

Universität Regensburg
Fakultät für Wirtschaftswissenschaften
Lehrstuhl für Wirtschaftsinformatik I - Informationssysteme

Harnessing Human Potential for Security Analytics



Dissertation

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft
eingereicht an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg

vorgelegt von
Manfred Vielberth, M.Sc.

Berichterstatter:
Prof. Dr. Günther Pernul
Prof. Dr. Doğan Kesdoğan

Disputation am: 25.05.2022

*To my parents, Franziska and Manfred,
and my Angelika*

Abstract

Humans are often considered the weakest link in cybersecurity. As a result, their potential has been continuously neglected. However, in recent years there is a contrasting development recognizing that humans can benefit the area of security analytics, especially in the case of security incidents that leave no technical traces. Therefore, the demand becomes apparent to see humans not only as a problem but also as part of the solution. In line with this shift in the perception of humans, the present dissertation pursues the research vision to evolve from a human-as-a-problem to a human-as-a-solution view in cybersecurity. A step in this direction is taken by exploring the research question of how humans can be integrated into security analytics to contribute to the improvement of the overall security posture. In addition to laying foundations in the field of security analytics, this question is approached from two directions. On the one hand, an approach in the context of the human-as-a-security-sensor paradigm is developed which harnesses the potential of security novices to detect security incidents while maintaining high data quality of human-provided information. On the other hand, contributions are made to better leverage the potential of security experts within a SOC. Besides elaborating the current state in research, a tool for determining the target state of a SOC in the form of a maturity model is developed. Based on this, the integration of security experts was improved by the innovative application of digital twins within SOCs. Accordingly, a framework is created that improves manual security analyses by simulating attacks within a digital twin. Furthermore, a cyber range was created, which offers a realistic training environment for security experts based on this digital twin.

Acknowledgement

A cumulative dissertation is by its nature characterized by cooperation among the authors of individual publications. Moreover, establishing an environment that provided the opportunities for the creation of this work lies in the commitment of many people to whom I owe my gratitude.

First of all, I would like to thank my supervisor Prof. Dr. Günther Pernul. I am grateful for giving me the opportunity to conduct my research at the Chair of Information Systems. Without this opportunity I would most likely not have started a research career. He created a positive research environment characterized by teamwork, which enabled the creative yet purposeful development of my research. As co-author of all publications within this dissertation, he has significantly shaped its content through his far-sighted and profound support. Whenever I had questions or uncertainties, I received prompt professional assistance. I also want to take this opportunity to thank my second supervisor Prof. Dr. Doğan Kesdoğan for his feedback and support while creating this dissertation.

Furthermore, I would like to thank my colleagues, starting with Petra and Werner for their organizational, technical and emotional support. I owe my gratitude to Fabian, who has accompanied me since my first day of work on this dissertation and was always a good partner for intensive professional discussions that deeply inspired my research. I would like to thank Ludwig, with whom I shared my office during the last years, for his awesome ideas. Furthermore, I want to mention Florian for the insights into the DINGfest project and the many scientific discussions we had. I owe my gratitude to my co-authors and colleagues Marietheres and Daniel, who contributed to this work by their valuable knowledge about digital twins and CTI. Subsequently, I want to thank Benedikt for the blockchain insights during our research and for the epic bike rides. Furthermore, I am grateful to Lena for her valuable contribution through her master thesis, our joint research and especially for the exciting cyber range sessions. I am looking forward to future collaborative research. I would also like to thank Philip and Sabrina and wish especially the newer colleagues a great time at the Chair of Information Systems. I want to thank Andreas from the Kronos AG for the practical insights that have ensured that my research is not only of scientific but also of practical relevance.

Finally, I am deeply grateful to my parents for their patience and assistance and to Angelika for her tireless support. They certainly created the necessary foundation, not only during my research career, for the creation of this dissertation. I would also like to thank my brothers Thomas and Tobias and my sister Stefanie.

Contents

Abstract	i
Acknowledgement	ii
List of Tables	iv
List of Figures	vi
I Overview of the Dissertation	1
1 Introduction	2
2 Research Questions	4
3 Research Methodology	7
4 Results	10
4.1 Overview of Research Papers	10
4.2 Focus Area 1: Security Analytics	12
4.3 Focus Area 2: Security Novices	19
4.4 Focus Area 3: Security Experts	24
4.5 Complementary publications	32
5 Conclusion and Future Work	33
II Research Papers	35
1 A Security Information and Event Management Pattern	35
2 Formalizing and Integrating User Knowledge into Security Analytics . .	48
3 Towards GDPR-compliant data processing in modern SIEM systems . .	77
4 Human-as-a-security-sensor for harvesting threat intelligence	97
5 Improving data quality for human-as-a-security-sensor	113
6 Security Operations Center: A Systematic Study and Open Challenges .	132
7 CTI-SOC2M2 – The quest for mature, intelligence-driven security opera- tions and incident response capabilities	157
8 Integrating Digital Twin Security Simulations in the SOC	178
9 A Digital Twin-Based Cyber Range for SOC Analysts	188
References	208
Appendix: Curriculum Vitae	211

List of Tables

1	Analysis of the initial state	10
2	Overview of research papers within this dissertation	11
3	Overview of complementary research papers	32

List of Figures

1	Overview of research papers and corresponding Focus Areas	12
2	Component diagram of the SIEM pattern as published in Paper P1	14
3	Knowledge Model from the perspective of the two personas as included in Paper P2.	16
4	Architecture for GDPR-compliant SIEM as published in Paper P3	18
5	The Human-as-a-Security-Sensor Incident Model and Taxonomy as pub- lished in Paper P4	20
6	A screenshot of the implemented tool for reporting incidents as published in Paper P4	21
7	Interviewing process with data quality improvement as published in Paper P5	22
8	People, processes, technology and governance & compliance as main building blocks for SOCs as published in Paper P6	25
9	Roles in a SOC with corresponding interconnections as published in Paper P6	26
10	Architecture of the CTI-SOC2M2 maturity model as published in Paper P7	27
11	Process-based framework for integrating the digital twin into a SOC as published in Paper P8	28
12	Micro-service architecture of the implemented digital twin and SIEM integration as published in Paper P8	29
13	Concept of the digital-twin-based cyber range as published in Paper P9 .	30
14	Measured increase of knowledge within each subject area as published in Paper P9	31

Part I

Overview of the Dissertation

1 Introduction

In recent decades, information systems have become an integral part of today's society. It can be seen that ever-larger areas of today's business and private landscape are increasingly distributed and connected over global networks. Currently, for example, operational technologies, such as industrial plants, are becoming increasingly interconnected. Besides the promised competitive advantage through increased connectedness, this causes a growing dependence on the correct and reliable operation of the involved information systems. However, this trend also leads to a growing heterogeneity of the underlying systems and an associated increase in complexity, which places special demands on information security that are currently not being met comprehensively. This is evidenced, among other things, by recent major security incidents such as WannaCry [24], Spectre [19] or the Kaseya attack [11].

To tackle this problem and enable the targeted detection of security incidents at an early stage, companies are increasingly applying data-driven approaches. The aim is to identify [23] and react [29] to security threats based on collected data originating from a wide range of sources. However, the collection of data alone does not solve the problem, but it is vital to analyze this data for being able to detect, in many cases very complex, incidents. This is where security analytics comes into play. Even though the term security analytics is used fairly often in the academic literature, mainly in an application-oriented way (e.g. [8, 9]), there is no generally accepted definition. Often, the term is used synonymously with big data security analytics [21], where the focus lies on the analysis of large amounts of data with the goal of enhancing security. However, since the term big data cannot be clearly delimited, security analytics is seen in a rather general way in this dissertation: *Security analytics deals with the analysis of data aiming to detect threats and improve security*. Within this context Security Information and Event Management (SIEM) systems have emerged as central security analytics applications in companies. The ENISA, for example, recommends the use of SIEM systems to counter current security threats [11]. Thus, the terms SIEM and security analytics are closely linked in the context of this dissertation, whereby SIEM rather refers to an application-related view and security analytics to the underlying activities and processes.

From a technical point of view, SIEM systems are becoming more and more sophisticated, and the practice-driven development lead to a quite mature market of SIEM products [18]. However, due to the increase in complex attacks, SIEM systems still need to be further developed and require continuous research. Advanced Persistent Threats (APTs) are to be particularly highlighted here, since most SIEMs have problems in detecting especially early attack stages [5]. This can be mainly explained by the fact that the first steps usually do not target any technical systems, but try to bypass security systems by gaining access to the organisations' infrastructure with methods like social engineering. Furthermore, APTs are very targeted, which means that it is hard to create general rules for automatic detection.

As social engineering is a key factor of most APTs and at the same time the attack

step that is most difficult to prevent, humans are commonly seen as the “weakest link” in information security [30, 1, 20]. As a result, the potential of the human factor for security analytics has been historically neglected and there is little consideration of how people can actively contribute to preventing incidents, even though a basic understanding of security exists in varying degrees among all people. However, in the past few years, a shift is recognizable and it is propagated to see people not only as a problem, but also as part of the solution [32]. The classical view where people are constrained by cybersecurity rules that employees must follow to avoid becoming victims of cyberattacks is being replaced by a view asking how people can actively contribute to improving security. This leads to the following research vision for this dissertation:

*Evolving from human-as-a-problem to human-as-a-solution
in cybersecurity.*

In order to be able to harness the human potential in security analytics, it must be taken into account that people have different levels of knowledge regarding cybersecurity. In knowledge research one distinguishes between novice knowledge and expert knowledge [26]. Thus, from a security analytics perspective, humans can be divided into security novices and security experts, depending on their knowledge within cybersecurity.

Security novices have comparatively little security-related knowledge. They have not had any security-oriented education and hardly deal with cybersecurity issues in their day-to-day business. However, to a certain extent, security novices have the ability to recognize conspicuous behavior and thus detect possible security incidents enabling them, in many cases, to contribute valuable information. The view of humans as the weakest link has led to a lot of effort being put into training employees to have a basic understanding of security issues. Thereby people are trained on what not to do in order to avoid becoming a victim of security incidents within awareness campaigns [22]. This knowledge can be harnessed by creating means that enable and support security novices to report security incidents in a structured way, which can be summarized under the human-as-a-security-sensor paradigm [13]. This paradigm figuratively considers humans as sensors that detect and report security incidents. Especially since APTs often target humans in the first stages and do not leave any technical traces, it is only the humans (and thus mostly security novices) who can recognize these steps. The problem here is that current systems for reporting or recording security incidents require a great deal of expert knowledge.

In contrast, security experts deal with cybersecurity issues on a daily basis and often have many years of training in the field. They have in-depth knowledge about security incidents and can detect and analyze them with appropriate tools and decide about targeted responses. In the context of security analytics, security experts are usually organized in the form of a Security Operations Center (SOC), which combines people, processes, and technologies. Within a SOC, security experts play a central role, as they are responsible for creating detection rules and analyzing security incidents. Thereby, an abandonment of security experts is almost impossible in the near future since APTs, and

previously unknown attacks, in particular, can hardly be detected automatically. In fact, the opposite is currently the case, as today's SOCs struggle with the high staff numbers needed for analyzing, detecting and handling incidents [7]. Additionally, the work of a security analyst within a SOC is very tedious, since analyzing incidents is in most cases monotonous resulting in high retention rates [28]. The underlying problem can be divided into two sub-problems. First, the processes in security analytics require too much manual work and are thus not sufficiently automated. Second, experts are not supported effectively in their work or are not trained well enough.

In summary, the integration of security novices and experts in security analytics is insufficient. Thus, connecting points have to be found, and innovative approaches have to be developed by which harnessing human potential can be improved.

The remainder of this dissertation is structured as follows. In Section 2, research questions are defined, and the considered problems are derived. Section 3 describes the methodology underlying the dissertation. Building on these sections, Section 4 describes the results. First, an overview of all included research papers is given, putting them in the overall context of the dissertation. Subsequently, the individual papers are summarized while highlighting their contributions. The section concludes with a brief description of complementary publications that are not directly assigned to the dissertation but were created in direct connection with it. Section 5 concludes the dissertation and discusses possible future research. Part II comprises all publications included in this dissertation in full length.

2 Research Questions

Derived from the problem that humans are not or not sufficiently well integrated into security analytics, the central research question of this dissertation arises:

RQ: *How can humans assist security analytics to improve an organization's security posture?*

The problem addressed by this research question is multifaceted and allows to define more specific sub-problems. The consideration of these sub-problems and the structure of the dissertation can be divided into three Focus Areas (FAs). In order to lay the foundations and find suitable starting points that allow improving the integration of humans into *Security Analytics*, it is first necessary to regard security analytics as a whole (FA 1). Thereby two categories for human expertise exist in the field of security analytics, distinguished by their knowledge in cybersecurity: *Security Novices* (FA 2) and *Security Experts* (FA 3). Thus, a consideration within separate focus areas is required to reflect the different initial situations regarding existing research and the different potentials emanating from the divergent human roles. In the following, the identified focus areas are examined in more detail.

Focus Area 1: Security Analytics

From a technological perspective, the field of security analytics has so far been very practice-driven. There are many security analytics applications in the form of SIEM systems on the market, which speaks for their great practical relevance. However, the initial situation in research shows the topic of security analytics being hardly considered holistically. Although some works on particular topics within security analytics exist, there is no abstract view of the subject area, making it challenging to identify starting points for subsequent research. Based on this problem, it can be concluded that there is still no integral view of how people can be effectively integrated into security analytics. This leads to the following research question:

RQ1: *What are the potentials of humans to security analytics, and how can their integration be facilitated?*

This dissertation addresses RQ1 by creating solution approaches to three underlying Practical Problems (PP). On the one hand, as mentioned earlier, there is a lack of a unified view on security analytics (**PP 1.1**). Since vendors of security software in this area are already one step ahead, and mature systems exist on the market, a first approach is to have an insight by analyzing existing systems. For enabling subsequent research, it is further necessary to describe this kind of system in an abstract way. Based on this, the next problem can be addressed (**PP 1.2**): The missing view on the connection points of humans to security analytics. The goal is to derive systematically, on the basis of the abstract view of security analytics, where and how people can be integrated. It is particularly relevant here that people can take on different roles and have varying levels of knowledge in the area of security analytics or security in general, which should be considered within the scope of this research question. Furthermore, this practical problem aims to form the link within the dissertation project between Focus Area 1 and Focus Area 2 by providing an integrating view of the two roles of security novices and security experts. The third practical problem (**PP 1.3**) is how to adapt security analytics systems for making them more human-friendly in order to facilitate their integration. Even though this problem is quite extensive and raises many follow-up questions, this point should not be neglected entirely since the willingness of people to participate in security analytics depends significantly on it. Thereby, especially data protection issues play an essential role in the case of human integration since the data produced by humans may be subject to special regulations.

Focus Area 2: Security Novices

Within security analytics, security novices are widely neglected, although they have the potential to detect security incidents and thus make a valuable contribution to security. This is especially valuable for attack steps taken in APTs that do not leave any technical traces. Within research, the topic area has so far been addressed by only a small number of researchers. The topic is known under the human-as-a-security-sensor paradigm

according to Heartfield and Loukas [13]. Existing research primarily relates to the reporting of semantic social engineering attacks. In other words, incidents where humans are potential victims or part of an attack. However, human perception of security incidents goes beyond this scope. They can also recognize irregularities that they observe which do not affect them personally, but represent a potential threat to the company. Thus, in order to make use of this potential, a more systematic integration of security novices into security analytics is needed, which raises the following research question:

RQ2: *How can security novices be supported to provide information about security incidents to security analytics?*

Two practical problems arise from this research question. The first practical problem (**PP 2.1**) is that there are no comprehensive approaches that integrate people into security analytics. It is necessary to create an approach that considers all facets of an incident and covers all identifiable and reportable incidents as far as possible. As mentioned, initial approaches for recording selected incidents do exist, but there is no consideration of the connection points to security analytics or where the contributed information adds value. Furthermore, it is important to create a framework, which enables the information to be used in further systems by recording the incidents in a structured and established data format. A second related practical problem (**PP 2.2**) is that data manually captured by humans is often of poor data quality. However, for the security incidents reported by humans, it is particularly important that they are correct and as complete as possible to provide added value for downstream security analytics. It is essential to ensure a high degree of data quality already during the collection of the information since subsequent correction is in many cases no longer possible, as for example people tend to forget relevant information over time.

Focus Area 3: Security Experts

In contrast to Focus Area 2, the current state of research and practice in the area of integrating security experts is more advanced. SOCs have emerged as organizational units in which security experts or security analysts are organized and integrated into security analytics by means of relevant technologies and processes. The central challenge in this area is the high demand for security experts, which can currently hardly be met by the job market. It can be deduced that the solution to this challenge lies less in finding new approaches for initially integrating experts into security analytics but more in analyzing and improving existing structures. It is therefore required to meet this challenge by improving integration approaches through the application of new technologies. These challenges lead to the following research question:

RQ3: *What are approaches to improve the potential of security experts for security analytics?*

This research question can be divided into four practical problems in the context of this dissertation. To improve the integration of security experts, it is first necessary to

assess the current state in this area (**PP 3.1**). Although SOCs are already a well-known concept in research, there is still no uniform view of them. Thus, it is necessary to integrate the various existing views. This provides an overarching representation of the subject area, which in turn facilitates subsequent research. Furthermore, it enables the identification of challenges for research and practice in the area through this analysis. Following on from this issue, the second practical problem (**PP 3.2**) arises. In addition to recording the current state, it is necessary to distinguish the desired target state of a SOC and what levers exist to achieve it. A maturity model not only facilitates the evaluation of research potential, but also serves as a basis for decision-making at management level when it comes to the concrete design of a SOC in a company. The main task of security experts within a SOC is to detect attacks and to create rules for security systems such as SIEM systems so that they recognize incidents automatically, leading to the third practical problem (**PP 3.3**). When security experts create detection rules, they have no way of testing these and there are no means to analyze potential incidents that have not yet occurred in the corporate environment. To create an environment to analyze potential incidents, the use of a digital twin is a promising option. Thereby the development of a systematic approach for leveraging this potential within a SOC is required. In addition to improving the integration, the main problem addressed by RQ3 can be approached by improving the training of security experts. This leads to the fourth practical problem (**PP 3.4**), how to train security experts using realistic scenarios that are tailored to the corporate infrastructure. By adapting training to the requirements arising from the corporate infrastructure, it is possible on the one hand, to train new security experts more quickly. On the other hand, better training enables security experts to work more efficiently, reducing the need for new personnel.

3 Research Methodology

The previously defined research questions have emerged from a practical relevance in the field of IT. This is emphasized by the practical problems, which illustrate the relevance arising from a practice-oriented perspective. Therefore, these questions and this dissertation can be assigned to the field of information systems research. Information systems research is at the intersection of technical IT research and the application of IT in organizations, and thus between natural, social, and behavioral science [4]. Thus it is necessary to apply research methods that consider both practical relevance and scientific rigor. In information systems research, the design science research paradigm can address both of these requirements [2], which has led to the design science paradigm being the central approach in information systems research over the last 20 years [14]. Consequently, this dissertation is significantly influenced by this paradigm.

According to Hevner et al. [15] the design science paradigm is a problem-oriented approach. It aims at producing innovative artifacts that do not depend on natural laws or behavioral theories. These artifacts are constructs, models, methods, and instances enabling researchers to understand and address the underlying problems. The artifacts

created in the context of this dissertation have been developed according to the requirements of the design science paradigm. Specifically, the design science research guidelines of Hevner et al. [15] were considered to ensure effective research. In combination with these guidelines, the design science research process of Österle et al. [27] was applied, which provides principles, objectives, and methods to conduct research in the context of the design science paradigm. Both methodologies are briefly described in the following while outlining their application in the context of this dissertation.

Guidelines for Design Science in Information Systems Research

It should be mentioned that the guidelines proposed by Hevner et al. [15] refer to a specific developed artifact and that several artifacts were created in the context of this dissertation. Therefore, a general description of the guidelines follows, whose specifications provide the framework for creating the individual artifacts. In addition, according to Hevner et al. [15], it is not mandatory to fully apply every single one of these guidelines, but rather, depending on the case, it may be sufficient to consider some of them in a more rudimentary way.

Guideline 1 – Design as an Artifact: The aim of any design science research project is to produce a useful artifact that addresses an important practical problem. An artifact can be either a construct, a model, a method, or an instantiation. Within this dissertation, each individual paper brings forth an artifact. Thereby, the origin of each artifact is a practical problem, which was defined in Section 2 based on a research question.

Guideline 2 – Problem Relevance: Relevance in design science research is understood as the development of a technology-based solution to an important and relevant practical problem. Relevance is ensured in this dissertation by tracing each artifact back to a practical problem. Relevance is also demonstrated within each paper and especially as the foundation for each focus area through appropriate studies. Thereby, the respective research area is analyzed (e.g., in the form of a literature survey) to identify the most relevant problems.

Guideline 3 – Design Evaluation: An artifact's usefulness, quality, and efficacy must be ensured by applying appropriate evaluation methods. Within the dissertation, all artifacts, for which it is appropriate, are evaluated by qualitative and quantitative approaches. The particularly extensive evaluation of the artifact designed in Paper P9 is worth mentioning here, in which an international user study was conducted.

Guideline 4 – Research Contributions: Every design science research must have clearly defined and verifiable contributions. These can be in the areas of design artifacts, design foundations and design methodologies. Within the dissertation, Section 4 presents the contributions of each paper. The contributions lie mostly in the design artifact itself, but also new foundations are created while improving existing approaches.

Guideline 5 – Research Rigor: This guideline demands the use of rigorous methods in both the design and the evaluation of the artifacts. However, Hevner et al. [15] highlight that an overemphasis on rigor can reduce relevance, which is why a good compromise must be found here. The guidelines and process presented here are the overarching methodologies within the dissertation. However, specific methods are applied for the individual artifacts to ensure scientific rigor.

Guideline 6 – Design as a Search Process: In design science research, it is often impossible to decide whether one solution is the best solution. Therefore, the design process is to be understood as a search for an effective solution using the given possibilities. Effectiveness of the solutions is ensured within the dissertation by the evaluation of each artifact. Furthermore, the artifacts are first designed abstractly to separate them from the actual implementation and thus enable a search for better solutions. In addition, the source codes and data generated in the course of the creation of each artifact are published to facilitate future search for alternative solutions or the improvement of the artifacts.

Guideline 7 – Communication of Research: It is crucial to present the research results to the research community and a management-oriented audience. This is ensured within the dissertation by publishing in peer-reviewed journals and presenting at conferences. Those were selected carefully ensuring that they address both researchers and practitioners.

Research Process

The research procedure followed in this dissertation is mainly based on the process proposed by Österle et al. [27] consisting of four phases, which are in accordance with the previously described guidelines. In the following, it is described how these phases are carried out within the framework of the dissertation, with special focus on the analysis phase, since the analysis approaches chosen in each focus area differ greatly:

1. Analysis: The goal of the first phase is to identify the problem to be solved. Thereby, research goals, questions, and gaps are specified. The initial state in research and practice as well as the chosen approach to analysis are shown in Table 1. In FA 1, practice was ahead of research (up-arrow) at the beginning of this dissertation project. A large number of SIEM systems existed on the market, while relatively few research papers existed (down-arrow). Therefore, the analysis approach taken was to derive a SIEM pattern from analyzing the state of the art and thus, providing a knowledge transfer to research. In FA 2, there has been relatively little research as well as few practical approaches, which is why an innovative approach was proposed that can serve as a basis for future research. FA 3 was already established in both research and practice. Hence a literature survey was conducted for analysis, considering research and practice.

2. Design: Research artifacts are designed based on the previously conducted analysis and the identified problems. In doing so, emphasis is placed on applying known methods, and the delimitation from known solution approaches.

	FA 1: Security Analytics	FA 2: Security Novices	FA 3: Security Experts
initial state	research ↓ practice ↑	research ↓ practice ↓	research ↑ practice ↑
approach	derive pattern	innovative	literature survey

Table 1: Analysis of the initial state

3. Evaluation: In this step, the artifact is evaluated, fulfilling the requirements of Guideline 3. Thereby it is validated whether the artifact meets the required objectives. Österle et al. [27] also see the review process before publication as part of the evaluation. Thus, in the context of the dissertation, all artifacts are evaluated individually. Among other things, case studies and user studies are conducted. Furthermore, each artifact is reviewed in a peer-review process, for ensuring its quality.

4. Diffusion: The goal of this phase is to communicate the knowledge generated by research to appropriate target groups. In addition to publishing in peer-reviewed journals, presenting at conferences, and publishing this dissertation, other diffusion instruments are used. For example, the results are directly incorporated into lectures and seminars. Furthermore, international cyber range trainings were used to pass on generated knowledge to students. Österle et al. [27] also mention funding applications as a possible diffusion instrument. In particular, the funding application for the INSIST project, which is partly based on research results of this dissertation, should be mentioned.

4 Results

In the context of this cumulative dissertation, each practical problem is addressed by one paper, answering the previously defined research questions. In the following section, an overview of the individual papers is given, with a focus on how they fit into the overall framework of the dissertation. Subsequently, to provide an overview, each paper is summarized within the context of the associated focus area.

4.1 Overview of Research Papers

Within this dissertation, a total of 9 research papers answer the research question defined in Section 2. Therefore, the methodology described in more detail in Section 3 was applied. Each paper can be assigned to a focus area while contributing to one practical problem (**PP1 - PP9**) answering the respective research question (**RQ1 - RQ3**). Thus, the numbering of the papers (compare Table 2) reflects the order of the related practical problems and is not in chronological order. Each of the papers was submitted to a peer-reviewed venue. Five papers have been published in scientific journals, and one was submitted to a journal where it is under review at the time of writing this dissertation. Three papers have been presented at scientific conferences. The relevance of the individual papers for research can be shown, in particular, by the total number of citations (56). Table

2 gives an overview of the papers that can be assigned to the dissertation. In addition to the paper number and the reference, it is indicated whether the paper has been published (pub.), whether it has been submitted (sub.) and whether it was submitted to a conference (C) or a journal (J). The information in the table represents the status at the reference date, 14 December 2021, shortly before submission of this dissertation.

No.	Publication	Status	Type	Cit. ¹
P1	VIELBERTH, M., AND PERNUL, G. A security information and event management pattern. In <i>12th Latin American Conference on Pattern Languages of Programs (SugarLoafPLOP)</i> . The Hillside Group, 2018, pp. 1–12	pub.	C	10
P2	BÖHM F., VIELBERTH, M., AND PERNUL, G. Formalizing and Integrating User Knowledge into Security Analytics. Submitted to <i>SN Computer Science</i> , (2021), 1–28	sub.	J	n/a
P3	MENGES, F., LATZO, T., VIELBERTH, M., SOBOLA, S., PÖHLS, H. C., TAUBMANN, B., KÖSTLER, J., PUCHTA, A., FREILING, F. C., REISER, H. P., AND PERNUL, G. Towards GDPR-compliant data processing in modern SIEM systems. <i>Computers & Security</i> 103, 102165 (2021), 1–19	pub.	J	9
P4	VIELBERTH, M., MENGES, F., AND PERNUL, G. Human-as-a-security-sensor for harvesting threat intelligence. <i>Cybersecurity</i> 2, 23 (2019), 1–15	pub.	J	17
P5	VIELBERTH, M., ENGLBRECHT, L., AND PERNUL, G. Improving data quality for human-as-a-security-sensor. A process driven quality improvement approach for user-provided incident information. <i>Information and Computer Security</i> 29, 2 (2021), 332–349	pub.	J	0
P6	VIELBERTH, M., BÖHM, F., FICHTINGER, I., AND PERNUL, G. Security Operations Center: A Systematic Study and Open Challenges. <i>IEEE Access</i> 8 (2020), 227756–227779	pub.	J	8
P7	SCHLETTE, D., VIELBERTH, M., AND PERNUL, G. CTI-SOC2M2 – the quest for mature, intelligence-driven security operations and incident response capabilities. <i>Computers & Security</i> 111, 102482 (2021), 1–20	pub.	J	0
P8	DIETZ, M., VIELBERTH, M., AND PERNUL, G. Integrating Digital Twin Security Simulations in the Security Operations Center. In <i>Proceedings of the 15th International Conference on Availability, Reliability and Security</i> (2020), ACM, pp. 1–9	pub.	C	11
P9	VIELBERTH, M., GLAS, M., DIETZ, M., KARAGIANIS, S., MAGKOS, E., AND PERNUL, G. A Digital Twin-Based Cyber Range for SOC Analysts. In <i>Data and Applications Security and Privacy XXXV</i> , vol. 12840 of <i>Lecture Notes in Computer Science</i> . Springer, Cham, 2021, pp. 293–311	pub.	C	1

Table 2: Overview of research papers within this dissertation

¹Citation count based on Google Scholar: <https://scholar.google.de/>

As explained in Section 2, this dissertation is divided into three focus areas. The categorization of the papers to the respective focus areas is illustrated in Figure 1. The exact arrangement is explained in more detail below.

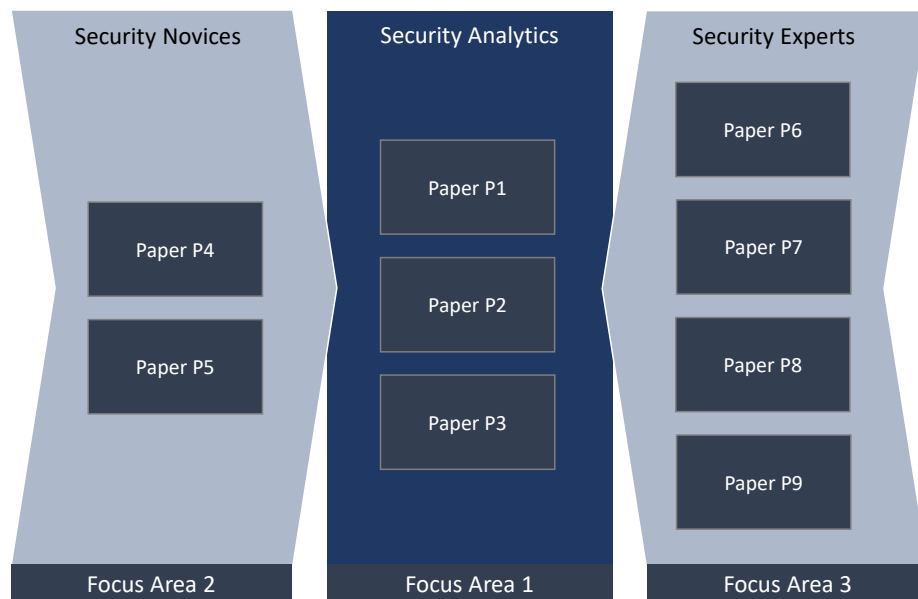


Figure 1: Overview of research papers and corresponding Focus Areas

Focus Area 1 as the central building block forms the foundation of this dissertation. On the one hand, it identifies what security analytics is (Paper P1) and, on the other, how it can be made more human-friendly (Paper P3). Paper P2 plays a special role here. It serves as a link between the three focus areas since it formally delineates the two personas of security novices and security experts within security analytics and shows how they can be integrated. Focus Area 2 addresses the integration of security novices into security analytics within the framework of the Human-as-a-security-sensor paradigm. In this context, Paper P4 develops an approach based on which employees can report security incidents. Paper P5 extends this approach and improves it by ensuring that the data quality of reported security incidents is as high as possible. On the other side, the involvement of security experts is addressed in Focus Area 3. Here, Paper P6 and Paper P7 lay the foundations by examining the current state and the desired state of a SOC. Based on this, Paper P8 and Paper P9 propose approaches that allow the potential of experts to be harnessed more effectively. Therefore, digital twins are used to simulate incidents in order to make the work in a SOC more efficient and effective (Paper P8). Also, a digital twin is used to create a realistic training environment in the form of a cyber range (Paper P9).

4.2 Focus Area 1: Security Analytics

Security analytics as the central focus area of this dissertation deals with the analysis of data to improve security. Although there is research on numerous sub-areas of security analytics, there is no uniform abstract understanding of this topic. Therefore, in the form

of three papers, the following section lays the foundation for further research by creating a unified understanding about security analytics. Furthermore, the types of knowledge in security analytics are formally defined, identifying the connecting points for the Focus Areas 2 and 3. Finally, an architecture for a GDPR compliant SIEM is presented to enable the processing of personal data. It should be mentioned here that Papers P1 and P3 described below use the term SIEM and Paper P2 uses the term security analytics. In the understanding of this dissertation, SIEM systems are, as identified in Paper P1, an implementation of a security analytics tool. Therefore, the two terms are to be understood as largely synonymous in the context of this dissertation.

P1: A Security Information and Event Management Pattern

At the beginning of the dissertation project, the field of SIEM, was very practice-driven. The market for SIEM products was relatively mature, as evidenced by a large number of products, with nearly every major security software vendor offering its own SIEM product [17]. While sub-parts of a SIEM system, such as log data normalization, were present in research, a holistic, abstract view of the structure and operation of a SIEM system was not available (**PP 1.1**). Therefore, the goal of this paper is to transfer the knowledge from practice to research to enable research efforts that build upon it. In this regard, the PLoP (Pattern Languages of Programs) community offers a platform that aims, among other things, to derive and describe patterns of software and publish them to a broad audience.

In Paper P1, a pattern for SIEM systems is described. Therefore, the pattern identification methodology of Fehling et al. [12] is applied. In addition to defining the domain to be analyzed and determining a format for describing the pattern, various sources of information were identified based on which a SIEM pattern can be derived. Available software products on the market serve as the central source of information. In addition, papers of various types, manuals, product documentation, and product websites were considered. For identifying the most important SIEM systems, the Gartner Magic Quadrant for Security Information and Event Management [17] is used, which analyzes the SIEM systems on the market according to various evaluation criteria and identifies the most mature systems.

The structure of describing patterns is quite rigid within the pattern community, which is why the following summary also adheres to this structure, with the individual subsections of the pattern description in italics. For describing the pattern, Paper P1 first defines synonyms for the term SIEM (*also known as*). The *intent* of a SIEM is the collection of security-relevant data in a central system to enable security analyses. The *context* in which a SIEM system operates is a corporate infrastructure that has grown historically from a large number of different systems. Various security systems such as firewalls or intrusion detection systems are in use. This leads to the *problem* that each of these systems individually collects security-relevant data. However, it is impossible to detect sophisticated attacks on the individual system in many cases because traces that

occur in different places must be correlated and jointly analyzed. Furthermore, the large amount of data complicates security analyses and overburdens human security analysts. The *solution* to these problems is a SIEM system that collects all security-relevant data in a central location and uses it to detect security incidents.

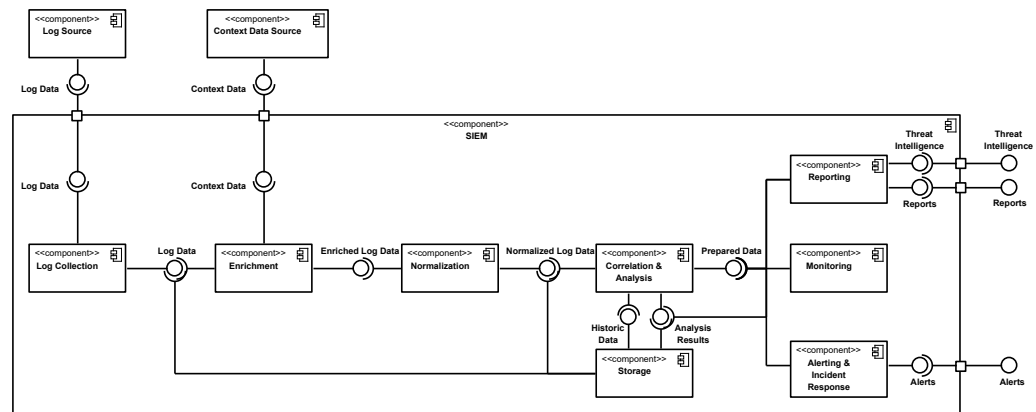


Figure 2: Component diagram of the SIEM pattern as published in Paper P1

The derived structure of a SIEM system is shown in Figure 2 in the form of a UML component diagram. A SIEM system consists of a log collection component that collects all security-relevant data from the various source systems. This data is enriched with context data within the enrichment component and translated into a uniform format in a normalization component. The actual detection of attacks follows in the correlation & analysis component. The results and the log data are stored to enable historical analyses, for example to identify and analyze initially undetected attacks at a later date. The reporting component generates reports and shares threat intelligence, whereas the monitoring component provides a user interface. Finally, the alerting & incident response component is used to notify stakeholders and trigger appropriate responses.

To show that the described SIEM pattern is utilized in practice, it is compared with a SIEM *implementation* on the market, IBM QRadar, and a SIEM system developed in a research project, DINGfest. With regard to IBM QRadar it was determined that essential components correspond to those of the pattern, whereby these are named differently with brand-specific terms. In comparison, it was shown that the SIEM system from the DINGfest project contains all components of the SIEM pattern, but goes beyond in some areas. For example, it places particular emphasis on data protection issues, which is why it contains a pseudonymization component. In addition, the visual analytics functionalities go beyond the identified capabilities within the SIEM pattern, whereas other selected components are neglected.

The paper concludes with a discussion of possible *variants* of the pattern found in the analyzed SIEM systems. In this respect, SIEM systems implementing the pattern in practice are listed, and the advantages and liabilities of the pattern are outlined.

Contribution of P1:

The paper's main contribution is to provide the basis for further research in the area of SIEM and security analytics by describing an abstract pattern of SIEM systems. This establishes a transfer of knowledge from practice to research, which is necessary due to the primarily practice-driven development in the field of SIEM. In addition, the pattern serves to provide a uniform understanding of SIEM, since it has not been defined and delimited so far.

P2: Formalizing and Integrating User Knowledge into Security Analytics

In the scope of this dissertation, people involved in security analytics are distinguished as security novices and security experts. The rationale for this and the underlying relationships are discussed in Paper P2. In this context, a knowledge model is created as a formal basis for subsequent publications to provide a structured understanding of what knowledge novices and what knowledge experts can contribute to enable or facilitate incident detection (**PP 1.2**).

After a brief introduction to the topic, the paper describes the different types of knowledge that exist in security analytics. It elaborates that knowledge is challenging to define based on the existing literature, which is why a formal definition of the major types of knowledge is provided. A distinction is made between explicit and implicit knowledge. Explicit knowledge is often referred to as machine-based knowledge. In the field of security analytics, explicit knowledge can take the form of models for machine learning approaches, rules or signatures for rule-based detection methods, or threat intelligence or forensic evidence resulting from in-depth investigations of incidents. Implicit knowledge, on the other hand, is the knowledge that people possess. In the context of security analytics, implicit knowledge can be broken down into three classes: Domain knowledge, situational knowledge, and operational knowledge. Domain knowledge is the knowledge that experts have about a specific topic. This domain can be security-related or knowledge from other areas (which does not mean that this knowledge is not relevant for incident detection). On the other hand, situational knowledge refers to the ability that each person has to recognize abnormal or conspicuous behavior or events. Operational knowledge is the knowledge required to operate a system. Therefore, in security analytics, this knowledge relates primarily to the operation of security-relevant systems, such as SIEM systems.

Based on this, knowledge conversion in security analytics, i.e., the processes that describe how knowledge types are converted into other knowledge types, is discussed. Four types of knowledge conversion have been identified: Internalization, externalization, combination, and collaboration. Internalization refers to the transition from explicit knowledge to implicit knowledge. An example would be when an analyst uses a SIEM system to analyze the data that led to an incident and thus obtains knowledge about how the incident occurred. Externalization refers to the transition from implicit knowledge to explicit knowledge. For example, in the area of security analytics, this occurs when an

employee uses his knowledge of a particular security incident to create a rule in a SIEM system so that it can be detected automatically, making his implicit knowledge available to the system. Combination means merging or exchanging different explicit knowledge sources. An example here would be when SIEM systems from different organizations exchange CTI (Cyber Threat Intelligence) and thus combine their knowledge about a security incident to get a more holistic picture. Collaboration, similar to combination, is the process of merging different knowledge bases, but in contrast, it combines implicit knowledge instead of explicit knowledge. This occurs primarily when multiple people work together, which can be supported by appropriate technologies. In security analytics, the collaboration of security experts and security novices, is of particular importance since their domain knowledge differs significantly, and security incidents can only be comprehensively detected by combining both.

The main contribution of the paper is the integration of the different knowledge types into the security analytics process. The underlying process is based on the incident detection process, which is in essence implemented by the SIEM pattern shown in Figure 2. In the context of this dissertation, implicit knowledge is especially important, as it enables a formal view of the knowledge that humans have.

