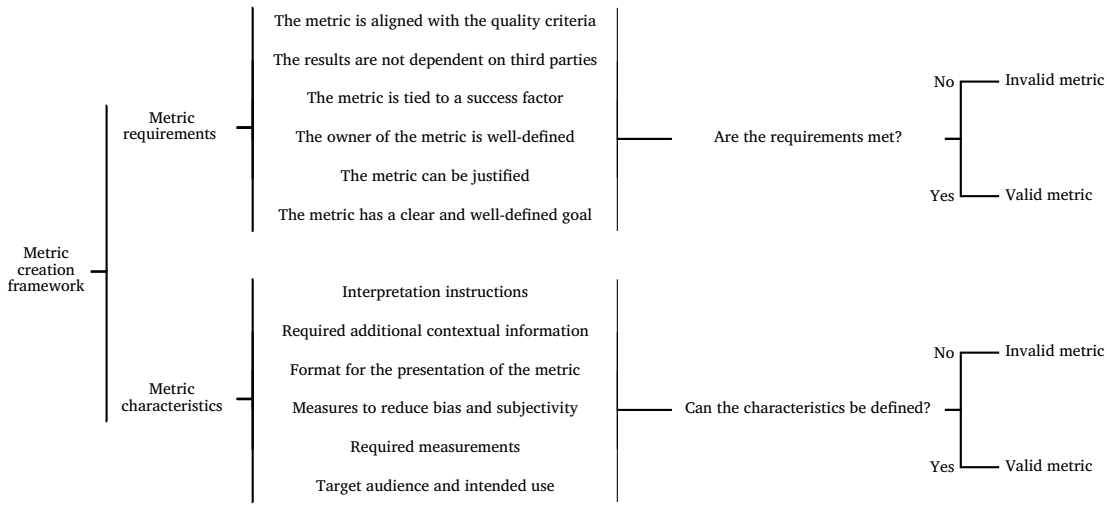

Fig. 4. The metric creation framework.

area where previous metrics are incomplete or missing. Based on the literature review, the requirements for the solution are the following:

1. A selection criteria for the creation of metrics should be well-defined.
2. A separate quality criteria for the metrics should be defined and the metrics created with the framework should conform with it.
3. The metrics should be directly associated with a specific SOC function.
4. The metrics should be universal and not tied to a specific technology or an organizational structure.
5. The metrics should be justified either by scientific research, industry standards, or through other means that considerably decrease the subjectivity of the metrics.

The justification for the objective 1 is that there does not appear to be an industry-standard framework for creating metrics for SOCs. There are selection criteria such as S.M.A.R.T. (Doran, 1981) and PRAGMATIC (Brotby and Hinson, 2013), but the criteria remain subjective at best. Creating a comprehensive framework for measuring the performance of a SOC is out of the scope of this research, and the focus is placed on enabling the creation of selection criteria that can be used as a part of a comprehensive framework.

The justification for the objective 2 is that low quality metrics can produce unwanted results, as discussed in section 4.2. Furthermore, based on the authors' first-hand experience, many of the metrics used for measuring SOC are of low quality, particularly regarding bias and objectivity. This viewpoint is partially supported by Kokulu et al. (2019). However, the quality aspect of the metrics was not discussed in detail in the SOC-related literature to form a valid conclusion about the insufficient quality. To ensure the resulting metrics are of high quality, quality criteria for SOC metrics should be defined.

The justification for the objectives 3 and 4 is the same, which is the universal applicability of the resulting metrics. In practice, this means the metrics should be usable by most SOCs out there, provided that they include the function to which the metric is tied. Finally, the justification for the objective 5 is that the metrics in some situations do not appear to be backed up by scientific research, particularly regarding commercial sources, which means the metrics are purely based on the authority of the author.

6. The metric creation framework

This section presents the creation of the design science artifact, *i.e.*, a metric creation framework. It can be used to construct metrics used to measure the performance of a SOC. The artifact is evaluated by utilizing it to create metrics to measure the technical performance of the SOC.

The metric construction model is based on requirements defined in two separate stages, as depicted in Fig. 3. The first stage consists of the solution objectives, and the second stage is the metric creation framework that consists of the requirements the metric should conform to and the characteristics that must be describable in the metric documentation, see Fig. 4.

A valid metric must fulfill both the metric requirements and the characteristics. The metric requirements are:

1. The metric has a clear and well-defined goal.
2. The owner of the metric is clear.
3. The results are not dependent on third parties.
4. The metric can be justified.
5. The metric is tied to a success factor.
6. The metric is aligned with the quality criteria.

A well-defined goal means the metric must be meaningful as per the quality criteria by Savola (2013) and the PRAGMATIC methodology by Brotby and Hinson (2013). The quality criteria within the context of a SOC are shown in Table 2.

The metric should also be assignable directly to a function within a SOC. If a third party influences the metric, the metric does not measure only the performance of the SOC but rather the entire chain related to threat detection and incident response.

The metrics should also be justifiable by scientific research, industry standards, or through other means documented in the metric descrip-

Table 2
The quality criteria for the SOC metrics.

Criteria	Characteristic	Description
Correctness	Granularity	The metric should provide the necessary granularity to tie the metric to a specific function or team within a SOC.
	Completeness	The metric should completely fulfill the goal defined in the metric documentation. If the metric cannot alone fill the goal, the metric should be coupled together with another metric.
	Objective and unbiased	The results should not be influenced by the activities performed by the person setting up the metric and the bias should be minimized to acceptable levels defined in the metric description.
Measurability	Availability	Measurements used to construct the metric must be automatically available in a reliable and consistent format.
	Reproducibility	The metric must be reproducible by different persons across multiple organizations given access to the same measurements.
Meaningfulness	Impactful	The metric must have an impact on the daily activities and it must be capable of showing the progression of development efforts.
	Clarity	The interpretation of the metric must be unambiguous and consistent across the entire lifecycle of the metric.
	Comparability	The result of the metric must aim to be comparable between multiple SOCs even between organizations.
Usability	Portability	The metric must be usable by multiple different SOCs and not be dependent on their size, structure, service model, or parent organization.
	Controllability	The team the metric is used to measure must be capable of keeping the metric value between the expected values.
	Scalability	The metric must be able to behave consistently with low and high volumes of measurements.
	Presentable	It must be possible to present the information the metric is expected to provide visually.

tion. If the metrics are justifiable, there is no need to rely on the authority of the source.

Establishing a connection between the indicator and the critical success factors (CSFs) is a fundamental requirement for key performance indicators (KPIs) as described in the book by Parmenter (2019). On a practical level, establishing a connection between the CSFs and the KPIs forces organizations to identify critical contributors to organizational performance and thus qualify to be measured with the KPIs. Finally, the metric should conform to the quality criteria, Table 2.