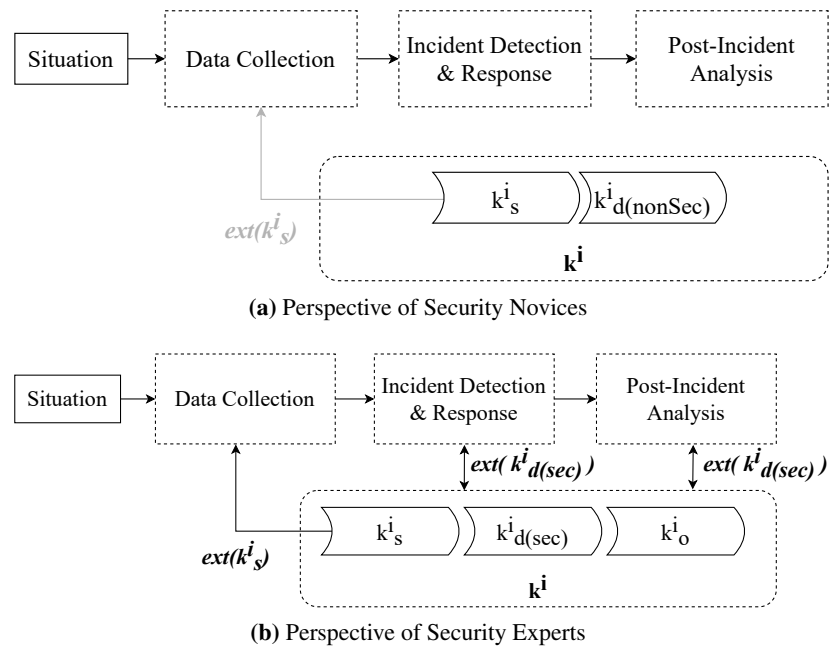


Figure 3: Knowledge Model from the perspective of the two personas as included in Paper P2.

As shown in Figure 3, a distinction can be made between two different personas - security novices and security experts - each of which has a different knowledge base with regard to the security analytics process. Based on this representation, it can further be shown that the knowledge needed for incident detection is distributed across the two personas, and only by merging the knowledge a comprehensive detection becomes possible. On the one hand, as shown in Figure 3a, security novices have situational knowledge (k_s^i) and non-security domain knowledge ($k_{d(nonSec)}^i$). As a result, security

novices can detect anomalies within their domain, but they do not have the necessary operational knowledge (k_o^i) to create SIEM rules or report possible incidents. On the other hand, security experts (Figure 3b) have the necessary security-related domain knowledge ($k_{d(sec)}^i$) and operational knowledge (k_o^i) to operate and configure a SIEM system. However, they lack non-security-related domain knowledge ($k_{d(nonSec)}^i$), which is why security incidents in these non-security-related domains often go undetected. This disparity of knowledge between the two personas is subsumed in Paper P2 as the dichotomy of security analytics.

The paper identifies knowledge gaps to outline where further research needs to be carried out in order to address this dichotomy. The first gap is that there is no way for security novices to contribute their observations (in the form of situational knowledge) to the detection process without having to operate complex systems, which require operational knowledge. The second gap is that non-security domain experts do not have the necessary knowledge to operate security systems, such as SIEM systems. The third gap describes the missing collaboration between security novices and security experts. Through the collaboration of the two personas, all relevant knowledge can be brought together, creating a central knowledge base for security analytics.

Contribution of P2:

The main contribution is the formal definition of knowledge types and related knowledge conversion processes in the area of security analytics. Thereby a unified, formal terminology for future research is established. Furthermore, based on this, security personas and research gaps are identified, which provide a framework for this dissertation.

P3: Towards GDPR-compliant data processing in modern SIEM systems

The General Data Protection Regulation (GDPR) forms a common data protection framework for all member states of the European Union [10]. Coming into effect in May 2018, the GDPR has raised the question whether it affects SIEM systems as well. That is the case, as depending on the data collected, personal data, mainly in the form of employee data, may be present in SIEM systems. In order to create the legal possibility to use data provided by humans in security analytics, it is necessary to create a SIEM system that complies with regulatory requirements, making it more human-friendly (**PP 1.3**). Therefore, based on the identified components of a SIEM system in Paper P1, locations, where personal data is collected or processed are delimited.

After a basic introduction to SIEM and the GDPR, a legal analysis of the requirements resulting from the GDPR for SIEM systems is performed. Thereby, it can be determined that it would be theoretically possible to remove or not collect any personal data in order to be GDPR compliant. However, from a security perspective, this can be problematic, as the removal of this data means that in some cases attacks might not be detected, or at least the complete scope of an attack is no longer apparent. One example would be insider

fields containing personal data. Furthermore, it is shown that all requirements stemming from the GDPR are met and thus, it can be assumed that a SIEM based on the presented architecture is GDPR compliant.

Contribution of P3:

The paper's contribution lies in an architecture that allows a SIEM to be GDPR compliant while having minimal impact on incident detection rates. This ensures that personal data is not exposed within the SIEM, thus protecting humans or, more specifically, employees. The proposed architecture meets requirements from both a legal and technical standpoint and therefore addresses interdisciplinary challenges.

4.3 Focus Area 2: Security Novices

Security novices are humans with limited knowledge related to cybersecurity. They do not deal with security issues in their day-to-day business but can still spot anomalies in many cases. This may be due to their intuition or their knowledge gained from participation in awareness campaigns. The following section addresses the question of how this knowledge can be used for security analytics within the framework of the human-as-a-security-sensor paradigm. Paper P4 develops an approach that enables the reporting of security incidents in a structured manner. Paper P5 extends this approach by a process-driven quality improvement approach, ensuring good data quality during reporting.

P4: Human-as-a-security-sensor for harvesting threat intelligence

The DarkVishnya case, which is presented based on the descriptions of a security expert from Kaspersky Labs, serves as a central example to motivate the problem tackled in the paper. During the course of this security incident, criminals attacked companies by entering the premises and physically connecting a device to the corporate network. This device was used to carry out further attack steps. What is dangerous about this attack is that the first steps left no technical traces and could therefore not be detected by security systems, which is why this attack was very successful from the attacker's point of view. The only way this attack can be detected and prevented at an early stage would be if humans see the attacker or the placed device and report it. However, since people have often been seen as the weakest link in security in the past, the idea of integrating security novices into security analytics is still relatively new. While first approaches for reporting simple social engineering attacks exist (e.g. [13]), there has not yet been a holistic view of how humans can be integrated into security analytics (**PP 2.1**). The goal of this paper is, therefore, on the one hand, to define connection points of human security sensors to SIEM systems based on the SIEM pattern presented in Paper P1. On the other hand, the goal is to develop an approach that enables to easily capture security incidents by security novices and translate them into a data format suitable for SIEM systems.

To integrate humans into the SIEM workflow, two approaches are identified: Push

and pull. The push method applies when an employee notices something suspicious and proactively reports the security incident. In this approach, it is of particular importance to offer guidance, as it must be assumed that the employee has little security knowledge. Therefore, it is presumably difficult for him to describe a security incident in every detail. The information generated by the push approach can be fed into a SIEM system during the initial event collection. The pull approach is used when a security incident is already detected, but information is missing to understand its full extent. This can occur in a SIEM system during the analysis of the security incident or the subsequent steps. The approach presented in the paper subsequently focuses on the push approach.

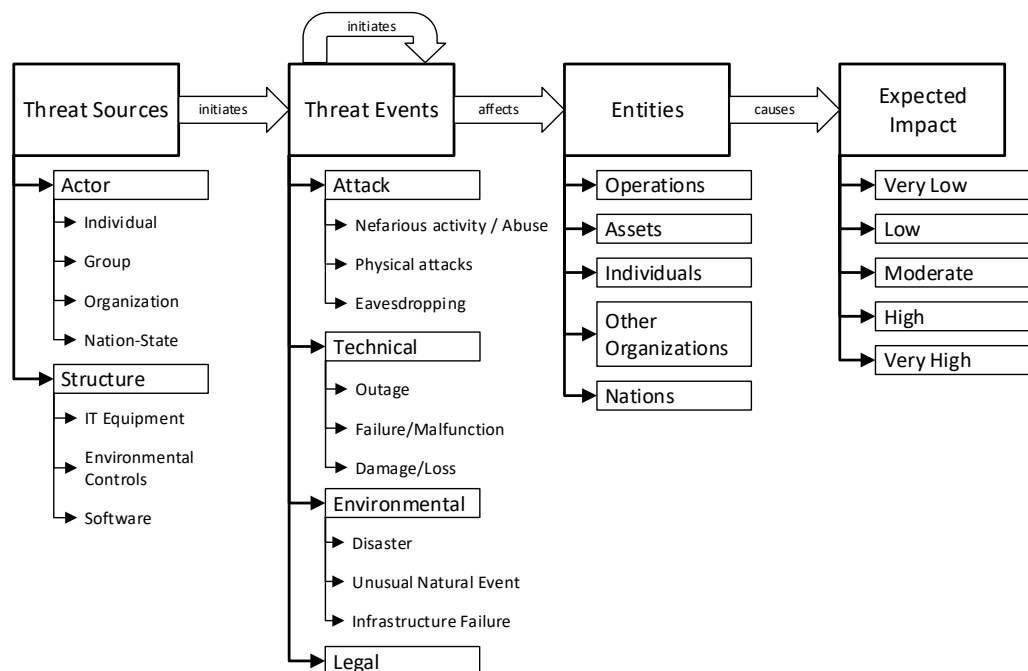


Figure 5: The Human-as-a-Security-Sensor Incident Model and Taxonomy as published in Paper P4

As a step towards a system for comprehensive reporting security incidents by humans, it is first necessary to determine which security incidents are recognizable by humans and can therefore be reported by a human. As a starting point, the NIST generic risk model [16] is used, from which the four entities *Threat Source*, *Threat Event*, *Entities*, and *Expected Impact* (compare Figure 5) are derived. To enable the description of an incident in more detail, a taxonomy is developed for each of the entities by applying the method for taxonomy development of Nickerson et al. [25]. The first two layers of the taxonomy are shown in Figure 5.

As a prerequisite for the input of the security incidents into a SIEM system, data formats are analyzed for suitability. A combination of STIX and CyBOX turned out to be the most appropriate. However, both data formats do not completely map the data elements defined in the taxonomy, which is why the formats are extended to include all data fields needed.

A prototypical implementation demonstrates the applicability of the concept. With

Human as a security sensor Wizard
REPORT INCIDENTS IN 4 SIMPLE STEPS.

Report Threat Sources

Threat Sources are entities, which initiate harmful events.
If the source is unknown or the harmful event was initiated by another event, you can skip this step.

Source 1

Source 2

Sourcetype
Actor

Detail 1
Individual

Detail 2
Outsider

Detailed description

+ add Source

Previous Next

Captured elements:

Sources:
Source 1: Software x
Source 2: Outsider x

Events:
Event 1: Outage x
Event 2: Unauthorised entry to pre... x

Entities:
Entity 1: Individuals x

Impact:
Impact: Low

Figure 6: A screenshot of the implemented tool for reporting incidents as published in Paper P4

the help of the prototype, a security novice is enabled to report security incidents in a structured way. As shown in Figure 6, users can enter the four entities Threat Source, Threat Event, Affected Entities and Expected Impact. When selecting the elements, a drop-down list is shown that allows to select the elements from the previously defined taxonomy. This ensures that users are not overstrained with too much information and can work their way from an abstract to a specific level of detail and stop when they can no longer contribute information at a more detailed level. Once the incident is reported, it is translated into STIX format on server-side for subsequent use in SIEM systems. Finally, a case study evaluates the proposed approach. For this purpose, the Dark-Vishnya case described above is recorded and reported with the help of the developed tool.

Contribution of P4:

The paper provides the foundation for the comprehensive reporting of security incidents by security novices. It goes beyond approaches that only allow the reporting of a restricted class of attacks and aims to enable capturing all incidents possible. Therefore, the main contribution of the paper is the development of a taxonomy that identifies elements of incidents that can be reported. In addition, the STIX and CybOX format is adapted to enable structured processing of the reported incidents. The feasibility is demonstrated with the help of a prototypical implementation and a subsequent case study.

P5: Improving data quality for human-as-a-security-sensor. A process driven quality improvement approach for user-provided incident information

The approach presented in Paper P4 offers the possibility for security novices to report security incidents in a structured and simple way. However, it does not yet ensure good data quality during the reporting process and does not support the user in providing the best possible data quality (**PP 2.2**). This is where Paper P5 sets in by presenting an approach that improves the data quality during the reporting process and enables the user to correct any errors or missing information by asking additional questions.

Laying the foundation for this approach, it is determined how data quality is defined in the context of the human-as-a-security-sensor paradigm in a first step. A vast number of data quality dimensions are known from the literature, which is why particularly relevant dimensions are selected. The selected data quality dimensions completeness, consistency, relevance, reliability, timeliness, and currency, are the most frequently used and considered the most critical dimensions in literature. Thus, these dimensions are generally defined based on well-known literature definitions and then reinterpreted for the human-as-a-security sensor context. In addition, these dimensions are supplemented to include character attributes that can be derived from the reporting human, so-called “predictors of detection efficacy” such as trustworthiness.

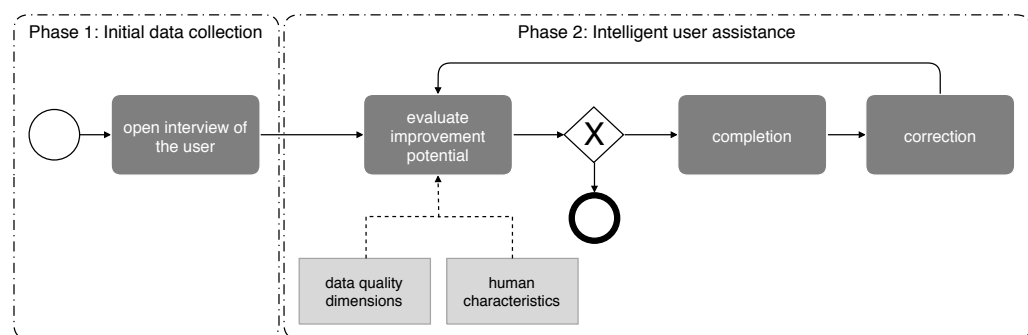


Figure 7: Interviewing process with data quality improvement as published in Paper P5

The main contribution of Paper P5 is a process that improves the data quality during the data collection, by extending the reporting process of Paper P4. The process shown in Figure 7 can be divided into two phases. The first phase is the *initial data collection*,

where the user can record security incidents without significant restrictions. Essentially, the wizard presented in Paper P4 is used here, with the help of which all elements of an incident can be recorded in a structured manner based on the incident taxonomy. Within the process, this step is referred to as the *open interview of the user*. As second phase follows the *intelligent user assistance*.

This phase begins with a review of whether the quality achieved in the first step can be further improved. Therefore, on the one hand, the data quality dimensions described above are taken into account, and on the other hand, the human characteristics of the reporting human are considered. Based on that, a decision is taken whether the data quality can be further improved. If this is the case, the two steps, *completion* and *correction*, are run through. Otherwise, the process is completed, and the incident can be passed on to downstream systems.

The basic idea of the *completion* and *correction* step is based on the principle of recommender systems building on the hypothesis that security incidents reported by security novices have already occurred in a similar form in the past. Under this hypothesis, the similarity of the security incident reported in the open interview phase, with all past security incidents in a database is calculated. The reported incident is then compared to the most similar incidents, whereby differences are identified. These differences are used for automated queries to the user. During the *completion* steps, the user is asked whether he or she has forgotten certain information that was present in similar incidents but not in the incident the user reported. In contrast, in the *correction* step the user is alerted to potentially incorrect information that was not present in similar incidents. It is important that the user remains involved in the decision-making process, as only he or she knows the ground truth and can decide whether the system's suggestions are really correct.

The approach is evaluated with the help of a use case analysis. For this purpose, the process was fully implemented and integrated into the prototype presented in Paper P4. In order to be able to use a database of past incidents for the calculation of similarity, the IBM X-Force² database was used, which provides a large number of known incidents as part of a threat exchange platform. The evaluation demonstrated that the quality of the data improves as the process is run through.

Contribution of P5:

The paper's contribution is twofold. On the one hand, data quality is defined in the context of the human-as-a-security-sensor paradigm, by identifying and describing relevant quality dimensions. On the other hand, a process is presented that enables the data quality to be improved during the reporting of security incidents. With the help of a prototypical implementation and a use case analysis, it is evaluated that the presented process achieves the desired results.

²<https://exchange.xforce.ibmcloud.com/>

4.4 Focus Area 3: Security Experts

Security experts are humans with profound security-related knowledge who deal with cybersecurity issues in their daily business. As the union between people, processes, and technologies, SOC's form the organizational unit in which security experts are integrated into security analytics. Therefore, the following section deals with improving the potential of security experts within a SOC. Paper P6 first establishes the basics by assessing the current state of research in the form of a literature survey. Building upon the results Paper P7 provides a tool for defining the desired target state of a SOC by creating a maturity model. Building on this groundwork, the subsequent two papers investigate the potentials of the digital twin for a SOC. Therefore, Paper P8 presents a framework integrating attack simulation within a digital twin into a SOC. Paper P9 builds on the results by using this digital twin to create a virtual training environment in the form of a cyber range.

P6: Security Operations Center: A Systematic Study and Open Challenges

The third research question of the dissertation addresses the integration of security experts into security analytics. In practice, security experts are organized in a SOC for this purpose. It can be seen in the literature that although the term SOC is used frequently and suggestions are made, for example, as how to improve individual elements of a SOC, there is no holistic view and no accepted definition (**PP 3.1**). Therefore, the main goal of Paper P6 is to capture the current research in the field of SOC within the framework of a structured literature review to create the foundation for further research.

From a methodological point of view, the literature review follows the well established three steps introduced by Tranfield et al. [31]. A total of 321 academic publications were considered in the review. After the removal of duplicates, 208 papers remained, of which 158 were classified as relevant. As a first insight, the number of publications has risen sharply since 2015 up to the date of the submission of the paper, indicating a solid increase in relevance of the topic.

As a result of the literature analysis, the paper initially describes general aspects of a SOC. Therefore, the term SOC is delimited from other terms, such as SIEM. Subsequently SOC is defined to provide a unified, abstract view that delineates the research area and is of central importance for subsequent research. Within the body of literature, a large number of papers deal with technical architectures and operating models of a SOC. Even if these often have only few commonalities, influencing factors can be synthesized, which are presented in a structured overview. Four main building blocks of a SOC can be derived from the analysis (see Figure 8): People, Processes, and Technology, which are expanded by Governance, and Compliance providing the framework within which a SOC operates. These four building blocks create the structure for the presentation of the further survey results.

Within the *People* building block, the main contribution is the identification of the different roles within a SOC (compare Figure 9). The aim here is to present a maximal configuration of a SOC, listing all the roles represented in the literature. Therefore, it

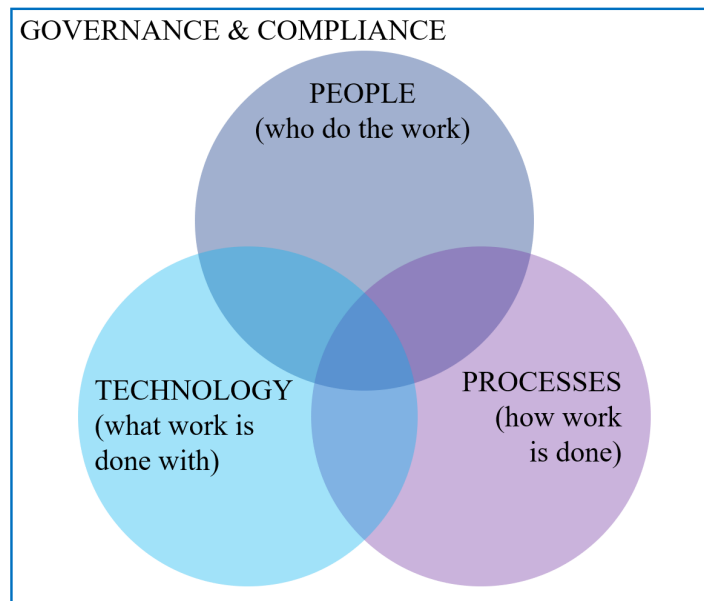


Figure 8: People, processes, technology and governance & compliance as main building blocks for SOC as published in Paper P6

must be considered that there may be SOC that occupy only a small portion of these roles. In the context of the *Processes* building block, it was found that literature only deals with SOC processes in a very limited way. It was possible to identify individual processes, which mostly relate to technical processes, such as the data collection process, but the overall picture is rather incomplete. Furthermore, it can be seen that many of the processes described are also used in other areas of security and are therefore not very SOC-specific. Within the *Technology* building block, the various data sources that are relevant for a SOC are examined. Essentially, this is log data that is supplemented by further intelligence, for example from external exchange platforms. A large part of the literature is dedicated to the topic of how incidents can be detected from a technical perspective. To understand which approaches are used, these are classified within the paper, whereby exclusively anomaly- or signature-based methodologies could be identified. Others however are hardly considered within the examined body of literature. *Governance and compliance* focuses on the standards relevant for SOC or to whose fulfillment it can contribute. It should be noted that there is no standard that addresses SOC as a whole. Instead, the standards can be assigned to a specific domain or a task that are at least in parts addressed by SOC.

One major contribution of the paper is the identification of challenges in the respective building blocks for which further research potential exists. Since the challenges from the people building block are particularly relevant for this dissertation, the remainder of this summary will focus primarily on them. The identified challenges can essentially be narrowed down to four main issues. Firstly, in particular SOC analysts, have a very monotonous job at the lower tiers. Their task is to analyze alerts triggered by SIEM systems and decide whether it is a real incident or a false positive. This is mainly done based on log data, which makes for a very monotonous workday. The second challenge is

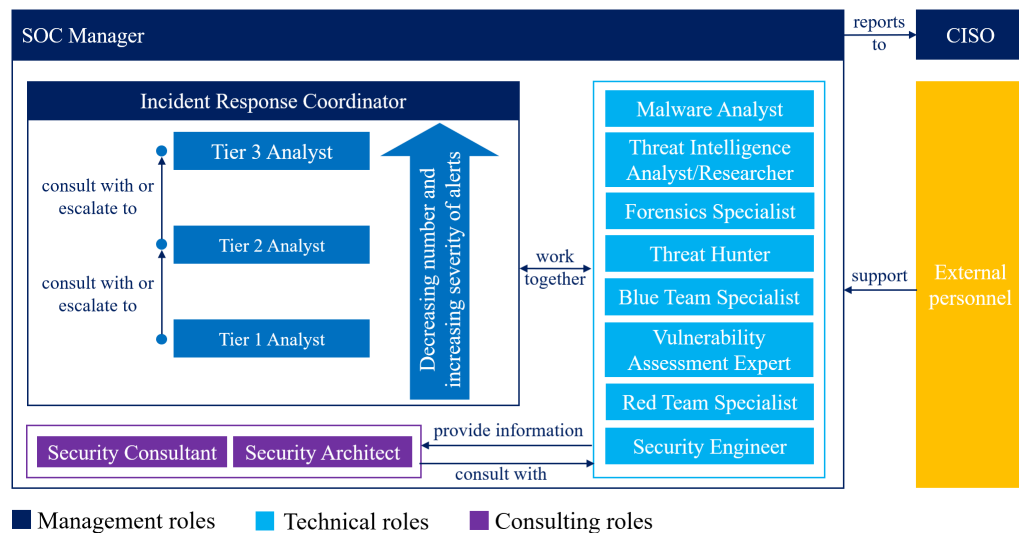


Figure 9: Roles in a SOC with corresponding interconnections as published in Paper P6

the neglected integration of domain knowledge. Although this knowledge would be vital for identifying attacks on complex industrial plants, as pure security knowledge is not sufficient there, but a deeper understanding of the functionality is needed. Third, there are hardly any approaches that enable collaboration between experts, although collaboration between security experts and domain experts would be crucial. The aforementioned challenges lead to the fourth challenge, which **PP 3.4** of this dissertation addresses: The lack of sufficiently well-trained personnel and the difficult retention of existing staff. On the one hand, staff turnover is very high due to the monotonous working conditions. On the other hand, there is not enough personal. The influencing factors are that existing personnel are not trained sufficiently well, and the working environment is inefficient, leading to a higher demand.

Contribution of P6:

The contribution of the paper is to present the current state of research on SOC and, based on that, derive challenges. On the one hand, this enables a quick entry into the research topic, especially for new researchers, and on the other hand, it creates a uniform integrated view of the topic. Furthermore, future research potential is identified, which enables a targeted improvement of the current status.

P7: CTI-SOC2M2 – The quest for mature, intelligence-driven security operations and incident response capabilities

In order to be able to identify improvement potential for SOC in a goal-oriented manner, it is necessary to define the desired target state (**PP 3.2**). Maturity models can be used for this purpose. Not only can the target state be defined, but these models also serve as a basis for decision-making at management level. For example, the benefit of a SOC can be measured within an individual organization to justify investment decisions. Building on the theoretical foundations developed in Paper P6, Paper P7 presents the

CTI-SOC2M2 maturity model. For developing the model, the paper adheres to the methodology of Becker et al. [3]. Therefore, in a first step the problem is defined. Based on that, the envisioned model is compared to and differentiated from existing models. After determining a development strategy, the maturity model is developed iteratively. Therefore, SOC services with common characteristics are grouped and mapped with the corresponding CTI data formats. Based on this, the corresponding maturity levels are defined, and a transfer medium is created through a prototypical implementation of an assessment tool. Finally, the model is evaluated.

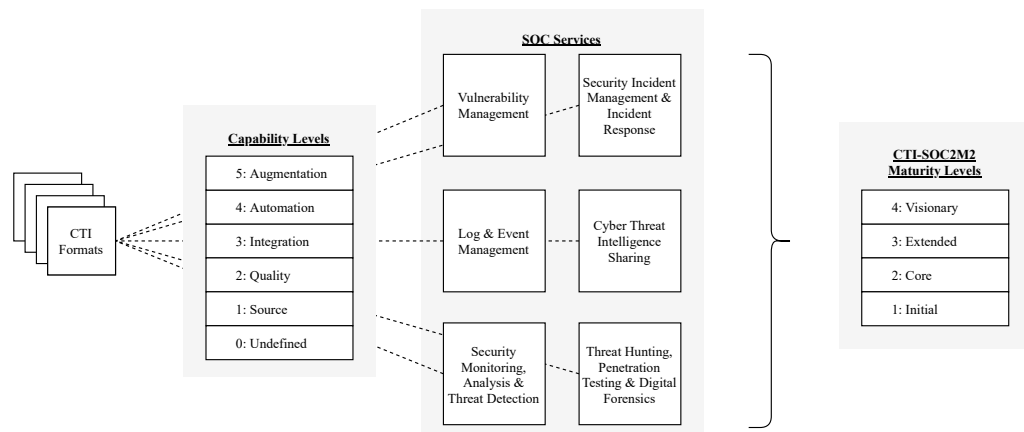


Figure 10: Architecture of the CTI-SOC2M2 maturity model as published in Paper P7

The developed maturity model follows the approach of measuring the maturity level of a SOC in a data-driven manner. Therefore it considers which data formats are integrated within the SOC as an indicator for maturity. The resulting architecture of the model is visualized in Figure 10. The derived SOC services were developed with the help of the results from Paper P6, meaning that the current state of research is represented within the model. The data formats serve as a reference for the capabilities present within a SOC. Thus, to measure the maturity, the data formats are linked to the respective SOC services. Within the paper, the individual services are described in detail while discussing, which data formats can support them.

For measuring the maturity of a SOC, capability levels are defined for the respective services. These measure how advanced the SOC services are based on the level of integration of the respective CTI formats. The capability levels range from “0: Undefined” to “5 Augmentation”. “Undefined” means that the corresponding CTI formats have not yet been considered. “Augmentation”, in contrast, indicates that the formats are not only obtained automatically but that there is a monitoring mechanism in place that ensures the handling of yet unknown CTI formats. The capability levels of the respective services then result in the overall maturity level of the entire SOC. Ranging from “1: Initial” to “4: Visionary” they indicate the maturity of a SOC. The NIST Incident Response Life Cycle [6] is used to translate the capability levels into a maturity level, which means that reaching a higher maturity level is accomplished by greater coverage of the Incident

Here, the experts in the SOC decide on the purpose they want to pursue with the simulation and thus determine the corresponding parameters and settings to use for this purpose. For example, they could decide in this step that they want to analyze a man-in-the-middle attack and simulate the network environment for reproducing the attack. Based on this, the actual *security simulation* is performed. The purpose of the simulation can be differentiated between “system security testing”, “pentesting” and “attack simulation”. In the first case, the goal is to explore whether the security systems are configured correctly. Pentesting tries to find vulnerabilities within the simulated system and attack simulation aims to examine the behavior of the system in the event of an attack. The information gathered during security simulation is used in the SOC for *incident analysis*. The aim is to gain knowledge that enables the incident to be prevented in the event of a real occurrence. This knowledge can then be used in the *incident detection and handling* step. A SIEM for testing purposes is used to detect and prevent the attack automatically. This is done, for example, by creating detection rules. If the detection was successful, the rules of the test-SIEM are transferred to the production SIEM in the *deployment*. If the attack was not detected appropriately, the process is rerun.

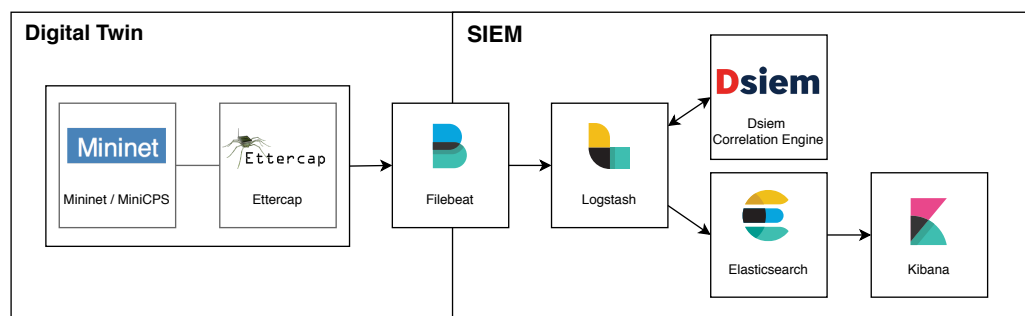


Figure 12: Micro-service architecture of the implemented digital twin and SIEM integration as published in Paper P8

For evaluation, the framework was implemented. Therefore, a real-world use case in the form of an industrial filling plant was simulated. Both the simulation component of the digital twin and the SIEM within the SOC were implemented and connected. Figure 12 shows the resulting micro-service architecture. The digital twin was implemented using *Mininet/MiniCPS*³. The man-in-the-middle attack was performed with the help of *Ettercap*⁴. The SIEM system was implemented based on the ELK stack⁵, with the correlation engine for detecting attacks being realized using *Dsiem*⁶. This implementation demonstrated the feasibility of the previously presented framework.

³<https://github.com/scy-phy/minicps>

⁴<https://www.ettercap-project.org/>

⁵<https://www.elastic.co/>

⁶<https://www.dsiem.org/>

Contribution of P8:

The contribution of the paper is a process-based framework that can be used to exploit simulations within a digital twin for SOCs. This provides a way to make the work of security experts both more efficient and effective. A technical implementation of the approach was used to evaluate its feasibility.

P9: A Digital Twin-Based Cyber Range for SOC Analysts

To cope with the problem of the high and ever-growing demand for security experts, it is necessary to educate more experts and enhance their training. Through improved training, not only can a larger number of experts be better educated, but their training also leads to lower turnover rates within the SOC [28]. Cyber ranges offer a promising way to do this. They make it possible to create a realistic environment for the training of security experts, thus providing a solution approach for **PP 3.4**. To this end, Paper P9 builds directly on the results of Paper P8 by using the simulation component of the digital twin to build a cyber range. This allows security experts to be trained using real security incidents without compromising the actual enterprise infrastructure and related real-world objects.

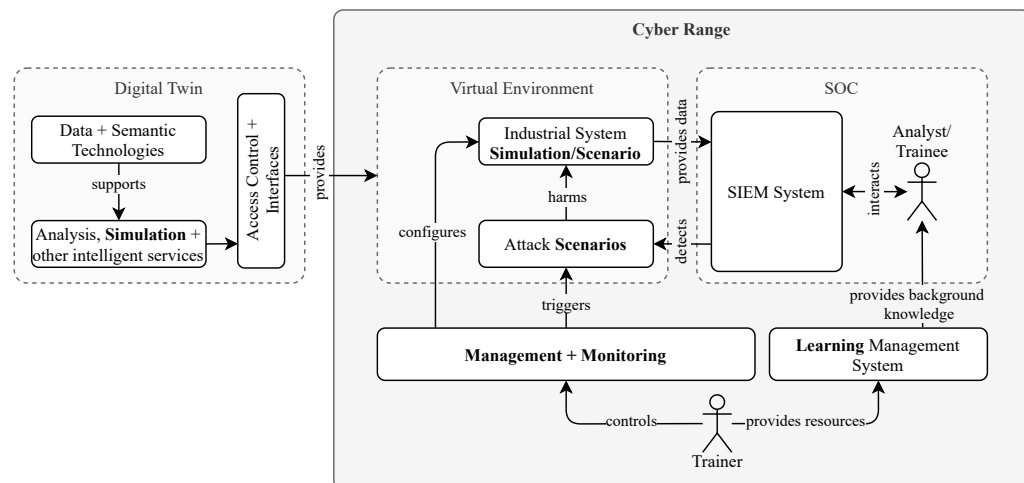


Figure 13: Concept of the digital-twin-based cyber range as published in Paper P9

As shown in Figure 13, a *virtual environment* is created using the simulation component of the *digital twin*. For this purpose, the bottling plant already simulated in Paper P8 is used as a scenario. For the purpose of the training, *attack scenarios* are automatically executed on this simulation. The analyst to be trained is located within a fictitious *SOC*. There, he or she is provided with a *SIEM system* that is actually used in practice so that a training environment is created that is as realistic as possible. The simulated industrial system produces log data, which is fed into the *SIEM system*. Based on this data, the trainee can use the *SIEM system* to detect the attacks manually or create detection rules that enable the *SIEM system* to detect the simulated attacks automatically. Furthermore, the cyber range contains a *management and monitoring* component, which triggers the attack, and enables the trainer to configure the simulation. Within the *learning manage-*

ment system, learning materials such as videos or texts are provided for the trainee, which provide the necessary background knowledge.

For enabling the actual learning with the help of the cyber range, a scenario and a learning concept were developed. The scenario consists of three programmable logic controllers (PLCs) controlling the industrial plant, which are connected to a human machine interface (HMI) via a switch. It is assumed that the attacker has access to the network and is directly connected to the switch. The learning concept consists of six tasks that must be solved by the trainee, which successively increase in difficulty. Each of the tasks includes a simulated attack, which the trainee must analyze with the SIEM system and create corresponding detection rules. Before each task, explanatory texts and videos are provided, which on the one hand, explain the scenario and, on the other hand, convey knowledge that is needed to solve the tasks.

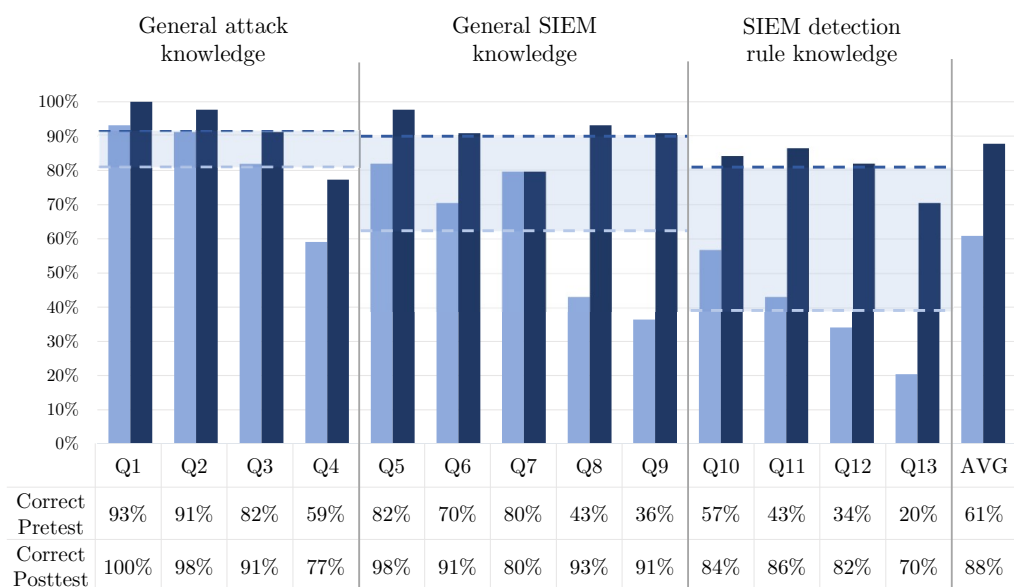


Figure 14: Measured increase of knowledge within each subject area as published in Paper P9

The cyber range was evaluated through an extensive international user study. For this purpose, it was carried out in collaboration with the Ionian University in several sessions with a total of 44 participants. To test whether the cyber range achieves the desired results, the knowledge gain was measured by questions that the participants had to answer before and after the cyber range training. Figure 14 shows the percentage of correctly answered questions divided into different knowledge classes. The results were statistically evaluated by applying a paired t-test. Among other things, a significant increase of 42.05% was found in the “SIEM detection rule knowledge” class. Thus it could be shown that the presented cyber range leads to learning success.

Contribution of P9:

The paper's contribution lies in the exploration of whether cyber ranges are suitable for the effective training of security experts. Thereby the potential of a digital twin was examined to create a realistic training environment. Furthermore, the cyber range was evaluated in an international user study whereby the collected data was provided publicly to facilitate subsequent research.

4.5 Complementary publications

In addition to Papers P1 - P9, further publications have been published in the context of this dissertation. These papers are not directly assigned to the dissertation, but were influenced by the included publications or contributed to the development of the scientific context. In order to get a comprehensive picture of this dissertation's frame, each additional paper will be briefly presented in the following. Table 3 provides an overview. The papers were published either at a conference (C) or in an encyclopedia (E).

No.	Publication	Status	Type	Cit. ⁷
A1	MENGES, F., BÖHM, F., VIELBERTH, M., PUCHTA, A., TAUBMANN, B., RAKOTONDRAVONY, N., AND LATZO, T. Introducing DINGfest: An architecture for next generation SIEM systems. In <i>Sicherheit</i> . 2018, pp. 257–260	pub.	C	11
A2	VIELBERTH, M. Security Operations Center (SOC). <i>Encyclopedia of Cryptography, Security and Privacy</i> , Springer Berlin Heidelberg, (2019), 1–3	pub.	E	0
A3	VIELBERTH, M. Security Information and Event Management (SIEM). <i>Encyclopedia of Cryptography, Security and Privacy</i> , Springer Berlin Heidelberg, (2019), 1–3	pub.	E	1
A4	BÖHM, F., VIELBERTH, M., AND PERNUL, G. Bridging Knowledge Gaps in Security Analytics. In <i>Proceedings of the 7th International Conference on Information Systems Security and Privacy</i> (Online Streaming, 2021), SciTePress, pp. 98–108	pub.	C	3

Table 3: Overview of complementary research papers

Paper **A1** is an extended abstract that presents a novel SIEM architecture. The focus is on creating a SIEM that meets forensic requirements. In addition, the architecture is extended with further data from virtual machine introspection, and the analysis capabilities are supplemented with identity behavior analytics capabilities. By presenting the general idea, it served as the basis for the SIEM architecture presented in Paper P3 and is thus directly related to this dissertation.

Both Papers **A2** and **A3** are contributions to an encyclopedia. Paper A2 deals with the term SOC and Paper A3 with the term SIEM. Both define the respective term and delimit it thereby, laying the foundation for research. In addition, the two papers present the use

⁷Citation count based on Google Scholar: <https://scholar.google.de/>

of the term in theory and practice as contextual information. The two definitions of terms are based on preliminary work done for this dissertation. Especially Paper P1 and Paper P6 are to be mentioned, which created the foundation for the definitions of the terms.

Paper **A4** is a conference paper, which was further elaborated by Paper P2 as an extended version. Therein, the different knowledge types and conversion processes are formally derived and delimited to create a uniform understanding. The focus is on the identification of research gaps. In order to take a first step towards closing those gaps, the paper presents a proof-of-concept implementation that uses visual programming approaches to enable security novices to contribute their knowledge to security analytics. The paper was extended in Paper P2 to include a classification of the different types of knowledge into the Incident Detection Lifecycle. In addition, the delimitation of the two personas, security novices, and security experts, was expanded.

5 Conclusion and Future Work

Since humans are, in many cases, easy victims of cyber attacks, the image of humans as the weakest link prevails within cybersecurity. This connotation is often communicated to employees as a mantra for ensuring enterprise-wide security. Thereby, people are told what they have to refrain from doing with a moralizing undertone in order not to open the door for attackers. This has led to the potential of humans for security being constantly neglected. This dissertation has addressed this problem in three focus areas with the central research question of how people can be integrated into security analytics in order to improve the overall security posture of organizations. Within this dissertation, nine research papers have been created. These are dedicated to answering the research question and provide a valuable contribution to the advancement of security analytics by moving a step from a human-as-a-problem towards a human-as-a-solution perspective.

The first focus area was dedicated to the assessment and improvement of the topic of security analytics. A transfer of knowledge from practice to research was created by analyzing SIEM systems as representatives of security analytics tools, which was published in the form of a software pattern. Based on this, a formal knowledge model in security analytics was defined. Thereby the two personas, security novices and security experts, were delineated and defined, creating the framework for this dissertation. Furthermore, a model for a SIEM system that meets the specifications of the GDPR and thus fulfills the requirements in terms of data protection was proposed.

Regarding the integration of humans into security analytics, on one side security novices were taken into account. In the context of the human-as-a-security-sensor paradigm, an approach was developed enabling the reporting of all incidents that can be observed. Furthermore, the STIX data format was extended so that the reported incidents can be integrated in a structured way into security analytics systems for further processing. Building on this, the reporting process has been further enhanced by developing a data quality improvement procedure. This in essence enables errors in the reported security

incident to be corrected or missing information to be added by comparing the incident with past incidents.

On the other side, the integration of security experts in security analytics was considered. In particular, the topic of SOC as an organizational aspect in which security experts are organized within the scope of security analytics was addressed. The first step was to create a theoretic foundation through a literature survey. Based on this, a maturity model for SOCs was developed, which can be used to measure the current status within a company. To enhance the creation of detection rules by experts, an approach was presented that leverages the potentials of a digital twin. This involves simulating attacks within the digital twin to enable security analyses without compromising the enterprise infrastructure. Building on this, the simulation was used to develop a cyber range that enables the training of security experts in an environment close to reality.

This dissertation started with the research vision of evolving from human-as-a-problem to human-as-a-solution. This vision is a goal that is almost impossible to achieve in its entirety in the near future. Thus, this dissertation, as any other research work, leaves potential for subsequent research. Therefore, in the following, connecting points for future research that were identified in the course of this dissertation are discussed.

The integration of the security novice into security analytics was addressed in this dissertation primarily from a technical or computer science perspective. This leaves some questions unanswered. For example, problems from psychology play an important role in this regard. For employees to report security incidents, it must be investigated how they can be motivated to do so. Since reporting may have negative consequences for the employee, for example, because an incident involves an error on his or her part, incentives must be created for reporting. The corporate culture also plays a decisive role here. Initial approaches already exist in this area for the creation of an anti-fraud culture, which could be adapted to the area of human-as-a-security-sensor.

In the area of security experts, the survey conducted in the course of Paper P6 already provides good starting points for future research by identifying open challenges. Nevertheless, a few major aspects should still be mentioned here. A holistic view of the processes is still missing within SOC research. Although there are some overarching security processes that can be related to the activities within a SOC, these are not very well tailored to the requirements of a SOC.

Furthermore, the problem of the lack of security experts is far from being solved. Thus, approaches for training security experts need to be further improved. The cyber range developed in the course of this dissertation offers a good starting point. However, research in the field of cyber ranges should focus on bringing forth more general methods to enable the development and adaptation of training environments to meet practical demands. For example, it is not yet known what constitutes a good cyber range or how its quality can be determined to facilitate the identification of potential for improvement.

Part II

Research Papers

1 A Security Information and Event Management Pattern

Current status:	Published
Conference:	12th Latin American Conference on Pattern Languages of Programs, SugarLoafPLOP 2018, Valparaíso, Chile, November 20-23, 2018
Date of acceptance:	24 October 2018
Full citation:	VIELBERTH, M., AND PERNUL, G. A security information and event management pattern. In <i>12th Latin American Conference on Pattern Languages of Programs (SugarLoaf-PLOP)</i> . The Hillside Group, 2018, pp. 1–12
Authors' contributions:	Manfred Vielberth 90% Günther Pernul 10%

Conference description: The goal of Sugarloaf PLoP is to bring together both researchers and practitioners, covering a broad range of topics. In doing so, the conference provides a platform where patterns within software systems can be presented and iteratively improved, aiming for close collaboration between pattern experts.

A Security Information and Event Management Pattern

MANFRED VIELBERTH, University of Regensburg

GÜNTHER PERNUL, University of Regensburg

In order to achieve a high level of cyber security awareness most mid to large sized companies use Security Information and Event Management (SIEM) embedded into a Security Operations Center. These systems enable the centralized collection and analysis of security relevant information generated by a variety of different systems, to detect advanced threats and to improve reaction time in case of an incident. In this paper, we derive a generic SIEM pattern by analyzing already existing tools on the market, among additional information. Thereby, we adhere to a bottom-up process for pattern identification and authoring. This article can serve as a foundation to understand SIEM in general and support developers of existing or new SIEM systems to increase reusability by defining and identifying general software modules inherent in SIEM.

Categories and Subject Descriptors: D.2.11 [Software Engineering]: Software Architectures—Patterns; K.6.5 [Management of Computing and Information Systems] Security and Protection

General Terms: Security patterns

Additional Key Words and Phrases: SIEM, Security Information and Event Management, Security Analytics

ACM Reference Format:

Vielberth, M. and Pernul, G. 2018. A Security Information and Event Management Pattern. 12th Latin American Conference on Pattern Languages of Programs (SugarLoafPLOP 2018), November 2018, 12 pages.

1. INTRODUCTION

The protection of corporate IT infrastructures against cyber attacks is becoming a more and more demanding task. Trends like Industry 4.0 and Internet of Things transform today's IT-landscapes into a complex and mazy structure with a growing amount of attack points. In most mid to large size companies, a Security Operations Center (SOC) is established to gain a holistic and centralized view on IT security and to enable fast reactions in case of an incident. According to the SANS 2017 Security Operations Center Survey [Crowley 2017], more than 80% of those SOCs are supported by a SIEM system in order to increase IT security awareness. Besides, SIEM systems enable the automation of incident detection and subsequent reactions in order to mitigate imminent damage or to preserve forensic evidence. Therefore, these systems collect security relevant data at a central point, to gain a holistic view of the organizations IT security. Additionally, historic and correlated analyses and further measures are enabled.

Although there are many SIEM systems on the market, there is no pattern to the best of our knowledge, which defines the generic structure of a general SIEM setting. To fill this gap and to create a foundation for understanding the basic components of a SIEM system, we propose a pattern for SIEM in this paper. Additionally this pattern can

This research was partly supported by the Federal Ministry of Education and Research, Germany, as part of the BMBF DINGfest project (<https://dingfest.ur.de>).

Authors' address: Lehrstuhl für Wirtschaftsinformatik I, Universitätsstraße 31, 93053 Regensburg, Germany; M. Vielberth email: manfred.vielberth@ur.de; G. Pernul email: guenther.pernul@ur.de;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission. A preliminary version of this paper was presented in a writers' workshop at the 12th Latin American Conference on Pattern Languages of Programs (SLPLOP). SLPLOP'18, NOVEMBER 20-23, Valparaíso, Chile. Copyright 2018 is held by the author(s). HILLSIDE 978-1-941652-11-4

help developers to delimit generic modules. This enables the enhancement of reusability and replaceability of selected modules and improves the general structure and compatibility of SIEM software.

The proposed pattern is deduced in conjunction with the DINGfest project. DINGfest is funded by the Federal Ministry of Education and Research in Germany and aims at developing an architecture for next generation SIEM systems. Especially collecting forensic evidence from virtual environments and leveraging data from Identity and Access Management is emphasized. Furthermore, the project is focusing on improving methods for Visual Security Analytics, by involving experts more efficiently.

One purpose of patterns is to show the current state of best practice for a recurrent problem in a specified context. Thus, it should be used in at least two different use cases [Gamma et al. 2011]. In this paper, we follow this approach and in order to identify the current state of best practice in SIEM we analyze established SIEM systems and related information. In the following, the SIEM pattern is deduced and explained oriented on the POSA (Pattern-Oriented Software Architecture) template [Buschmann et al. 2013].

The intended audience for this pattern includes developers and administrators of SIEM systems. Additionally, it can serve as a foundation for people who are new to this topic in order to gain a basic understanding. We assume our readers to have basic knowledge about UML and a fundamental awareness of IT security issues.

In this paper, we give a theoretical foundation by explaining SIEM in general in Section 2. Subsequent, the derived pattern is described in Section 3. In Section 4, we describe the pattern mining / identification process, which we followed. In Section 5 we finally conclude our work and give hints for future directions, to gain a more detailed view on SIEM. Appended an explanation of domain specific terminology can be found.

2. SECURITY INFORMATION AND EVENT MANAGEMENT

The first notion of Security Information and Event Management is attributed to a report of Gartner Inc. [Williams and Nicolett 2005]. According to this work the expression is composed of two terms: Security Information Management (SIM) and Security Event Management (SEM) [Goldstein et al. 2013]. While SIM deals with centralized management, collection, preservation of historic log data and the generation of reports for compliance purposes, SEM covers threat management, real-time monitoring of security incidents and triggering proper reactions in case of an incident. Thereby, the collected data is aggregated to reduce the amount of data and facilitate the usage for appropriately reacting to security events.

Nevertheless, nowadays' SIEM has evolved from a sole combination of those two technologies to a more holistic and integrated security solution and thus combines the advantages of SIM and SEM in one single centralized system. Due to the numerous functionalities a SIEM has to fulfill, an appropriate academic definition is hard to accomplish. However, the Gartner IT Glossary [Gartner Inc. 2018] outlines the expression quite well. According to their definition, SIEM systems collect relevant data and conduct historic analyses on security events from a broad range of different types of event or contextual data sources. Additionally, it supports reporting for compliance purposes and forensic investigation by analyzing the stored historic data from the same sources. The main functionalities of SIEM are its broad scope on event sources as well as its ability to correlate and analyze these events across heterogeneous sources.

The general purpose of collecting and analyzing such a big amount of data in a central place is the identification of anomalies and incidents, which would not be visible considering data of only one single system or device. Additionally, it facilitates the monitoring and management of the organizations' state of IT security.

However, SIEM is not an isolated software, which runs untouched for a long period of time, but it is in most cases embedded into a Security Operations Center (SOC). A SOC as described by Bhatt et al. [Bhatt et al. 2014] is a centralized organizational unit, with the goal to monitor all relevant events related to security of an IT infrastructure. A SOC basically consists of software, processes and a team of security analysts and experts [Radu 2016]. The

utilized software for collecting data and log files from all relevant assets and for further analyses is a SIEM system. After the detection of a potential incident, the SOC staff determines the measures to be taken, like informing an incident response team or the stakeholders of affected systems. In most cases SOCs are organized hierarchically in multiple layers, whereby the lowest layer depends on a large amount of employees. In large organizations they are organized in multiple shifts for being able to react accordingly to security incidents around the clock.

Due to the growing amount of different sources for log data and other security relevant events, and the subsequent growing need for well trained security analysts, the demand for a higher level of automation rises. Thereby, SIEM can help to automatically analyze and react to security incidents in a centralized manner. It further helps analysts by better visualizing the big amount of data to leverage the expert knowledge of the analysts more efficiently.

Further explanations of relevant terminology is given in appendix A.

3. SIEM PATTERN

3.1 Also Known As

SIEM, SIEM system, (Big Data) Security Analytics System, Cyber Threat Intelligence Tool/System

3.2 Intent

Collect all security relevant data in a central point in order to identify threats or incidents. Therefore, provide the opportunity to deal with different data formats and to analyze the data in real-time and historically. In addition, offer the possibility to interact with human experts.

3.3 Context

In today's organizations, a big number of different security systems and devices are part of IT-infrastructures. In most cases they have been grown historically and thus it is hard to keep track of the overall state of corporate IT-security. Additionally, isolated security systems, like firewalls can only detect or prevent very specific attacks. To gain a holistic picture of the state of security and to uncover more advanced cyber attacks collecting all security relevant information in a centralized system is required. Therefore, real time threat analysis is essential for being able to react quickly, whereby historic analyses have to be performed in addition.

3.4 Problem

The big amount of different security relevant devices and software makes it hard to gain a holistic view of an organizations security. Especially, sophisticated attacks affecting multiple systems (commonly referred to as advanced persistent threats) often remain undetected. Moreover, the tremendous amount of generated log data makes the situation even more challenging, because it complicates security related analyses and makes it more difficult to find appropriate reactions in case of an incident. The forces associated with this problem are as follows.

Forces:

- *Detection rate and false positives:* We want to decrease the manual workload of experts and analysts as much as possible by automatically detecting incidents and anomalies. In order to unburden them, the false positive rate has to be very low, as staff has to analyze the wrongly detected incident and decide for further steps. Additionally, the detection rate should be as high as possible, to provide a high level of security. Every incident that was not detected in time might end up causing more work hours or additional damage.
- *Time of discovery and reaction:* We want to be able to detect and react to an attack as fast as possible, because the time a company needs to detect and react to an attack directly impacts the financial damage [Kaspersky Lab 2016]. In order to mitigate the losses, it is important either to react automatically to an incident or to inform the right people in time, who ideally carry out a reaction plan.

A Security Information and Event Management Pattern — Page 3

- *Usability*: We need a system, that is easy to use by analysts and experts. The effort to connect new devices and configure the detection of new incidents (e.g. create new detection rules) has to be as easy and as little time consuming as possible. Furthermore, the necessary expert knowledge should be as low as possible.
- *Visual preparation of data and integration of expert knowledge*: In the future, it will not be possible to replace experts who manually analyze the data so quickly. Especially the automatic detection of incidents that have never occurred before is challenging. Thus, visual preparation of the data and enrichment with context data is very important for optimally integrating expert knowledge, not only at the beginning and the end of the analytics process, but also in the middle [Ropinski et al. 2017; Gates and Engle 2013].
- *Degree of automation*: Well trained staff in the security domain is quite rare. However, in big organizations SOCs have to be staffed around the clock. We want a system, that can help to reduce the number of people needed by automating certain steps. On the one hand, the analysis should be further automated and on the other hand, reactions should be triggered or carried out automatically wherever possible.
- *Number of connectible devices*: In order to provide a holistic view on corporate security, we want to exploit all relevant data. Therefore, every device pertinent for security has to be connected. This poses two requirements: First, it must be possible to process a large number of log data. Second, nearly every different kind of device or software has to be able to be connected. Thus, we must be capable of handling many different log formats.
- *Analytics*: For being able to detect and identify attacks, a large variety of events and additional data has to be analyzed. Therefore, the possibility to create rules or other detection mechanisms have to be provided. The analysis mechanism should be able to correlate multiple events, which enables the detection of even more incidents or threats. Thereby, capabilities for automatic as well as manual or human-supported analyses must be offered.
- *Reusability and intelligence sharing*: We want to provide a high degree of reusability. Therefore, the pattern should be split into different modules, which are easy to exchange and to integrate by other similar systems. Thus, they should be loosely coupled and interfaces have to be defined. Additionally, interfaces for sharing analyses results with other organizations should be provided. Moreover, reporting for compliance reasons should be possible as it becomes increasingly important especially for critical infrastructure providers [European Parliament 2016; Congress of the United States of America 2014].
- *Costs*: Costs play an important role for every system, intended for application in a company. Therefore, we want to reduce the time needed for implementation and staff necessary for running and monitoring the software. However, it is hard to make a statement about profitability for most cyber security topics [Weishäupl et al. 2018].

3.5 Solution

Implement a SIEM system that collects, enriches, normalizes and stores all relevant log data in a centralized manner. This system automatically detects and reacts to as many incidents as possible, provides interfaces for reporting those and enables the integration of expert knowledge.

Structure:

The resulting SIEM Pattern can be split into eight components, as shown in Fig. 1. The first module is responsible for *log collection* and for providing the collected *log data* to further modules. The input data can be provided by various *log sources*.

The *enrichment* component uses *context data* for providing additional information for bare *log data* in order to improve further processing by providing *enriched log data*. Various *context data sources* supply the needed data.

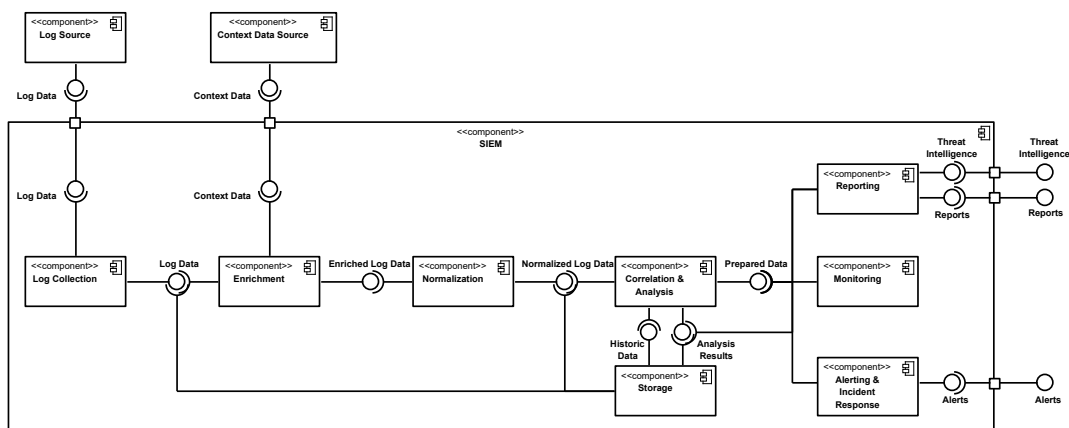


Fig. 1. SIEM Pattern as UML component diagram

For enabling and facilitating further processing, the data is restructured into a standardized format (*normalized log data*) in the *normalization* component.

The main part is the *correlation and analysis* component, as there the incidents are finally recognized based on the incoming *normalized log data*. Additionally, analyses are conducted based on *historic data* provided by the *storage* component. *Analysis results* and raw *log data* are also stored herein. In order to enable further processing by other modules, *prepared data* is provided.

Finally, the findings can either be *reported* or shared with other systems or organizations in form of *threat intelligence*. The *monitoring* component provides an interface for human interaction and the *alerting & incident response* component can propagate *alerts* to responsible stakeholders or react automatically in case of an incident.

3.6 Implementation

In this section we present additional information for implementing the proposed pattern and corresponding components. Additionally, we describe two examples for SIEM systems and show that their pattern fits the one presented in this paper for the most part. Therefore, we first describe a SIEM, which is quite popular in industry and then introduce DINGfest, a SIEM proposed by a research project.

The *log collection* component can either pull the logs from a variety of *log sources* or the log generating source pushes the log data into the *log collection* module. In most cases, the pull based method is implemented by simply accessing the file system of the log generating source by well known standard protocols like FTP (file transfer protocol) or SCP (secure copy). In order to push the logs to the SIEM, the *log source* has to use supported protocols like syslog or a so called agent has to be installed on the source, which is responsible sending logs to the SIEM. Logs can also be integrated as a data stream.

An example for *enrichment* of log data is to resolve the IP address to a geolocation, which can enable the detection of some specific threats. Furthermore, asset discovery tools provide information about systems or sensitive data which requires more protection. The data from vulnerability assessment can help to relate threats to specific data sources. Additionally, enriched data is in most cases much easier to read and interpret by a human analyst in subsequent modules.

The *normalization* component is needed as the various connected log generating sources use many different formats for storing and sending logs in practice. In addition, varying storage technologies are applied. For example, some systems utilize relational databases and others use plain text files or proprietary technologies like syslog. Thus, it is necessary to transform them into a single uniform log format. Additionally, it is much easier for experts to create detection rules for standardized log formats in further steps.

The *correlation & analysis* is conducted in real time on incoming *normalized log data* for recognizing incidents as fast as possible. Additionally, historic analyses are performed for forensic purposes or in order to detect incidents overlooked before, as, for example, the pattern of an attack was not yet known at this time. A multitude of different methods for correlating and analyzing logs exist. However, the most common and very simple approach is detecting incidents on the basis of rules. Such rules can be derived automatically, created by human analysts or retrieved from other systems or organizations. The structure and content of the provided *prepared data* depends on the requirements of consuming components.

Reports can be generated for multiple internal reasons as well as for meeting regulatory compliance obligations. For example, the USA [Congress of the United States of America 2014] and the EU [European Parliament 2016] have laws in place that demand the reporting of information about occurred incidents in certain cases. Second, *threat intelligence* can be exchanged with other SIEM systems. Therefore, threat exchange platforms like the Alien Vault Open Threat Exchange¹ can be applied.

In the *monitoring* component not automatically detected incidents can be recognized by analysts and a decision can be made, whether a recognized threat is a false positive. Additionally, after an incident was detected, experts need additional information for determining further steps in most cases. Furthermore, detection rules of the *correlation and analysis* module can be extended or created manually.

According to the Gartner magic quadrant for security information and event management [Kavanagh and Bussa 2017], IBM QRadar is the most advanced tool. Thus, we compare our proposed pattern to the one used by QRadar. The information therefore can be found in the official documentation [IBM Corporation 2017]. Therein, the QRadar architecture is structured in the layers "data collection", "data processing" and "data searches".

The first layer collects logs ("QRadar Event Collectors") and data streams ("QRadar QFlow Collector") and normalizes them by first parsing and then restructuring the data into a usable format.

The second layer enriches the logs with additional data from sources like the "QRadar Risk Manager", which provides a map of the organizations' network topology and the "QRadar Vulnerability Manager", which identifies security risks in the network. Additionally, the "Custom Rules Engine" analyses the data for incidents and offenses. After the data is analyzed, it is written to the storage, while the "QRadar Incident Forensics" provides historic in-depth investigation by storing collected raw data.

The "data searches" layer provides interfaces for human interaction. Thereby, reports can be created and the user has the possibility to search and analyze the collected data with the help of the "QRadar Console". Additionally, alerts are triggered and presented or forwarded to analysts.

To sum up, IBM QRadar covers all major SIEM modules presented in Fig. 1, although they are named differently with brand specific terms.

The DINGfest project as introduced by [Menges et al. 2018] aims at developing an architecture for next generation SIEM systems. The architecture is divided into the layers "Data Acquisition", "Data Analysis" and "Digital Forensics & Incident Reporting".

The "Data Acquisition" layer provides basic log collection with the help of Logstash², an open source tool for managing logs. In addition, virtual machine introspection [Jain et al. 2014] is used in order to monitor a system in a

¹ <https://www.alienvault.com/open-threat-exchange>

² <https://www.elastic.co/products/logstash>

virtualized environment from the outside without the need for installing an agent on the system. Both collection methods normalize their data into a semi-structured JSON format and push it into a data stream implemented with Apache Kafka³.

Within the "Data Analysis" layer the collected data gets enriched with information from "Identity Behavior Analytics". Thereby additional details are provided for logs containing user data, such as granted permissions, assigned roles and statistics. Occurred incidents can be recognized by the "Event Processing" module by matching pre-defined events stored in a fingerprint database. The "Visual Security Analytics" module provides monitoring and alerting capabilities by presenting occurred incidents and the collected data to the user.

The analysis results and relevant raw data are stored in a so called "IoC (Indicators of Compromise) Vault" in the "Digital Forensics & Incident Reporting" layer. Within this layer incidents can be shared in the STIX⁴ format, a representation for cyber threat intelligence.

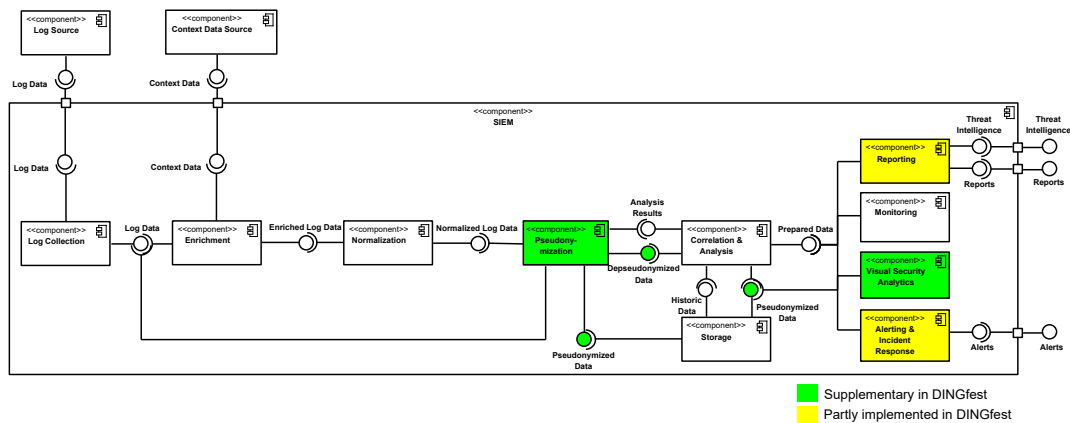


Fig. 2. Pattern of the Dingfest SIEM as UML component diagram

Although DINGfest fits the SIEM pattern shown in Fig. 1, it lacks some capabilities, but at the same time goes beyond certain aspects. The comparison between the DINGfest SIEM and the abstract SIEM pattern is portrayed in Fig. 2 and further elaborated in the following.

First, the ability to generate custom reports for internal purposes is not contained. Second, incident response capabilities are missing, though they could be appended through third party software.

On the contrary, the DINGfest project provides aspects beyond a standard SIEM. In order to preserve forensic evidence it can collect snapshots of a whole virtual machine, which gives an untampered view on the state of a compromised system. Furthermore, experts are involved more closely into the analysis process by applying advanced *visual security analytics* methods. Additionally, *pseudonymization* plays a central role in the DINGfest SIEM for providing a higher level of privacy. Therefore, all data is pseudonymized before it is stored, analyzed or further processed. Only on reasonable grounds of suspicion this data can get depseudonymized.

³<http://kafka.apache.org/>

⁴<https://stixproject.github.io/>

3.7 Variants

Of course not all current SIEM systems on the market fit into the previously defined pattern with all their parts. Especially the *enrichment* and *normalization* modules are in some cases switched. Thereby, the logs are first normalized and afterwards enriched with relevant context data. However, the input data for the *correlation and analysis* module remains the same.

Furthermore, our research revealed that most SIEM vendors provide the possibility to outsource certain parts in the cloud (e.g. storage) or similar environments. However, the overall structure is mostly not affected thereby.

An additional variant is to extend a SIEM system by active defense capabilities as described by [Docking et al. 2015]. This can improve security by mitigating attacks before they even happen. It also relies on centralized threat intelligence, which can be provided by a SIEM.

3.8 Known Uses

- *IBM QRadar*: IBM QRadar is a modular system and therefore is applicable for medium to large size companies. Furthermore it can be deployed as a standalone system, in a distribute architecture or in the cloud. [IBM Corporation 2018]
A more detailed comparison of IBM QRadar with our pattern can be found in section 3.6.
- *Splunk*: Splunk Enterprise is the core of their SIEM solution and provides collection and storage of data. It can be extended with different packages, which for example extend it by user behavior analytic capabilities. [Splunk Inc. 2018]
- *LogRhythm*: LogRhythm provides host and network monitoring abilities in addition to core SIEM functionalities. Furthermore, an AI engine aims at automating certain operations. [LogRhythm Inc. 2018]
- *McAfee ESM*: McAfee ESM can be connected with a threat intelligence feed, which provides additional information and signatures. Furthermore, it supports big data technologies like Elasticsearch and Kafka. [McAfee LLC 2018]
- *AlienVault USM & OSSIM*: Compared to the previously mentioned systems, AlienVault OSSIM is distributed as open source software. Thereby, it is the most widely used SIEM system of this kind. AlienVault USM is a paid version of OSSIM hosted in a cloud environment extended by additional features. [AlienVault Inc. 2018]

These examples are just a small selection of SIEM systems on the market fitting our pattern at a large extent. Though, they are in our estimation the best-known ones. A detailed market analysis of SIEM systems can be found in [Kavanagh and Bussa 2017].

3.9 Consequences

The SIEM pattern has the following *advantages*:

- *Detection rate and false positives*: By analyzing the data in a correlated manner, the detection rate should be much better compared to systems, which analyze only data of a single system. Additionally, context data can improve the detection rate even more. In addition, the false positive rate can be lowered by this approach.
- *Time of discovery and reaction*: The proposed solution can improve the time of discovery and reaction significantly, as it is much easier to react to events which are collected centrally instead of reacting to incidents, which could happen in any system located anywhere in the organization. Additionally, our approach enables automatic incident response.
- *Usability*: The usability is improved by providing a central user interface for monitoring different connected systems and by providing an interface for easily connecting various data sources.
- *Visual preparation of data and integration of expert knowledge*: In order to integrate expert knowledge, a monitoring module is provided. Thereby, experts can analyze the visually prepared collected data and

customize the automatic analysis. Furthermore, the interaction with humans is in general enhanced by enriching the log data with context data. As a result, for example hard to interpret information is transformed into more useful intelligence (e.g. IP-addresses get translated into the geolocation of its origin). Moreover, a high level of usability enables experts to work efficiently.

- *Degree of automation*: A high level of automation is maintained by normalizing, correlating and analyzing the data automatically. Additionally, reports can be generated without human interaction and incident response processes can be conducted autonomously. To reduce time and effort, the data is visually prepared and presented bundled at a single place for the whole companies IT infrastructure.
- *Number of connectible devices*: An interface for connecting log sources and context data is provided in order to easily connect any data generating device. Additionally, the data gets normalized into a standard format for being able to further process it independently of its initial representation.
- *Analytics*: The central component of the pattern is the correlation and analysis module, which can detect incidents or threats automatically. Therefore, the incoming data is correlated. Additionally, the monitoring module provides human analysts with the possibility to analyze data manually.
- *Reusability and intelligence sharing*: In order to provide a high level of reusability, interfaces and modules are identified. Furthermore, sharing of threat intelligence and generating reports for compliance reasons is taken into account. Especially the definition of standardized reporting and sharing formats for detected incidents plays an important role and is a very active topic in current research [Menges and Pernul 2018].
- *Costs*: In order to reduce costs, our approach aims at a high level of automation. Additionally, it supports staff by providing a single interfaces for monitoring the whole organizations' security. The modular design can help to reduce costs even further.

In contrary to the advantages the SIEM pattern has the following *liabilities*:

- *High effort for introduction*: Due to the complexity of historically grown organizations' IT infrastructures, integrating a SIEM demands high effort. Additionally, the heterogeneity of demands to the system makes it hard to create a standard solution, usable by multiple companies.
- *Demand for experts*: Although the demand for staff is reduced, there is still a need for experts for introducing and monitoring the system.
- *Lack of information from observations made by humans*: The pattern only uses log and context data generated by machines. However, humans could also provide valuable information for a SIEM. For example if an employee receives a malicious phone call, which could be the start of a large-scale attack, there is no possibility for him to feed it into the SIEM system.

3.10 Related Patterns

- *Security Patterns for Intrusion Detection Systems (IDS)*: The security pattern for IDS [Kumar and Fernandez 2012] is quite similar to the SIEM pattern proposed in this paper. However an IDS does not analyze log data, but requests from clients. Additionally, it protects certain points in a network and is thus not a centralized security system. In the context of SIEM, an IDS can serve as log source as it generates security relevant events, when detecting attacks.
- *Centralized systems logging*: The centralized systems logging pattern uses the factory pattern to create loggers for different applications to gather logs in one single place [Bijvank et al. 2013]. However, it does not address any kind of analysis or security related processing.

- *Adapter Pattern*: In order to be able to add modules with incompatible interfaces, after the SIEM system was installed, Adapter patterns [Gamma et al. 2011] are needed. Additionally, this can enable or at least simplify the connection of external data sources or consumers.

4. PATTERN MINING / IDENTIFICATION PROCESS

The process for identifying the SIEM pattern is derived from the process for pattern identification, authoring, and application introduced by [Fehling et al. 2014]. However, we will mainly focus on the identification of a SIEM pattern. In the following, we describe how those processes are applied. The pattern identification process contains five phases, which we followed accordingly:

Domain Definition: In the first step the domain has to be defined in order to gain a common understanding and knowledge about it. The domain of our pattern is SIEM, as described in chapter 2. In addition, we describe the used terminology in the appendix in order to gain a deeper understanding of domain specific terms.

Coverage Consideration: As our domain is a quite broad research area, we narrow down the coverage of our pattern. Specifically, we limit our pattern to one level of abstraction. In addition, only the software part of SIEM is considered. We do not focus on surrounding processes, which for example are performed by staff in a SOC.

Information Format Design: The information format, we applied in our research is UML. Specifically, we decided to use a component diagram, as it perfectly fits the intended level of abstraction and is well known by most researchers in the field of software engineering.

Information Collection: To gain a holistic view of the relevant aspects we considered several sources of information:

- *Software products*: The most important source of information for our research are products that are already well established in the market of SIEM. In this way the outlined pattern represents the state of the art of SIEM systems and is as close to reality as possible.
- *Research papers*: In order to integrate the current view on SIEM in research we also used scientific papers as source of information. However, there are only few papers, which specifically address SIEM as a whole.
- *Whitepapers*: In most cases companies, developing SIEM systems provide whitepapers, elaborating the functionalities of their products in more detail. In some cases even their architecture is displayed.
- *Manuals*: With the help of the corresponding manuals of the SIEM products additional information can be gained. However, manuals are written for the users of the product and thereby only give few information about the functionality of the product in depth.
- *Documentations*: SIEM products usually provide several kinds of documentations. They mostly target developers, who for example want to extend the software with plug-ins. In addition, documentations for implementing SIEM in a company are a valuable source of information to gain insight into the structure of the software.
- *Product websites*: Websites of the SIEM product serve as additional source of information. However, they normally provide a very shallow insight into the software. Furthermore, they mostly serve marketing purposes. Thus, this information should be treated with caution.

Information Review: To reduce the amount of different sources of information, we have only analyzed chosen SIEM systems. For choosing the most advanced systems, the Gartner Magic Quadrant for Security Information and Event Management [Kavanagh and Bussa 2017] was utilized. Thereby, available SIEM systems are classified regarding the two dimensions *ability to execute* and *completeness of vision*. The first dimension essentially describes the usefulness of the system in daily usage inside an organization and the second one the strategic orientation. We analyzed all systems, classified as leaders: "IBM QRadar", "Splunk Enterprise Security", "LogRhythm Enterprise", "HPE ArcSight" and "McAfee Enterprise Security Manager". In addition, the open source SIEM called "OSSIM" is considered, as it gives a deeper insight into its functionality and is frequently utilized in research projects.

5. CONCLUSION AND FUTURE WORK

In this paper we have deduced a generic SIEM pattern by among other things, analyzing the most advanced SIEM systems on the market. This was done by adhering to a pattern identification process published in literature. In order to derive the pattern, forces that SIEM systems depend on were identified. The pattern was visualized as a UML component diagram and its advantages and liabilities were discussed. In order to give deeper insights, the pattern was compared to a commercially used SIEM and a SIEM resulting from a research project. This paper can be applied by developers of SIEM tools for delimiting modules in order to improve the general structure. Furthermore, it can serve as a foundation for understanding the functional principles and assembly of SIEM.

However, the proposed pattern is not enough to depict SIEM in detail. Thus, it is necessary to continue our research by creating further patterns, explaining the corresponding modules which we have identified in more detail and divide them each into modules themselves.

APPENDIX

A. Terminology

In this section the relevant terminology is outlined. Thereby definitions for specific terms are given in order to establish a common understanding. We deduce the definitions valid in the context of SIEM whereby they may not be applicable in other contexts.

Event / Log: In the context of SIEM the nouns event and log are frequently used synonymously. However, a log is the record of an occurred event. An event can simply be defined as something that happens and can come in different sizes and levels of abstraction (e.g. opening a text file compared to processing an order of a customer) [Luckham 2012]. Thus, multiple logs can describe one event. Furthermore, logs can be generated from a multitude of devices or programs and their representation and structure can vary greatly. Additionally, logs can also be emitted as a data stream. Logs most pertinent for SIEM systems originate from security-relevant devices, like firewalls and routers, but are not restricted to those.

Incident: According to [Cichonski et al. 2013] an incident in the context of IT security is a special form of an event. Thereby it is restricted to adverse events which include loss of data confidentiality or disrupt integrity or availability of systems or data. Additionally, the violation of security policies and standard security practices can be an incident.

IT security analyst: Security analysts in the context of SIEM are usually organized in a SOC. An IT security analyst is an expert in the domain of cyber security and thus is trained to identify vulnerabilities and attacks on an organizations' IT infrastructure. Thereby, the expert analyzes different kinds of data and uses several tools in order to detect anomalies. In connection with SIEM, this data usually consists of logs, which are visually prepared and presented to facilitate its analysis.

ACKNOWLEDGMENTS

We thank our shepherd, Eduardo B. Fernandez, for his valuable comments that have significantly helped to improve this paper.

REFERENCES

- AlienVault Inc. 2018. OSSIM: The Open Source SIEM. (2018). Retrieved 24.09.2018 from <https://www.alienvault.com/products/ossim>
- Sandeep Bhatt, Pratyusa K. Manadhata, and Loai Zomlot. 2014. The Operational Role of Security Information and Event Management Systems. *IEEE Security & Privacy* 12, 5 (2014), 35–41. DOI:<http://dx.doi.org/10.1109/MSP.2014.103>
- Roland Bijvank, Wiebe Wiersema, and Christian Köppe. 2013. Software architecture patterns for system administration support. In *Proceedings of the 20th Conference on Pattern Languages of Programs*. 1.

- Frank Buschmann, Regine Meunier, Hans Rohnert, Peter Sommerlad, and Michael Stal. 2013. *Pattern-Oriented Software Architecture, A System of Patterns* (1. Aufl. ed.). Wiley, s.l.
- Paul Cichonski, Tom Millar, Tim Grance, and Karen Scarfone. 2013. Computer security incident handling guide. *International Journal of Computer Research* 20, 4 (2013), 459.
- Congress of the United States of America. 2014. National Cybersecurity and Critical Infrastructure Protection Act of 2014. (2014).
- Christopher Crowley. 2017. Future SOC: SANS 2017 Security Operations Center Survey. (2017).
- Michael Docking, Anton V. Uzunov, Chris Fiddymont, Richard Brain, Scott Hewett, and Lee Blucher. 2015. UNISON: Towards a Middleware Architecture for Autonomous Cyber Defence. In *2015 24th Australasian Software Engineering Conference*. IEEE, 203–212. DOI:<http://dx.doi.org/10.1109/ASWEC.2015.29>
- European Parliament. 2016. DIRECTIVE (EU) 2016/1148 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. (2016).
- Christoph Fehling, Johanna Barzen, Uwe Breitenbücher, and Frank Leymann. 2014. A process for pattern identification, authoring, and application. In *Proceedings of the 19th European Conference on Pattern Languages of Programs - EuroPLoP '14*, Veli-Pekka Eloranta and Uwe van Heesch (Eds.). ACM Press, New York, USA, 1–9. DOI:<http://dx.doi.org/10.1145/2721956.2721976>
- Erich Gamma, Richard Helm, Ralph E. Johnson, and John Vlissides. 2011. *Design patterns: Elements of reusable object-oriented software* (39. printing ed.). Addison-Wesley, Boston.
- Gartner Inc. 2018. Security Information and Event Management (SIEM). (2018). Retrieved 20.06.2018 from <https://www.gartner.com/it-glossary/security-information-and-event-management-siem>
- Carrie Gates and Sophie Engle. 2013. Reflecting on visualization for cyber security. In *2013 IEEE International Conference on Intelligence and Security Informatics*. IEEE, 275–277. DOI:<http://dx.doi.org/10.1109/ISI.2013.6578842>
- Markus Goldstein, Stefan Asanger, Matthias Reif, and Andrew Hutchison. 2013. Enhancing Security Event Management Systems with Unsupervised Anomaly Detection. In *ICPRAM*. 530–538.
- IBM Corporation. 2017. IBM Security QRadar: Architecture and Deployment Guide. (2017). Retrieved 24.09.2018 from https://www.ibm.com/support/knowledgecenter/SS42VS_7.3.1/com.ibm.qradar.doc/b_siem_deployment.pdf
- IBM Corporation. 2018. IBM QRadar SIEM. (2018). Retrieved 24.09.2018 from <https://www.ibm.com/us-en/marketplace/ibm-qradar-siem>
- Bhushan Jain, Mirza Basim Baig, Dongli Zhang, Donald E. Porter, and Radu Sion. 2014. SoK: Introspections on Trust and the Semantic Gap. In *2014 IEEE Symposium on Security and Privacy*. IEEE, 605–620. DOI:<http://dx.doi.org/10.1109/SP.2014.45>
- Kaspersky Lab. 2016. Measuring financial impact of it security on businesses: IT Security Risks Report 2016. (2016).
- Kelly M. Kavanagh and Toby Bussa. 2017. Magic Quadrant for Security Information and Event Management. *Technical Report - Gartner Inc.* (2017).
- Ajoy Kumar and Eduardo B. Fernandez. 2012. Security patterns for intrusion detection systems. In *1st LACCEI International Symposium on Software Architecture and Patterns (LACCEI-ISAP-MiniPLoP'2012)*, Panama City, Panama.
- LogRhythm Inc. 2018. LogRhythm, The Security Intelligence Company. (2018). Retrieved 24.09.2018 from <https://logrhythm.com/>
- David C. Luckham. 2012. *Event processing for business: Organizing the real time strategy enterprise*. Wiley, Hoboken, N.J. <http://onlinelibrary.wiley.com/book/10.1002/9781119198697>
- McAfee LLC. 2018. McAfee Enterprise Security Manager. (2018). Retrieved 24.09.2018 from <https://www.mcafee.com/enterprise/en-us/products/enterprise-security-manager.html>
- Florian Menges, Fabian Böhm, Manfred Vielberth, Alexander Puchta, Benjamin Taubmann, Noëlle Rakotondravony, and Tobias Latzo. 2018. Introducing DINGfest: An architecture for next generation SIEM systems. In *SICHERHEIT 2018*, Hanno Langweg, Michael Meier, Bernhard C. Witt, and Delphine Reinhardt (Eds.). Gesellschaft für Informatik e.V, Bonn, 257–260.
- Florian Menges and Günther Pernul. 2018. A comparative analysis of incident reporting formats. *Computers & Security* 73 (2018), 87–101. DOI:<http://dx.doi.org/10.1016/j.cose.2017.10.009>
- Sabina Georgiana Radu. 2016. Comparative Analysis of Security Operations Centre Architectures; Proposals and Architectural Considerations for Frameworks and Operating Models. In *Innovative Security Solutions for Information Technology and Communications*, Ion Bica and Reza Reyhanitabar (Eds.). Springer International Publishing, Cham, 248–260.
- Timo Ropinski, Daniel Archambault, Min Chen, Ross Maciejewski, Klaus Mueller, Alexandru Telea, and Martin Wattenberg. 2017. How do Recent Machine Learning Advances Impact the Data Visualization Research Agenda?. In *IEEE VIS Panel*. Phoenix.
- Splunk Inc. 2018. Splunk. (2018). Retrieved 24.09.2018 from <https://www.splunk.com/>
- Eva Weishäupl, Emrah Yasasin, and Guido Schryen. 2018. Information security investments: An exploratory multiple case study on decision-making, evaluation and learning. *Computers & Security* (2018). DOI:<http://dx.doi.org/10.1016/j.cose.2018.02.001>
- Amrit T. Williams and Mark Nicolett. 2005. Improve IT Security With Vulnerability Management. *Technical Report - Gartner Inc.* (2005).