If the metric has passed the requirements, it can be constructed. The metric can have different characteristics depending on variables, such as which SOC function it relates to or the stakeholders the metric targets. The fundamental characteristics that should always be defined are the following:

1. Target audience and intended use
2. Measures to reduce bias and subjectivity
3. Required additional contextual information
4. Required measurements
5. Format for the presentation of the metric
6. Interpretation instructions

The target audience of the metrics and the intended use must be well defined because there is a mismatch between the evaluation metrics when it comes to SOC managers, SOC analysts, and other technical personnel (Kokulu et al., 2019).

If we can reduce bias to a minimum and ensure the metric is objective, we can fulfill the most important quality criterion, correctness.

There can also be situations where a counter-metric or other additional contextual information is needed. For example, a SOC may have a metric to measure the number of distinct monitoring rules, which could be used to measure the detection potential of the SOC. However, the high number of monitoring rules does not directly correlate with the performance of the SOC, since if a majority of the security incidents resulting from the monitoring rules are false-positive, the SOC is unlikely to be able to handle them effectively (Alahmadi et al., 2022). Therefore, the metric for the number of monitoring rules should be coupled with the false-positive rate to measure the effective detection potential.

As the metrics consist of measurements, the source for the measurement data must be defined along with the format of measurement, the measurement interval, and any other information that affects the measurements or metrics in any way. For example, a metric measuring the mean time to resolution would require each security incident to have two measurements, one to measure the time when the incident was opened and another one to measure the time when the incident has been resolved.

Metrics must also be presented in such a way that they clearly and consistently depict the information the metric is supposed to deliver. For example, the metric can be presented numerically, by various charts or time series graphs, or in a text format within a table. The way the metric is presented should provide the person interpreting the metric with the necessary information to make decisions based on the data seen.

Interpreting the metrics is important in ensuring the metrics provide valuable insights for the stakeholders. Although the fundamental idea is that the metrics themselves should be self-explanatory, in the practical sense, some metrics can be hard to interpret. The metric documentation should include the expected way to interpret the results.

7. Metrics for security operations center

This section describes how the metrics meet the objectives defined in the metric creation framework. Measurements used to construct the metrics have been programmatically generated and presented using open-source tools and Python modules, most prominently Jupyter Notebooks[4] and a Python graphing library Plotly.[5] Additionally, scikit-learn, which is a Python module that can be used for data analysis and machine learning algorithm development (Pedregosa et al., 2011), was used to calculate the statistical fitting of a trend line in the form of linear regression. The parameters for creating the data points are adjusted every 100 steps to create variation in the results over time. The source code for the metrics is published in Mendeley Data (Forsberg, 2023).

7.1. Distribution of detections among the UKC

Distribution of detections among the Unified Kill Chain (UKC) measures how effective the SOC is in detecting threats in the early stages of the UKC and thus decreases the impact of a security incident. The metric is tied to a function responsible for custom analytics and detection creation, as per the definition by Knerler et al. (2022). If the SOC does not have a team responsible for the function, the metric is not measuring the performance of a SOC but rather the tools the SOC is using. In that case, the metric depends on third parties and thus is not a valid metric for a SOC.

Fig. 5 depicts a visual representation of the metric with a detection strategy focused on the initial foothold stage. This means that a larger

[4] https://jupyter.org/about.
[5] https://github.com/plotly/plotly.py.

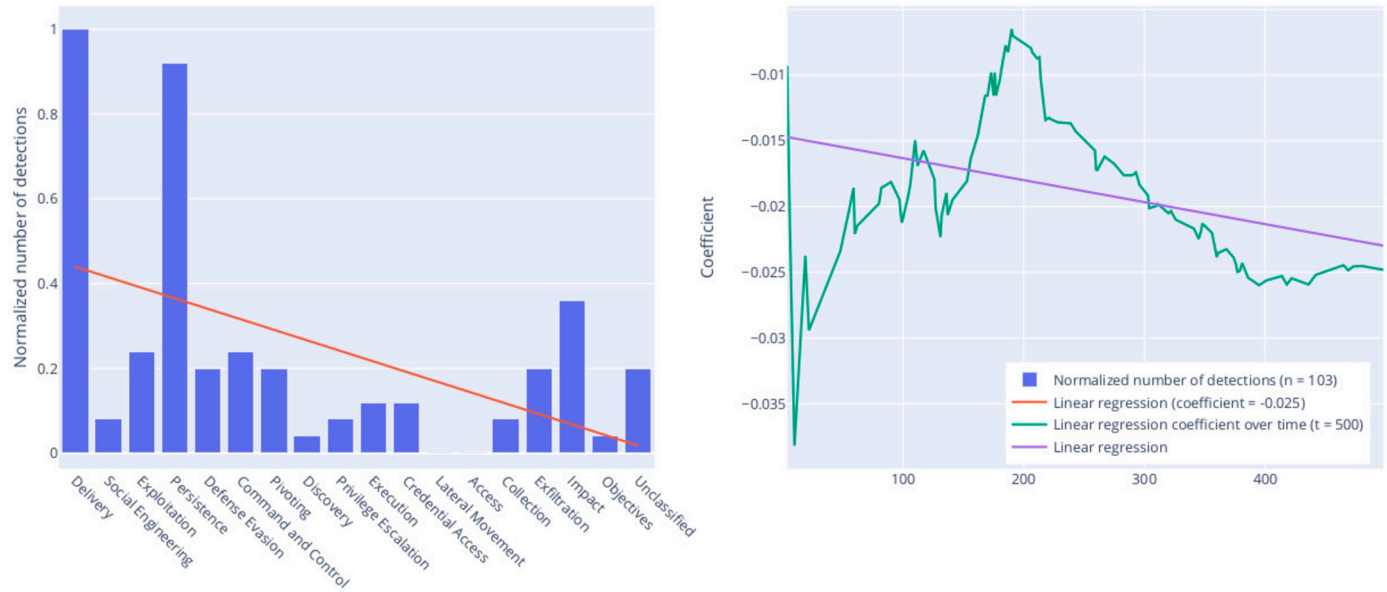

Fig. 5. Initial Foothold focused detection strategy. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