SLPLoP'18, NOVEMBER 20-23, Valparaíso, Chile. Copyright 2018 is held by the author(s). HILLSIDE 978-1-941652-11-4

A Security Information and Event Management Pattern — Page 12

2 Formalizing and Integrating User Knowledge into Security Analytics

Current status:	Under Review
Journal:	SN Computer Science
Date of submission:	30 September 2021
Full citation:	BÖHM F., VIELBERTH, M., AND PERNUL, G. Formalizing and Integrating User Knowledge into Security Analytics. Submitted to <i>SN Computer Science</i> , (2021), 1–28
Authors' contributions:	Fabian Böhm 45% Manfred Vielberth 45% Günther Pernul 10%

Journal description: SN Computer Science is a broad-based, peer reviewed journal that publishes original research in all the disciplines of computer science including various inter-disciplinary aspects. The journal aims to be a global forum of, for, and by the community.

Formalizing and Integrating User Knowledge into Security Analytics

Fabian Böhm^{1*}, Manfred Vielberth¹ and Günther Pernul¹

^{1*}Chair of Information Systems, University of Regensburg,
Universitätstr. 31, Regensburg, 93053, Bavaria, Germany.

*Corresponding author(s). E-mail(s): fabian.boehm@ur.com;
Contributing authors: manfred.vielberth@ur.com;
guenther.pernul@ur.com;

Abstract

In our cyber-physical world, an ever-increasing number of enterprise assets is interconnected, leading to increasingly complex infrastructures within organizations. Due to these and similar developments, companies are becoming increasingly vulnerable to cyber attacks and cyber-physical attacks. In addition, many current attacks not only exploit technical vulnerabilities but try to gain access through phishing or social engineering. As traditional security measures and systems repeatedly prove to be unreliable in effectively detecting such attacks, people and their knowledge prove to be a critical factor for cyber security. Therefore, an organization needs to maintain an overview of the security knowledge distributed throughout the enterprise. However, there is no uniform understanding of the concept of knowledge in the security analytics environment. Our research contributes to filling this gap by formalizing the concept of knowledge in the context of cybersecurity and establishing a corresponding conceptual knowledge model. This enables a better classification of existing related research and the identification of potentials for future work. In particular, improved collaboration among domain experts and stronger cooperation between humans and machines could leverage previously untapped but essential knowledge. For example, this knowledge is of extraordinary importance in creating policies and security rules in existing security analytics systems. For this purpose, we present a proof of concept that uses visual programming methods to show how security novices can easily contribute their domain knowledge to improve an organization's security posture.

2 *Formalizing and Integrating User Knowledge into Security Analytics*

Keywords: Security Analytics, Domain Knowledge, Visual Analytics, Security Awareness, Security Operations

1 Introduction

Although a lot of money and effort is invested into awareness campaigns and professional training, humans within cybersecurity are still widely considered the weakest link of an organization's cyber defenses [1]. However, this simplification in no way does justice to the role of humans in modern Security Analytics ¹ (SA), because their domain knowledge is invaluable for any effective and efficient SA operation [2, 3]. So far, SA approaches have essentially been limited to integrating the knowledge of security experts to decide, for example, whether identified indicators actually represent malicious incidents or just unusual but benign activities. From our point of view, this is a major shortcoming of existing SA approaches, as it is equally important to include the knowledge of non-security domains in SA processes.

This shortcoming becomes evident in the context of the ever-growing Internet-of-Things (IoT), Industry 4.0, and ubiquitous Cyber-Physical Systems (CPS). All of these trends are leading to increased connectivity of a company's internal and external physical assets. Quite apart from the already skyrocketing number of cyberattacks, the attack surface for cyber-physical attacks is significantly increasing due to this trend. The cyber-physical attacks specifically use the connection between information (cyber) systems and physical systems within CPSs or the IoT to cause actual physical harm to machines or people [4]. Detecting and averting, or mitigating, such multidimensional attacks poses a challenge to existing security measures. To achieve comprehensive security, they must monitor all assets of an organization, which are connected in some way to cyberspace. With the progressive implementation of the IoT and Industry 4.0, these systems range from firewalls or individual sensors to cyber-physical systems such as complete manufacturing lanes. The problem in the context of these CPS is that traditional security measures and systems, such as a Security Information and Event Management (SIEM) system used in a Security Operations Center (SOC), are not able to sufficiently and effectively protect the CPS due to a lack of knowledge and capabilities [5, 6].

Security experts can make well-informed decisions in this context to identify incidents in cyberspace. However, they lack crucial knowledge about the physical domain. For this reason, they often cannot effectively decide whether, for example, a turbine used to generate electricity is operating within its specification or may have a problem [7]. However, engineers and appropriately trained staff have that knowledge of physical operations to decide whether or

¹Since security analytics has not yet been universally defined we interpret this term as a collection of methods for proactively identifying attacks and threats by analyzing and correlating collected data.

not the turbine is behaving normally. In turn, however, these employees lack the know-how to contribute to effective SA [6].

This imbalance limits an organization's ability to implement holistic SA methods that could reliably detect indicators of both cyber and cyber-physical attacks. For this reason, it is necessary to integrate the knowledge of engineers and the like into security operations. Only then can incidents related to physical assets also be effectively detected and prevented [8].

In recent years, neither research nor practice has been able to establish effective means to integrate the knowledge of employees away from security experts into their security analyses. With this work, we make a twofold contribution to address this problem. To establish a unified and fundamental vocabulary for this research domain, we first define the different types of knowledge and knowledge conversions relevant to cybersecurity. We then present a model for knowledge-based SA which integrates knowledge into core processes of SA. This is a valuable contribution, as no security-specific definition of different knowledge aspects exists so far. Their formal definition can also build future research on a well-defined foundation. Our second contribution addresses quite explicitly the lack of integration of domain knowledge as an open issue within knowledge-based SA by presenting a research prototype that allows experts to integrate their knowledge into active security measures.

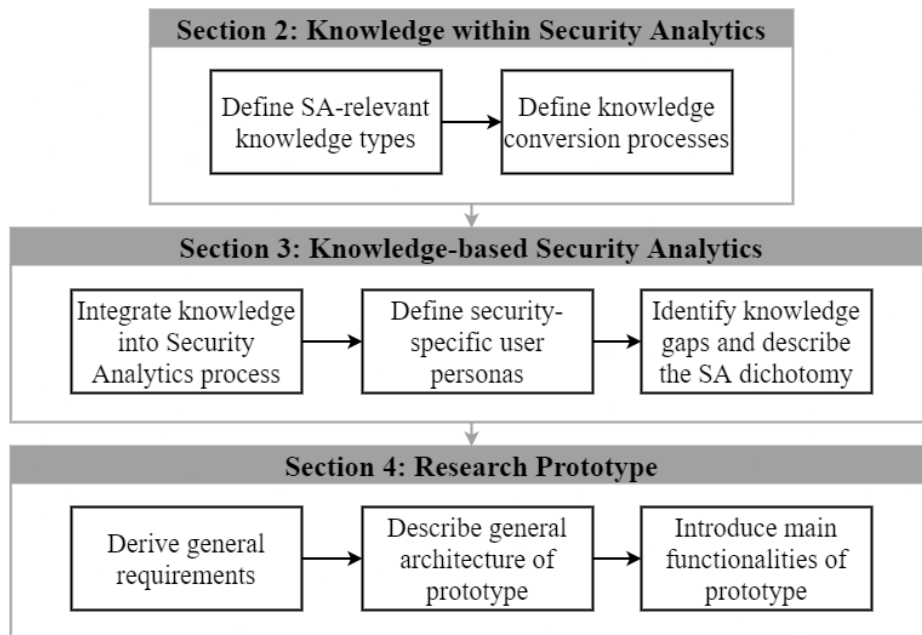


Fig. 1 Schematic of this work's main contribution.

The main contributions of this work are structured according to Figure 1. The remainder of this paper is organized as follows. We formally describe relevant notions of knowledge and conversion processes for Security Analytics in Section 2. These formal definitions allow us and any future work to have

4 *Formalizing and Integrating User Knowledge into Security Analytics*

a well-defined, precise vocabulary. In the next step, this vocabulary is integrated into an Incident Detection Lifecycle for a cohesive picture of what we call knowledge-based SA within Section 3. Besides several knowledge gaps, the resulting model reveals a significant dichotomy in current SA approaches, which is not yet appropriately addressed. Thus, we present a research prototype in Section 4 showcasing a possible approach to integrate security novice's domain knowledge into an exemplary SA solution, i.e., a signature-based incident detection component. The prototype highlights that the implementation of knowledge-based SA requires innovative approaches but can drastically improve cybersecurity. Finally, Section 5 concludes our work and points out possible directions for future work.

This article is an extended version of our work presented at the 7th International Conference on Information Systems Security and Privacy 2021 (ICISSP, February 2021) [9], kindly invited for consideration in this journal. The first difference to the initial work is a more in-depth and better-structured preparation of the formal definitions of integrating knowledge into security measures. The most fundamental change is an extended and adapted version of the model for knowledge-based security analytics. This adaptation has been based on feedback and new insights since the publication of our original paper. We also provide a more detailed insight into the dichotomy of security analytics derived from the improved model. In a final step, we defined common requirements as a basis for our prototype.

2 Knowledge within Security Analytics

In this chapter, we provide a detailed insight into the different knowledge aspects that play a crucial role in the context of current SA operations. For this purpose, we establish a formal understanding of the types of knowledge and the processes for knowledge conversion. While Sallos et al. [10] present the importance of cybersecurity-related knowledge on an abstract and management-oriented level, we aim at the implications of integrating knowledge into security measures in the following sections.

2.1 Knowledge Types

Scientific literature describes a variety of different, sometimes even contradictory, definitions of the term “knowledge” and the different sub-aspects related to it. A frequently cited definition that provides a clear starting point for opening up the concept of knowledge is the data-information-knowledge-wisdom (DIKW) hierarchy, which defines “knowledge” as the application of data and derived information to answer “how” questions [11]. However, information systems research often criticized the DIKW hierarchy as unsound and undesirable [12]. A more human-oriented definition by Davenport describes knowledge as a mixture of experience, intuition, values, contextual information, and expert insight [13].

This definition of Davenport appropriately explains the concept of knowledge from a human point of view; however, knowledge is not bound to humans. Instead, corresponding research emphasizes that knowledge can also be captured in documents, memos, and the like [14]. Following this route, it is only logical to conclude that knowledge can also be stored within IT. This knowledge within IT is different from human knowledge, especially if it is also generated by IT through some kind of automatic analysis [15, 16]. Based on this line of thought, it is an established and accepted procedure to distinguish between two basic types of knowledge: explicit knowledge and tacit (or implicit) knowledge [14].

All these aspects clearly show that the term “knowledge” is difficult to define in a generally valid way. Instead, different facets of knowledge must be distinguished and embedded in the relevant context. This is because even the notions of explicit and tacit knowledge are still too abstract in their basic form to be incorporated into processes of SA. For this reason, we define and formalize below different notions of knowledge that are of central importance in the field of cybersecurity and even more specifically in the field of SA.

2.1.1 Explicit Knowledge

Explicit knowledge K^e is mainly referred to as machine-based knowledge. Accordingly, this term denotes knowledge that machines can read, process and store [14]. In the context of SA, we distinguish three types of explicit knowledge in the further course of the work, which can be distinguished from each other by their intended use for SA. The transitions between these different types of explicit knowledge are fluid, i.e., a given machine-readable object can also be assigned to a different expression depending on the current context and use case. Equation 1 defines explicit knowledge formally as a union of the three sub-aspects defined in the following paragraphs.

$$K^e = K_m^e \cup K_s^e \cup K_i^e \quad (1)$$

Models (K_m^e): Models for machine learning approaches, neural networks, and the like are primarily used for anomaly-based detection mechanisms. This knowledge allows a machine to detect outliers and evaluate them to some extent as to whether they indicate malicious or undesirable behavior.

Signatures & Rules (K_s^e): Like models for machine learning approaches, signatures and rules are also to be valued as explicit knowledge, especially in signature-based security analytics methods. They are the basis for more traditional SA approaches such as SIEM systems and their correlation engines for detecting indicators of compromise (IoC).

Threat Intelligence & Forensic Evidence (K_i^e): Threat Intelligence and forensic evidence describe the results of primarily manual, in-depth analysis of suspected or actual incidents and include extensive information on the attackers’ modus operandi, identifiable traces, suspect groups or individual

6 *Formalizing and Integrating User Knowledge into Security Analytics*

perpetrators, and many other details. Because of their level of detail, Threat Intelligence and Forensic Reports allow answering “how” questions.

2.1.2 Implicit Knowledge

After contextualizing explicit knowledge in SA, we turn to so-called tacit knowledge in the following paragraphs. This kind of knowledge can only be possessed by humans and is very specific to each individual [17]. Although “tacit knowledge” would be a more commonly used term, we will use “implicit knowledge” K^i in this paper to clarify the distinction from the explicit knowledge of a machine.

Humans improve their K^i by combining new insights with existing knowledge. The existing knowledge itself can in turn be divided into, on the one hand, domain knowledge and, on the other hand, operational knowledge [18]. However, in the context of SA, we consider this differentiation too vague. To describe the problem at hand concisely, a more fine-granular and contextualized view on K^i is necessary. In the domain of SA, we also consider another new type of tacit knowledge to be highly relevant: situational knowledge. As for explicit knowledge, we also define implicit knowledge as a union of its three main facets (c.f. Equation 2). We go into more detail about these three aspects of tacit knowledge in the following paragraphs.

$$K^i = K_d^i \cup K_s^i \cup K_o^i \quad (2)$$

Domain Knowledge (K_d^i): Generally speaking, domain knowledge describes what people know about a particular context or on a specific topic (the “domain”) [2, 6]. For SA, we define K_d^i in Equation 3 in a more detailed way as a combination of two disjoint subdomains $K_{d(sec)}^i$ and $K_{d(nonSec)}^i$. $K_{d(sec)}^i$ comprises security-related domain knowledge, which is mainly part of the tacit knowledge of security experts. The components of $K_{d(sec)}^i$ are all safety and security aspects considered from a cybersecurity perspective. For example, this includes knowledge about firewall rules in use, the ability to identify suspicious network connections or unauthorized access to classified information. This facet of domain knowledge is to some extent already considered in several SA means [19]. In contrast, there is a lack of integration of $K_{d(nonSec)}^i$. Under this second aspect of general domain knowledge, we summarize non-security domain knowledge. This type includes domains such as manufacturing or engineering. The knowledge from these domains is of high importance to detect incidents on cyber-physical systems [5]. An example of domain knowledge not directly related to security is the expected Rounds per Minute (RPM) of a power turbine or the maximum temperature for a blast furnace. However, in the context of SA, this knowledge is necessary to assess the CPS’s security posture cohesively. For SA, both components of domain knowledge are necessary to build and operate comprehensive security operations. Especially in light of the challenges associated with CPS and the rise of cyber-physical attacks,

Formalizing and Integrating User Knowledge into Security Analytics 7

the integration of $K_{d(nonSec)}^i$, in particular, is one of the biggest challenges currently faced by SA research.

$$K_d^i = K_{d(sec)}^i \cup K_{d(nonSec)}^i \quad (3)$$

Situational Knowledge (K_s^i): Situational knowledge is a new type of knowledge previously not acknowledged, which we consider crucial in the SA environment. In SA, this type mainly encompasses the concept of situational awareness, which also plays a vital role in cybersecurity in recent years, especially in the form of Cyber Situational Awareness [20, 21]. K_s^i describes the ability of any employee of an organization to perceive unusual events or suspicious behavior. The relevant events range from receiving suspicious mail, which represents a possible phishing attempt, to identifying a private storage medium connected to a corporate device. With the appropriate situational security knowledge, which has been imparted, for example, through security awareness training or campaigns [22], employees can evaluate the e-mail or the storage medium from a security perspective and deduce that these events could pose a threat to the company. However, specific domain knowledge K_d^i about the SA of the enterprise is not required to make these inferences.

Operational Knowledge (K_o^i): Operational knowledge in the context of SA refers to the ability of a human to operate specific systems. Specifically, employees with SA-related operational knowledge can adequately operate a company's security systems. This ability can relate to a wide variety of systems. For example, employees may have the experience to define correlation rules for a SIEM, fine-tune models for anomaly- or behavior-based SA approaches, or create new threat intelligence. It is important to note here that K_o^i does not refer to expertise, such as the syntax of the threat intelligence format used, but rather to the ability to deal with the corresponding IT system.

These three different subsets of tacit knowledge are necessary to detect and resolve both cyber and cyber-physical attacks as completely as possible. K_d^i and K_s^i would, in a perfect world, need to be comprehensively integrated into an organization's SA systems. They are the pre-requisite to cohesive security operations, especially in the context of CPS and IoT. However, operational knowledge K_o^i represents the barrier to entry for this integration. Only with the necessary K_o^i can employees, for example, define an appropriate SIEM correlation rule based on their K_d^i .

2.2 Knowledge Conversion

The different knowledge types can be converted into each other. Nonaka and Takeuchi define the knowledge conversions between explicit and tacit knowledge in terms of four different knowledge conversion processes [14]. Various research directions have picked up upon this formalization to formally describe the exchange and interaction between humans and machines. Especially research in the field of information visualization and human-machine

8 *Formalizing and Integrating User Knowledge into Security Analytics*

interaction work frequently and intensively with these concepts [19, 23, 24]. In SA, corresponding knowledge exchange is also desirable in the underlying approaches since effective security operations require both automated discovery processes (involving explicit knowledge) and the expertise of different human experts (and their tacit knowledge). To provide the necessary foundation and common vocabulary regarding knowledge conversion in SA after defining the aspects of knowledge, we formalize the four key knowledge conversion processes for SA in the following paragraphs.

Internalization (int): Internalization describes the process of making explicit knowledge available to users, who can then perceive this knowledge using the K_o^i available to them and convert it into $K_{d(sec)}^i$ or K_s^i (Eq. 4). How efficient this process is and how significant the increase in implicit knowledge is, depends strongly on the respective user's level of K_o^i . We have implied this dependence in the formal definition in Equation 4 by defining operational knowledge as a catalyst for the *int* conversion process. This notation is adopted from the formal descriptions of chemical reactions. Effective internalization *int* of K^e is supported primarily by any kind of visual representation of the K^e . For security-related domain knowledge, this includes examples like visualizing the raw data that led to the triggering of a SIEM rule, the rule itself, and the components of the data that were conducive to the decision.

$$int : (K^e \xrightarrow{K_o^i} K_{d(sec)}^i \cap K_s^i) \quad (4)$$

Externalization (ext): When tacit knowledge, especially K_d^i or K_s^i , is transferred into a form that can be processed by computers, we refer to this as the process of externalization (Eq. 5). Externalized tacit knowledge can thus be read, persisted, and eventually processed by computers. A variety of examples for externalization can be found in the context of modern security analytics. For example, this process includes the direct adaptation of model parameters (i.e., K_m^e) and the formulation of rules for signature-based analysis (i.e., K_s^e). Structuring and formalizing indicators, incidents, and corresponding evidence into CTI (i.e., K_i^e) also represents a form of externalization. Here, direct access to explicit knowledge and possibilities for active processing of the same are of primary importance. Thus, the corresponding operational knowledge K_o^i is again a fundamental prerequisite for enabling and performing externalization. Only if the human being can operate a system (e.g., SIEM system with the corresponding correlation rules), the possibility to externalize implicit knowledge can be retained. The process described here for translating tacit to explicit knowledge is also necessary for avoiding the loss of any K^i due to, for example, the retirement of a security analyst from the company. If there is no possibility to keep the knowledge of this security analyst in the company, this poses a risk for the company [25].

$$ext : (K_d^i \cap K_s^i \xrightarrow{K_o^i} K^e) \quad (5)$$

Formalizing and Integrating User Knowledge into Security Analytics 9

Combination (comb): The conversion process of combination describes the exchange of knowledge from two or more explicit knowledge bases (Eq. 6). At the same time, it can also mean the exchange of knowledge between corresponding K^e . Concerning the explicit knowledge types defined in Section 2.1.1, *comb* can describe both the exchange of knowledge within a constant knowledge type but also the transfer of knowledge from one type to another. An example of the first process is the exchange of cyber threat intelligence (CTI) and forensic evidence between different actors ($K_i^e \mapsto K_i^e$). The second process may be, for example, using CTI to define new or adapt existing rules for signature-based incident detection ($K_i^e \mapsto K_s^e$).

$$comb : K^e \mapsto K^e \quad (6)$$

Collaboration (coll): In the context of collaboration, multiple individuals work together and combine their K^i (Eq. 7). Less formally, this knowledge conversion specifies that people can learn from each other (i.e., increase their K^i) by collaborating. This process is difficult to capture and formally define because it is purely implicit without any direct indication that it is happening. Even a simple conversation between two people can correspond to a knowledge conversion. However, we interpret collaboration in the context of SA as a process that is supported by technology. Accordingly, operational knowledge of collaborators is again required to enable collaboration, as also indicated in Equation 7 by K_o^i as the catalyst of collaboration. Collaborators can thus work together, for example, in correlating various indicators of compromise to determine which indicators genuinely represent a threat. To do this, they could use an appropriate analysis tool designed for just such a purpose. On the one hand, the tool support enables remote collaboration among the employees, but at the same time, they need the operational knowledge to be able to operate this tool. Collaboration supported in this technological way enables users to share knowledge and learn from each other. With respect to SA and the need to involve $K_{d(sec)}^i$ and $K_{d(nonSec)}^i$, appropriate knowledge sharing is vital between collaborators to enable the most comprehensive SA possible. Also, for example, tool-based training or workshops can be defined as a type of collaboration. These workshops often impart domain knowledge to improve the situational knowledge of other collaborators ($K_d^i \mapsto K_s^i$).

$$coll : K^i \xrightarrow{K_o^i} K^i \quad (7)$$

3 Knowledge-based Security Analytics

After a basic understanding of the formal knowledge types and the processes describing the conversion among these knowledge types is established in Chapter 2, the following chapter is dedicated to embedding these concepts into the core activities of security analytics. In this context, we interpret the detection of security incidents, i.e., attacks on an organization's assets, to be

10 *Formalizing and Integrating User Knowledge into Security Analytics*

the the essential task of SA [26]. A cohesive approach to implementing this task requires comprehensive data collection combined with powerful analytical capabilities and the integration of any available knowledge base. Thus, we introduce our model for knowledge-based SA based on an extended, general process for SA and the critical role that knowledge plays in this context. Based on this, we identify different personas of users that play a role in knowledge-based SA. Finally, we explain the central problem faced by SA in the context of current developments such as the Internet of Things and Industry 4.0, which we refer to as the “Dichotomy of Security Analytics”.

3.1 Incident Detection Lifecycle

The starting point for our model of knowledge-based security analytics is the incident detection process defined by Menges and Pernul [27]. This process describes four basic steps involved in incident detection: *Data*, *Observables*, *Indicators*, and *Incidents*. *Data* of a system under consideration are collected and normalized, resulting in so-called *Observables*. The authors refer to detected anomalies in these observables as Indicators of Compromise or just *Indicators*. Only the combination of several indicators finally confirms a recognized textitIncident. While this simple model describes the core activities for detecting security incidents, it neglects two central aspects of modern security analytics. First, an incident detection is usually followed by a post-incident analysis to extract and secure forensic evidence. Second, the subsequent analysis of an incident can also serve to generate threat intelligence, which can again be used to detect indicators or specific incidents. Incident detection is thus an iterative process in which the output (threat intelligence) can be used as an input in further detection runs. For this reason, we are extending the original Incident Detection Process to an Incident Detection Lifecycle, which more appropriately reflects the processes within modern security analytics.

Figure 2 represents this adapted and extended lifecycle. The boxes highlighted in gray represent the central results of the activities. The SA activities themselves are annotated at the edges of the model. The starting point of the Incident Detection Lifecycle is some real event within an organization – that is, something that “happens” –, which can be physical or digital. Examples of such a *situation* are the authentication of a user at an IT system or the use of a private USB stick at company computers. We refer to these events as *situations* in the further course. The Incident Detection Lifecycle is divided into three overarching phases, which are executed in order to detect, resolve, and understand incidents. Overall, the Incident Detection Lifycycle provides a more detailed view on the Detect and Respond phases of the established NIST Cybersecurity Framework [28].

The first of these phases is the *Data Collection*. Each *situation* produces raw data which could be relevant for the detection of possible attacks. These *data* are normalized (and sometimes standardized) in the first phase of the lifecycle, producing so-called *observables*. Observables can thus be understood as normalized representations of the raw data available about the situation.

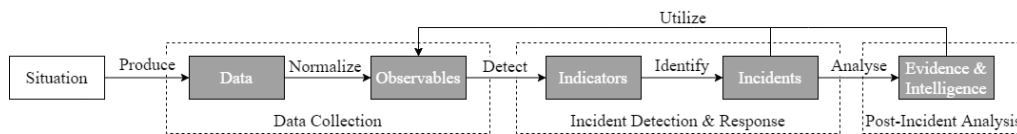


Fig. 2 The Incident Detection Lifecycle.

They are not yet attributed and thus have no significance for why something happened or who might be responsible for it. Observables only serve as input for the second phase of the Incident Detection Lifecycle.

This second phase of the lifecycle can be summarized under the terms *Incident Detection & Response*. This phase aims to detect actual incidents, capture the impact, and contain the incident as quickly as possible. The first step is the detection of *indicators*, which are often also referred to as Indicators of Compromise (IoC). These indicate potentially suspicious activities and behaviors within the observables. However, IoCs can also indicate unusual but not malicious behavior. For this reason, a further step is necessary to identify actual *incidents* from detected indicators. For this purpose, it is necessary to correlate indicators with each other and possibly to include additional data or observables in the analysis process. However, if an incident is identified, direct measures for defense and containment must be initiated in this lifecycle phase.

After the initiation and implementation of countermeasures and containment actions, the third phase of the Incident Detection Lifecycle, the *Post-Incident Analysis*, is carried out. In this phase, careful and intensive analyses of an incident produce further vital artifacts. On the one hand, *evidence* which can be used in possible judicial proceedings is collected in this step through forensic analysis. On the other hand, *threat intelligence* is generated through the attribution of the identified incident. Since the gained intelligence can also be crucial to detect new indicators or identify similar incidents, it feeds into the previous phases, creating an iterative lifecycle.

3.2 Knowledge Model

The Incident Detection Lifecycle can now be extended to a model for knowledge-based SA in the next step. In the course of this extension, the knowledge terms and conversion processes introduced in Section 2 are integrated into the lifecycle to obtain a comprehensive picture of the stages in the lifecycle at which knowledge and knowledge exchange play a central role. The extended model is shown in Figure 3. In the following paragraphs, we will go through this knowledge model in detail to highlight the significant adjustments made compared to the original Incident Detection Lifecycle.

To recognize indicators in a considerable amount of observables and, above all, to derive correct indicators is an enormously challenging task. Due to the sheer amount of observables that need to be monitored, this task is primarily automated in modern SA systems [26]. The corresponding processes in the *Incident Detection & Response* phase of the lifecycle use explicit knowledge K^e in the form of signatures or rules (K_s^e) for signature-based detection, but also

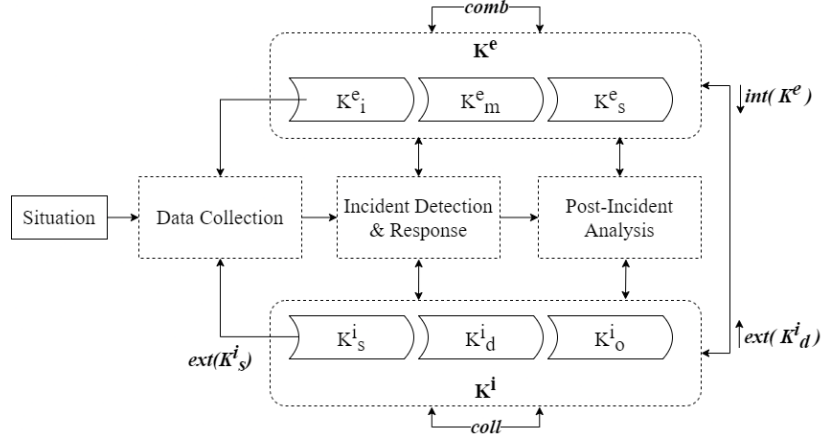
12 *Formalizing and Integrating User Knowledge into Security Analytics*

Fig. 3 The Incident Detection Lifecycle extended with Knowledge Types and Conversions.

models (K^e_m) for behavior-based procedures. Thus, explicit knowledge plays a central role, especially for incident detection. Nevertheless, a pure focus of incident detection on K^e is not purposeful and can even be associated with direct limitations. First of all, with the use of K^e_s only indicators and incidents that were known apriori and whose signatures were integrated into the system, can be detected. Behavior-based methods are better at classifying unknown indicators but often tend to generate a large number of false positives. By incorporating human domain experts, these two fundamental problems can be eliminated or at least mitigated to some extent. On the one hand, experts can analyze parts of the available observables to discover new, previously unknown indicators. On the other hand, humans can use their domain knowledge K^i_d to decide whether an indicator ultimately describes a destructive action or not. For this reason, integrating K^i_d at this stage is highly beneficial and can even be considered inevitable for any approach of effective SA.

Within this phase of the Incident Detection Lifecycle, the next step is to identify incidents within the previously recognized indicators. In this step, the involvement of K^i is even more critical than for the detection of indicators. The main reason for this is that utilizing K^e in this step can only detect previously known attacks for which the corresponding K^e_s has already been defined. For new, unknown attack procedures, K^e_s cannot contribute, and also, K^e_m can hardly detect more than indicators that point to potentially malicious activity. In this context of actual attack detection, K^e cannot capture an incident to its full extent. Again, the involvement of human domain experts is necessary. Only this human component with K^i_d can analyze various indicators in their context, correlate them, and ultimately distinguish between malicious and regular activity. In summary, K^e in its various forms can contribute significantly to detecting indicators in the observables and thus reduce information overload. However, a final interpretation and classification of the indicators and the associated indicating of concrete incidents is not effectively possible in the vast majority of cases without direct integration of K^i_d .

Formalizing and Integrating User Knowledge into Security Analytics 13

While automatic analysis using K^e plays a major role in the first two phases of the Incident Detection Lifecycle, this focus shifts in the final phase, the *Post-Incident Analysis*. In this step, almost exclusively manual work steps take place in the context of forensic investigations and the attribution of incidents. Thus, the influence of K^e is rather low compared to K^i and the integration of K^i into automated workflows is stronger.

In addition to the inclusion of both K^e and K^i in the Incident Detection Lifecycle, we have indicated several other knowledge conversion processes in Figure 3. We identify all these additional processes as relevant and necessary for a comprehensive and effective implementation of SA, which covers both the cyber domain and the cyber-physical domain. Some of the processes plotted have already been presented in detail in Section 2.2: $int(K^e)$ (Eq. 4), $comb$ (Eq. 6), and $coll$ (Eq. 7). For this reason, we focus on the two remaining processes: $ext(K_s^i)$ and $ext(K_d^i)$. They are each an instance of ext , but require a closer, contextualized look.

Externalization of situational knowledge K_s^i ($ext(K_s^i)$) fundamentally allows employees to feed events (i.e., situations) they have observed or experienced into the SA system as observables. This allows the semantic information transformed from K_s^i into observables by $ext(K_s^i)$ to be used in the further steps of the Incident Detection Lifecycle. If this possibility is exploited efficiently, it significantly expands the availability of observables for SA because many aspects of targeted attacks are not detected in automatically collected data. Examples are social engineering attacks or direct physical access attempts. Information about these and a multitude of other attack vectors cannot be collected through automated data collection mechanisms. With the ability to externalize $ext(K_s^i)$, virtually every employee turns into an extremely valuable source of observables for incident detection when, for example, the employee reports a phone call attempting to discover critical access privileges. A unique feature of this conversion process is that it does not build on domain knowledge K_d^i , but a more general knowledge that comes primarily from situational awareness K_s^i . The externalization of situational knowledge is particularly relevant because it is the only way to fully capture possible attack vectors involving the physical aspects of modern attacks.

It is also necessary, to take a closer look at the process $ext(K_d^i)$. This activity basically comprises two processes: $ext(K_{d(sec)}^i)$ and $ext(K_{d(nonSec)}^i)$. It thus describes the interaction of humans with the analysis processes sustained by K^e . The fundamental goal of this interaction is to integrate K_d^i into security analytics, thus making human domain knowledge available to improve the overall incident detection lifecycle. In the era of cyber-physical systems, domain knowledge, specifically $ext(K_{d(nonSec)}^i)$, is widely distributed across enterprises. At the same time, however, the entirety of domain knowledge is necessary for comprehensive and effective security analytics. For this reason, the $ext(K_d^i)$ process is critical as it is the only way to translate human knowledge into SIEM correlation rules, attack signatures, or improved behavior models for the organization's assets.

14 *Formalizing and Integrating User Knowledge into Security Analytics*

Another aspect that stands out in Figure 3 is the exclusion of the *Utilize* loop, which illustrates the iterative nature of the Incident Detection Lifecycle in Figure 2. However, a closer look at Figure 3 reveals that this process step is by no means missing but has only been made more precise by integrating K^e and the corresponding conversion processes into the representation. Through the bi-directional connections between the lifecycle phases *Incident Detection & Response* and *Post-Incident Analysis* as well as K^e , our knowledge model makes clear that in these phases, K^e can be used and at the same time also generated. The K^e generated in these phases can be defined more precisely as the K_i^e described in previous sections. K_i^e serves as input to the *Data collection* phase, thus preserving the iterative nature of the life cycle.

3.3 Knowledge-based Security Personas

Based on the various security-related knowledge types, different groups of users can be distinguished. As shown in Equation 8 the knowledge of users can be seen in this context as different instances of K^i .

$$k_{d(nonSec)}^i, k_{d(sec)}^i \in K_d^i, k_s \in K_s^i, k_o^i \in K_o^i \quad (8)$$

In an organizational context, employees can essentially be assigned to two roles from an SA perspective, which can be referred to as security personas: security novices S_n and security experts S_e .

- Security novices: In general, a novice is a user without profound knowledge and experience within a specific domain. In our case, S_n are employees without deeper knowledge in security. However in practice, a clear differentiation is not always easy, since almost everyone has a basic sense of security. It is easier to make a distinction by taking an employee's areas of activity into account. Security novices can be defined as persons who do not deal with security in their daily activities, or only to a very limited extent (like for example gained from participating in awareness programs). From a knowledge perspective, S_n have domain knowledge in a domain other than security: $k_{d(nonSec)}^i$. This domain knowledge can be very individually pronounced from user to user. An example would be engineering knowledge, if a user is responsible for maintaining a turbine and thus knows precisely how it works. From a security perspective, situational knowledge k_s^i of S_n is particularly relevant, as it enables them to judge a situation in combination with their unique $k_{d(nonSec)}^i$. Thus, they can contribute to the Incident Detection Lifecycle by observing and reporting possible attack vectors. As shown in Fig. 4(a), however, S_n have very few connection points with the Incident Detection Lifecycle, which is mainly due to the lack of k_o^i . $ext(K_s^i)$ is possible if the respective system is sufficiently simple to use, thus, this process is present in the figure but grayed out.

$$S_n = \{k_{d(nonSec)}^i, k_s^i\} \quad (9)$$

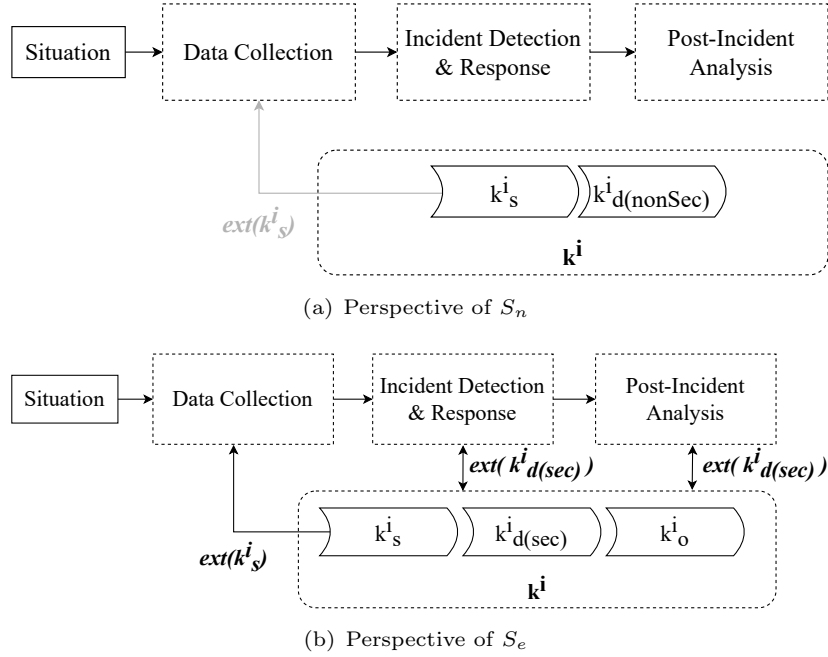


Fig. 4 Knowledge Model from the perspective of the two personas.

- Security experts S_e , in contrast, are employees with in-depth security-related domain knowledge $k_{d(sec)}^i$. This usually results from the significant involvement with security issues in their day-to-day business (for example as an employee within a Security Operations Center). Their $k_{d(sec)}^i$ in combination with k_s^i enables them to identify security incidents at a high level of detail and to realistically assess its extent and severity. In addition, they have the necessary operational knowledge k_o^i to operate security systems (such as SIEM systems) that are used for automated analyses within the Incident Detection Lifecycle. This results in a S_e being the main gateway to the Incident Detection Lifecycle (see Fig. 4(b)). Both $ext(K_s^i)$ and $ext(K_d(sec)^i)$ are possible, since the expert has the necessary k_o^i to comprehensively operate the systems involved.

$$S_e = \{k_{d(sec)}^i, k_s^i, k_o^i\} \quad (10)$$

3.4 Dichotomy of Security Analytics

The previous breakdown of the two security personas already highlights the different types of knowledge that are divided between the two personas. It is particularly noticeable here that neither of the two combines all knowledge and thus the knowledge required for the Incident Detection Lifecycle in one person. This circumstance indicates what we call the dichotomy of SA. When comparing the knowledge sets of S_n and S_e , it is noticeable that the differences can essentially be broken down to two knowledge types: Domain knowledge k_d^i

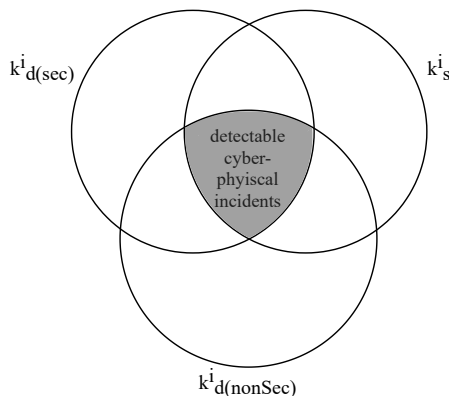
16 *Formalizing and Integrating User Knowledge into Security Analytics*

Fig. 5 Required knowledge types for incident detection.

differs ($k_{d(sec)}^i$ vs. $k_{d(nonSec)}^i$) and S_n has no or very little operational knowledge k_o^i .

As already defined in Equation 5 the externalization of implicit knowledge is the intersection of k_d^i and k_s^i . However, when considering cohesive incident detection in the era of cyber-physical attacks, the necessary domain knowledge k_d^i is distributed between S_n and S_e in the form of $k_{d(sec)}^i$ and $k_{d(nonSec)}^i$. Cyber-physical incidents are only detectable if knowledge about security incidents in general ($k_{d(sec)}^i$) and knowledge about the physical aspects in particular ($k_{d(nonSec)}^i$) are combined. In addition, situational knowledge k_s^i is necessary for the incident to be recognized in the first place. Therefore, only incidents for which all three types of knowledge are combined can be detected. Fig. 5 shows this relationship as an intersection. All incidents that do not reside on the intersection cannot be detected by humans, which is why these areas potentially constitute a blind spot in the Incident Detection Lifecycle and thus have to be minimized.

Operational knowledge k_o^i takes on a special role in the context of the dichotomy, since it is highly dependent on the security systems in use. Since these are usually expert systems, it is assumed that only security experts have the necessary knowledge to operate them properly. However, these systems should aim to be so easy to use that they require only little operational knowledge to empower S_n to contribute to security operations. Therefore, operational knowledge ideally is kept to a minimum in practice, in contrast to the other types of knowledge.

3.5 Knowledge Gaps

The dichotomy in SA creates some knowledge gaps, some of which have already been alluded to in Section 3.4. Essentially, three knowledge gaps limit the Incident Detection Lifecycle or prevent security incidents from being detected. In the following, these gaps are described in detail to highlight a path to close them:

1. $\text{ext}(K_S^i)$: The first gap that can be identified is the lack of possibilities to externalize K_S^i . The main difficulty here is how S_n can be incorporated appropriately or to create the means to do so. For example, if an employee notices a security incident, they need to be able to contribute their observations to the Data Collection phase of the Incident Detection Lifecycle. Initial approaches to this already exist in the form of the human-as-a-security-sensor paradigm [29, 30]. However, further research is needed in this direction to solve this problem in an applicable way.
2. $K_{d(\text{nonSec})}^i$: The next gap stems from the aforementioned k_o^i , which is not held by S_n in necessary amounts. Therefore, it must be ensured that the required k_o^i is reduced so that people without expert knowledge can operate security mechanisms. For example, it should be possible to involve engineers who know precisely how a turbine works and what security incidents can look like in the Incident Detection Lifecycle. However, it is unlikely that this problem will be solved entirely. For example, even with a great deal of effort, it will hardly be possible for engineers to create correlation rules for SIEM systems, as these are relatively complex by nature. Therefore, these systems must be simplified to the extent that S_n can at least contribute their knowledge in a simplified manner to contribute to the definition of meaningful rules.
3. *coll*: Collaboration between the actors, especially between S_n and S_e , within the Incident Detection Lifecycle, is vital because, as elaborated in Section 5, knowledge is not concentrated on individual persons but is distributed among several personas. Collaboration between the various personas can help create a central knowledge base in the Incident Detection Lifecycle in which as much relevant information as possible is brought together. The knowledge gaps mentioned in 1. ($\text{ext}(K_S^i)$) and 2. ($K_{d(\text{nonSec})}^i$) can help to enable or at least simplify collaboration. Collaboration has not yet been considered much in SA research, although it plays a significant role within the Incident Detection Lifecycle.

4 Research Prototype

The gaps described in Section 3.5 are not yet addressed explicitly in existing work. For this reason, in the following section of our paper, we present the second part of our contribution: a research prototype for a signature-based incident detection system that supports the two above-mentioned conversion processes. The concept and structure of the prototype are built according to the model of knowledge-based SA (see Figure 3). In order to detect indicators and identify incidents from their context, we apply a Complex Event Processing approach, which can be based on an arbitrarily complex pattern hierarchy. This hierarchical approach initially allows the detection of indicators based on observables. Additional and more advanced patterns are then used to identify actual incidents by correlating IoCs. The patterns, i.e., signatures needed for

18 *Formalizing and Integrating User Knowledge into Security Analytics*

this purpose, correspond to K_s^e in the context of knowledge-based SA and are made accessible to humans by the prototype.

In the following sections, we first derive general requirements. We then present our prototype’s system architecture and detail two essential components that are central to address the knowledge conversion processes. For the sake of clarity, we use the term “event” whenever it is not necessary to distinguish specifically between observable, indicator, or incident.

4.1 Requirements Analysis

In the age of CPS and IoT, one of the most pressing obstacles to overcome in the quest for holistic Security Analytics is to minimize the tremendous amount of K_o^i necessary to implement $ext(K_{d(nonSec)}^i)$. This can be achieved by providing centralized, interactive visual access to the K^e underlying the lifecycle. In addition, the next step is to provide better technical support for collaboration, or *coll*, between people. These two problems are summarized by the following paragraphs in their immediate context:

1. *Reduce the needed \mathbf{K}_o^i for $ext(\mathbf{K}_d^i)$:* K_o^i is required for all sub-aspects of the *int* and *ext* processes. However, since it has no direct impact or purpose for cybersecurity itself, the K_o^i required to operate security systems should be reduced as much as possible, especially for S_n . Besides offering training for S_n , this is the only possible way towards better integration of $K_{d(nonSec)}^i$. Novices are not skilled in dealing with security solutions, such as a company’s SIEM system. Therefore, for the integration of their $K_{d(nonSec)}^i$, the entry barrier to these systems (i.e., $K_{d(nonSec)}^i$) should be kept as low as possible. Thus, concerning the chosen notation of K_o^i as a catalyst for knowledge conversion, it is necessary to reduce the “need” for the catalyst as much as possible.
2. *Enable **coll** between \mathbf{S}_e and \mathbf{S}_n :* While security experts own knowledge of a variety of possible attack vectors, the knowledge of adapting these attack vectors for a particular context is often within the scope of activity and knowledge of non-security experts. In order to build up comprehensive security analytics from this perspective, technical support for collaboration should be improved. Only with a well-developed infrastructure for collaboration between experts from different domains can the broadest possible protection against a wide variety of attack vectors be successful.

These problems form the starting point for the basic idea of our prototype. We aim to simplify the creation and processing of signatures (patterns), which can be used to detect attacks or at least indicators of compromise. This central concept is supported by an approach for visual programming, which is already established in education. With the help of visual programming, the entry barriers for complex systems can be successfully lowered [31, 32]. The objective is to make a complex, text-based syntax for defining patterns for attack detection easier to understand and use. To achieve this goal, we define the following requirements:

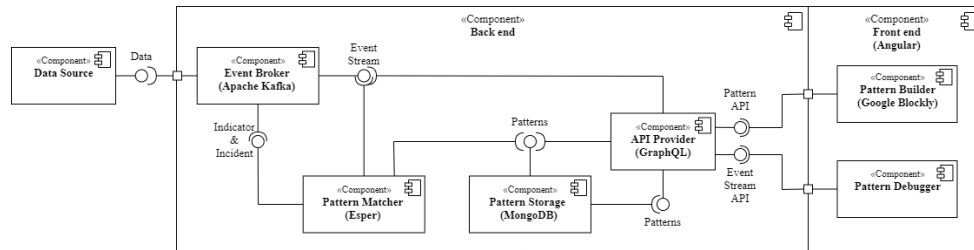


Fig. 6 Component diagram of the architecture for visual collaborative pattern definition [9].

R1 - Overview of currently deployed patterns: For users to get a quick overview of patterns that are currently already in use, the prototype must enable a corresponding display. All essential functions (such as editing a pattern) should be directly accessible from this overview view.

R2 - Visual abstraction for complex pattern definition syntax: A selected visual programming approach should make the complex syntactic structure more accessible to users with little operational knowledge. It is essential that users can externalize their knowledge in a semantically simplified way. At the same time, the prototype has to ensure the correct, necessary syntax for the mechanism used for incident detection.

R3 - Details for deployed patterns including situational context: If necessary, all details of a defined attack pattern should be available via the prototype. These details include the processing timestamps, the final pattern statement, and an insight into the activities or events associated with this pattern.

R4 - Debugging mechanism for patterns: To promote an understanding of how the patterns work, the prototype should at least provide an easy way to debug the statements. Such debugging should clarify the relationships between individual events that have led to the triggering of the attack detection. In addition, it is also desirable that debugging can represent hierarchies of patterns of varying complexity.

R5 - Centralized pattern storage and detection mechanism: To provide technical support for (remote) collaboration between different users, the prototype must centrally store and manage the defined signatures. The detection mechanism that uses the patterns must also be located in the center. Accordingly, architectures corresponding to a client-server structure should be aimed for.

4.2 System Architecture

The prototype we developed is built on a client-server architecture, which is shown in Figure 6. The server is the backend responsible for detecting indicators and identifying incidents based on predefined patterns and signatures. On the other hand, the frontend provides a user interface that enables the creation, editing, and debugging of these patterns. The entire system architecture

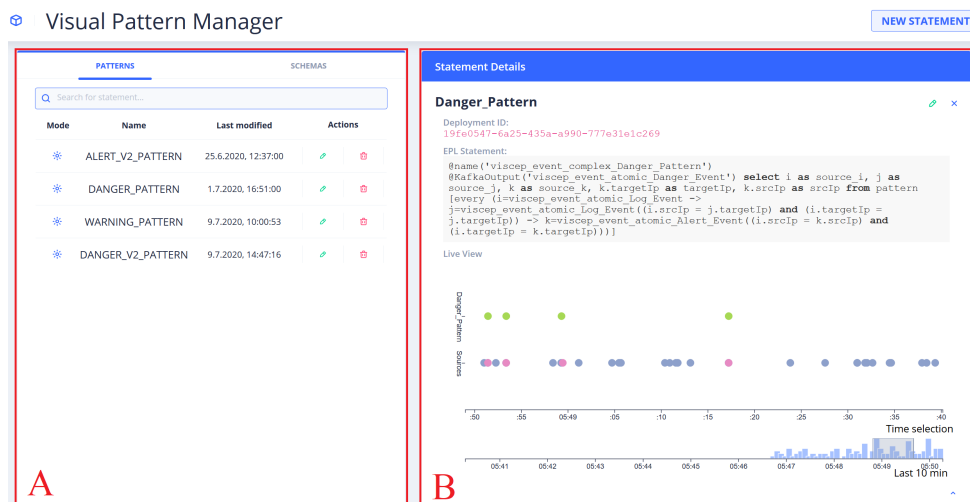
20 *Formalizing and Integrating User Knowledge into Security Analytics*

Fig. 7 Screenshot of the front end's landing page with a selected statement [9].

of our prototype is based on open source technology. The source code of the application itself is also available as open source on GitHub ².

4.2.1 Back End

In the following, the individual components of the back end are outlined, whereby their interconnection is shown in Figure 6. In the backend, the actual rule-based event correlation takes place with the help of the Complex Event Processing Engine Esper³. The actual events are provided by various data sources that reflect the current situation of the Incident Detection Lifecycle. With the help of Apache Kafka⁴, a central message broker is provided that manages and passes on the events generated by the various backend components. Patterns created in the front end are persisted in the pattern storage (MongoDB⁵) to make them available to Esper on demand (cf. R5). An API Provider implements the connection between back end and front end using a modern GraphQL⁶ interface.

4.2.2 Front End

The front end of our prototype consists of three basic views, which are embedded in an overarching user interface (UI). The UI is based on Angular ⁷. The first view, the landing page, is divided into two components. These components are marked with two red boxes (A) and (B) in Figure 7. The left component (A) provides an overview of all currently defined patterns and related details such as the name of the pattern, the time of the last change of the pattern, and its current deployment mode (R1). This deployment mode indicates whether a

² <https://github.com/Knowledge-based-Security-Analytics>

³ <http://www.esper.tech.com/esper/>

⁴ <https://kafka.apache.org/>

⁵ <https://www.mongodb.com>

⁶ <https://graphql.org/>

⁷ <https://angular.io/>

Formalizing and Integrating User Knowledge into Security Analytics 21

pattern is still under development (i.e., whether work is currently being done on it) or whether it has already been integrated into the back end’s incident detection operations. In addition, the pattern overview in component (A) can be used to initialize the editing of a pattern or to delete the corresponding pattern. By clicking on the “pencil” icon (i.e., editing a pattern), a user opens the current definition of the pattern in the *Visual Pattern Builder*, which is described in more detail in Section 4.3. A new pattern is created via the “New Statement” button in the navigation bar. This action opens the Visual Pattern Builder without an already existing pattern definition, only with an empty editing area. In addition to the patterns, the overview also contains a tab for schemas. The event types defined there can be used to use the pattern. However, since their structure is very straightforward and event types can also be defined using the Visual Pattern Builder, we will focus on defining the patterns in the remainder of this section.

Respective for R3, the second component (B) from Figure 7 of the landing page contains further information about a pattern selected in component (A). This includes the ID and the EPL statement, which formally describes the pattern and which is used in the pattern matcher. In addition to this information, component (B) presents a *Live Event Chart*, which provides a quick overview of the activities to be assigned to the pattern within the last ten minutes. The bar chart in the lower part shows the entire time window (10 minutes) and the number of events registered to the pattern. The upper part of the event chart represents an interactively selectable time frame from the last minutes and the events generated by the pattern matcher after a match was identified within a set of source events and the source events themselves. Herein, circles with the same colors correspond to the same event type.

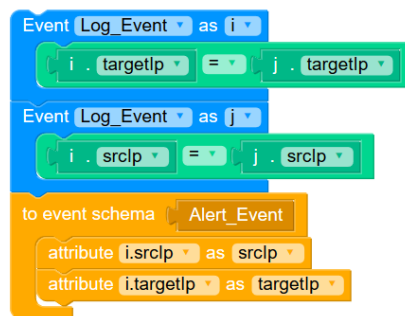


Fig. 8 Screenshot of EPL statement built with Blockly [9].

4.3 Visual Pattern Builder

This component of our prototype is used to create new statements or edit existing statements. For this purpose, we use the visual code editor Google

22 *Formalizing and Integrating User Knowledge into Security Analytics*

Blockly⁸. Blockly has so far been used primarily in the educational environment, for example, to teach the basic principles and concepts of programming. The approach of Google Blockly is catchy and straightforward. It allows the definition of specific, logical building blocks, which the users can then assemble and parameterize. In the background, these blocks are compiled into executable source code. Blockly has proven its ability to lower entry barriers for novice users in many places. Therefore, we consider it suitable for abstracting the complex syntax and logical flow of the EPL expressions used within the pattern matcher (R2).

In our prototype, we implemented building blocks based on Google Blockly to create and edit Esper EPL expressions. Figure 8 shows a simple EPL statement defined with Blockly. The main components of these statements are event patterns (blue blocks), conditions (green blocks), and actions (yellow blocks). The pattern shown in Figure 8 instructs the Pattern Matcher to emit an “Alert_Event” with the corresponding attributes after detecting two consecutive “Log_Event” instances with matching “srcIp” and “targetIp” attributes. An example of an Esper EPL expression generated by corresponding Google Blockly modeling can be seen in the gray box in component (B) in Figure 7.

Please refer to our open-source implementation linked above for the full range of different Esper EPL statements supported. Among others, our implementation includes a logical combination of event sequences (including “and”, “or”, “not”), counted event sequences, and logical conditions. Although we cannot yet express all possible Esper EPL expressions as Blockly building blocks, the concept is promising so far. In subsequent iterations of our work, we expect to achieve near-complete coverage of Esper EPL.

4.4 Pattern Debugger

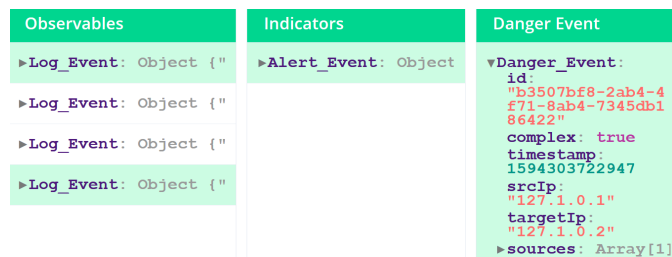


Fig. 9 Screenshot of the Pattern Debugger [9].

Using the arrow on the lower right side of component B, the *Pattern Debugger* can be opened for the respective pattern. It allows testing of the created patterns by providing a detailed view of the event’s data and its relationships (R4). As mentioned before, observables are assigned to an indicator and indicators to an incident in a hierarchical way. This hierarchy is visualized with the

⁸ <https://developers.google.com/blockly>

help of the pattern debugger to make relations easily recognizable. For displaying the hierarchy in a structured way, observables, indicators, and incidents are arranged next to each other in columns. The elements above or below in the tree are highlighted when hovering over them with the cursor to highlight the elements' hierarchical structure further.

The individual elements are represented as JSON. In order to maintain an overview, they are initially displayed in collapsed form. Only by selecting an element the complete JSON tree expands, whereby besides the overview, the option for displaying details is provided.

4.5 Discussion

To conclude the description of the prototype, it is discussed subsequently, emphasizing the implementation of the requirements and the approach to the underlying problems.

R1 is implemented on the landing page. An overview of all patterns is given here. The direct accessibility of the activities for the patterns is also available here in the form of action icons.

R2 is implemented with the help of Google Blockly. Using this technology makes it possible to create patterns without having to use complex text-based syntax. The visual programming approach additionally ensures that no erroneous patterns can be created. In our implementation, the syntax of the pattern is abstracted by the visual programming language while avoiding to cut the functionality of the pattern language. Additional research would be needed here to determine which level of abstraction is most appropriate.

R3 is implemented using a detailed view when a pattern was selected. The events that are affected by this pattern are visualized in the form of a live view to present the relationships in a comprehensible way.

R4 is implemented with the help of the pattern debugger. Within the debugger, the user has the possibility to highlight the events that have led to the triggering of an alarm. The respective events are hierarchically divided into Observables, Indicators, and Incident.

R5 is mainly implemented on the backend side. There, all created patterns are stored in the pattern storage to enable multiple users to work on them collaboratively. Furthermore, with the help of the Esper-based pattern matcher, event correlation is performed centrally.

The implemented requirements contribute to solving the two underlying problems, reducing the required K_o^i and enabling *coll* between S_n and S_e . The problem of reducing the needed K_o^i was solved with the help of the visual programming approach. This way, it is possible to create patterns without requiring in-depth expert knowledge of pattern syntax. Above all S_n is enabled to contribute its $K_{d(nonSec)}^i$. In addition, the complexity of pattern debugging has been reduced. The user does not have to work through various log files but can visualize the events in an easy-to-understand way. To not overwhelm the user, only a very abstract view of the events is given in the form of a life chart. However, if the user has the necessary expert knowledge, he can display the

24 *Formalizing and Integrating User Knowledge into Security Analytics*

details of the events. If an even more comprehensive view is desired, the pattern debugger can be used, which shows the relationships between the individual events. The abstract representation of the events also facilitates *int*.

The reduction of the required K_o^i already contributes to enabling *coll* between S_n and S_e , as it reduces entry barriers especially for S_n and thus enables him to participate in creating detection patterns. This, for example, gives S_n the possibility to adapt rules created by S_e and enrich them with their $K_{d(nonSec)}^i$ and thus refine them. In addition, this is achieved because patterns are stored in a central location, and all actors have access to them. The combination of *int* and *ext* is thus combined to allow *coll*.

Like any research work, the prototype presented has its limitations and points to future research potential. For example, it would be worth evaluating whether rule creation can be further simplified. Even if the underlying syntax is much more accessible through our visual approach, rule creation could be even more straightforward. Debugging can also be enhanced to provide even deeper insight into how the Pattern Matcher works. Furthermore, it needs to be empirically evaluated to what degree the prototype actually reduces the required K_o^i . This can be done in a user study that examines what effect *coll* has on detection rates.

5 Conclusion

This article presents and formalizes the concept of knowledge, its facets, and the concept of knowledge conversion in the context of security analytics. Building on this formalization, we present a model for knowledge-based security analytics based on the incident detection lifecycle. Our structuring and conceptualization makes it possible to raise the mostly inconsistent and informal descriptions to a formal and consistent level. With this contribution, we lay a sound foundation for future research in the field of security analytics.

Several sub-areas and activities within the knowledge-based SA model could be identified as not sufficiently considered in academic research. We presented a research prototype to demonstrate the first possible approach for externalizing human domain knowledge and collaboration between security experts and security novices. This prototype leverages the power of modern visual programming approaches to reduce the operational knowledge required to interact with security analytics systems, thereby lowering the barrier to entry for security novices. This also allows these domain experts to better provide their knowledge, which is especially important for incident detection in the CPS and IoT context in the form of signatures.

Although we were able to present a first research prototype that addresses the first open challenges in security analytics, there is still room for future research. First, we need to develop further technical support for collaboration between security experts and security innovators. Our prototype shows first possibilities here, but the approach needs to be improved together with users. A corresponding evaluation of the prototype to empirically confirm its suitability

is also necessary. Furthermore, approaches are needed to integrate situational knowledge into SA better. Although initial approaches to this exist in the human-as-a-security-sensor environment, they must be improved and further developed.

Acknowledgments. This research was partly supported by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (BayStMWi), as part of the INSIST project.

Declarations

Funding

Not applicable.

Conflicts of interest

The authors declare that they have no conflict of interest.

Availability of data and material

Not applicable.

Code availability

Github Repositories: <https://github.com/Knowledge-based-Security-Analytics>

References

- [1] Schneier, B.: *Secrets and Lies: Digital Security in a Networked World*, 15. edn. John Wiley & Sons, Hoboken, NJ, USA (2015)
- [2] Ben-Asher, N., Gonzalez, C.: Effects of cyber security knowledge on attack detection. *Computers in Human Behavior* **48**, 51–61 (2015). <https://doi.org/10.1016/j.chb.2015.01.039>
- [3] Zimmermann, V., Renaud, K.: Moving from a “human-as-problem” to a “human-as-solution” cybersecurity mindset. *International Journal of Human-Computer Studies* **131**, 169–187 (2019). <https://doi.org/10.1016/j.ijhcs.2019.05.005>
- [4] Loukas, G.: *Cyber-Physical Attacks*. Butterworth-Heinemann, Oxford, United Kingdom (2015). <https://doi.org/10.1016/C2013-0-19393-2>
- [5] Dietz, M., Vielberth, M., Pernul, G.: Integrating digital twin security simulations in the security operations center. In: *Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES)*, pp. 1–9. ACM, New York, NY, USA (2020). <https://doi.org/10.1145/3407023.3407039>

26 *Formalizing and Integrating User Knowledge into Security Analytics*

- [6] Eckhart, M., Ekelhart, A.: Towards security-aware virtual environments for digital twins. In: Proceedings of the 4th ACM Workshop on Cyber-Physical System Security - CPSS '18, pp. 61–72. ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3198458.3198464>
- [7] Schneier, B.: Click Here to Kill Everybody: Security and Survival in a Hyper-connected World, 1. edn. W.W. Norton & Company, New York (2018)
- [8] Chen, T.M., Sanchez-Aarnoutse, J.C., Buford, J.: Petri net modeling of cyber-physical attacks on smart grid. *IEEE Transactions on Smart Grid* **2**(4), 741–749 (2011). <https://doi.org/10.1109/TSG.2011.2160000>
- [9] Böhm, F., Vielberth, M., Pernul, G.: Bridging Knowledge Gaps in Security Analytics:. In: Proceedings of the 7th International Conference on Information Systems Security and Privacy, pp. 98–108. SCITEPRESS - Science and Technology Publications, Online Streaming (2021). <https://doi.org/10.5220/0010225400980108>
- [10] Sallos, M.P., Garcia-Perez, A., Bedford, D., Orlando, B.: Strategy and organisational cybersecurity: a knowledge-problem perspective. *Journal of Intellectual Capital* **20**(4), 581–597 (2019). <https://doi.org/10.1108/JIC-03-2019-0041>
- [11] Ackoff, R.L.: From data to wisdom. *Journal of Applied System Analysis* (16), 3–9 (1989)
- [12] Frické, M.: The knowledge pyramid: a critique of the dikw hierarchy. *Journal of Information Science* **35**(2), 131–142 (2009). <https://doi.org/10.1177/0165551508094050>
- [13] Davenport, T.H., Prusak, L.: Working Knowledge: How Organizations Manage What They Know. Harvard Business School Press, Boston, Mass. (2000)
- [14] Nonaka, I., Takeuchi, H.: The Knowledge Creating Company. Oxford University Press, Oxford, United Kingdom (1995)
- [15] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI Magazine* **17**(3), 37 (1996). <https://doi.org/10.1609/aimag.v17i3.1230>
- [16] Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G., Keim, D.: Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics* **20**(12), 1604–1613 (2014)
- [17] Polanyi, M.: The Tacit Dimension. University of Chicago Press, Chicago

Formalizing and Integrating User Knowledge into Security Analytics 27

(2009)

- [18] Chen, M., Ebert, D., Hagen, H., Laramee, R.S., van Liere, R., Ma, K.-L., Ribarsky, W., Scheuermann, G., Silver, D.: Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications* **1**(29), 12–19 (2009)
- [19] Wagner, M., Rind, A., Thür, N., Aigner, W.: A knowledge-assisted visual malware analysis system: Design, validation, and reflection of kamas. *Computers & Security* **67**, 1–15 (2017). <https://doi.org/10.1016/j.cose.2017.02.003>
- [20] Jaeger, L.: Information security awareness: Literature review and integrative framework. In: Bui, T. (ed.) *Proceedings of the 51st Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, Honolulu, HI, USA (2018). <https://doi.org/10.24251/HICSS.2018.593>
- [21] Vasileiou, I., Furnell, S.: Personalising security education: Factors influencing individual awareness and compliance. In: *Information Systems Security and Privacy. Communications in Computer and Information Science*, vol. 977, pp. 189–200. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25109-3_10
- [22] Ponsard, C., Grandelaudon, J.: Guidelines and tool support for building a cybersecurity awareness program for smes. In: *Information Systems Security and Privacy. Communications in Computer and Information Science*, vol. 1221, pp. 335–357. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49443-8_16
- [23] Wang, X., Jeong, D.H., Dou, W., Lee, S.-W., Ribarsky, W., Chang, R.: Defining and applying knowledge conversion processes to a visual analytics system. *Computers & Graphics* **33**(5), 616–623 (2009). <https://doi.org/10.1016/j.cag.2009.06.004>
- [24] Federico, P., Wagner, M., Rind, A., Amor-Amorós, A., Miksch, S., Aigner, W.: The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In: *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)* (2017)
- [25] Thalmann, S., Ilvonen, I.: Why should we investigate knowledge risks incidents? - lessons from four cases. In: Bui, T. (ed.) *Proceedings of the 53rd Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, Honolulu, HI, USA (2020). <https://doi.org/10.24251/HICSS.2020.607>

28 *Formalizing and Integrating User Knowledge into Security Analytics*

- [26] Mahmood, T., Afzal, U.: Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools. In: 2013 2nd National Conference on Information Assurance (NCIA), pp. 129–134. IEEE, New York, NY, USA (2013). <https://doi.org/10.1109/NCIA.2013.6725337>
- [27] Menges, F., Pernul, G.: A comparative analysis of incident reporting formats. *Computers & Security* **73**, 87–101 (2018). <https://doi.org/10.1016/j.cose.2017.10.009>
- [28] National Institute of Standards and Technology: Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1 (2018). <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf> Accessed 14.09.2021
- [29] Vielberth, M., Englbrecht, L., Pernul, G.: Improving data quality for human-as-a-security-sensor. a process driven quality improvement approach for user-provided incident information. *Information & Computer Security* (2021)
- [30] Vielberth, M., Menges, F., Pernul, G.: Human-as-a-security-sensor for harvesting threat intelligence. *Cybersecurity* **2**(1) (2019). <https://doi.org/10.1186/s42400-019-0040-0>
- [31] Chao, P.-Y.: Exploring students’ computational practice, design and performance of problem-solving through a visual programming environment. *Computers & Education* **95**, 202–215 (2016). <https://doi.org/10.1016/j.compedu.2016.01.010>
- [32] Sáez-López, J.-M., Román-González, M., Vázquez-Cano, E.: Visual programming languages integrated across the curriculum in elementary school. *Computers & Education* **97**, 129–141 (2016). <https://doi.org/10.1016/j.compedu.2016.03.003>

3 Towards GDPR-compliant data processing in modern SIEM systems

Current status:	Published
Journal:	Computers & Security
Date of acceptance:	27 December 2020
Full citation:	MENGES, F., LATZO, T., VIELBERTH, M., SOBOLA, S., PÖHLS, H. C., TAUBMANN, B., KÖSTLER, J., PUCHTA, A., FREILING, F. C., REISER, H. P., AND PERNUL, G. Towards GDPR-compliant data processing in modern SIEM systems. <i>Computers & Security</i> 103, 102165 (2021), 1–19
Authors' contributions:	Florian Menges 20% Tobias Latzo 15% Manfred Vielberth 15% Sabine Sobola 10% Henrich C. Pöhls 10% Benjamin Taubmann 5% Johannes Köstler 5% Alexander Puchta 5% Hans P. Reiser 5% Felix Freiling 5% Günther Pernul 5%

Journal description: Computers & Security is the most respected technical journal in the IT security field. With its high-profile editorial board and informative regular features and columns, the journal is essential reading for IT security professionals around the world.

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/coseComputers
&
Security

TC 11 Briefing Papers



Towards GDPR-compliant data processing in modern SIEM systems



Florian Menges^{a,*}, Tobias Latzo^b, Manfred Vielberth^a, Sabine Sobola^d,
Henrich C. Pöhls^c, Benjamin Taubmann^c, Johannes Köstler^c,
Alexander Puchta^a, Felix Freiling^b, Hans P. Reiser^c, Günther Pernul^a

^aDepartment of Information Systems, Universität Regensburg, Germany^bDepartment of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany^cInstitute of IT-Security and Security Law, Universität Passau, Germany^dPaluka Sobola Loibl & Partner attorneys at law, Regensburg, Germany

ARTICLE INFO

Article history:

Received 14 November 2019

Revised 14 March 2020

Accepted 27 December 2020

Available online 31 December 2020

Keywords:

Security information and event management

SIEM

GDPR

Threat intelligence

DINGfest

ABSTRACT

The introduction of the General Data Protection Regulation (GDPR) in Europe raises a whole series of issues and implications on the handling of corporate data. We consider the case of security-relevant data analyses in companies, such as those carried out by Security Information and Event Management (SIEM) systems. It is often argued that the processing of personal data is necessary to achieve service quality. However, at present existing systems arguably are in conflict with the GDPR since they often process personal data without taking data protection principles into account. In this work, we first examine the GDPR regarding the resulting requirements for SIEM systems. On this basis, we propose a SIEM architecture that meets the privacy requirements of the GDPR and show the effects of pseudonymization on the detectability of incidents.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Motivation

The security of the modern information infrastructure is of high importance. In order to detect misuse and attacks at an early stage a lot of information about the events inside IT-infrastructures, e.g. inside computer networks and software applications also across many systems, is required to detect or post-mortem report and document attacks. Security

Information and Event Management (SIEM) systems help organizations to keep up with the ever increasing complexity by providing a holistic view on IT-infrastructures. Naturally, SIEM systems process enormous amounts of data about security related events, e.g., when specific users login or certain users perform critical actions. It is often argued, that generally the quality of service depends critically on the quality and detail of the data collected and processed within the system Wang and Strong (1996), which has been shown for different domains such as threat intelligence Schlette et al. (2020).

Events like those just described that are processed within the SIEM system are clearly related to concrete users and therefore must be treated as personal information, which require protection under Europe's General Data Protection

* Corresponding author.

E-mail address: florian.menges@ur.de (F. Menges).
<https://doi.org/10.1016/j.cose.2020.102165>

0167-4048/© 2020 Elsevier Ltd. All rights reserved.

Regulation (GDPR) [European Parliament and the Council of the European Union \(2016\)](#). Adopted in May 2018, it regulates and harmonizes the protection of personal data in the processing and transfer of data within and between private companies and/or public bodies in the European member states. Although, the GDPR is only compulsory for EU member states, it has evolved into a blueprint for data protection all over the world, as discussions between the US Congress and Mark Zuckerberg in the aftermath of the Cambridge Analytica case indicate¹.

Hence, SIEM systems must also comply to the regulations themselves, which leads to conflicting interests. On the one hand, SIEM systems rely on personal data such as information from the identity and access management (IAM) for providing high detection rates of incidents and thus a high level of protection. On the other hand, the requirements of the GDPR suggest that investigations of data streams as carried out in current SIEM systems may no longer be legally compliant. To complicate things even further, regulations regarding the handling of digital evidence mandate that authenticity and integrity of the data related to an incident should be guaranteed at all times in order to maintain its high legal probative value. It is therefore necessary to find the best trade-off between those two demands. With this work we attempt to fill the resulting research gap and to harmonize legal GDPR requirements with the technical architecture for SIEM systems. To bridge the gap between the disciplines of computer science and law and to produce the most reliable results possible, this paper was written by IT security researchers in collaboration with a lawyer. A central idea is the integration of *anonymization* and *pseudonymization* into threat analytics mechanisms. While this makes it necessary to change the original data, it is possible to maintain legal integrity and authenticity by using *redactable and sanitizable signatures*, a cryptographic concept to retain a level of authenticity useful to retain a suitable level of legal evidence even when data gets obfuscated or if certain parts of it are missing. We deploy cryptography to enable balancing authenticity proofs for the collected security-related events with the confidentiality requirements of the information about commercially-relevant internals (trade secrets) and employees' as well as customers' privacy (personal data). Thus, our goal is to minimize the amount of data which is being made accessible to third-parties in every step of the SIEM process. By enforcing this with cryptography the proposed system adheres to the security-by-design principle of least privilege as well as the privacy-by-design principle of data minimization. At the same time we aim to keep the impact on detection as low as possible and thus we provide an audit-able process to gain access to more details if security analysis is needing it. For the reason of being able to reconstruct original data, leaving a trace in an audit log, we focus on cryptographic methods and support pseudonymization rather than anonymization. Technically, we encrypt and sign events early and store the decryption keys with a party trusted for logging access to stored keys; moreover we employ signatures that allow to slice or redact data.

¹ <https://www.theverge.com/2018/4/11/17224492/zuckerberg-facebook-congress-gdpr-data-protection>.

1.2. Related work

When looking at the application of privacy mechanisms to threat analytics (e.g. SIEM systems), literature can be divided into a pre-GDPR and a post-GDPR phase, as this regulation still has a big impact on the integration of privacy. In the former phase there are not many results to be found regarding applying privacy to SIEM systems, however the challenges in integrating privacy in forensic and threat analyses has been identified [Jensen \(2013\)](#); [Stahlberg et al. \(2007\)](#). Although the challenges were not solved for SIEM systems, selected works in the IT security domain address it. For example [Burkhart et al. \(2010\)](#) describe a privacy preserving solution for secure multi-party computation. Furthermore, a main focus during this era was the application of privacy to intrusion detection systems (IDS), which could be declared as the predecessors of modern SIEM systems and thus in our context are worth a closer look: [Sobirey et al. \(1997\)](#) propose an approach for pseudonymizing user related data in IDS and closely examined, which records need to be pseudonymized in audit records. Based on this work, [Biskup and Flegel \(2000\)](#) and [Park et al. \(2007\)](#) propose an approach which is quite similar to the one presented in this paper as it uses cryptographic methods to pseudonymize personal data, though these are closely tailored to IDS and not completely adaptable to SIEM. In addition, they were issued before the publication of the GDPR and thus did not have all the requirements in mind and respectively were not evaluated against the new requirements. Our approach also differs, as we have a more abstract view of the whole system and do not focus largely on cryptographic details. Furthermore, [Buschkes and Kesdoğan \(1999\)](#) discuss requirements such as data avoidance and reduction of personal data.

Although privacy preserving methods were widely discussed in the past, recently the application of GDPR received an increased amount of attention and new works were published. In relation to SIEM, some work was published covering GDPR compliant data processing. [Sgaglione and Mazzeo \(2019\)](#) and [Coppolino et al. \(2018\)](#) introduce the COMPACT project, which is a GDPR compliant SIEM. However, they do not go into detail, how this is realised technically. Current research for SIEM systems mainly focuses on the architecture and improvement of such systems and not on the integration of privacy [Miloslavskaya and Tolstoy \(2019\)](#); [Mokalled et al. \(2019\)](#); [Nespoli and Gómez Mármol \(2018\)](#).

In cryptography, digital signatures are used to ensure authenticity and integrity of data, i.e., they guarantee that upon inspection data is unchanged and comes from an attributable source. Special techniques of *redactable signature schemes* (RSS) by [Steinfeld et al. \(2002\)](#); [Steinfeld et al. \(2002\)](#) allow subsequent deletions in the data, while *sanitizable signature schemes* (SSS) as proposed by [Atieniese et al. \(2005\)](#) even allow subsequent edits by dedicated authorised parties while maintaining authenticity of the remaining data. Both RSS and SSS allow to balance authenticity with privacy protection, because they allow retaining the integrity and authenticity protection for the unedited or not-removed parts of the document and at the same time keep the confidentiality protection for the overwritten parts of the document.

In cryptography the latter property is intuitively termed privacy. While many schemes have appeared in the literature [Bilzhaue et al. \(2017\)](#), only some of which uphold privacy and only those schemes that additionally fulfil detectability, known as non interactive public accountability [Brzuska et al. \(2012\)](#) can be used for eIDAS² compliant signatures [Pöhls \(2018\)](#); [Pöhls and Höhne \(2011a\)](#); [van Geelkerken, F.W.J. et al. \(2015\)](#). While the legal compliance of such signatures has been subject to research, the integration of such schemes into privacy protection of SIEM have not yet been investigated. In particular, in the application scenario of SIEM we want to be able to later reveal previously not-shared content. For this, a special form of digital signatures is needed which has the property of *mergeability*, i.e., the ability to re-add signed content to previously redacted but still signed content and re-generate a valid signature over the merged content [Pöhls et al. \(2012\)](#).

1.3. Contribution and outline

To the best of our knowledge we are not aware of an approach, that integrated GDPR into SIEM in a comprehensive way. Given the fact that these regulations need to be applied by all companies that operate within the European Union, there appears to be high demand for systems that are GDPR compliant. In this paper we present the first privacy-friendly – and thus GDPR-compliant – SIEM architecture that protects the confidentiality as well as the authenticity of security-relevant events starting at their collection, keeping the protection during the analysis and finally sending an incident report.

The presented architecture allows the deployment of a SIEM that meets the regulatory requirements under the EU data protection. It protects personal information in the data sets from unnecessary visibility using pseudonymization and encryption techniques without a significant reduction in detectability. Hence, we balance data-quality (detection of incidents) with legal obligations from privacy legislation and thus also protect trade secret by sharing only the minimum necessary information in any step of the SIEM process. Thus we strongly adhere to the GDPR's data minimization principle. Still, we achieve the highest level of confidence that the security-relevant events initially recorded and reported into the SIEM process are protected from tampering by using redactable and sanitizable signature schemes to proof authenticity. This allows us to balance the need for generating data with a high legal evidence with the need to protect privacy (and trade-secrets).

The legal analysis carried out for this architecture and presented in this paper shows that even potentially invasive data can be collected in a GDPR-compliant manner as our proposed system balances the necessity of the collection (detection and reporting of actual security incidents) with the protection of users privacy and customer's trade-secret needs. The paper shows the actual influence of pseudonymization on incident

detection mechanisms and the results of the performed legal evaluation.

The remainder of this paper is structured as follows. First, we give some background on the GDPR (legal) and SIEM (technical) in [Section 2](#). In [Section 3](#) we develop the research questions that arise from integrating GDPR into SIEM. On this basis, we describe our GDPR-compliant architecture for SIEM systems in [Section 4](#). The architecture is the evaluated on both technical and legal level in [Section 5](#). The paper concludes in [Section 6](#).

2. Background

This section provides the background information that is needed to understand the approach presented in this paper. Thus, we first give an overview of the functionality and properties of SIEM systems, as this serves as a basis for our architecture. Subsequently, we give an overview of the requirements the GDPR defines with special attention to the processing of personal data.

2.1. Security information and event management (SIEM)

In general, SIEM was first mentioned by Gartner [Williams and Nicolett \(2005\)](#). It originated from the initially separate systems Security Information Management (SIM) and Security Event Management (SEM) [Goldstein et al. \(2013\)](#). SIEM must fulfill several requirements, which are all connected: Log collection, enrichment with context data, log normalization, event correlation, and analysis as well as long- and short-term storage of log data, reporting, monitoring, alerting, and incident response [Gartner Inc. \(2018\)](#); [Miller et al. \(2011\)](#); [Vielberth and Pernul \(2018\)](#).