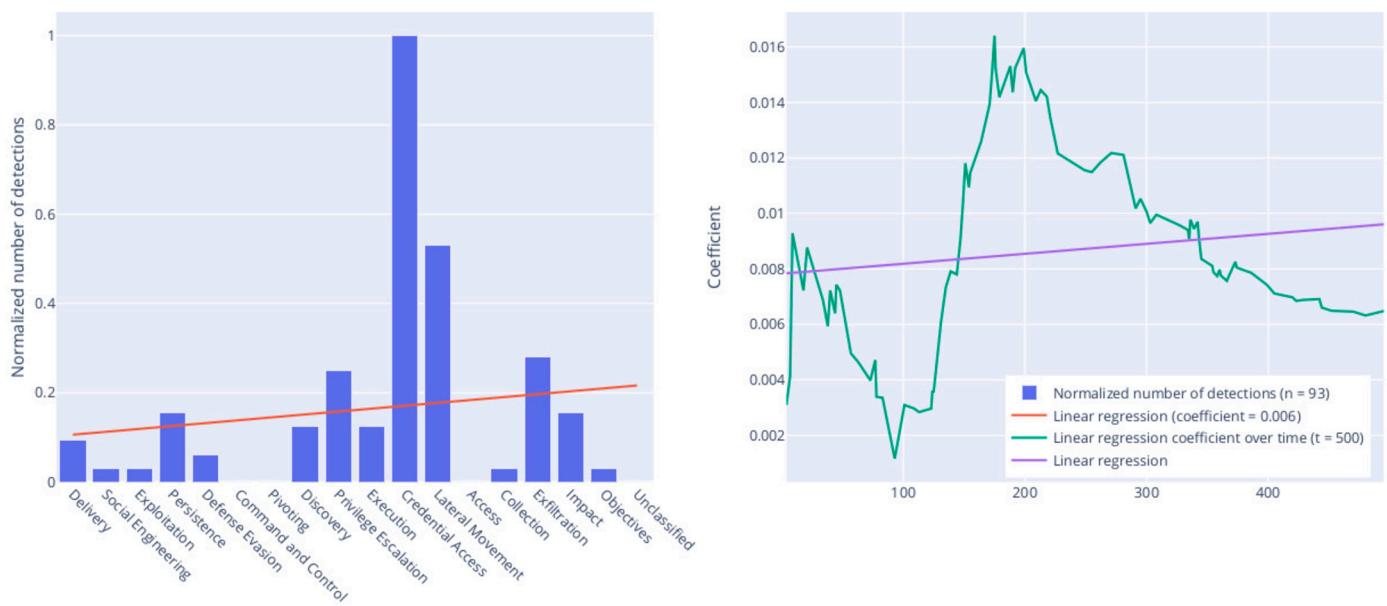

Fig. 6. Network Propagation focused detection strategy.

portion of the detections is within the initial foothold stage, compared to network propagation or the action on objectives stages. Fig. 6 depicts the metric with a detection distribution focused on network propagation. Both strategies statistically have a similar amount of detections in the final phase of the kill chain. Thus, both are equally effective in preventing adversaries from achieving their objectives. However, the strategy focusing on the initial foothold stage is better in terms of the metric, as it is more effective at preventing the incidents from traversing forward within the UKC.

Fig. 5 also shows that the value of the linear regression coefficient decreases over time, meaning the detection strategy shows signs of improvement. By contrast, the trend is increasing in Fig. 6, showing signs of deterioration. Both figures display a large variation at the beginning of the metric, which means the metric is not accurate until there is a sufficient quantity of true-positive detections, but otherwise, the metric conforms with the requirements defined by the framework.

7.2. Number of verifiable monitoring rules

The metric *Number of verifiable monitoring rules* measures the share of monitoring rules that can be automatically or manually verified by executing actions that trigger the rules. The goal of the metric is to improve the verification rate of the monitoring rules, encourage detection engineers, who are responsible for the creation of the detection capabilities, to better utilize threat intelligence as a part of their daily routines, and provide a means for a SOC to demonstrate what attacks they can detect. The metric is tied to a function responsible for custom analytics and detection creation, and it is valid only if the SOC has a team responsible for it.

The metric output is not affected by external conditions and meets the clarity and scalability characteristics of the quality criteria. The metric is usable by multiple SOCs, although the interpretation can vary slightly between SOCs. The team responsible for threat detection can

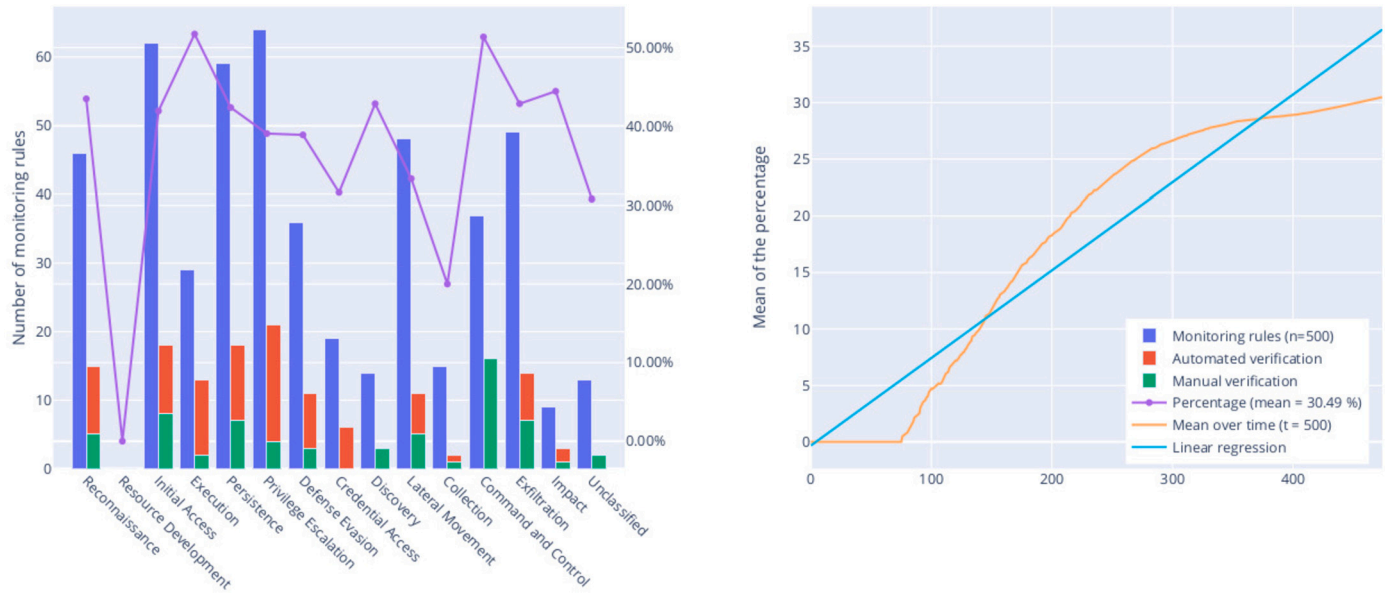

Fig. 7. Number of verifiable monitoring rules.

control the outcome of the metric, and it can be presented in terms of the current stage and historical progress, as shown in Fig. 7.

Fig. 7 depicts two ways to present the metric, one as a combined bar and line chart and another as a line chart, depicting the evolution of the metric over time. The bar chart displays the number of distinct monitoring rules and the sum of verifiable monitoring rules per MITRE ATT&CK tactic. Next to the bars is a trend line displaying the percentage of verifiable monitoring rules. The value of the metric is the mean of the percentage of the monitoring rules covered per MITRE ATT&CK tactic at a given time. The intended way to interpret the metric is to follow the trend of the mean to determine whether the SOC is improving over time.

For the metric to provide meaningful information, it needs a counter-metric that displays the number of distinct monitoring rules, as shown in the blue bars in Fig. 7. If the SOC has a low number of monitoring rules, the metric value can be abnormally high, increasing bias and decreasing the comparability of results between different SOCs. The number of monitoring rules can also be misleading, as there is a possibility of creating many low-fidelity monitoring rules.

As the metric measures only the number of verifiable monitoring rules resulting from the SOC development efforts, it fails to demonstrate the verification capabilities of native vendor detections. This makes the metric subjective, as some SOCs are likely to rely more on native capabilities for threat detection than others, making the metric inefficient for comparing metric values from one SOC to one another, even if their operative models would be similar. A dimension that displays the number of vendor-native detections that can be tested could be added to the metric. However, doing so would increase the bias of the metric. A better solution would be to create a separate metric for measuring the number of verifiable vendor-native scenarios and construct another metric that considers both metrics. The metric is not considered to be valid in terms of the metric creation framework, as the metric is highly subjective and overly biased.

7.3. Distribution of detections by source

Distribution of detections by source metric determines to what extent the development efforts of a SOC can contribute to the detection of security incidents. On a practical level, if a large portion of detections originates from the native capabilities of the technologies in use, the detection engineering function may not be able to provide additional value in the form of new monitoring rules. The metric can be tied to a specific function, which is responsible for creating custom analytics within the SOC. The metric results are somewhat dependent on third parties, as the technologies selected to protect the environment have an impact on the metric results. However, as the fundamental purpose is to compare the custom capabilities against the native capabilities, the metric is not dependent on third parties but rather influenced by them, making the metric pass the requirement.

Knerler et al. (2022) state that custom analytics and custom capability development are the functional areas of the SOC. Hence, it can be argued that the capability to augment the detection capabilities provided natively by the technologies is a success factor for the SOC. Overall, SOC-related literature does not succeed well in defining why custom analytics should be created in the first place. For example, Knerler et al. (2022); Ahlm (2021); Van Os (2016); Vielberth et al. (2020) mentioned the creation of monitoring rules as a fundamental part of SOC capabilities. However, none are discussing in detail whether the creation of monitoring rules is something the SOC should focus on or not. Moreover, as the literature appears to agree that the creation of monitoring rules is something SOCs should be doing, it acts as a justification for the metric. The metric can also be justified by displaying whether it makes sense to invest in the development of custom capabilities or not.

With MITRE ATT&CK tactic as the primary dimension for the metric, see Fig. 8, the metric provides a method for the detection engineering function to align and focus their development efforts on specific tactics. Furthermore, SOC management can also benefit from the metric, as the value of the metric over time can be used to determine the direction and the impact of changes made within the detection engineering team and subsequently demonstrate the value the detection engineering function provides.

Certain limitations must be imposed on the metric to reduce subjectivity and bias. Within the context of this metric, detections originating from native capabilities are something that some technology, for example, SIEM, EDR, or an IDS, has provided the first indication of compromise with out-of-the-box capabilities and without correlation to other data sources. This means that if the source of the detection is an alert from an IDS or anti-virus program, it constitutes a native detection. However, if the anti-virus alert correlates with other endpoint-related logs before generating a detection, it is a custom capability. All detections that have been contained automatically and do not require further actions from the SOC, such as an anti-virus program or an IDS preventing an infection or the delivery of a malicious payload, are out-of-scope

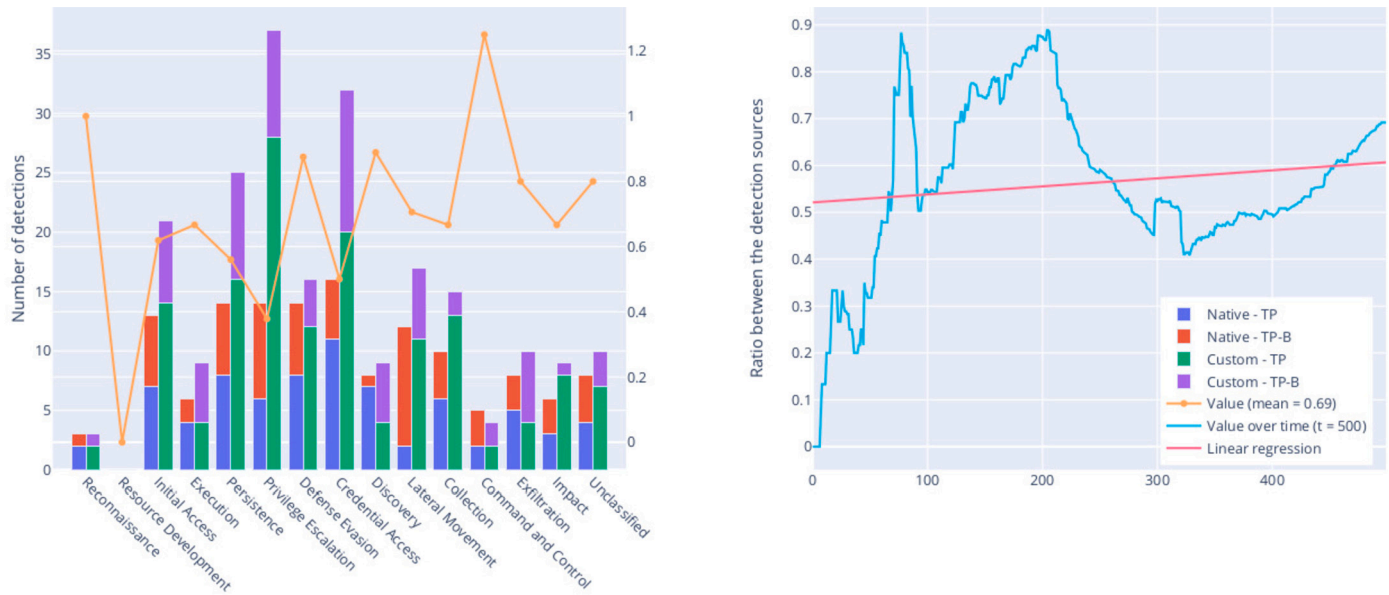

Fig. 8. Distribution of detections per source (TP = true-positive and TP-B = benign true-positive).

for this metric. Furthermore, all false-positive detections and detections not originating either from a technology or a monitoring rule, such as reports from end-users, are also removed from the measurements of the metric, as they are unnecessary for achieving the goal of the metric.