A SIEM system as described in [Vielberth and Pernul \(2018\)](#) is essentially designed for collecting relevant log data in a central place from arbitrary systems such as network devices or operating systems. This among other things enables the detection of incidents and in this way gaining situational security awareness. On a high level of abstraction, a SIEM system consists of the three main steps *data acquisition, processing and reporting*, which are elaborated in the following in more detail.

Data acquisition: Hereby, it first collects relevant event information, in most cases in the form of log data, which gets enriched with additional context data. There are basically two approaches for data acquisition: First, the data can be pushed into the SIEM by the data generating system. Thereby, the SIEM does not influence the generated data. Second, the data can be pulled by the SIEM from the observed system, which grants more control over the generated data enabling for example the assurance of integrity. This data then is translated into a uniform representation during the normalization step.

Processing: The core of the system is the correlation and analysis component, wherein information from various sources is correlated and incidents get detected by methods such as pattern matching. Real-time threat detection enables fast reactions in case of an incident,

² eIDAS is short for the current legislation which defines the technical functionalities to allow electronic signatures to be legally equivalent to handwritten signatures within the EU [European Parliament and the Council of the European Union \(2014\)](#).

whereas forensic analysis pursues the goal of analyzing the whole extent of the event in the aftermath in order to secure evidence. A distinction can, therefore, be made between short-term and long-term storage of relevant data. For long-term storage, it is particularly important to preserve the data in a tamper-proof way in order to be able to use it as evidence in court. Monitoring and visual security analytics enable security analysts to be actively involved in the analysis process. In the case of a detected incident, alerting and incident response triggers necessary reactions to mitigate further harm.

Reporting: An essential part of modern SIEM systems is reporting occurred incidents for compliance reasons (e.g. critical infrastructure providers) or enables the participation in established threat intelligence sharing platforms between participating organizations.

2.2. GDPR

In the following we present some background on general and SIEM-specific GDPR demands and later (see Sect.3) detail, which problems arise for SIEM systems to be built to comply.

Since 25 May 2018, the GDPR has been in force throughout the European Union (EU) to ensure the protection of the “natural person” in data processing. This regulation is directly applicable to all member states of the EU. The GDPR does not contain any immediate legal requirements for software developers but if they intend to sell their product to customers, they must be aware of the legal requirements.

Data protection is the protection of the natural person from privacy impairments through the processing of data concerning the person. Everyone should be free to decide who, when and how their data should be accessible. The term of personal data is therefore defined as “all information relating to an identified or identifiable natural person”, Art. 4 (1) GDPR.

2.2.1. General principles of data processing

In order to achieve the goal of high personal protection, the GDPR pursues the regulation of all basic principles which are regulated in Art. 5 (1) GDPR. According to the GDPR, only lawful, fair and transparent data processing is permitted in a transparent manner, Art. 5 (1) lit. a GDPR, in order to serve the principle of *good faith*. Furthermore, data processing shall be lawful only if and to the extent that it is applied by Art. 6 GDPR. But even then, it must be done in a transparent way which is traceable for the data subject. Another important principle is the *purpose limitation*, Art. 5 (1) lit. b GDPR. Thus, the clear purpose of processing must be previously established and legitimate. Furthermore, the *principle of data minimization* is applicable, Art. 5 (1) lit. c GDPR. This means that data collection is only allowed within limits for specified, explicit and legitimate purposes and not further. A further principle is *accuracy*, Art. 5 (1) lit. d GDPR. Only correct data may be collected. Even after processing it must be ensured that personal data is accurate. If this is not the case, data must be erased or rectified with delay. One further fundamental principle is the *storage limitation*, Art. 5 (1) lit. e GDPR. In the course of data processing it must be ensured that an identification is only possible in case of it being necessary for the purpose. *Integrity and confidentiality* have recently become further important principles,

Art. 5 (1) lit. f GDPR. This means that processing must occur in compliance with general safety standards. Compliance with these principles must be proven at any time by the controller, Art. 5 (2) GDPR.

2.2.2. Processing on a legal basis and transparency obligations

Generally, First of all, processing is lawful if the data subject has given *consent* to data processing, Art. 6 (1) point (a) GDPR. Processing is also allowed for the performance of contracts, Art. 6 (1) point (b) GDPR. If the processor is subjected to a legal obligation, data processing is also lawful without further requirements, Art. 6 (1) point (c) GDPR. This also applies if the protection of vital interests is pursued, Art. 6 (1) point (d) GDPR. The processing is also possible for the performance of a task carried out in the *public interest* or in the *exercise of official authority*, Art. 6 (1) point (e) GDPR. Lastly, processing is lawful for the purposes of the *legitimate interests pursued by the controller or by a third party*, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, Art. 6 (1) point (f) GDPR. While a lot of the exceptions might be triggered by the need and want to have a SIEM to protect from security breaches, the collection must meet a balance test, e.g. collection must not overshoot the goal, and must be transparent, e.g. clearly communicated to the data subjects.

2.2.3. Data security

In addition, the GDPR regulates a close interconnection of data protection with data security (especially Art. 32 GDPR) to the effect that technical and organizational measures in the data-processing company enable the highest degree of data security (availability, confidentiality, integrity, and resilience). In our scenario, the two security goals confidentiality and integrity are particularly relevant. We must ensure that information about events that could relate to incidents is transferred from the source to the sink, and is not altered or made accessible to unauthorised persons. To protect integrity an unauthorized modification must be detected if it happened, which is especially important to use non-tampered recorded data as evidence. Thus, protecting integrity and authenticating the data's origin provides legal value. Further, confidentiality protection guarantees that no unauthorized party is able to obtain information not intended for them, e.g. we must securely communicate the personal data to have them reach only the right recipients.

2.2.4. Redaction

The term “redaction” itself is not found in the GDPR directly; it refers to the irreversible removal of the information [The National Archives \(2011\)](#). This process is explicitly mentioned in guidance documents that explain how to remove information that is not subject to the information to be released under laws for the freedom of access to information, e.g. [UK FOIA Parliament of the United Kingdom \(2000\)](#) and thus is also applicable as a technique in the context of GDPR's data minimization [The National Archives \(2011\)](#); [United Kingdom Ministry of Justice \(2009\)](#).

2.2.5. Pseudonymization

The term “pseudonymisation” of the GDPR means the processing of personal data in such a manner that the personal

data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person, Art. 4 (5) GDPR. In order to make a pseudonymization, the data subject must first be assigned to pseudonyms. These can be user IDs. Thereafter the necessary data for identification must be kept separately. It must be ensured that they are strictly separated from the pseudonyms. It is important that features that can only indirectly lead to identification must be removed in the event of pseudonymization. The pseudonymization can be made by the data subject himself, the controller or an independent third party, a data trustee. An assignment rule must be created, e.g. through a reference table. The pseudonymization must be performed without the knowledge of the data subject. The data subject must be informed about the pseudonymization and it must be clarified who generates the pseudonym, who owns the assignment rule and under what circumstances an identification may take place. This is because pseudonymized data continues to be personal data. Particularly in the area of monitoring software development compliance with the GDPR is mandatory because it extracts its results from a data stream that contains personal data. The pseudonymization protects the data of the data subjects. At the same time, it is an opportunity to still being able to detect security incidents effectively and identify the responsible user in this regard.

3. Problem statement and research questions

In SIEM systems very large amounts of data are processed from various sources, while at the same time a GDPR compliant data protection must be guaranteed. This can be achieved by protecting all data relevant to data protection against unauthorized access using techniques such as encryption or pseudonymization. Working on protected data, however, brings different additional problems with it. On the one hand, it needs to be ensured that incident recognition is still possible despite the data protection. On the other hand it also needs to be possible to remove the protection in case of an actual incident. These aspects, which we have identified as essential for a GDPR-compliant analysis process, translate into the following three specific research questions. In this work we use the pseudonymization for the realization of the data protection, since this represents a valid procedure according to both GDPR and different reporting regulatory environments.

3.1. Data protection considerations and attacker model:

SIEM systems work with data from highly heterogeneous sources. As a result, different requirements need to be met in order to enable data protection in accordance with the GDPR. Establish data protection through the full encryption of all data would be the most intuitive and legally compliant way to process the data. However, since the GDPR only requires the protection of personal data, the data can also be classified according to protection requirements and partially pseudonymized in this context. In this way it can be achieved to

still be legally compliant, while more meaningful data is available for analysis at the same time. To achieve this, all acquired data needs to be available in a standardized form to allow the identification of information that needs to be protected. More specifically, the data acquired can be differentiated into information that is not relevant for data protection, data that may be relevant for data protection (e.g. path information in folder structures) and specific information relevant for data protection (e.g. e-mail addresses or contents of e-mails). In addition to this, the pseudonymization mechanism also needs to be protected. It must be ensured in a technical and organizational way for each data processing step within the SIEM system.

For being able to design a compliant and secure system, it is conducive, to define an attacker model, that determines the necessary measures. Thereby, the role, the goal, behaviour and the resources of the attacker are delimited:

- **Role:** The attackers role against which we consider our system protected can either be an outsider or an insider. An outsider is any person who has only access to interfaces of the system, which are open to the public. The outsider can however utilize a breach to gain access to certain parts or data of the system. Any third party who is involved in the SIEM system can also be referred to as an outsider. In contrast, an insider is any person who is directly involved into the system, such as analysts or server-admins.
- **Goal and behaviour:** The attacker can be either passive or active. The passive attacker only lists to the data without any intervention, whereas the active attacker tries to gain access to the data or the system by actively interacting with the system. For our approach, the considered goal of the attacker is to gain access to private data, since we design a SIEM system which is GDPR compliant.
- **Resources:** Since we utilize measures which are based on common asymmetric or symmetric cryptography, we can only consider attackers, with limited resources.

In summary, this raises the first question:

Q1: How must data that is processed in SIEM systems be protected to be GDPR compliant?

3.2. Impairment of incident identification through data concealment:

The GDPR stipulates that data protection must be applied as early as possible within the analysis process. Considering the data management of SIEM systems, this translates into a data protection obligation at the time of data acquisition. As a result, incident detection always needs to be performed on pseudonymization data. In this context, the relationship between pseudonymized data and plain text data within the data stream is a significant factor influencing possible analyses. This may impair both automated and manual analyses due to possible losses in the meaningfulness of the data analyzed. This leads to the second question:

Q2: Does the recognition of security incidents function properly despite of data pseudonymization or may losses and trade-offs be expected here?

3.3. Lifting the pseudonymization while retaining data authenticity:

To enable the utilization of information about detected security incidents, while being compliant with the legislation, two main conditions need to be met. On the one hand, information about incidents must be available in the long term in an integrity preserving manner. On the other hand, a de-pseudonymization of the data needs to be possible at any time after detection. This warrants that the information can be used as a reliable means of evidence in trials that might take place in the future. Please note that the GDPR proposes both anonymization and pseudonymization techniques as possible data protection measures. However, since the use of anonymization would prevent the data from being used as evidence, this technique will not be considered further in this paper.

To achieve this, appropriate technical and organizational measures need to be in place ensuring that de-pseudonymization is only possible in case of actual incidents. Additionally, legal compliance also needs to be ensured for the data after lifting the pseudonymization for further processing. This requires protecting the data's integrity including origin authenticity.

Thus, the principle of data minimization complicates especially the goal of integrity. This principle has always been at the center of data protection and can be found in European and member state legal texts, e.g. already in the former Directive 95/46/EC and thus also in the GDPR. In detail Art. 5 GDPR describes that personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

In general, there are two ways to conform with data minimization: (a) not collect or not forward personal data if it unnecessary, e.g. deleting it from data sets or blacken it out, i.e. redact it, or (b) making it harder to restore the personal data. For the latter, an important measure in this regard is the pseudonymization of data because it reduces the risks for the data subject and simultaneously it helps the controller to fulfil his data protection obligations (see also recital 28). By, for example, ensuring that only pseudonymized data is used during the incident analysis and that the person is only revealed if an anomaly is detected, would make a legally compliant processing of personal data in a SIEM conceivable. For the former, personal data, e.g. fields that contain this information, could just be removed before forwarding them.

On the other hand, both mechanisms for data minimization (pseudonymization or removal) result in a modification of the initially gathered data; an intended modification but a modification nevertheless. This results in standard cryptographic mechanisms to protect the data's integrity, such as digital signatures or message authentication codes, to fail. Thus, they are unsuitable to protect the integrity end-to-end.

The more recent cryptographic mechanisms, known as redactable Johnson et al. (2002); Steinfeld et al. (2002) or sanitizable signatures Ateniese et al. (2005) are capable of allowing our architecture to authorize modifications such as removal of unnecessary data points from authentic data set gathered by the SIEM. From their initial versions these algorithms have evolved (see Bilzhaue et al. (2017) for an overview). Most re-

cently they undergo the process of becoming an internationally recognized signature standard³. While this process takes time, the current status shows that the cryptographic mechanisms have the needed maturity to be backed and accepted in the cryptographic community. Once becoming recognised through such a standard, legal argumentation for compliance becomes a lot easier as legislators and judges will find the algorithms in lists of known mechanisms. Even if not standardised (or not yet) the provided authenticity offerings are technically equivalent to normal signatures Höhne et al. (2012); Pöhls and Höhne (2011b) and in any case much better than having none and also non-standard algorithms are suitable to win legal arguments in court cases – bearing the need for technical expertise appointed by court. This leads to the third question:

Q3: Which conditions need to be met to ensure that incident information can be de-pseudonymized in case of an incident and how can it be used as means of evidence?

4. Conceptualizing a GDPR-compliant SIEM system

Although SIEM systems have grown to mature security tools, privacy has largely been neglected in this area. Thus, we have previously defined central research questions that arise when applying the GDPR regulation. To answer these questions, we propose a SIEM architecture that is compliant with GDPR, while largely preserving its functionalities in this section. Therefore, we propose concrete solutions for each of the individual research questions based on an extended, GDPR-compliant architecture.

4.1. DINGfest base architecture

This section gives an overview on our general security monitoring architecture and assigns the previously defined research questions to the respective areas of the architecture. The presented architecture is based on the general DINGfest architecture as presented in Menges et al. (2018) and extends it by data protection measures and the resulting GDPR compliance. DINGfest is a research project that aims at improving the detection, forensic analysis and the reporting of detected incidents. The project started June 2016 and will finalize at the end of 2019 and is funded by the German Federal Ministry of Education and Research. The general system monitoring architecture is illustrated in Fig. 1. It consists of three main modules – namely data acquisition, data analysis and incident reporting located within an organization and shows an external authority possible counterpart for the receipt of detected incidents. The counterpart is intentionally included in the architecture design, since its role and the management of the data flows generated there are one central factor in ensuring the legal compliance of the system. This concerns data protection requirements according to the GDPR on the one hand and may also concern existing statutory reporting obligations of the organization.

³ ISO/IEC 23,264 Redaction of Authentic Data <https://www.iso.org/standard/78341.html> [last accessed: Jan. 2020].

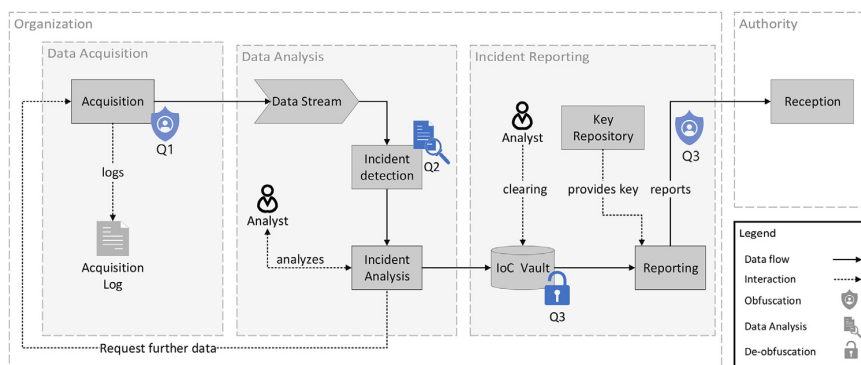


Fig. 1 – DINGfest Base Architecture.

The **data acquisition** module collects data from all monitored computing resources in the company. This data may contain personal data of employees and customers that needs to be protected. The monitored resources are not only computing devices like workstations, servers and mobile devices, but also network devices like routers and switches. The actual data is obtained from various sources. This includes, for example, data extracted with the help of Virtual Machine Introspection (VMI). This also includes data obtained from system log files or incident information provided by human sensors [Vielberth et al. \(2019\)](#). Moreover, all data extractions within the acquisition area are stored in the acquisition log. This enables a later auditability of all the information obtained. The extracted data is finally pushed into a larger data stream, which serves as data basis for the data analysis section. Data acquisition is the starting point at which all data (including personal data) is transferred to the system. As a result, research question Q1 must be addressed within this module to show how data must be protected or pseudonymized to ensure GDPR-compliant data handling. This additionally generates the required prudential value for the gathered evidence.

The **data analysis** module analyzes the whole data stream and tries to detect security incidents using a combination of fingerprinting and pattern recognition. If the detection engine discovers a potential security violation it generates an incident alert that contains a description of the assumed violation and the related data records. The alert is then received and analyzed by a forensic analyst. The analyst can use a visual analysis interface and request additional data from the data acquisition module. Should the suspicion be confirmed, the incident is forwarded to the reporting module. Otherwise, the incident alert is deleted right away. As shown above, the data acquired during data acquisition must be protected. This makes data analysis more difficult, since information is lost as a result of pseudonymization. Therefore, the research question Q2 will be addressed within this module to show how far incident detection with disguised data is still possible.

The tasks of the **reporting module** include the long-term storage of analyzed incidents (usually between several weeks and several years depending on the local legislation) and the reporting to local authorities in accordance with the legislation in force. Arriving incidents are therefore stored in a database and processed by an incident reporter. During this process the

reporter might query the database to contrast the current incident with past incidents. Eventually a report is generated and forwarded to local authorities, in order to inform them or comply with regional regulations. Such reports may contain information about innocent individuals or company assets which also need to be protected. Within this module the research question Q3 will be addressed. The aim is to ensure that information can be de-pseudonymized in the event of an actual incident, while preserving its integrity. This is necessary to enable the use of the data as evidence in possible later court cases. Furthermore, it must also be ensured that the data can be reported to the appropriate authorities in compliance with the law and data protection regulations.

4.2. GDPR compliant data processing

In the following we present our approach in more detail, especially relevant parts, which enable GDPR compliant data processing inside SIEM (Q1). To this end, we propose an approach that pseudonymizes personal data at relevant points and at the same time allows to de-pseudonymize this data in case of a detected incident (Q3) in compliance with GDPR regulations. [Fig. 2](#) shows the basic structure of this approach. To achieve a pseudonymization of the information, cryptographic methods are used. These on the one hand prevent access to personal data by encrypting it and on the other hand allow decrypting it under certain constraints specified by the GDPR. However, the decisive question is where the data must be encrypted and how to implement the key management for encryption.

The GDPR demands the protection of personal data as it is processed. Thus, we argue that personal data must be pseudonymized as soon as possible in the system. In the case of SIEM this is the case directly after or ideally during the data is acquired. To achieve this, a public key (B) is provided by a TTP (Trusted Third Party), which is responsible for key management. With this key, the fields containing personal data are encrypted asymmetrically and the unencrypted personal data is deleted. In order to be able to comprehend and proof, that all personal data has been pseudonymized, an Acquisition Log is kept. For this purpose, we propose to use a tamper proof logging scheme, which synchronizes all logged data with an external blockchain as presented by [Putz et al. \(2019\)](#).

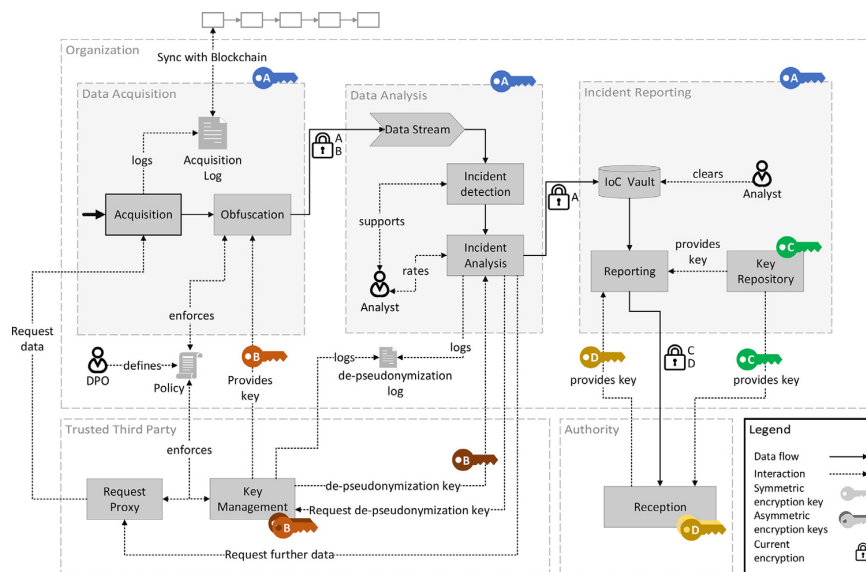


Fig. 2 – DINGfest GDPR Architecture.

For proper use of public key cryptography, we refer to López et al. López et al. (2005).

In order to determine, which fields must be encrypted, a Policy is followed. For each logging system type an individual mapping must be defined that specifies the fields containing personal data. This policy is defined by the *Data Protection Officer (DPO)* or an equivalent position inside the organization. A DPO is responsible for compliance and data protection within an organization and should have the necessary expertise to make the required decisions.

In addition to the data protection aspects shown above, the model presented is also intended to provide protection against attacks resulting from the attacker model defined in Section 3. According to the assumptions made, possible dangers from insiders and outsiders are examined more detailed in the following.

Outsider: According to the model, outsiders can be divided into two main groups. Common outsiders, which have no specific reference to the system, and the TTP as an outsider, which is partially involved in the analysis process. The authority as the third participant is not considered in detail here, as it is supposed to have access to the data it receives in the context of a report.

- **Common outsider:** The major problem is to prevent data flows to outsiders. Specifically, the areas of data acquisition, data analysis and incident reporting must be protected. This is essentially guaranteed by a consistent use of the internal, symmetric key A. This ensures that the data remains protected even in the event of an unwanted extraction. The data may only be possibly unprotected in the case of an extraction during the data acquisition process. The security of this data mainly depends on the level of protection of the underlying source system.

- **TTP:** The TTP used in the present data model also represents a specific outsider, who is integrated into the SIEM process. However, the TTP only receives the key B from the data flow for custody, but does not have access to data from the data stream at any time. It is also worth noting that the TTP does not have access to the key A at any time. Accordingly, the TTP must be considered the equivalent of other outsiders in the case of data leaks.

During data analysis, an incident detection approach is followed. This approach is mainly automated but can also be supported by human analysts. Thereby, the incident detection is conducted solely on pseudonymized data and thus is GDPR compliant. The thereby used event detection approach is elaborated further in the following chapter.

Insider: In the present model, only two groups of people have access to the internal data. These are analysts in the areas of data analysis and incident reporting on the one hand and data protection officers who define the corresponding policies on the other hand.

- **Analyst:** The analysts involved are only provided with specific data extracts and personal data under certain circumstances. For this purpose, an approval for specific data components must be granted according to the policy defined by the data protection officer. If such an approval does not exist, analysts always work only with pseudonymized data.
- **DPO:** The data protection officer is never given access to the data within the data stream and thus has access to resources that are equivalent to an outsider. On the other hand, the DPO has a protective influence on the data stream by defining the respective policy. This ensures that the protection of the data stream is always split between two different roles within the company.

Table 1 – The unified log message format Latzo and Freiling (2019).

Name	Description
source	The source from where the message comes from.
type_id	Describes the type of the message, e.g., the system call number.
date	Timestamp of the message when it was generated.
path	A path, e.g., which path was opened.
user	The user who performs the event.
process_name	The name of the process that performs the event.
...	...
misc	Can be used for random things (no personal data) that do not fit into that format.

```
{
  "source"      : "syscalls",
  "type_id"    : 59,
  "process_name": "ls",
  "user"       : "alice",
  "pid"        : 103,
  "path"       : "/home/alice/topsecret/"
  ...
}
```

Listing 1 – Example of a unified message.

4.3. Event detection on protected data

Different software usually comes with different log formats that is often loose text. In our case, we use standard Linux logs like *syslog* and *auth.log* that usually come with a Linux distribution. Furthermore, we use *access.log* of Apache's HTTP Server [The Apache Software Foundation \(2019\)](#). Since system call traces are a rich source of behavioral information [Lanzi et al. \(2010\)](#); [Rieck et al. \(2008\)](#), we also use system calls traces that are obtained via virtual machine introspection. System call tracing has a negative impact on performance, but especially enterprise environments can benefit since some events cannot be detected using common logs.

Log messages are transformed into a unified structured log format. An excerpt of the message format that we use in DINGfest can be seen in [Table 1 Latzo and Freiling \(2019\)](#): The entry *source* specifies from which of our sources log the message comes from. Thereby, we assume that it is not possible to deduce from the source to the user, i.e., in server scenarios. One of the most important attributes of the unified log message is *type_id*. This ID specifies what kind of message it is. In case of system calls the *type_id* is the system call number. Misc may contain arbitrary information that does not fit into the unified message format, e.g. command line option. We assume, that this field does not contain personal data. For the evaluation we checked manually that this field does not contain personal data. Another useful feature is *path*. However, the path may contain personal data such as the user name.

An example of a unified log message is shown in [Listing 1](#)

So we can distinguish between three kinds of log file entries:

1. Those that definitely contain personal data (e.g., user),
2. those that may contain personal data (e.g., path), and
3. those that do definitely not contain personal data (e.g., source, type_id, process_name, misc)

The classification may vary from system to system. For example, it is also possible that in a specific scenario a process name or a source name may also contain personal data. The classification, however, determines which features can be used for privacy friendly event detection, namely only features from the third category. We use this idea in our evaluation in [Section 5.1.2](#) to assess the impact of privacy protection on event detection quality.

4.4. De-pseudonymization in case of an incident

In the previous section we presented an approach that allows us to perform incident detection on data that is pseudonymized according to the GDPR regulation. On this basis, this section describes how security incidents can be de-pseudonymized after detection in order to analyze and process them further and to prepare them for a legally compliant report. The complete process for these descriptions is additionally shown in [Fig. 3](#).

When the automated data analysis found indications for a possible incident, it is first necessary to verify the result. For this purpose, the data is revised by an analyst to ensure that it is an actual incident and to avoid false positives. Once the analyst approves the incident within the *Data Analysis* module, the *Trusted Third Party (TTP)* is contacted, which initially provided the public key B for the data pseudonymization within the *Data Acquisition* module. The TTP receives the signature of the data packet concerned in order to be able to identify the appropriate key and checks the request against the decryption policy specified by the DPO. If the check is negative, the request is denied. If the check is positive, the TTP determines the correct private key B for the signature provided and sends it to the *Data Analysis* module to allow the de-pseudonymization of the data. When de-pseudonymizing incident data, it is also important to enable auditing of de-pseudonymizations. Therefore, we store every de-pseudonymization request and sign it, in order to be able to provide proof of data access afterwards.

After the analysis module has received the private key B from the TTP, the initially appended data markings are used to identify all fields that contain pseudonymized information within the data package and to de-pseudonymize it. The resulting data package is then transferred to the Incident Reporting module where it is decrypted using the key A and stored within the *IoC Vault*, which is an integrity proof, long term storage for incident data as shown by Boehm et al. [Böhm et al. \(2018\)](#). This data can then be used for further analyses and incident reports. If an incident must be reported, the data needs to be cleared by an analyst first. This is important to prevent both privacy violations of personal data and the publication of confidential company data. More specifically, the analyst must decide, which data is to be excluded from the

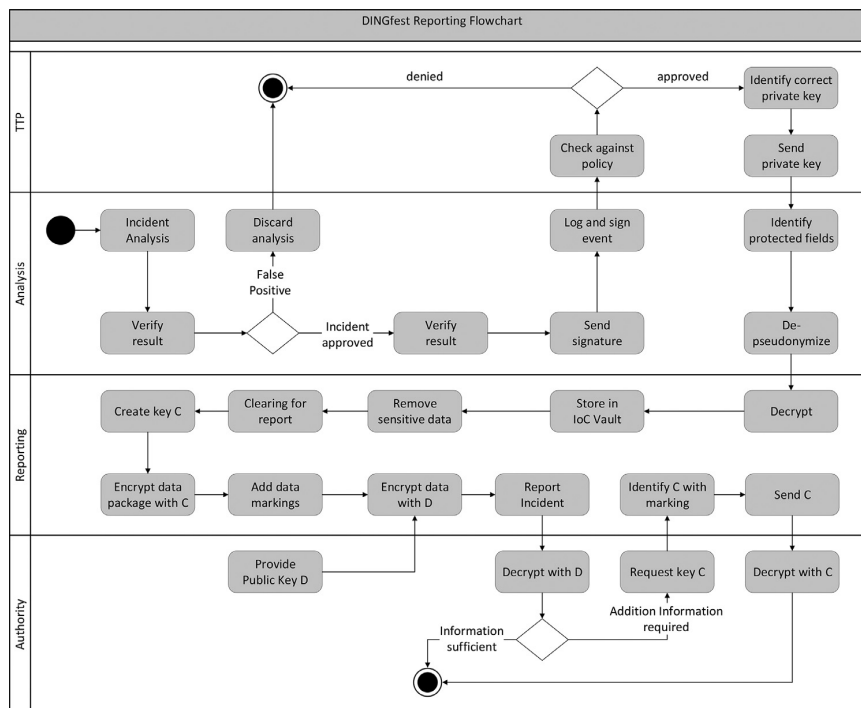


Fig. 3 – DINGfest Reporting Flowchart.

report and which is to be secured. This additional information about performed pseudonymizations and exclusions is appended to the data using further data marking definitions. The actual extent of data protection, however, may strongly depend on the use-case of the report. While in the case of reports within the scope of a reporting obligation, the statutory requirements must be complied with, in the case of voluntary reports significantly more data can be concealed or removed.

After an analyst has cleared an incident and prepared it for a report within the *IoC Vault*, it is transferred to the *Reporting* module in the next step. A symmetrical key C is created in the key repository and assigned to this very data package to pseudonymize the previously chosen contents. Furthermore, the package is extended by an additional data marking that contains the signature of the utilized key C key to enable later attributions. The key C is then used to pseudonymize the data according to the analyst's specifications. In addition to this, the data is also encrypted with the public key D, provided by the recipient (for example, an authority). This ensures that the reported data can only be opened by the correct recipient, i.e., a legal authority. In a final step, the data (secured with keys C and D) is transferred to the recipient. If the data was reported due to a reporting obligation, such as European NIST directive [European Commission \(2016\)](#) or German IT-security law [Deutscher Bundestag \(2015\)](#), the receiving authority may request a decryption afterwards under certain conditions. In this case, the authority needs to be able to request the key C from the organization. In order to receive the key C, the authority transmits the key signature contained in the data package to the organization. This enables the assignment of

the correct key. If the correct key C is assigned, it can be transferred to the authority.

4.5. Evidence generation using malleable signatures which withstands pseudonymizations

We can positively answer the second half of Q3, i.e. we can ensure that we are able to use the pseudonymized or partial reported events as means of evidence. Assume we add protection of integrity and origin authentication during the data gathering inside the data acquisition module. Inside DINGfest's base architecture, the evidence could be protected by standard electronic signature schemes, e.g. the different acquisition modules would sign the data they gathered. When a cryptographic signature algorithm complies with the requirements of common legal frameworks for electronic signatures its signature provides a high probative value for the data being signed. This said, any subsequent modification for data-protection compliance would destroy any evidence guarantees for the remaining data, e.g. removing the full path from a signed full file name of a malicious executable as it contains a personally-identifiable user name also removes the evidence protection for the name of the executable. Thus, we propose use redactable or sanitizable (malleable) signature schemes to retain the authenticity and integrity of the data gathered from the data acquisition module towards the final report. This means that, if wanted, the digital signature protects the authenticity of the data provided even till the incident report, i.e. so in the final report one can verify that the event data has

not been modified in unauthorized ways –not tampered with– and that it originated from a trusted data acquisition module.

By omitting the cryptographic details of other malleable signature schemes, the privacy statement describes, which parts of the data could be removed or pseudonymized. This allows to control the signature scheme for which subsequent changes are made and parts are authorized. Thus, removing the original data or encrypting these parts, would allow the subsequent steps to always verify the authenticity of the remaining data. When data is de-pseudonymized the best-suitable malleable signature schemes are those that offer mergeability Pöhls and Samelin (2015). This allows to put data parts back into the signed data set and thus the original malleable signature would now verify over all remaining parts, the ones previously readable plus those added by the de-pseudonymization. If not added, private malleable signature schemes Bilzhause et al. (2017) retain the confidentiality of all those parts that have been removed, i.e. even though one can successfully verify the signature on partial data, the information contained in the signature itself does not allow an attacker to gain information on the data parts removed and thus not shared. Hence, the added value of a retain-able private malleable signature, like Pöhls and Samelin (2015), does not violate any GDPR requirements Pöhls (2018).

The legal analysis of these private accountable redactable signature scheme shows that they increase the legal probative value for the signed reported data as eIDAS compliant electronic signatures could provide Höhne et al. (2012); Pöhls (2018); Pöhls and Höhne (2011a); van Geelkerken, F.W.J. et al. (2015). Hence, the DINGfest GDPR architecture protects the records such that the remaining information can be used as means of evidence; further after the de-pseudonomization steps at any later time the data's origin and originality is at-tested.

5. Evaluation

In the previous sections we presented an architecture for a GDPR compliant SIEM system. The central elements of our approach are to guarantee a GDPR compliant data processing, to enable the recognition of security incidents on pseudonymized data as well as the de-pseudonymization of the data in the case of an incident. In this section we evaluate the validity of our approach in two ways. First, we conduct a technical evaluation of the impact of pseudonymization on the detectability of events. Subsequently, we carry out a legal evaluation for our proposed solution. To achieve this, we investigate the individual components of our architecture presented in Section 4.2 on conformity with the specifications of the GDPR as shown in Section 2.2.

5.1. Impact on detectability

5.1.1. evaluation methodology

The evaluation of the impact of privacy protection on the quality of SIEM is performed based on the theory for forensic fingerprint calculation of Dewald Dewald (2015).

In this theory, all interactions of interest with the system (e.g., by users) are called *events*. An example event is the login

of a user. Many events either directly or indirectly leave digital traces within the system (e.g., in log files on disc or in main memory). These traces are formalized as *feature vectors*. Generally, a feature is a quantifiable attribute of a system that can be observed by the SIEM. In our study we concentrate on feature vectors that can be extracted from log files system call traces Latzo and Freiling (2019). Obviously, tracing all system calls is very expensive in terms of performance, there are ways to get rid of most overhead caused by system call tracing. The theory Dewald (2015), which we now explain, defines conditions under which an event is detectable based on the features traces it leaves in the log files of a system.

The set of features that we consider in our evaluation is based on an abstract representation of log file entries and attempts to harmonize many log files in modern systems. Our format represents every log file entry using the following four *features*:

- a *source* from what log the message comes from,
- a generic *type_id* that describes the kind of log message,
- a *path*, and
- a *misc* field that may contain arbitrary content (e.g., the name of a network adapter).

A *feature vector* is a vector of values for these features. Depending on the system, there can be many different features vectors consisting of these four features. Since an event can cause multiple entries in multiple log files, we define the set of feature vectors that are generated as the *evidence set* of that event.

More formally, let Σ be the set of all possible events that can happen in the system and are of interest to the SIEM. When some event $\sigma \in \Sigma$ happens, log entries are generated. The *evidence set* $E(\sigma)$ of event σ is the set of all subsets of feature vectors that are thereby generated by σ . It is technically necessary, that the evidence set is closed under subsets. Intuitively, it can be interpreted as the fact that partial evidence is also evidence of the event.

It is obvious that the evidence sets of different events may overlap. To be able to detect an event, it is necessary to calculate the *characteristic evidence set* $CE(\sigma)$ Dewald (2015) of an event σ , which is defined as the set that contains only feature vectors that are caused by σ and not by any other event $\sigma' \in \Sigma$. Formally, the set of characteristic evidence of an event σ with respect to a set of other events Σ' is defined as follows:

$$CE(\sigma, \Sigma') = E(\sigma) \setminus \bigcup_{\sigma' \in \Sigma'} E(\sigma')$$

The set of characteristic evidence of an event is also called *characteristic fingerprint* of that event.

As one can see in the formula above, a characteristic fingerprint is defined for a specific reference set Σ' . All feature vectors that are caused by $\sigma' \in \Sigma'$ are not in $CE(\sigma, \Sigma')$. One can say, that $CE(\sigma, \Sigma')$ is the evidence set $E(\sigma)$ minus all other evidence sets of events in Σ' . Let $|CE(\sigma, \Sigma')|$ and $|\Sigma'|$ be sufficiently large, then a match of the feature vector with the log files of a system is a clear indication that σ happened and not σ' . The size of CE is an indication for the discriminative power of the evidence. The larger the set, the higher is the probability that

Table 2 – The events used for the evaluation.

Class	Name	Description	$ CE(\sigma, \Sigma') $		Loss Factor
			F_1	F_2	
CLI	ls	Lists files	1	1	0.0
	cp	Copies file	4	1	0.75
	mv	Moves file	2	1	0.5
	cat	Cats file	0	0	0
	vmstat	Virtual memory statistics	6	1	0.833
	netstat	Network statistics	15	1	0.933
	tar	Creates compressed tar archive	5	4	0.2
	rm	Removes file	1	1	0.0
	shred	Shreds file	2	1	0.5
	curl	Downloads file	1	0	1
CLI Root	tailShadow	Reads /etc/shadow	7	2	0.714
	catCredentials	Reads Wordpress config file	4	2	0.5
	vimHosts	Opens /etc/hosts in Vim	220	3	0.986
	rmSudo	Removes file with sudo	2	2	0.0
Web	shredSudo	Shreds file with sudo	9	3	0.667
	wordpressLogin	Wordpress Login	63	10	0.841
	wordpressSearch	Wordpress Search	3	0	1
Service	wordpressOpen	Opens Wordpress website	0	0	0
	sshLogin	SSH login (server side)	2219	466	0.79
	apacheStop	Stops apache web server	1712	15	0.991
Kernel Modules	mysqlWp	Login into Wordpress DB via command line	47	1	0.979
	lsmod	Lists loaded kernel modules	251	1	0.996
	insmod	Loads kernel module	10	3	0.7
Docker	rmmod	Unloads kernel module	12	3	0.75
	dockerHelloWorld	Starts docker hello world example	28	3	0.893
	dockerUbuntuLog	Starts docker ubuntu and show log	23	5	0.783
	dockerImages	Lists all docker images	1	1	0.0
	dockerPs	Lists all running dockers	0	0	0
	dockerPSA	Lists all dockers container	0	0	0
	dockerUbuntuSleep	Starts docker in background	2	2	0.0
	dockerRm	Removes all docker containers	0	0	0
	dockerNginx	Runs nginx docker and curl it	65	8	0.877
	dockerUbuntuBash	Attaches bash of container	0	0	0
	dockerPrune	Removes unused container	1	1	0.0
	dockerPruneVolumes	Removes unused objects and volumes	1	1	0.0
	dockerRmImages	Removes all images	2	2	0.0
	dockerUbuntuBashCp	Attaches container and runs cp	0	0	0
	dockerUbuntuBashMv	Attaches container and runs mv	18	1	0.944
	dockerUbuntuBashRm	Attaches container and runs rm	3	1	0.667
	dockerUbuntuBashCat	Attaches container and runs cat	24	0	1
Nextcloud	nextcloudStatus	Shows Nextcloud status	3	2	0.333
	nextcloudAppList	Lists Nextcloud apps	44	2	0.955
	nextcloudUserList	Lists Nextcloud user	3	2	0.333
	nextcloudUserAdd	Adds new Nextcloud user	103	16	0.845
	nextcloudGroupList	List Nextcloud groups	5	2	0.6
Average					0.508

the event may be detected no matter what the reference set Σ' looks like. However, it is also possible, that $CE(\sigma, \Sigma')$ is empty, i.e., one cannot detect reliably the occurrence of σ .

5.1.2. Characteristic evidence without personal data

We now evaluate the impact of pseudonymization on the existence of characteristic evidence that is needed for event detection. The evaluation setting is based on the DINGfest architecture as described by Latzo and Freiling [Latzo and Freiling \(2019\)](#) [Latzo \(2020\)](#). We calculated evidence and characteristic evidence sets for 45 different events (see also [Table 2](#) that typically appear in Linux server environments as one typically finds them in small and medium-sized enterprises. In

our threat model, we consider an adversary with root privileges that were either gained via a privilege escalation attack or by having them anyway (i.e., a malicious insider). Basically, it is not possible to determine the intention of an administrator's input. Hence, most events can also be used maliciously, e.g., for information retrieval, covering traces, etc. So, basically all events might be interesting during a forensic analysis or event detection.