To construct the metric, the source for the detection, in addition to MITRE ATT&CK categorization, must be recorded for all true-positive and benign true-positive detections. One way to present the metric is seen in Fig. 8, which displays two charts. A bar chart displaying, for each MITRE ATT&CK tactic and an unclassified category, the current state of the metric in terms of the number of detections per category, and the ratio between native and custom detections. Additionally, the metric displays a line graph on a secondary y-axis depicting the score for each tactic and the overall value of the metric, which is the mean of the scores for each category. The line graph displays the evolution of the value of the metric over time, along with a linear regression depicting the trend of it.

Table 3
Grouping of activities per NPS category.

Category	Activity
Promoters	True-positive incident was escalated to third-party. Escalated incident was not returned to SOC for further investigation. Original priority was correct throughout the incident lifecycle. No unknown entities before escalation.
Passives	Benign true-positive incident was escalated. Escalated incident was returned to SOC for further investigation. Priority of the security incident was adjusted after the initial analysis. Unknown entities before escalation. The initial conclusion on the returned incident was correct.
Detractors	False-positive incident was escalated. False-negative detection. The initial conclusion on a returned incident was incorrect.

7.4. Technical accuracy of the analysis

The quality of the analysis work is subjective, and it is unlikely to be possible to measure with a general-purpose metric. However, measuring the accuracy of the analysis could provide some hints about the quality. One possible way to calculate the accuracy of the analysis is to adopt an approach similar to the calculation of the net-promoter score (NPS), which can be used to measure customer satisfaction (Reichheld, 2004). In NPS, customers score a service on a range from 0 to 10, where scores 0-6 are detractors, 7-8 are passive, and 9-10 are promoters. The NPS is calculated by subtracting the percentage of detractors from the percentage of promoters, providing a score between -100 and 100. The formula is $NPS = (\frac{P_1 - D}{P_1 + P_2 + D}) * 100$ where P_1 are number of promoters, P_2 are the number of passives and D are the number of detractors. Although in the academic sense, the NPS methodology has some issues for what it is being used for (Bendle et al., 2019), the model succeeds in producing a value of the relationship of discouraged (detractors), neutral (passives), and encouraged (promoters) activities, and as such, it is a valid and relatively simple approach to take. Identifying proper methods to measure the quality of the security analysis would be a valuable topic for additional research.

Table 3 summarizes the activities. Promoters are activities that should be encouraged to be performed continuously and are signs of a well-performing SOC, passives are expected from a SOC under normal operations and detractors are activities that the SOC should attempt to avoid, as they can harm the overall situation.

The metric can be justified by the idea that if the analysis is of low quality or the quality decreases over time, the SOC might not be able to combat the challenges produced by a modern-day adversary, or they might not have sufficient knowledge of the monitored environment. The same idea could be considered to be a success factor for the SOC. The metric is more intended for the management of the SOC as they are likely to be more interested in the overall situation rather than focus on specific metrics. The measurements required to construct the metric vary depending on the activities chosen for the metric.

The metric can be presented similarly to the other metrics; see Fig. 9. The bar chart depicts the count of activity occurrences for each category (promoters, passives, detractors) per MITRE ATT&CK tactic. If there are zero items, the tactic is omitted. The NPS is also displayed individually for each MITRE ATT&CK tactic to demonstrate the difference in the technical accuracy of the analysis between tactics. Due to the way the NPS is calculated, the actual value of the metric must be calculated from all occurrences of the activities, rather than taking the average of the individual NPS. The secondary graph in Fig. 9 depicts the evolution of the NPS over time and a linear regression that demonstrates the trend of the evolution of the metric. If the NPS is above 0, it means there are more promoters than detractors, and as such, the higher the score, the better the metric value is. The metric works as a standalone metric. It

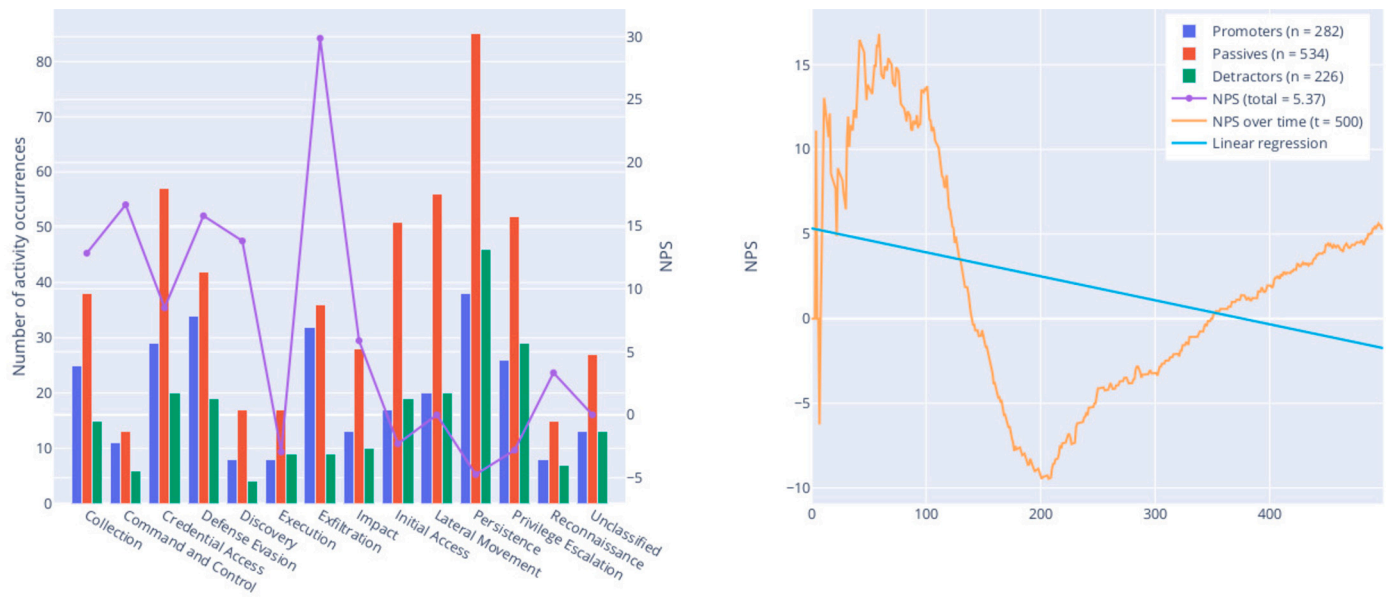

Fig. 9. Detection accuracy NPS.

cannot be used alone to compare different SOCs as the value is affected by the selection of the activities.

8. Model validation

Metrics are evaluated within a SOC that provides managed SOC services to large and medium enterprises. The metrics are judged by evaluation criteria: *"Can the measurements required to construct the metric be collected within the SOC used as the test subject?"*.

The demonstration of metrics with live data could also have been a valid evaluation metric for the metrics, but due to the metrics requiring historical data to provide information other than a snapshot of the current state, the evaluation would require a longer period to be properly evaluated. In this research, the metrics have been tested with simulated data during the creation process and the expectation is that the metrics behave similarly when real-world measurements are used as a source for the metrics.

Before implementing the metrics, it is necessary to understand which measurements are required to construct the metrics and whether they can be collected or not. Out of the metrics generated, the metrics that were determined to be valid were *distribution of detections among the unified kill chain*, *distribution of detections by source*, and *technical accuracy of the analysis*. The metric *number of verifiable monitoring rules* was considered to be invalid by the criteria defined by the metric creation framework. The following section briefly discusses the measurements for the metrics that are considered to be valid.

The SOC is utilizing a Security Orchestration, Automation, and Response (SOAR) platform to manage the security incident management workflows. In this particular SOC, the SOAR platform collects all detections into a single pane of glass in the form of cases, which unifies the analyst workflows across different environments and security products. The SOAR platform also produces a significant portion of the measurements needed to construct metrics used to measure the performance of the SOC.

Detections have MITRE ATT&CK tactic classification, which means that the MITRE ATT&CK tactic can be successfully utilized in multiple metrics. Furthermore, as the UKC is closely associated with the MITRE ATT&CK framework, the UKC stage can be partially derived from the MITRE ATT&CK tactic. The detection source is also available in the platform but is currently not extracted. All detections and resulting security incidents are classified on a high level either as false-positive, true-positive, or benign true-positive. All cases have a default priority, which depends on the priority of the associated detections, but the priority can change as the investigation process moves forward. The original and the final priority are also recorded in the SOAR platform.

A pivotal part of the daily SOC operations is the escalation of true-positive or benign true-positive detection as security incidents to various stakeholders, for example, a customer or a dedicated incident response team. The escalation status can be derived from the escalation target group, but it is not explicitly recorded in the SOAR platform.

As a part of the investigation workflows, the SOC is attempting to resolve unknown entities that are associated with the detection. For example, an IP address without a hostname associated is an unknown entity. Known entities are extracted to cases from the detections, but the unknown entities identified by the SOC are not recorded in any particular field other than free-text case notes. This causes the state of unknown entities to not be reliably available in the SOAR platform. It could be possible to construct a workflow that reminds analysts to record the unknown entities on the case. However, as there is a human reliance on the activity, if the data cannot be reliably and automatically collected, it can violate the availability characteristic of the quality criteria and cause the metric to become biased.

Information about the accuracy of the initial conclusion and whether the SOC had to re-investigate the security incident can be produced as a part of the workflows used to manage the security incidents. If the incident has to be re-opened, it is an indication that the SOC had to re-investigate the incident. The accuracy of the initial conclusion is confirmed by the analyst when the case is closed. Neither of these measurements are currently recorded in the SOAR platform.

The metrics, their required measurements, and the outcome of the measurement validation are summarized in Table 4. The validation of the measurement is considered to be a success if the measurement is already recorded, a partial success if it is not recorded at the moment but can be made available, and a failure if the measurement cannot be collected without significant changes to the ways of working or the technical solutions in use.

9. Results and discussion

The metric creation framework, which is the design science artifact constructed in this research can be successfully used to construct metrics that can be used to measure the technical performance of a SOC, as was demonstrated by the metrics outlined in the section 7. Out of the four metrics evaluated, one was considered invalid, as bias remained

Table 4
Results of the measurement collection.

Metric	Validity	Measurements	Outcome
Distribution of detections among the UKC	Valid	UKC stage	Success (partial)
		MITRE ATT&CK tactic	Success
Distribution of detections by source	Valid	Original detection source	Success (partial)
		MITRE ATT&CK tactic	Success
		Detection classification	Success
Technical accuracy of the analysis	Valid	Security incident classification	Success
		Security incident original priority	Success
		Security incident final priority	Success
		Escalation status	Success (partial)
		Escalation target	Success
		Unknown entities	Failure
		Initial conclusion accuracy	Success (partial)
		Re-investigation required	Success (partial)
Number of verifiable monitoring rules	Invalid	(not evaluated)	(not evaluated)

high, and as a result, the metric was not valid as per the requirements of the framework. On an overall level, the remaining metrics suffered from similar issues, as reducing bias and subjectivity was difficult to perform, and the entire definition of what is an acceptable level of bias or subjectivity remained relatively subjective as well.

The valid metrics created with the framework can help SOCs to push their detection capabilities more towards the earlier stages of the unified cyber kill chain to decrease the potential impact of the security incidents (Distribution of detections among the UKC), quantify the value of their detection engineering function (Distribution of detections by source) and provide insights on the activities performed as a part of the analysis process (Technical accuracy of the analysis). While the metrics are by no means comprehensive, they can be used to measure the technical performance of a SOC within the respective areas, and as such, can be used to enhance the reporting capabilities related to the technical performance of the SOC, for as long as the required measurements can be made available.

One of the limitations of this research is that the design science methodology is not followed rigorously, as the feedback loop between the “Evaluation” and the “Design and development” activities are not completely enforced as per the design science methodology outlined by Peffers et al. (2007). The decision to limit the number of iterations was made due to constraints related to the research schedule and the fact that to properly evaluate the metrics, they would require years of measurements to be collected.

Although the artifact was able to produce metrics that can be used to enhance the reporting capabilities of the technical performance of a SOC, the link between the technical performance and the metric creation framework could have been slightly more concrete. The framework itself does not directly enforce the relationship between the metrics and technical performance. This does not necessarily make the criteria to be less useful, but it leaves the determination of whether the resulting metric measures technical performance or not up to the user of the artifact to decide. As an upside, the framework can also be used to create non-technical metrics.

The literature review did not establish a clear pattern when it comes to the availability of SOC-related metrics, and as such, a conclusion was reached that no such framework currently exists. Many of the more commonly used metrics as summarized in Table 1 are operative, such as *number of security incidents*, *mean time to reaction* and *mean time to resolution*. As a result, they cannot be directly used to measure the technical performance of a SOC. Although it could be argued that the technical performance has a direct impact on the operational performance and as such, the operative metrics are also an indication of the technical performance. However, as the operational metrics are also influenced by the process and people aspects of SOC, the metrics provide inconclusive results when attempting to be used to measure the technical performance.