The higher the number of feature vectors in a characteristic fingerprint, the better the quality of that fingerprint. This is intuitive since a feature vector in a fingerprint is basically an indicator of an event. In the evaluation, we compare the size of characteristic evidence sets with and without taking personal

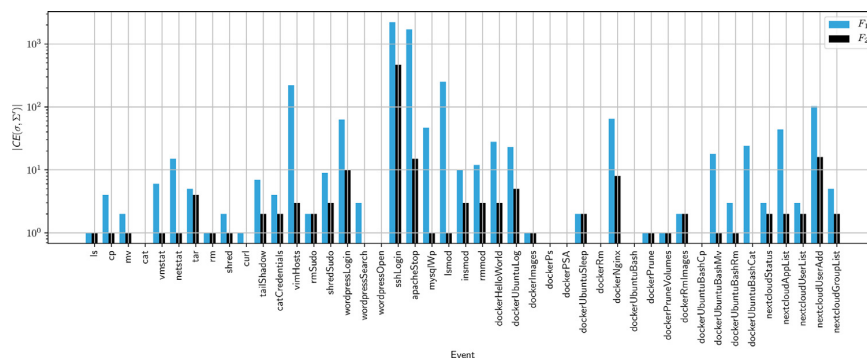


Fig. 4 – Comparison of powers of characteristic evidence sets with personal data (blue) and without personal data (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data into account. More concretely, we consider the following two feature sets:

- $F_1 = \{source, type_id, path, misc\}$
- $F_2 = \{source, type_id, misc\}$

First, F_1 is a feature set that has turned out to be reasonable for our events. However, F_1 includes the *path* feature that may contain personal data. Features of F_2 do not contain personal data. For calculating the fingerprints, an event was executed 40 times (trainings set). Furthermore, the reference set Σ' for a characteristic fingerprint of σ are always all other events. Table 2 compares sizes of the characteristic evidence set using the two feature sets including the decrease rate when using F_2 instead of F_1 . As one can see, omitting the path as a feature has a huge impact on the size of characteristic evidence. On the average, using F_2 reduces a characteristic fingerprint by half of its feature vectors. For three events, there is no characteristic fingerprint, anymore. Fig. 4 shows the absolute number of feature vectors of the characteristic evidence sets using F_1 and F_2 . “Big” events, that come originally with big characteristic fingerprints are in general more affected than smaller events.

We have shown that it is possible to calculate characteristic evidence for all events even if features that contain personal data are not used. However, the fingerprints that we generated had a lower quality, i.e., the size of the characteristic evidence set was reduced by an average of about 50%. By extending the feature set and the set of traces acquired from the SIEM, we conjecture that fingerprints can also be calculated for this action even if data is pseudonymized.

In the following we want to compare the matching results using characteristic fingerprints with F_1 and F_2 . For matching, we calculate a score that indicates what proportion of feature vectors in event traces are matched by a characteristic fingerprint. Fig. 5 (F_1) and Fig. 6 (F_2) show the corresponding matching matrices. The values there are average values of 10 traces of the event (test set). It stands out that the matching matrices are quite similar. In Fig. 6 there is much less matched noise. The characteristic fingerprints in Fig. 6 are much smaller than in Fig. 5, though. For that events for which we could calculate

a characteristic fingerprint, the matching results are similar good. This is also confirmed by the Receiver Operating Characteristic (ROC) curve in Fig. 7. There, the true positive rate (sensitivity) and the false positive rate are plotted against each other with different thresholds. It shows, that the sensitivity with F_2 is only a little smaller with about 78% versus the sensitivity of F_1 with about 84%.

In this section we showed that it is possible to calculate characteristic fingerprints with a reduced feature set that does not contain personal data. While the size of the characteristic fingerprints decreased when using that feature set, the matching results were very similar. It showed, that when there is a characteristic fingerprint, matching usually also works. So, to improve this approach, future work should focus in extracting more features from logs that help to increase the size of characteristic fingerprints.

5.2. Legal evaluation

As we have seen above, to generate fingerprints of high quality, data must be obtained by processing previously collected data during the data acquisition which clearly relates to a personal user’s actions and thus is considered as personal data. Hence, the general principles mentioned in Section 2.2 relating to processing of personal data and especially a lawful processing are important. Processing shall be lawful only if one of the Art. 6 GDPR included reasons applies. Applicable and best suited for the SIEM case is Art. 6 (1) lit. f, when processing is necessary for the purposes of the legitimate interests. This clause is different to the other lawful bases as it is not centered around a particular purpose and it is not necessary that the individual has specifically agreed to (consent). Legitimate interests are more flexible and could in principle apply to any type of processing for any reasonable purpose. Art. 6 (1) lit. f states:

“1. Processing shall be lawful only if and to the extent that at least one of the following applies: (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, [...]”.

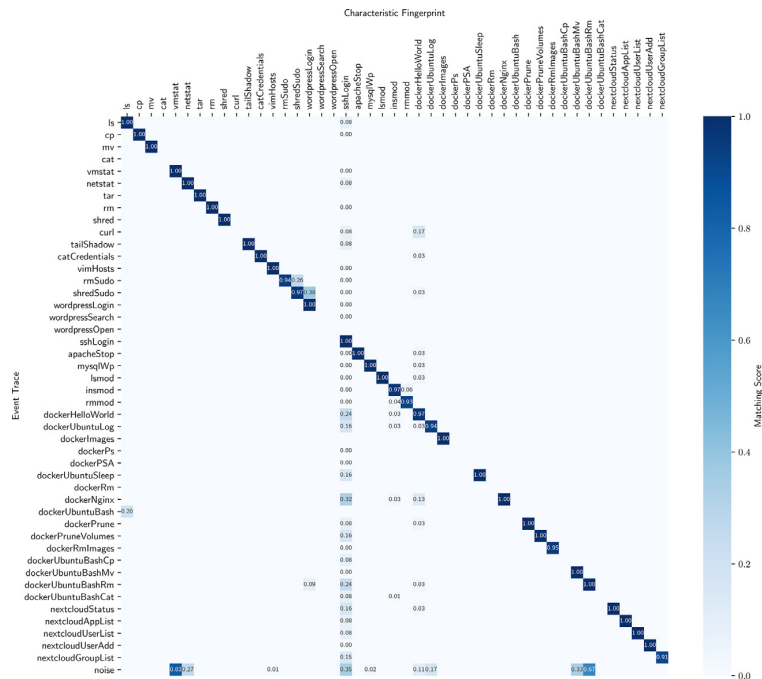


Fig. 5 – Matching matrix using F_1 . The events listed on the y-axis are the ground truth, the events on the x-axis correspond to the characteristic fingerprints [Latzó \(2020\)](#).

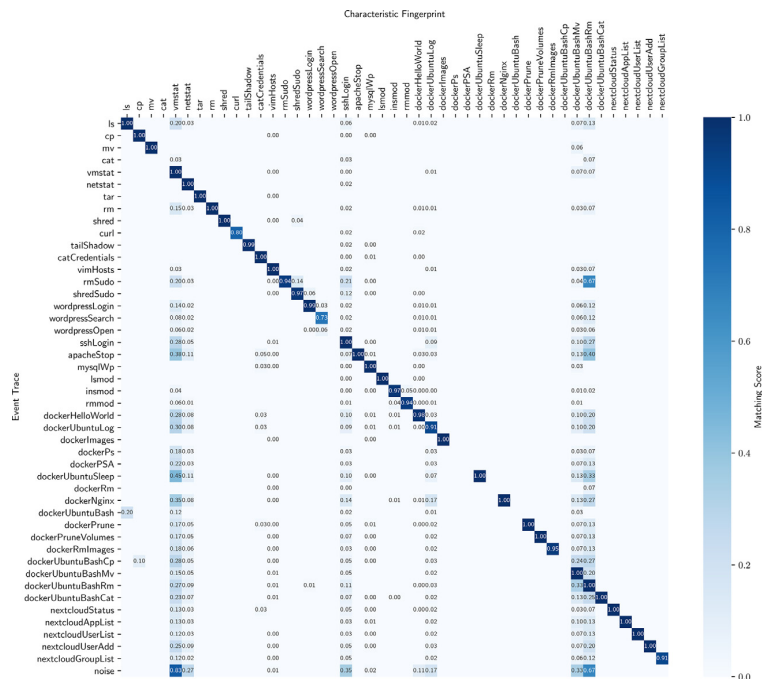


Fig. 6 – Matching matrix using F_2 . The events listed on the y-axis are the ground truth, the events on the x-axis correspond to the characteristic fingerprints.

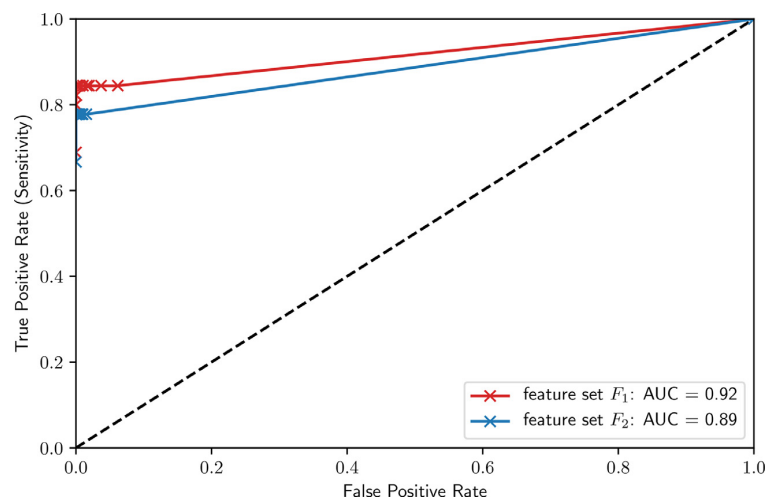


Fig. 7 – ROC curve with F_1 in red and F_2 in blue. The differences are quite small. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.2.1. Legitimate interest is balanced with personal data protection

Since legitimate interests can apply in a wide range of circumstances, it is mandatory that the controlling party puts its legitimate interests and the necessity of processing the personal data to the interests, rights and freedoms of the individual in balance. To provide a balance-test, the key elements of the legitimate interests provision is contained in a so-called 'three-part test'. Whereas this test is not explicitly named in the GDPR, the legitimate interests provision does incorporate three key elements:

- Purpose test: there must be a legitimate interest behind the processing.
- Necessity test: the processing must be necessary for that purpose.
- Balancing test: the legitimate interest must be balanced with the individuals interests, rights or freedoms.

This concept of a three-part test for legitimate interests has been confirmed by the Court of Justice of the European Union in the *Rigas* case (C-13/16, 4 May 2017) in the context of the Data Protection Directive 95/46/EC, which contained a very similar provision. This means, the controller must be able to meet all three requirements of the test prior to commencing the processing of personal data.

Firstly, **purpose** is clearly given as the whole purpose of the SIEM architecture as given in Section 4 is to detect unlawful use of information systems and their data, it is important to make clear that the European Parliament has already considered the legitimate interest of processing personal data necessary for the purposes of preventing fraud. This is explicitly backed by recital 47: "[...] The processing of personal data strictly necessary for the purposes of preventing fraud also constitutes a legitimate interest of the data controller concerned. [...]"

Secondly, with regards to the condition relating to the **necessity** of processing personal data, it is important that deroga-

tions and limitations in relation to the protection of personal data must apply only in so far as is strictly necessary. In that regard, communication of features which do not contain personal data, does not make it possible to identify a person with enough precision in order to be able to bring an action against him. Accordingly, for that purpose, it is necessary for the SIEM system to obtain also the possibility of full identification of that person, i.e. allow to de-anonymize and retain authenticity proofs in order to construct substantial and reliable evidence of an unlawful use of the system against that person.

Thirdly, it is necessary to make a '**balancing test**' to justify any impact on individuals. During the test the controller takes into account "the interests or fundamental rights and freedoms of the data subject which require the protection of personal data", and makes sure they don't override his interests. In recital 75 speaks of the risks of the rights and freedoms of natural persons, of varying likelihood and severity, may result from personal data processing which could lead to physical, material or non-material damage, in particular: where the processing may give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorized reversal of pseudonymisation, or any other significant economic or social disadvantage; where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data; [...]; where personal aspects are evaluated, in particular analyzing or predicting aspects concerning performance at work, [...]; or where processing involves a large amount of personal data and affects a large number of data subjects."

Since the data acquisition module collects data from all monitored computing resources in the company, one can assume a great danger for personal data of employees and customers. Also, the analysis of personal data in the data stream by fingerprinting and pattern recognition and especially the merging of data is in general - interfering with the privacy

rights of a natural person. And finally, is the reporting module and the included long-term storage of analyzed incidents as well as the reporting to the authorities itself a potential risk for personal data. Since the complete monitoring of the users without cause is not compliant with the GDPR, especially since the user does not have any possibility to intervene, fundamental rights would be violated, if the controller is not implementing appropriate technical and organizational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate, for example the pseudonymisation and encryption of personal data.

5.2.2. Mechanisms for GDPR-compliance in DINGfest architecture

DINGfest's GDPR architecture counters the above-mentioned problems by implementing special steps as part of their work flow for a GDPR-compliant SIEM system:

First of all, by a continuous pseudonymisation through obfuscation during the data acquisition. The suspension will only be carried out under certain conditions determined by controller and, in particular, in case of suspicion of a criminal offence. Since the public key is always provided by a trusted third party (TTP) and policies provide the organizational background before a special field gets encrypted, the balance between the rights of the controller and the user should be met. All technical steps in which personal data is processed are accompanied by a special pseudonymization method through obfuscation. If data analysis has then found indications for an (possible) incident, the data protection officer has to approve this case as an "incident case" within the data analysis module. Only then the TTP receives a key identifier for the data packet - not the packet's contents, not even in pseudonymous form -, in order to find the appropriate key. The critical point is the policy (see the "Policy" defined by the Data protection office (DPO) in Fig. 2), which is being consulted by the TTP before sending the decryption key to the data analysis module. This organizational measure ensures a level of security appropriate to the risk, which is to reveal private data to the controller. By providing a log of every request to de-pseudonymize data fields the architecture enables to comply with transparency requirements, like the right of access.

Only after passing this safety measure the DINGfest architecture allows to reveal data to the controller (the data analyst) using the private key B, to de-pseudonymize all necessary fields that contain pseudonymized information within the data package. Only if the analyst decides to include this in the report the resulting data package is then transferred to the incident reporting module, during transfer and storage it gets again encrypted under key A. Again, the access to the encrypted long-term storage of the data within the IoC vault, including the use of data for further analysis and incident reports, only applies for cases that deserve an attention because of potential unlawful behavior. This is clearly a legitimate interest which is not overridden by recital 47 of the GDPR. Furthermore, it depends on the individual use case, which data is to be excluded from the report and what data must be removed or stays pseudonymized. This extra step, in which the analyst balances the rights of both parties, is the very essence of the balance test, and here the DINGfest GDPR architecture implements this check to balance the legitimate inter-

est with the individual's interests. As said, if the data analysis module decides to keep pseudonymization of fields then only pseudonymized data of this incident is transferred to the incident reporting module. When the report is generated, the data is again encrypted with the public key of the recipient in most cases an authority. This way, the reported data can only be opened by the correct recipient.

5.2.3. DINGfest's GDPR architecture reduces the risks for personal data

By using the methods of pseudonymisation and encryption of personal data, it is to be concluded that scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller has implemented appropriate technical and organizational measures to ensure a level of security appropriate to the risk.

6. Conclusion

In this paper we presented an architecture for a GDPR compliant SIEM system, as implemented in the DINGfest prototype SIEM system. We first identified central questions that must be answered for the development of such a platform. The questions affected the necessary conditions for a GDPR compliant data processing, techniques for the incident recognition on pseudonymized data as well as a lawful de-pseudonymization techniques in the case of occurred incidents. We then answered these questions with the help of our architectural design and evaluated them both from a technical and legal perspective. Using this evaluation, we have shown that it is possible to comply with the legal requirements for pseudonymization, while at the same time keeping detectability. Altogether we presented a base architecture for a GDPR compliant SIEM system with this work. Although it was developed based on our underlying system DINGfest, it may also serve as a draft for other security systems that have to be adapted to GDPR specifications.

In the context of this work, it was revealed that the performance of the recognition mechanisms used can be impaired using pseudonymization. This is one challenge that could be addressed in future work. Beyond that we defined the fundamental boundaries of a GDPR conform architecture with this work. However, various details were not considered. An example for this is to transfer our architecture to already established SIEM systems. Each system is tailored to its infrastructure and thus, it is necessary to define, which of the collected data sets needs to be protected. This applies both to data that is collected during initial data acquisition and to data that is prepared for a report. To support this process, it would be helpful to develop a central repository that defines the data points relevant to data protection for frequently used data sources. Furthermore, it will also be necessary to develop the needed details for the data protection policy within SIEM systems in future works. It would be conceivable to develop a generally applicable basic policy and specific implementations of this policy adapted to individual systems.

Regarding the legal probative value DINGfest using malleable signatures allows to balance integrity protection for

evidence and GDPR-compliant removal or pseudonymization of the gathered data. To achieve this the data acquisition module emits malleable signed data –instead of simply digitally signed data– and hence any subsequent modification due to GDPR-compliant processing does not inhibit the verification of the integrity and origin of the remaining data. With a scheme that is accountable and private and supports mergeability, previously obfuscated parts of an entry can be subsequently de-obfuscated and the signature still verifies and provide means of evidence.

CRedit author statement for the paper

Towards GDPR-compliant data processing in modern SIEM systems

The conceptualisation of our GDPR concept was jointly developed by all participating authors within the framework of the DINGfest research project. This includes the development of the methodology and software as well as the validation of the results and the creation of the manuscript. The specific contributions of this work are grouped according to the participating institutions in the following: The implementation of this work was coordinated and administered by the authors of the University of Regensburg. They also contributed significantly to the development of the problem definition and the SIEM basics. Nexis GmbH was able to contribute important insights into the practical application of SIEM systems. The conceptualization of the presented architecture was mainly developed by the authors of the University of Regensburg and Passau. The University of Passau has also contributed extensive knowledge about redactable signatures to this work. The detection of events as well as the technical evaluation was carried out by the authors of the University of Erlangen-Nuremberg. The legal analysis of the work was conducted by the law firm Paluka Sobola Loibl & Partner in cooperation with the authors of the University of Passau.

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.

Acknowledgement

This research was supported by the Federal Ministry of Education and Research, Germany, as part of the BMBF DINGfest project (<https://dingfest.ur.de/>). The research of H.C.Pöhls was carried out in the project SEMIOTICS funded by EU's H2020 grant no. 780315.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.cose.2020.102165](https://doi.org/10.1016/j.cose.2020.102165).

REFERENCES

- Ateniese G, Chou DH, de Medeiros B, Tsudik G. Sanitizable Signatures. In: Proc. of European Symposium on Research in Computer Security (ESORICS 2005). Springer; 2005. p. 159–77.
- Bilzhaue A, Pöhls HC, Samelin K. Position Paper: The Past, Present, and Future of Sanitizable and Redactable Signatures. Proc. of International Conference on Availability, Reliability and Security (ARES 2017). ACM, 2017. 87:1–87:9.
- Biskup J, Flegel U. Transaction-based pseudonyms in audit data for privacy respecting intrusion detection. In: International Workshop on Recent Advances in Intrusion Detection. Springer; 2000. p. 28–48.
- Böhm F, Menges F, Permül G. Graph-based visual analytics for cyber threat intelligence. *Cybersecurity* 2018;1(1):16.
- Brzuska C, Pöhls HC, Samelin K. Non-Interactive Public Accountability for Sanitizable Signatures. In: Revised Selected Papers of European PKI Workshop: Research and Applications (EuroPKI 2012). Springer; 2012. p. 178–93.
- Burkhardt M, Strasser M, Many D, Dimitropoulos X. Sepia: privacy-preserving aggregation of multi-domain network events and statistics. *Network* 2010;1(101101).
- Büschkes R, Kesdogan D. Privacy enhanced intrusion detection. *Multilateral security in communications, information security* 1999;187–204.
- Coppolino L, D'Antonio S, Mazzeo G, Romano L, Sgaglione L. How to Protect Public Administration from Cybersecurity Threats: The COMPACT Project. In: 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA); 2018. p. 573–8. doi:[10.1109/WAINA.2018.00147](https://doi.org/10.1109/WAINA.2018.00147).
- Deutscher Bundestag, 2015. Gesetz zur Erhöhung der Sicherheit informationstechnischer Systeme. <https://www.bmi.bund.de/SharedDocs/Downloads/DE/Gesetztexte/it-sicherheitsgesetz.pdf>.
- Dewald A. Characteristic evidence, counter evidence and reconstruction problems in forensic computing. *it - Information Technology* 2015;57(6):339–46.
- European Commission, 2016. NIS Directive 2016/1148 (EU) of the European Parliament and of the Council. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L1148>.
- European Parliament and the Council of the European Union. Regulation (EU) no 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing directive 1999/93/EC. *Official Journal* 2014;OJ L 257 of 28.8.2014:73–114.
- European Parliament and the Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation). *Official Journal* 2016;OJ L 119 of 4.5.2016:1–88.
- Gartner Inc., 2018. Security information and event management (SIEM). <https://www.gartner.com/it-glossary/security-information-and-event-management-siem>.
- Goldstein M, Asanger S, Reif M, Hutchison A. Enhancing security event management systems with unsupervised anomaly detection. In: ICPRAM; 2013. p. 530–8.
- Höhne F, Pöhls HC, Samelin K. Rechtsfolgen editierbarer Signaturen. *Datenschutz und Datensicherheit - DuD* 2012;36(7):485–91. doi:[10.1007/s11623-012-0165-8](https://doi.org/10.1007/s11623-012-0165-8).
- Jensen M. Challenges of privacy protection in big data analytics. In: 2013 IEEE International Congress on Big Data. IEEE; 2013. p. 235–8.

- Johnson R, Molnar D, Song D, Wagner D. Homomorphic signature schemes. In: Proc. of the RSA Security Conference - Cryptographers Track. Springer; 2002. p. 244–62.
- Lanzi A, Balzarotti D, Kruegel C, Christodorescu M, Kirda E. Accessminer: using system-centric models for malware protection. In: Al-Shaer E, Keromytis AD, Shmatikov V, editors. In: Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS 2010, Chicago, Illinois, USA, October 4–8, 2010. ACM; 2010. p. 399–412.
- Latzo, T., 2020. Efficient fingerprint matching for forensic event reconstruction. Under submission.
- Latzo T, Freiling F. Characterizing the limitations of forensic event reconstruction based on log files. 2019 IEEE Trustcom/BigDataSE. IEEE, 2019.
- López J, Oppliger R, Pernul G. Why have public key infrastructures failed so far? Internet Research 2005;15(5):544–56.
- Menges F, Böhm F, Vielberth M, Puchta A, Taubmann B, Rakotoniravony N, Latzo T. Introducing dingfest: An architecture for next generation siem systems. In: Langweg H, Meier M, Witt BC, Reinhardt D, editors. In: SICHERHEIT 2018. Bonn: Gesellschaft für Informatik e.V.; 2018. p. 257–60.
- Miller D, Harris S, Harper A, VanDyke S, Blask C. Security information and event management (SIEM) implementation. New York, NY: McGraw-Hill; 2011.
- Miloslavskaya N, Tolstoy A. New siem system for the internet of things. In: World Conference on Information Systems and Technologies. Springer; 2019. p. 317–27.
- Mokalled H, Catelli R, Casola V, Debertol D, Meda E, Zunino R. The applicability of a siem solution: Requirements and evaluation. In: 2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). IEEE; 2019. p. 132–7.
- Nespoli P, Gómez Mármol F. e-health wireless ids with siem integration. In: Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC18), Barcelona, Spain; 2018. p. 15–18.
- Park H-A, Lee DH, Lim J, Cho SH. Ppids: privacy preserving intrusion detection system. In: Pacific-Asia Workshop on Intelligence and Security Informatics. Springer; 2007. p. 269–74.
- Parliament of the United Kingdom, 2000. Freedom of Information Act 2000. URL http://www.legislation.gov.uk/ukpga/2000/36/pdfs/ukpga_20000036_en.pdf.
- Pöhls HC. Increasing the Legal Probative Value of Cryptographically Private Malleable Signatures. University of Passau; 2018.
- Pöhls HC, Höhne F. The Role of Data Integrity in EU Digital Signature Legislation - Achieving Statutory Trust for Sanitizable Signature Schemes. In: Meadows C, Fernandez-Gago C, editors. In: 7th International Workshop, STM 2011, Copenhagen, Denmark, June 27–28, 2011, Revised Selected Papers. Springer Berlin Heidelberg; 2011. p. 175–92. doi:10.1007/978-3-642-29963-613.
- Pöhls HC, Höhne F. The Role of Data Integrity in EU Digital Signature Legislation - Achieving Statutory Trust for Sanitizable Signature Schemes. In: Revised Selected Papers from the 7th International Workshop on Security and Trust Management (STM 2011). Springer; 2011. p. 175–92.
- Pöhls HC, Samelin K. Accountable Redactable Signatures. In: Proc. of International Conference on Availability, Reliability and Security (ARES 2015). IEEE; 2015. p. 60–69.
- Pöhls HC, Samelin K, Posegga J, de Meer H. In: Technical Report. Transparent Mergeable Redactable Signatures with Signer Commitment and Applications (MIP-1206). Faculty of Computer Science and Mathematics (FIM), University of Passau; 2012.
- Putz B, Menges F, Pernul G. A secure and auditable logging infrastructure based on a permissioned blockchain. Computers and Security 2019;101602.
- Rieck K, Holz T, Willems C, Düssel P, Laskov P. Learning and classification of malware behavior. In: International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. Springer; 2008. p. 108–25.
- Schlette D, Böhm F, Caselli M, Pernul G. Measuring and visualizing cyber threat intelligence quality. Int. J. Inf. Secur. 2020. doi:10.1007/s10207-020-00490-y.
- Sgaglione L, Mazzeo G. A GDPR-Compliant Approach to Real-Time Processing of Sensitive Data. In: Intelligent Interactive Multimedia Systems and Services. Cham: Springer International Publishing; 2019. p. 43–52.
- Sobirey M, Fischer-Hübner S, Rannenber K. Pseudonymous Audit for Privacy Enhanced Intrusion Detection. In: Information Security in Research and Business. Springer; 1997. p. 151–63.
- Stahlberg P, Miklau G, Levine BN. Threats to privacy in the forensic analysis of database systems. In: Proceedings of the 2007 ACM SIGMOD international conference on Management of data. ACM; 2007. p. 91–102.
- Steinfeld R, Bull L, Zheng Y. Content extraction signatures. In: Proc. of International Conference on Information Security and Cryptology (ICISC 2001). Springer; 2002. p. 163–205.
- The Apache Software Foundation, 2019. Apache http server project. <https://httpd.apache.org/>.
- The National Archives, 2011. Redaction toolkit – editing exempt information from paper and electronic documents prior to release. URL http://www.nationalarchives.gov.uk/documents/information-management/redaction_toolkit.pdf [last accessed: Nov. 2019].
- United Kingdom Ministry of Justice, 2009. Lord Chancellor's Code of Practice on the management of records issued under section 46 of the Freedom of Information Act 2000. URL <http://www.nationalarchives.gov.uk/documents/foi-section-46-code-of-practice.pdf> [last accessed: Sep. 2019].
- van GeelkerkenFWJ, Pöhls HC, Fischer-Hübner S. The legal status of malleable- and functional signatures in light of Regulation (EU) No 910/2014. In: Proc. of the 3rd International Academic Conference of Young Scientists on Law & Psychology 2015 (LPS 2015). L'viv Polytechnic Publishing House; 2015. p. 404–10. <https://drive.google.com/file/d/0B-Yu3Ni9z3PXM2lBajhCXzhoWk0/view>.
- Vielberth M, Menges F, Pernul G. Human-as-a-security-sensor for harvesting threat intelligence. Cybersecurity 2019;2(23).
- Vielberth M, Pernul G. A Security Information and Event Management Pattern. 12th Latin American Conference on Pattern Languages of Programs (SugarLoafLOP 2018), 2018.
- Wang RY, Strong DM. Beyond accuracy : what data quality means to data consumers. Journal of Management Information Systems 1996;12(4):5–34. <http://w3.cyu.edu.tw/ccwei/PAPER/ERP/dataquality%28JMS%29.pdf>.
- Williams AT, Nicolett M. Improve it security with vulnerability management. Technical Report - Gartner Inc. 2005.
- Florian Menges** received both the Bachelor of Science and Master of Science degree from the University of Regensburg, Germany. Currently he is research assistant at the Department of Information Systems at the University of Regensburg, Germany. His research interests include threat intelligence with a focus on sharing and reporting intelligence data, storage strategies for intelligence data as well as anonymization techniques and incentivizing the sharing and reporting of incident data.
- Tobias Latzo** received both the Bachelor of Science and Master of Science degree from the Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany. Currently he is PhD student at the Department of Computer Science at the FAU. His research interests are memory forensics and system security.
- Manfred Vielberth** studied Management Information Systems (Wirtschaftsinformatik) at the University of Regensburg. His major

fields of study were Financial Computing and Information Security during his Bachelor's degree and IT-Security during his Masters degree. Since February 2017, he is a research assistant at the Chair of Information Systems I. There, he is responsible for the procedure of student seminars and projects and the modules "Corporate Databases" and "Information Systems - Developments and Trends".

Sabine Sobola studied at the University of Regensburg and graduated with her 2nd state examination in law in 2000. From the very beginning, she has specialized in IT-Law and related fields and she started working as a lawyer in the year of her graduation. Since 2005 she has been partner of the law firm Paluka Sobola Loibl & Partner Rechtsanwälte. In 2016 she also received the Master of Arts degree in Philosophy, Politics and Economics (PPW) from the Ludwigs-Maximilian-University Munich, Germany. In addition to her work as a lawyer she has been working at the University of Regensburg and other Universities in Bavaria since 2002. Sabine Sobola specializes in IT law, IT security, data protection, media and copyright law and intellectual property rights.

Henrich C. Pöhls received his Ph.D. from University of Passau for his work interdisciplinary at the intersection of applied cryptography and law. His research currently focuses on practical applications of advanced cryptography to foster the exchange of authentic data while upholding data-minimisation for increased privacy and legal compliance. He has authored many academic publications, especially on the topic of tailoring cryptographic primitives for legally compliant applications in various domains, like supply chain, Internet-of-Things, and the cloud. He is keen on interdisciplinary work especially in the field of cryptography, software development and law, as he thinks the more gaps between those three worlds can be bridged the more sound (=safe, secure and legally compliant) ICT-enhanced products and environments like smart homes or smart cities become. Henrich C. Pöhls also holds a graduate diploma in computer science (Dipl. Inf.) from the University of Hamburg and an M.Sc. in Information Security from Royal Holloway University of London.

Benjamin Taubmann received both the Bachelor of Science and Master of Science degree from the Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. Currently, he is a research assistant at the faculty of computer science and mathematics at the University of Passau. His research interests include main memory forensics, virtual machine introspection, and the

question of how to apply these techniques to production environments to enhance their security level.

Johannes Köstler holds a master's degree in IT-Security and is currently a Ph.D. student at the University of Passau, where he works at the Institute of IT-Security and Security Law. His main research area is the reliability of distributed systems, but he also did some work on privacy issues in information systems.

Alexander Puchta studied Business Informatics within the Honors Elite Program at the University of Regensburg and at the University of Technology in Sydney. Since February 2018, Alexander Puchta is a research assistant at the Chair of Information Systems I (Prof. Dr. Pernul). In the context of the research project DINGFEST he is working on real-time analysis technologies for Identity and Access Management. His research is supplemented by his practical experience within Identity and Access Management as he is parallelly working as a consultant at the Nexis GmbH.

Felix Freiling is a professor of computer science at Friedrich-Alexander-Universität Erlangen-Nürnberg. His research interests lie in computer security and forensic computing.

Hans P. Reiser is an assistant professor at University of Passau, where he joined the Institute of IT Security and Security Law in 2011. He holds a PhD in the area of middleware for fault-tolerant systems from Ulm University. Since 2007 he worked as an assistant professor at LaSIGE, University of Lisbon, and in 2010 he spent one semester at the Carnegie Mellon University, Pittsburgh, USA as a visiting professor. Hans P. Reiser's research focus is on technical aspects of reliability and security in distributed systems, including intrusion-tolerant algorithms, adaptivity and self-optimization in resilient distributed system and methods for incident analysis and digital forensics in virtualised environments.

Günther Pernul received both the diploma degree and the doctorate degree (with honors) from the University of Vienna, Austria. Currently he is full professor at the Department of Information Systems at the University of Regensburg, Germany. Prior he held positions with the University of Duisburg-Essen, Germany and with University of Vienna, Austria, and visiting positions at the University of Florida and the College of Computing at the Georgia Institute of Technology, Atlanta. His research interests are manifold, covering data and information security aspects, data protection and privacy, data analytics, and advanced data centric applications.

4 Human-as-a-security-sensor for harvesting threat intelligence

Current status:	Published
Journal:	Cybersecurity, Volume 2, Number 1, December 2019
Date of acceptance:	29 August 2019
Full citation:	VIELBERTH, M., MENGES, F., AND PERNUL, G. Human-as-a-security-sensor for harvesting threat intelligence. <i>Cybersecurity</i> 2, 23 (2019), 1–15
Authors' contributions:	Manfred Vielberth 45%
	Florian Menges 45%
	Günther Pernul 10%

Journal description: The *Cybersecurity* journal aims to systematically cover all essential aspects of cybersecurity, with a focus on reporting on cyberspace security issues, the latest research results, and real-world deployment of security technologies.

RESEARCH

Open Access

Human-as-a-security-sensor for harvesting threat intelligence

Manfred Vielberth* , Florian Menges and Günther Pernul

Abstract

Humans are commonly seen as the weakest link in corporate information security. This led to a lot of effort being put into security training and awareness campaigns, which resulted in employees being less likely the target of successful attacks. Existing approaches, however, do not tap the full potential that can be gained through these campaigns. On the one hand, human perception offers an additional source of contextual information for detected incidents, on the other hand it serves as information source for incidents that may not be detectable by automated procedures. These approaches only allow a text-based reporting of basic incident information. A structured recording of human delivered information that also provides compatibility with existing SIEM systems is still missing. In this work, we propose an approach, which allows humans to systematically report perceived anomalies or incidents in a structured way. Our approach furthermore supports the integration of such reports into analytics systems. Thereby, we identify connecting points to SIEM systems, develop a taxonomy for structuring elements reportable by humans acting as a security sensor and develop a structured data format to record data delivered by humans. A prototypical human-as-a-security-sensor wizard applied to a real-world use-case shows our proof of concept.

Keywords: Cyber threat intelligence, Human awareness, Human-as-a-security-sensor, Security information and event management (SIEM)

1 Introduction

Today's security analytics solutions like Security Information and Event Management (SIEM) systems heavily rely on a huge amount of data in order to reliably detect incidents in organizations (Bhatt et al. 2014). New sources providing security-relevant data, such as knowledge about occurred incidents observed by human individuals, can therefore significantly enlarge the data basis for incident detection.

During past years, humans or employees were generally seen as the weakest link in corporate IT security (Lineberry 2007). To mitigate the risk of humans for IT security, a lot of effort is put into awareness campaigns and training of employees (Mello 2017) to ensure that they receive a basic understanding of this topic. This also enables them to distinguish between "normal" events and events harming the organization. However, the ability to recognize malicious events is not harnessed to its full extent. Information about potential incidents might be

hidden in the minds of humans and could be the missing link for attack detection or for forensic reconstruction of adverse events. Especially when it comes to nontechnical traces. Therefore, we argue that the connection of digital events with non-digital events observed by people is crucial to IT security.

In this paper, we describe an approach that integrates the human data source to further processing in security analytics systems (e.g. SIEM systems). Therefore, we illustrate the problem with a motivating example in Section 2. Subsequently, related work in the area of human-as-a-security-sensor is portrayed within Section 3. In Section 4, we present the problem and research question tackled and show how to integrate human sensors into SIEM systems in Section 4.1. In Section 4.2, a risk model and a taxonomy for human threat reporting are proposed. On this basis we develop a CTI base data structure for human sensor information in section 4.3 and a data format for the representation of this data in Section 4.4. Finally, the proposed approach is evaluated in Section 5 and concluded in Section 6.

*Correspondence: manfred.vielberth@ur.de
Universität Regensburg, Universitätsstr. 31, 93053 Regensburg, DE, Germany

2 Motivating example

In the following section, we use a real-world attack to illustrate the main problem tackled in this work. The example underlines benefits that may arise from integrating the human factor into threat detection mechanisms, including improved threat detection and additional context information.

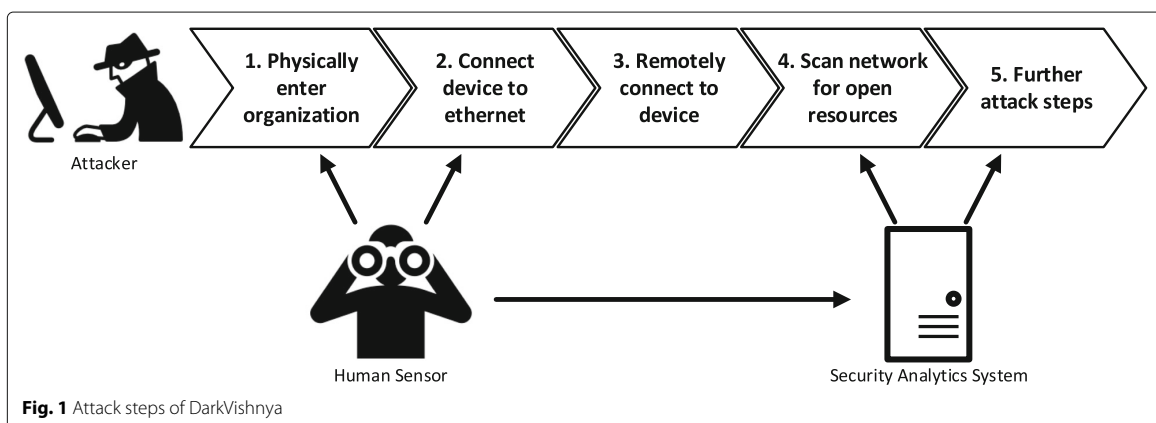
Between 2017 and 2018, Kaspersky Lab (Golovanov 2018) investigated several cybersecurity incidents that go by the name of DarkVishnya. Malicious devices were directly connected to organizations' local networks, causing damage estimated to multiple millions of dollars. As shown in Fig. 1, the attack was conducted in the following essential steps:

- 1 The attacker tries to physically enter the premises of the attacked organization, claiming to be a person with legitimate interest (e.g. being an applicant or a courier).
- 2 After the successful entrance, the attacker tries to place a network device unobtrusively and hides it by blending it into the surrounding area. Moreover, the device is connected to the local network infrastructure in order to enable further attack steps.
- 3 After the attacker has left the organization, the placed device is remotely accessed by utilizing standard mobile technologies like GPRS, 3G or LTE to control it for further attack steps.
- 4 The attacker scans the network for usable information and for accessible resources in the local network. This may include shared folders, servers or other systems that execute critical actions. Additionally, brute-force attacks or network sniffing is used to gain access to login credentials.
- 5 The attacker tries to exploit the previously gained access e.g. by installing malware to retain access and to execute malicious services.

The crux of the attack is that the first three steps are nearly impossible to detect with technical security systems like SIEM, or Intrusion Detection Systems (IDS), as neither the attacker entering the building, nor the placing of a hardware device leave any digital traces. The first digital traces that may be detected by security systems are left at the beginning of the network access. Unlike automated analyses, employees have the ability to detect and report such anomalies before technical traces and potential damages occur. If, for example, a suspicious person walks around the office building, the employee might already categorize this event as an anomaly. Additionally, context information, such as a description of a person, enhances this first perception. However, employees are often not able to recognize technical traces, such as network scans. The example demonstrates that it is hardly possible to capture the full extent of an attack, when collecting technical or human traces independently or if one of them is not considered at all. Therefore, we propose an approach that enables the acquisition of anomalies or potential attacks detected by employees, to translate them into machine readable language and thus to create the basis for combining these two types of data.

3 Related work

The first IT security related approaches for threat reporting by humans are systems that handle malicious or unwanted emails. These can be narrowed down to spam and phishing emails. There are several examples available in practice that allow to report such threats. These are in most cases integrated into email software, where emails can be marked (Google LLC; Microsoft Corporation) or a standalone web interface is provided (Anti-Phishing Working Group). In most cases, these reports are used to train phishing or spam filters of the provider.