Some of the metrics can be used to partially measure the technical performance, such as the *false-positive rate* or *number of incidents handled automatically*, but they are at best imperfect when evaluated with the metric creation framework, due to lack of additional context that causes the metric to become overly biased. For example, a high true-positive rate, and subsequently a low false-positive rate, can be an indication of a well-performing SOC but without understanding the detection strategy or level of automation, the metric alone can be misleading. The SOC could automatically close a majority of the false-positive and benign true-positive detections, and instead use them for threat hunting or to provide additional context in the form of potentially related low-fidelity detections when investigating high-severity detections. The framework can be used to enhance the reporting capabilities provided by these technical metrics, as it can be used to reduce both the bias and subjectivity as well as take into account the additional contextual information required to properly interpret the metric. In-depth validation of the metrics in Table 1 against the framework could be a suitable topic for future research.

Additional research should be performed to validate the framework to reach a definitive conclusion about the applicability of the framework in real-world SOC deployments. Furthermore, the academic research related to modern SOC operations appears to also be limited and would require additional research, especially when it comes to the following questions: (1) What are the technical success factors for SOCs? (2) What makes an effective cyber defense program?, and (3) Should SOCs focus more on proactive and preventative capabilities, rather than being mostly reactive? Answers to these questions would help with the creation of better metrics for measuring the technical performance of a SOC. Both academia and practitioners are encouraged to share metrics and ways to describe them, with the security community to enhance the methods by which organizations can measure the effectiveness of their SOC deployment. In support of this initiative, the authors have published the source code for the metrics used to validate the framework (Forsberg, 2023).

10. Conclusions

The outcome of this research further emphasizes the need for better capabilities to measure the technical performance of SOCs. The commonly used metrics focus on operational activities and are inadequate to measure the technical performance of a SOC. Furthermore, the metrics observed in the literature do not appear to be a result of a systematic development but rather be loosely based on generic security metrics or otherwise based on industry best practices without significant scientific justification.

This research resulted in a design science artifact. A novel metric creation framework that can be used to construct relevant metrics to measure both the technical and non-technical performance of a SOC. The literature review uncovered certain metrics that held the potential

for partially measuring the technical performance of a SOC. However, when examined in light of the developed metric creation framework, these metrics were found to benefit from additional context to be more effectively utilized.

As a part of the demonstration of the artifact, four unique metrics were created. Three of the four metrics were considered valid in terms of the metric creation framework. The metrics were validated by verifying that the measurements can be collected within a SOC service provider.

One of the key limitations of this study is the relatively narrow selection of metrics that were used to demonstrate the framework. The metrics created with the framework are mostly related to the detection capabilities of the SOC, and other functions, such as threat hunting or cyber threat intelligence, have not been included in the metrics chosen for the demonstration. Therefore, whether the framework is suitable for creating technical metrics for functions other than those that work closely with the security incident management process remains unknown. The wider demonstration was left to be researched in the future.

Despite the minor limitations of the research, the framework and the metrics used for demonstration can be adopted by SOCs to construct metrics they can use to measure and demonstrate their technical capabilities. Due to the lack of industry-standard reporting schema for the technical performance of SOCs, the SOC industry as a whole is encouraged to enable industry-driven development of the measurement capabilities, be open, and share the metrics they use to measure their technical capabilities with the wider community. In addition to industry-backed development of technical performance measurement, additional academic research is needed on the subject.

CRediT authorship contribution statement

Joonas Forsberg: Researching, Writing, Reviewing and Editing. **Tapio Frantti:** Supervision, Writing, Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

See ref. Forsberg (2023).

Acknowledgement

The research was supported by Business Finland (grant number 10/31/2022) and the University of Jyväskylä.

References

Agyepong, E., Cherdantseva, Y., Reinecke, P., Burnap, P., 2020. Towards a framework for measuring the performance of a security operations center analyst. In: 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), pp. 1–8.

Agyepong, E., Cherdantseva, Y., Reinecke, P., Burnap, P., 2023. A systematic method for measuring the performance of a cyber security operations centre analyst. Comput. Secur. 124, 102959. https://doi.org/10.1016/j.cose.2022.102959.

Ahlm, E., 2021. How to Build and Operate a Modern Security Operations Center. Technical Report. Gartner Inc.

Alahmadi, B.A., Axon, L., Martinovic, I., 2022. 99% false positives: a qualitative study of soc analysts' perspectives on security alarms. In: 31st USENIX Security Symposium (USENIX Security 22), pp. 2783–2800.

Bendle, N.T., Bagga, C.K., Nastasoiu, A., 2019. Forging a stronger academic-practitioner partnership–the case of net promoter score (NPS). J. Mark. Theory Pract. 27, 210–226. https://doi.org/10.1080/10696679.2019.1577689.

Böhme, R., 2010. Security metrics and security investment models. In: Echizen, I., Kunihiro, N., Sasaki, R. (Eds.), Advances in Information and Computer Security. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 10–24.

Brotby, W.K., Hinson, G., 2013. PRAGMATIC Security Metrics. Auerbach Publishers, Incorporated.

Chew, E., Swanson, M., Stine, K., Bartol, N., Brown, A., Robinson, W., 2008. Performance Measurement Guide for Information Security. Technical Report NIST Special Publication (SP) 800-55, Rev. 1. National Institute of Standards and Technology, Gaithersburg, MD.

Cichonski, P., Millar, T., Grance, T., Scarfone, K., 2012. Computer Security Incident Handling Guide. Technical Report NIST Special Publication (SP) 800-61, Rev. 2. National Institute of Standards and Technology, Gaithersburg, MD.

Crowley, C., Pescatore, J., 2019. Common and Best Practices for Security Operations Centers: Results of the 2019 SOC Survey. Technical Report. SANS Institute. https://www.sans.org/media/analyst-program/common-practices-security-operations-centers-results-2019-soc-survey-39060.pdf.

Doran, G.T., 1981. There's a S.M.A.R.T way to write management's goals and objectives. Manag. Rev. 70, 35–36.

European Union Agency for Cybersecurity, 2021. In: Lella, Ifigeneia, Theocharidou, Marianthi, Tsekmezoglou, Eleni, Malatras, Apostolos (Eds.), ENISA Threat Landscape 2021: April 2020 to Mid-July 2021. European Network and Information Security Agency.

Forsberg, J., 2023. Data for Technical Performance Metrics of a Security Operations Center. Technical Report. Mendeley Data, V1. https://doi.org/10.17632/c48syxjdzn.1.

Hauser, J., Katz, G., 1998. Metrics: you are what you measure! Eur. Manag. J. 16, 517–528. https://doi.org/10.1016/S0263-2373(98)00029-2.

Hevner, A.R., March, S.T., Park, J., Ram, S., 2004. Design science in information systems research. MIS Q. 28, 77–105. https://doi.org/10.2307/25148625.

ISO/IEC 27004:2016, 2016. Information technology — Security techniques — Information security management — Monitoring, measurement, analysis and evaluation. Technical Report. International Organization for Standardization. https://www.iso.org/standard/64120.html.

Jacobs, P., Arnab, A., Irwin, B., 2013. Classification of security operation centers. In: 2013 Information Security for South Africa, pp. 1–7.

Keltanen, P., 2019. Measuring outsourced Cyber Security Operations Center. Master's thesis. South-Eastern Finland University of Applied Sciences.

Knerler, K., Parker, I., Zimmerman, C., 2022. 11 Strategies of a World-Class Cybersecurity Operations Center. The MITRE Corporation.

Kokulu, F.B., Soneji, A., Bao, T., Shoshitaishvili, Y., Zhao, Z., Doupé, A., Ahn, G.J., 2019. Matched and mismatched SOCs: a qualitative study on security operations center issues. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, Association for Computing Machinery, pp. 1955–1970.

Logsign, 2020. Guide for security operations metrics. https://www.logsign.com/uploads/Guide_for_Security_Operations_Metrics_Whitepaper_2f999f27cc.pdf.

Nathans, D., 2014. Designing and Building a Security Operations Center. Elsevier Science & Technology Books. Editor Steve Elliot.

Onwubiko, C., 2015. Cyber security operations centre: security monitoring for protecting business and supporting cyber defense strategy. In: 2015 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), pp. 1–10.

Parmenter, D., 2019. Key Performance Indicators, 4 ed. John Wiley & Sons, Incorporated.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830. https://doi.org/10.48550/arXiv.1201.0490.

Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S., 2007. A design science research methodology for information systems research. J. Manag. Inf. Syst. 24, 45–77. https://doi.org/10.2753/MIS0742-1222240302.

Pendleton, M., Garcia-Lebron, R., Cho, J.H., Xu, S., 2016. A survey on systems security metrics. ACM Comput. Surv. 49. https://doi.org/10.1145/3005714.

Reichheld, F., 2004. The one number you need to grow. Harv. Bus. Rev. 81, 46–54. 124.

Rosso, M., Campobasso, M., Gankhuyag, G., Allodi, L., 2022. Saibersoc: a methodology and tool for experimenting with security operation centers. Digit. Treats Res. Pract. 3, 1–29.

Salmi, N., 2018. The present state of information security metrics. Master's thesis. University of Jyväskylä.

Savola, R.M., 2007. Towards a taxonomy for information security metrics. In: Proceedings of the 2007 ACM Workshop on Quality of Protection. Association for Computing Machinery, New York, NY, USA, pp. 28–30.

Savola, R.M., 2013. Quality of security metrics and measurements. Comput. Secur. 37, 78–90. https://doi.org/10.1016/j.cose.2013.05.002.

Schinagl, S., Schoon, K., Paans, R., 2015. A framework for designing a security operations centre (soc). In: 2015 48th Hawaii International Conference on System Sciences, pp. 2253–2262.

Schlette, D., Vielberth, M., Pernul, G., 2021. CTI-SOC2M2 - the quest for mature, intelligence-driven security operations and incident response capabilities. Comput. Secur. 111, 102482. https://doi.org/10.1016/j.cose.2021.102482.

Shah, A., Ganesan, R., Jajodia, S., Cam, H., 2018. A methodology to measure and monitor level of operational effectiveness of a csoc. Int. J. Inf. Secur. 17. https://doi.org/10.1007/s10207-017-0365-1.

Simos, M., Dellinger, J., 2019. CISO series: lessons learned from the Microsoft SOC—part 1: organization. https://www.microsoft.com/security/blog/2019/02/21/lessons-learned-from-the-microsoft-soc-part-1-organization/.

Sundaramurthy, S.C., Bardas, A.G., Case, J., Ou, X., Wesch, M., McHugh, J., Rajagopalan, S.R., 2015. A human capital model for mitigating security analyst burnout. In: Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security. USENIX Association, USA, pp. 347–359.

Van Os, R., 2016. SOC-CMM: Designing and Evaluating a Tool for Measurement of Capability Maturity in Security Operations Centers. Master's thesis. Lulea University of Technology.

Vielberth, M., Böhm, F., Fichtinger, I., Pernul, G., 2020. Security operations center: a systematic study and open challenges. IEEE Access 8, 227756–227779. https://doi.org/10.1109/ACCESS.2020.3045514.

Zimmerman, C., Crowley, C., 2019. Practical SOC metrics. https://www.fireeye.com/content/dam/fireeye-www/summit/cds-2019/presentations/cds19-executive-s03b-practical-soc-metrics.pdf. fireEye Cyber Defense Summit 2019.

Joonas Forsberg is a holder of a Master of Science (M.Sc) degree awarded by the Faculty of Information Technology at the University of Jyväskylä. Presently, he acts as the Lead Cyber Defense Architect at Nixu Corporation. In this role, he lends his expertise to assist both clients and internal stakeholders in devising and implementing cyber defense solutions dedicated to safeguarding the information technology (IT) landscapes of major corporations in the Nordics. Although he is just embarking on his research journey and academic qualifications are pending, he has accumulated over a decade of experience in diverse capacities within the IT sector. Joonas has fulfilled roles including cybersecurity architect, security consultant, IT manager, and IT system specialist.

Tapio Frantti holds degrees of MSc, LicTech and Dr. Tech. from the Department of Automation and Information Technology, University of Oulu. He is also an Adjunct Professor in the University of Oulu. He has worked at the Outokumpu Polarit Oy, University of Oulu, Nokia Telecommunications, and Nokia Mobile Phones as a researcher, senior researcher, chief engineer, and research manager. He has also worked in Technical Research Centre of Finland as a chief research scientist and Research Professor and as Visiting Professor in Tokyo Denki University. Lately he has worked in Renesas Mobile Europe and Broadcom Communications Finland as a distinguished researcher and research leader, in IoLiving Ltd. as a CTO and as a Co-Director of Security and Software Engineering Research Center. Currently he works as a cybersecurity professor in the University of Jyväskylä. He also works in FRE company doing security, communication and control engineering consultation. He has been on the field about 30 years and he has published +100 scientific and technical papers in journals, magazines, books and international conferences. He has also authored several patents. Tapio Frantti also acts as a regular reviewer in international conferences and 12 scientific top rated journals and he is a member of Technical Program Committees in tens of international conferences. His research interest is in the adaptive and intelligent control theory, networking technologies and cyber security.