A second approach commonly applied in practice is human-to-human reporting. A central contact point (e.g. the help desk of an organization) is set up. Especially when implementing an information security management system (e.g. control A.13.1 of the ISO 27001 standard demands the reporting of security events or weaknesses from all employees (ISO/IEC 27001: Information technology – Security techniques – Information security management systems – Requirements 2013)) this is common practice for reporting security issues by employees (Hintzbergen et al. 2015). However, this approach entails some disadvantages. For example, the human point of contact has to interpret the received information and decide how to proceed. This might result in wrong decision-making, especially as help desk personnel are commonly no security specialists. Additionally, the collected data is poorly structured and not utilizable for technical analyses in most cases. Although not security related, the idea of using humans as sensors has been a topic of interest for a while. For example, Wang et al. (2014) pursue the idea that social networks might be the largest existing human sensor networks. Furthermore, Kostakos et al. (2017) investigate several scenarios, where humans can act as sensors. They consider, among others, crowdsourcing markets, social media and the collection of citizen opinions.

Heartfield and Loukas (2018) recently proposed a more general approach focused on semantic social engineering attacks. In their work, they develop and prototypically implement a framework for reporting semantic social engineering attacks. They propose a model for predicting the reliability of reports generated by humans and show, that human sensors can outperform technical security systems in their considered context. In addition, they implement a backend application, which is mainly responsible for incident response and dashboard capabilities. In one of their previous works (Heartfield et al. 2016), they also coined the term human-as-a-security-sensor, which refers to the "paradigm of leveraging the ability of human users to act as sensors that can detect and report information security threats". For our work, we adopt the meaning of the paradigm. This capability is strongly influenced by the security training the person received in advance. In addition, an approach for scoring the trustworthiness of human sensors was introduced by Rahman et al. (2017). They especially monitor features of the mobile device, utilized for conducting the report, for predicting the reliability of the provided data.

To sum up the developments in this area, platforms for reporting potential malicious or unwanted emails were implemented at first. This was followed by the development of processes for human-to-human reporting and succeeded by more sophisticated approaches for detecting semantic social engineering attacks with the help of a

human-as-a-security-sensor framework. However, to the best of our knowledge, there are no approaches that support reporting a wide range of possible attacks detectable by humans. Additionally, there are no concepts for integrating reported incidents into existing, and in many organizations already established, security systems (e.g. SIEM systems). Moreover, the participation of people with different knowledge in the field of cybersecurity, is currently neglected.

4 Integrated human-as-a-Security-Sensor (IHaaS)

Resulting from the explanations in Section 2 and Section 3 we tackle the issue, that **observations of humans are either poorly or not at all integrated into the automatic security analytics process**. This raises the following research questions:

- Q1: What are the connection points of a human-as-a-sensor to the data flow of a SIEM system?
- Q2: How can human-provided information be structured (data format) in order to facilitate further technical processing?
- Q3: How can incident information be systematically acquired from people?

To answer these research questions, we applied the following approach:

1. To answer Q1, we illustrate how to integrate human-as-a-security-sensors into security analytics in Section 4.1. This is based on existing data collection approaches and the generic data flow of SIEM systems identified in literature (Vielberth and Pernul 2018).
2. To answer Q2 and Q3 it is in a first step necessary to identify all possibilities a human sensor can report. This is carried out by developing a risk model and taxonomy, adhering to the method for taxonomy development by Nickerson et al. (2013) in Section 4.2.
3. To answer research question Q2, we first conceptualize a CTI base data structure for the representation of human sensor data in Section 4.3. On this basis we then identify suitable CTI data format standards to realize this base data structure and extend them for the capturing of human sensor data in Section 4.4. This allows the integration of human-generated reports into SIEM systems for further processing.
4. Finally, the incident information can systematically be acquired (Q3) following the risk model and taxonomy, which is restricted by constraints identified in Section 4.5.

Thereby, we see the main contributions of this paper in the identification of connection points, the development

of the taxonomy, the extension of well-established data formats and the identification of constraints for a systematic data acquisition. We also show the practicability of our approach using a prototypical implementation and an exemplary real-world use case.

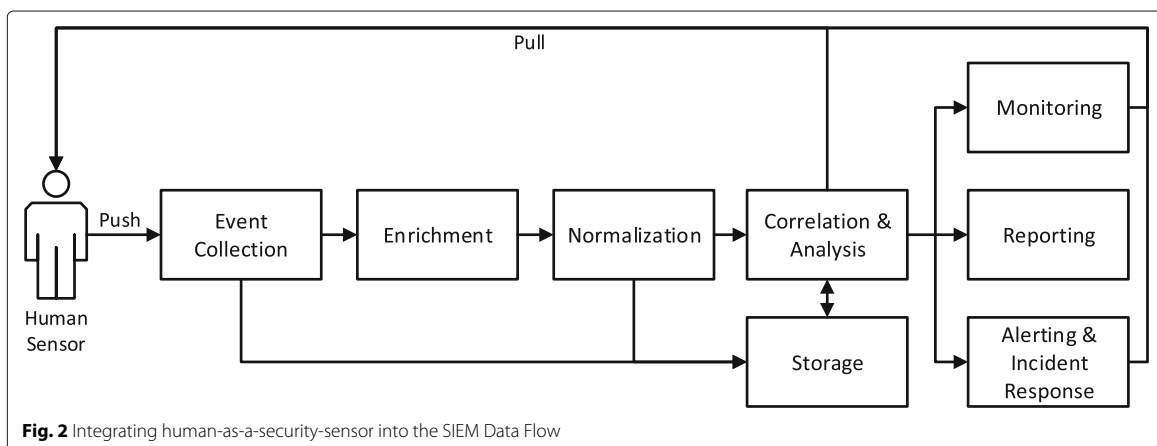
4.1 Integrating human sensors into SIEM

To connect technical data with human-generated traces, both need to be brought together in one single system. One way to achieve this is to integrate human knowledge into SIEM systems that are already in place in most organizations within a security operations center (SOC) (Crowley and Pescatore 2018). Apart from that, the presented approach can easily be adapted to other security monitoring tools.

A SIEM system is essentially designed for the collection of relevant log data to detect incidents and gain situational security awareness. In Fig. 2 we extended the basic SIEM structure as proposed by Vielberth and Pernul (2018) with the data flow of an integrated human-as-a-security-sensor. Hereby, the SIEM first *collects* relevant event information, in most cases in the form of log data. This data gets *enriched* with additional context data and translated in a uniform representation during the *normalization* step. The core of the system lies in the *correlation and analysis* component, where information from various sources is connected and incidents are detected using methods such as pattern matching. *Monitoring* enables security analysts to be actively involved in the analysis, whereas *reporting* delivers compliance reports or enables the participation in established threat intelligence sharing platforms between organizations. In case of a detected incident, *alerting and incident response* triggers necessary reactions to mitigate further harm. Finally, the *storage* module is responsible for both, short- and long-term storage of event data and analysis results.

For integrating human sensors into SIEM (Q1), we extend the basic SIEM data collection approach. According to Holik et al. (2015) and Turnbull (2019), two fundamental approaches can be applied. They both distinguish between push- and pull- based log collection. Since we do not collect log data, but human-generated incident information, these two approaches require adaptation. In the following, both approaches are described in the context of this paper:

- **Push:** The push method applies when an employee initially detects an incident and actively delivers the gathered information to the system. It is important to offer guidance for enabling humans to provide information in a structured way, especially if their knowledge about security is limited. Additionally, employees might report information in different levels of detail, depending on how much they know about the incident. The push approach is similar to systems pushing log data into SIEM systems as described in literature (Holik et al. 2015). Thus, the connection point of the push approach is the *event collection* (compare Fig. 2).
- **Pull:** In traditional SIEM systems, the pull approach basically refers to polling-based systems (Turnbull 2019), which query the data periodically, generally in fixed time intervals. Since periodically polling information from human sensors is hardly feasible, we only pull information in certain cases. These cases occur during certain steps of the SIEM data flow (as described subsequently), which are the connection points for the pull approach. The pull approach is applied if important information is missing during the *monitoring* or *analysis* of incidents. Presumably, this happens in case an incident is reported by people with little knowledge about IT security or about the context of the incident. The lack of information can



either be detected automatically during the *correlation and analysis* phase, or by human experts monitoring the system or during *incident response* steps. Furthermore, needed information might be missing in case technical indications about an incident occur, but previous attack steps were not reported. For instance, technical traces from step four of the attack in Fig. 1 could be identified in the system, while previous attack steps were not reported. Therefore, it is necessary to advise employees to report missing hints. In order to gain more information, an expert can interview the reporting person and guide him to contribute further or more detailed information.

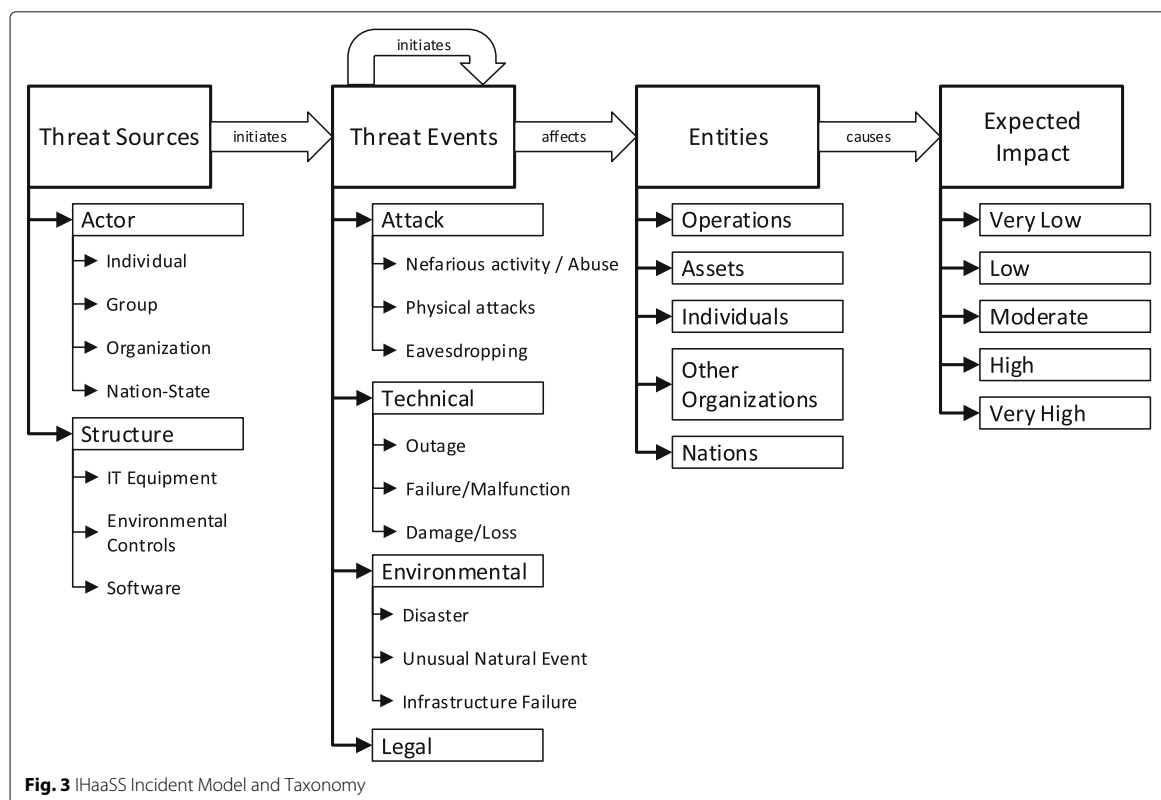
4.2 IHaaS incident model and taxonomy

For being able to develop a format (Q2) and structure the acquisition of information (Q3), it is necessary to capture everything that human-security-sensors can perceive. Information security management standards and its associated resources provide a good basis by providing risk assessment guidelines. These consider and evaluate mostly future risks. However, in our approach we want to report past incidents, requiring to adjustment for

some elements. Regarding the NIST Guide for Conducting Risk Assessments (Joint Task Force Transformation Initiative 2012) and Juliadotter and Choo (2015), the key risk factors are *Threat Sources*, *Threat Events/Vector*, *Targets/Vulnerabilities* and *Impact*. All four risk factors are observable or can at least be assessed by human sensors and thus, have to be dealt with. The resulting threat model can be seen in Fig. 3.

In the proposed Integrated Human-as-a-Security-Sensor (IHaaS) incident model, there are two types of threat events: threat events caused by humans or technical sources (commonly security events) and events which are not necessarily assignable to a source (especially safety events). Threat events can be either initiated by threat sources or by previous events. Furthermore, it is possible that no entities are affected, or the affected entities are not (yet) known. The same applies to the expected impacts. This leads to the conclusion that only threat events are mandatory elements, as without threat events there is no need to report.

In order to get a deeper insight into human sensor reports, we examine the four risk factors in more detail, thereby create a taxonomy for human-as-a-security-sensor threat reporting. This taxonomy classifies and



structures the security-related artifacts a human sensor can observe. Thereby, we loosely adhere to the method for taxonomy development by Nickerson et al. (2013). However, we did not develop a completely new taxonomy, but rather combined and adapted existing taxonomies to fit the purpose. Thereby we followed the “conceptual-to-empirical approach” (Nickerson et al. 2013), because of the existing foundations and a well-established knowledge base in this area. The identified objects are described in more detail in the following:

- **Threat Sources:** Threat sources are the starting point of the incident and can initiate subsequent threat events. This part of the taxonomy is based on the NIST Taxonomy of Threat Sources (Joint Task Force Transformation Initiative 2012). However, to avoid overlaps with subcategories of threat events, we narrow the scope. In the context of our paper, a threat source is an entity, which can decide and initiates events. Thus, the environment defined as a threat source by NIST is equal to a threat event in our taxonomy as we argue that environmental factors cannot take decisions. Additionally, environmental events might be initiated by sources and are therefore better classified as events (e.g. a fire can be set by a person). However, the risk model process remains unaffected, because events can initiate other events. As a result, environmental events can still trigger structural events such as outages.
- **Threat Events:** Threat Events are the processes actually causing harm to an organization and thus are the key component of an IHaaS report. Our taxonomy of threat events is based on the ENISA Threat Taxonomy (Marinos 2016) with some changes in order to fit in the rest of our model. We have defined more general categories, which allow the distinction between intentional (Attack) and potentially unintentional (Technical, Environmental, Legal) events. This is especially important in order to form dependencies in Section 4.5. Furthermore, the environmental events are merged with environmental threat sources from the NIST Taxonomy of Threat Sources (Joint Task Force Transformation Initiative 2012).
- **Entities:** The identification of relevant assets (asset inventory) is much discussed in academic literature and by industry, due to its importance to risk management (Fenz et al. 2014). Our approach, however, is somewhat broader, which is why we talk about entities (e.g. other organizations may be affected, which are not necessarily an asset for the company). The entity taxonomy is taken from (Joint

Task Force Transformation Initiative 2012), wherein it is called adverse impact.

- **Expected Impact:** The expectation of possible impacts is usually quite hard to classify for humans. Therefore, the human sensor commonly provides qualitative estimations, especially when the IT security knowledge is low. Nevertheless, this estimation can be very helpful for evaluating further actions and reactions. Very Low to Very High is a rating of the effect of the event as described by the NIST (Joint Task Force Transformation Initiative 2012). It ranges from “negligible” to “multiple severe or catastrophic effects”.

4.3 Conceptualizing a CTI data structure for human sensor data

In the previous sections, we introduced connecting points for the integration of human knowledge into SIEM systems and developed a taxonomy that serves as an information basis for the acquisition of threats detected by humans. In this section, we lay the theoretical foundations for the integration of human-provided information into SIEM data processing. The central factor for this integration is the harmonization of data structures to ensure compatibility of information. As shown in 4.1, SIEM systems work with both normalized raw data and enriched context data, which can be summarized under the term Cyber Threat Intelligence (CTI). To enable the integration of these types of information, we propose an approach of translating the human provided information into the existing CTI data structures in this section. To this end, we first discuss the types of information that can be provided by human sensors and classify them in the context of CTI information. On this basis, we then propose a CTI data structure that allows to fully capture information provided by human sensors to answer the research question Q2 on a general level. Finally, Table 1 summarizes the results of this section

The work of Burger et al. (2014) serves as a basis for the allocation of human sensor information to CTI data structures. It divides CTI into the three main categories *Intelligence*, *Attribution* and *Indicator*. *Intelligence* refers to rather complex issues such as concrete procedures of attackers or methods for mitigating security incidents, which cannot be fully acquired from automated analyses. Although a deeper expert knowledge is necessary for the final evaluation of intelligence information, untrained employees can contribute valuable information, which may make an incident detection possible in the first place. An example of this would be the detection of unauthorized physical access to protected resources. The *Attribution* category describes various types of additional contextual information about a security incident. These include, for example, information on attackers or affected devices.

Table 1 CTI base model extensions

Classification	Taxonomy		UPSIDE Model	Changes
Intelligence	Threat Events	Attack	Attack Event	-
		Technical	-	Technical Event
		Environmental	-	Environmental Event
		Legal	-	Legal Event
	Expected Impact		Result	Result
Attribution	Threat Sources	Actor	Attacker	Actor
		Structure	-	Structural Source
	Entities	Assets	Attack Target	Affected Entity
		Persons	Attack Target	Affected Entity
Indicator	Threat Events		Indicator	-

This data is also only recognizable to a limited extent through automated analyses. Since attribution information usually does not require specific specialist knowledge, employees can also make a valuable contribution here. For example, employees can help identifying a potential attacker and point out potentially affected devices. In contrast to these categories, *Indicator* describes specific system events that can, for example, be obtained from system logs. Since log files contain extensive information, they are usually evaluated using automated analyses and can only be used to a limited extent within a human sensor platform. However, when an incident is captured, additional fine-granular information may also be provided. For example, a malicious email provides information about a potential attack or an attacker, but also provides fine-grained information within its source code. As a result, indicator information is not primary information that is obtained from human observations, but secondary information that is collected when entities are created and populated. Summarizing, it can be stated that human sensors can mainly contribute to analyses with context information from the categories *Intelligence* and *Attribution* whereas *Indicator* information is only used to a very limited extent.

After performing a classification of human sensor data in the context of CTI data structures, we propose a CTI data structure that is able to cover the full range of human sensor information in the following. To achieve this, we utilize the previously introduced categories *Intelligence*, *Attribution* and *Indicator* to describe the individual changes necessary. More specifically, we use the UPSIDE model that describes CTI base entities by Menges and Pernul (2018) to determine and discuss entities that can be mapped by CTI data structures and those that are still missing for the representation of human provided information. On this basis, we propose conceptual adaptations to existing CTI data structures to support human sensor data as described in our taxonomy.

- **Intelligence:** The *Intelligence* category describes the attack patterns used, countermeasures taken and additional information on incidents such as the expected impact. The Threat Events and Expected Impact sections of the taxonomy can be assigned to this category. Threat events are divided into active (attack) and passive (technical, environmental and legal) incidents. According to the CTI base model, the description of active attacks is possible by defining attack events and the underlying procedure. Incidents without an active component are not supported so far. In addition, the model offers the possibility to define the result of an attack as result entity. This allows "Expected Impact" from our taxonomy to be mapped, however, this also only applies for active attacks.
- **Attribution:** The *Attribution* category defines various contextual information, such as information about attackers and targets. The sections Threat Sources and Entities from the taxonomy can both be assigned to this category. In the area of threat sources, the CTI base model can represent active attackers. Although, an unintentionally involved actor and other threat sources cannot be defined yet. In the taxonomy section entities, both assets and persons can be represented within the CTI base model. However, these can only be represented as targets in connection with an attack. It is not possible to represent any other kind of participation of these entities.
- **Indicator:** The *Indicator* category is used to display detailed information within threat events. The entity indicator from the CTI base model defines a generic representation within a security incident that can be assigned to any other entity. Accordingly, the requirements of the taxonomy are basically fulfilled in this area.

After comparing our taxonomy with the capabilities of the CTI base model, we discuss necessary adjustments for

the integration of human sensor information in the following. Several adjustments are necessary within the *Intelligence* section. Since only attack events are supported, it is necessary to introduce additional entities to be able to map passive events. This includes technical events, environmental events and legal events. In addition, the result of an event must be adapted in such a way that the result of passive events can also be represented. The *Attribution* area also requires several adjustments. On the one hand, the attacker element must be extended in such a way that a passive participant can also be represented. In addition, it is also necessary to introduce an additional entity to represent a structural source for incidents. Finally, entities can be represented completely, but only in the context of an attack. Here an appropriate extension is necessary so that entities can also be affected by passive events. The indicator area does not require any adjustments at the conceptual level. Summarizing, Table 1 gives an overview of the results of this section. Column Classification assigns the results to the respective CTI category, while column Taxonomy shows the elements of the taxonomy under consideration. The UPSIDE Model column shows the assignment to the CTI base model and column Changes shows the necessary adjustments to the base model to support human sensor information.

4.4 A structured representation for threat intelligence reported by humans

In the previous sections, we introduced connection points for integrating human knowledge into SIEM systems and a taxonomy that defines the information basis for the acquisition of threat information detected by humans. Subsequently, we introduced the theoretical foundation for a CTI data structure that is able to represent human sensor data. Based on these findings, we develop a CTI data format in this section that allows to capture information provided by human sensors and enables further technical processing according to research question Q2. In developing the data format we pursue two main objectives. On the one hand, we aim to achieve a high compatibility to existing SIEM systems to allow a direct integration of additional information into the system. On the other hand, we aim to create a format that allows a complete representation of human sensor data. More specifically, the full scope of the taxonomy shown in Section 4.2 needs to be covered. In order to meet these requirements as completely as possible, we first select existing and well supported CTI data format standards as development basis in the following. Subsequently, we propose a specification of necessary extensions for the integration of human sensor data according to Section 4.3.

Event collection modules within SIEM systems handle heterogeneous raw data from different log sources. This data is then translated into homogeneous indicator data

structures to allow further processing. Literature provides different standards for the structured representation of indicators, such as CybOX¹ or openIoC². These data structures are commonly referred to as Indicators of Compromise (IoC) as they depict a set of observations associated with a threat (Appala et al. 2015). These basic incident data can furthermore be enriched using intelligence and attribution data, such as information about attackers, utilized attack patterns or attackers' objectives as shown by Burger et al. (2014). Together, they allow the representation of complex security incident information as shown in Section 4.3. Literature also offers different standards for representing enriched incidents information, such as STIX, IODEF, VERIS and X-ARF (Barnum 2014; Dandurand et al. 2015; Menges and Pernul 2018). In order to allow the representation of human delivered information, we chose the combination of the existing formats CybOX and STIX as development basis. Both formats are issued together by MITRE³ and a combined usage is explicitly intended. Since these formats are most commonly applied to represent comprehensive threat intelligence information (Shackleford and SANS Institute 2015; Sauerwein et al. 2017), high compatibility to existing systems can be assumed. Moreover, they offer broader representation capabilities in their basic configuration than comparable formats as shown by Menges and Pernul (2018) and therefore, represent a solid foundation for the integration of human delivered information. Both CybOX and STIX are briefly introduced in the following and examined for necessary extensions to represent human delivered information afterwards.

CybOX provides an extensive catalog of object types for the description of the indicator layer. Each object represents individual components of log files, such as files, processes or network packets and offers description options at a detailed level. For example, the object type *file* allows the description of basic file properties such as path, extension or file name but also additional information such as permissions, compression procedures or creation date. STIX is the most extensive and widespread format for the structured representation of cyber threat intelligence information available today (Burger et al. 2014). It provides flexible data structures, such as non-structured free-text attributes, built-in controlled vocabularies using predefined values (vocabs) as well as integrated references to external data sources such as platform or vulnerability databases (enumerations). STIX uses indicators provided by CybOX as information basis and a wide range of well-defined data definitions to express the intelligence and attribution information for threats. The data model consists of the following core concepts. Incident is the central

¹<https://cyboxproject.github.io>

²https://github.com/mandiant/OpenIOC_1.1

³<https://www.mitre.org>

entity for structuring the incident information. TTP (Tactics, Techniques and Procedures) and Course of Action to describe the Intelligence layer. Campaign, Threat Actor and Exploit Target describe the Attribution layer. Indicator, Observable serves as interface to the Indication layer that is essentially provided by CyBOX. Moreover, numerous attributes for a detailed expression of these concepts are provided by the data model (Barnum 2014; Menges and Pernul 2018; Fransen et al. 2015).

After this short introduction of the data formats STIX and CyBOX, we develop adjustments for these formats to represent human delivered incident information following the IHaaS taxonomy (see Section 4.2) and CTI data structure (see Section 4.3) in the following. For this purpose, we first discuss the missing elements within the data formats based on the CTI basic data structure. On this basis, we propose the following changes to the formats to allow the integration of human sensors.

- Intelligence:** Previously, it was shown that attack events can be mapped within the CTI base model, whereas other events are not available yet. Using the taxonomy, we are able to limit these additional events to the categories *Structural*, *Environmental* and *Legal*. In order to also support these events within the data format, we have defined the additional entities "Technical Event", "Environmental Event" and "Legal Event". All these entities are derived from the basic entity TTP, which describes tactics, techniques and procedures used in the course of an attack. An essential property of TTP objects is the structured representation of attack patterns. For this purpose, STIX uses the CAPEC Enumeration, a freely available data set of known attack patterns for the unambiguous description of specific attacks. In order to achieve a comparable functionality for the additionally defined events, we defined a corresponding vocabularies for structural, legal and environmental events. Each vocabulary offers predefined event definitions according to our taxonomy. In addition to the event definitions, the area of intelligence also offer possibilities for describing the expected impact of an incident. For this purpose it was previously shown that the base model only provides impact definitions that emerge from active attacks. Although this is basically also true for the data format, its data definitions do not explicitly restrict the representation of incident results to an underlying attack. As a result, no changes are necessary to enable the definition of specific event results.
- Attribution:** It was shown that an integration of structural sources is necessary for addressing passive threats within the CTI base format. In addition, it was

shown that the entities are limited to the expression of active attacks. The data format already provides elements such as Threat Actor, Exploit Target to represent active attacks and attackers, and Asset Vocabulary to define assets. To enable the integration of passive threats, we extend STIX with the definition of an additional entity "structural source" as intended in the CTI base format. Since this is an alternative threat source, the object is derived from the existing Threat Actor object and exists on the same level. This object is extended by an additional vocabulary "StructuralSourceTypeVocab" to be able to represent structural threat sources in a structured way. Since this extension of threat sources also extends the scope of attribution, we additionally defined an extension of the asset vocabulary. This makes it possible to define additional assets that can occur in connection with passive threats.

- Indicator:** The indicator category is used to represent incident event information on a high level of detail, which are basically able cover the event information that may be delivered by humans. However, humans are usually not capable of delivering information on this level of detail and will rather provide unstructured data fragments. Consequently, such data fragments must be evaluated afterwards and the format must allow the unstructured data to be recorded at the time of acquisition. For this reason, we have also added an extension to the Observable object that allows to include unstructured data, which can later be translated into structured CyBOX information.

In addition to these specific extensions, all objects were equipped with specific IHaaS IDs and to enable additional references between the objects. This allows employees to express their perception by establishing links between objects. These additional connections can then be separately evaluated by analysts and integrated into the analysis results. In summary, it was shown in this section that STIX already fulfills numerous requirements for the implementation of an IHaaS platform. However, the format requires different extensions to fully match the taxonomy according to the CTI base model. To achieve this, additional entities to represent structural threat sources as well as environmental, structural and legal events are defined within the data model. Moreover, different vocabularies are introduced to unambiguously represent these entities. Finally, the Observable object is extended by an attribute for the unstructured capture of event data. Table 2 gives an overview of all these adjustments to the data format. A detailed overview of the specific extensions integrated as well as the actual object specifications can be found in the repository published together with this

Table 2 STIX extensions

Classification	Base entity	Additional Entity	Additional attribute
Intelligence	TTP	Structural Event	StructuralEventTypeVocab
	TTP	Environmental Event	EnvironmentalEventTypeVocab
	TTP	Legal Event	LegalEventTypeVocab
Attribution	Threat Actor	Structural Source	StructuralSourceTypeVocab
	Incident		
Indicator	Observable		Observation

work⁴. The repository includes XML-schema definitions for the STIX schema extension types and vocabularies that are developed with this work.

4.5 Structured acquisition of human-as-a-security-sensor information

To implement a system harvesting incident information from a human sensor, it is necessary to develop a systematic approach to guide the user through the acquisition (Q3). This supports the structured input into a data format 4.4 and encourages human sensors to provide as much information as possible. The process for guiding the user is basically given by the IHaaS Incident Model and Taxonomy as shown in Fig. 3. Thereby, multiple threat sources, threat events, and entities can be specified consecutively. The expected impact is estimated for the whole incident and thus recorded only once. The respective subtypes for sources, events or entities are also gathered in hierarchical sequence to avoid overstraining of the user. Each event is assigned a cause (either a threat source or another threat event), which leads to a chain of events. However, the process is subject to some constraints. More precisely, threat events cannot be initiated by some threat sources or preceding threat events. The constraints for our acquisition process are defined as follows and explained in more detail subsequently. The notation is based on the formal model of Klingner and Becker (2012):

$$\text{prohibits}(\text{Attack}) = \text{Environmental} \vee \text{Legal} \quad (1)$$

$$\text{prohibits}(\text{Technical}) = \text{Legal} \quad (2)$$

$$\text{prohibits}(\text{Environmental}) = \text{Legal} \quad (3)$$

$$\begin{aligned} \text{prohibits}(\text{UnusualNaturalEvent}) \\ = \text{Actor} \vee \text{Structure} \vee \text{Attack} \\ \vee \text{Technical} \vee \text{Legal} \end{aligned} \quad (4)$$

Equation 1 defines that an attack cannot be initiated by an environmental or a legal event. The reason for this

is that an attack requires action by a human being or at least some technical device and thus cannot be initiated by nonhuman events or sources. Furthermore, the cause of a technical security event cannot be a legal event (Eq. 2), *technical events* can only follow *physical events* or *sources*. The same applies to environmental events (Eq. 3). *Unusual natural events* (e.g. sunspots) cannot be caused by any other events or sources except *Environmental* ones as stated in Eq. 4, because they have a natural cause.

These constraints are the most explicit ones. It would be possible to define additional constraints considering more detailed layers of the underlying taxonomy. However, the constraints would depend on the organization where they are implemented and would not be unambiguous.

5 Evaluation

In the previous sections, we presented an approach for integrating human sensor information into SIEM systems. Therefore, we first discussed possible connecting points for the interaction between human sensors and SIEM systems. We also developed an incident model that extends the scope of SIEM threat detection by incidents that are additionally detectable by human sensors. Based on these findings, we extended the STIX data model to create data structures capable of capturing this information and proposed a concept for the structured acquisition of human sensor information. In design science research, demonstration is like a light-weight evaluation, to show that the artifact works to solve instances of a given problem (Venable et al. 2012; Peffers et al. 2007). To evaluate that our approach achieves its purpose in our context, we demonstrate it threefold: First, we explain our prototypical implementation, which shows that it is realizable in practice. Thereafter, we use the example from chapter 2 to show that it can be mapped to the IHaaS Incident Model and Taxonomy presented in Section 4.2. Finally, we demonstrate how this example would be represented in the STIX based format presented in 4.4. Hereby it is worth mentioning, that a taxonomy is never perfect and has to be shaped and extended as the field of its purpose advances (Nickerson et al. 2013). Furthermore, it is hardly possible to evaluate the taxonomy going beyond a demonstration, since it can only

⁴<http://tinyurl.com/y3h5k25t>

be shown exemplary, that it fits its intended purpose. This is especially true for the context of this paper, as there are almost no limits to the variety of cyberattacks and incidents. To the best of our knowledge, there is no similar taxonomy describing the artifacts that can be recognized by human security sensors. Thus it is not possible, to compare the performance of our taxonomy to others.

5.1 Prototypical implementation

Our application prototype realizes the rendering of information delivered by human security sensors into the structured threat intelligence information. A working example of the IHaaS prototype is available online⁵. The prototype pursues two different goals. On the one hand, it demonstrates the use of IHaaS in a possible scenario for the structured acquisition of incident information to show the overall validity of our approach. On the other hand, it illustrates the value of information delivered by human security sensors and the combination possibilities with data from existing analytics processes. The application consists of two major components: First, a wizard component that allows the reception of incident information delivered by humans. Second, a server component that translates the acquired incident information into the structured format to be further processed afterwards. The frontend is implemented by using Angular⁶ and Typescript⁷. Java EE in combination with a Glassfish⁸ application server was used to implement the STIX conversion logic and the database access.

Figure 4 shows a screenshot of the first step in the wizard component. The wizard is divided into two components. In the first component, the information can be entered by the user. The second part (Captured elements) gives an overview of already declared incident elements so that the user can see what has been previously entered. The wizard is structured in four steps as specified by the taxonomy. At first, the threat sources can be reported. Thereby, an arbitrary number of sources can be added. For selecting a source, the user is presented a drop-down list containing the elements of the first layer of the taxonomy (Actor and Structure). When an element is selected, a second drop-down list with the elements of the next layer is displayed. This continues until there are no sub-elements left. The same selection mechanism is implemented for event types and entities in subsequent steps. Only for events a "triggered by" input field is added to specify the previously reported threat source or threat event that initiated the event. There the

selectable events get filtered according to the constraints defined in chapter 4.5. In the fourth and final step, the estimated impact of the whole incident can be entered. Furthermore, the following additional information is requested:

- **Email:** The email is used to enable follow-up contact to the user who reported an incident for example when additional information is required.
- **Date:** The date on which the incident occurred. The current date is used as default value.
- **General description of the incident:** A free text explanation of the incident enables the statement of additional context information.
- **Technical data:** This input field is used for providing technical information like log data or the content of a phishing mail. This information could also be gathered partially automatically as described by Heartfield and Lukas (2018) depending on the incident and the organizations' infrastructure.

After the incident information was acquired by the wizard component, the data is transferred to the backend component. The backend provides the conversion logic, which translates the information collected by the wizard into corresponding STIX objects. It also provides the underlying data storage for persisting the translated STIX objects for later use.

5.2 Case study

In order to evaluate the wizard in combination with the underlying taxonomy and constraints we show how an employee could report an incident using the wizard. We used the DarkVishnya incident as shown in Section 2 as an exemplary use-case, which we iterate through below. Note that we take the role of a fictional employee that could have observed the incident. Thus, we only consider occurrences that may have been observed by a non-technical staff member for this example. The potential selection steps within the wizard are subsequently shown in brackets. For this incident, we identified the following two threat sources:

- 1 An unknown person is observed inside the premises (Actor → Individual → Outsider)
- 2 A suspicious hardware device is seen in an office room (Structure → IT Equipment → Processing)

Moreover, two threat events can be identified:

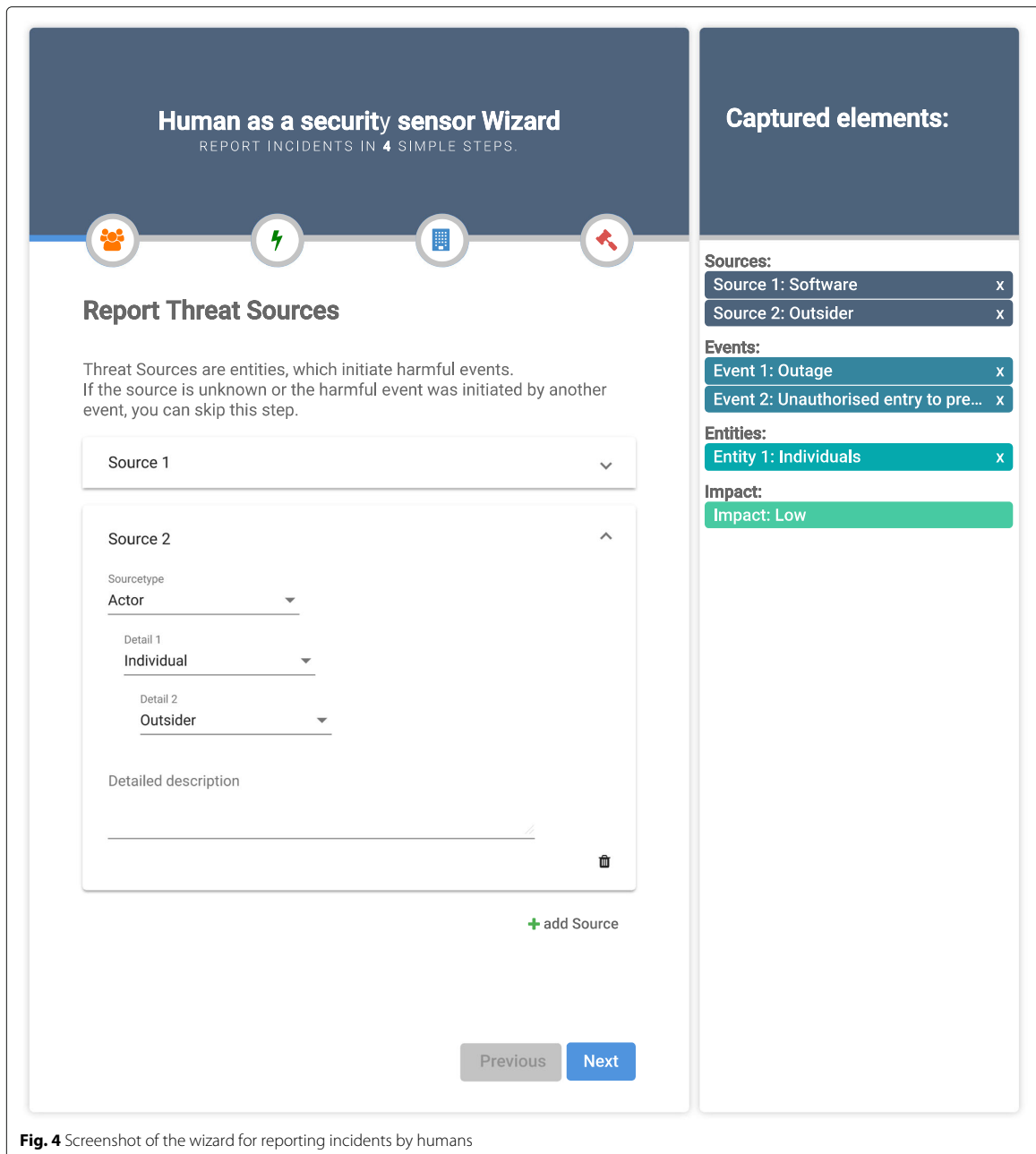
- 1 The person falsely claims to have legitimate access and enters the premises (Attack → Physical attacks → Unauthorized entry to premises)

⁵<http://tinyurl.com/yyqqjgg7>

⁶<https://angular.io/>

⁷<https://www.typescriptlang.org/>

⁸<https://javaee.github.io/glassfish/>



- 2 The hardware device is placed in an office room and connected to internal network infrastructure (Attack → Nefarious activity/Abuse → Manipulation of hardware and software)

In addition, a network device was identified as a negatively affected entity. Thus, assets are selected from the

wizard. Finally, the impact is estimated as low, since the employee may not be able to judge the whole extent of the incident. After the data is collected from the human sensor, it is translated into the corresponding STIX data objects by the server component as described in the following. The outsider (1) who falsely claimed to have legitimate access to the premises is translated

into a Threat Actor object. Its specific properties are mapped to the internal vocab "ThreatActorTypeVocab" that was extended within this work. The technique of gaining unauthorized access to the premises is translated into a TTP object and matching attack patterns from the CAPEC enumeration are mapped. The suspicious hardware device (2) attached to the internal network is then mapped to a structural source object and its specifics are mapped using the "StructuralSourceTypeVocab" created with this work. The action of planting a malicious device is described using a further TTP object and the corresponding CAPEC attack patterns analogous to the first TTP object. After creating these specific entities, the general descriptions of the incident as well as the time of the occurrence, affected assets, and the expected impact are recorded using an Incident object. All these objects are then finally wrapped using a Report object. The complete STIX report for this exemplary use-case is appended to this work as supplementary material. Moreover, it can be viewed under the past incidents overview section within the wizard prototype⁹.

Considering the results of this incident, there are different possible connecting points to automated analyses within a SIEM system. Firstly, the generated report delivers information about the approximate time of the occurrence, the exact location as well as the affected network device and possibly even the used network port. This data can then be enriched with the corresponding log information from the SIEM system in order to clarify the findings. Furthermore, if an electronic access control has been circumvented in any way, the log data available can also be used as further evidence and to enrich the incident information gathered from the employee.

5.3 Discussion

The prototypical implementation has shown three key aspects: First, it was demonstrated, that it is possible to represent the beforehand theoretically defined IHaaS incident model and taxonomy (Section 4.2) as a wizard-like application. This application guides the user through the taxonomy and enables him to select and report all possible elements. Second, the acquisition can be conducted in a structured way since the constraints defined in Section 4.5 were all implemented within the prototype. Nevertheless, practical usage over a longer period of time will reveal whether these constraints are exhaustive. Third, the acquired data can be translated into a STIX representation, which could be further used for security analytics systems, despite the volume of possible user input.

⁹<http://tinyurl.com/y5tsoxo3>

The case study has shown that it is generally possible to apply the prototype for a real-world incident. Therefore, it was validated with an expert who analyzed the attack as a member of the incident response team. However, only a broad long-term study can show the usability, which we will address in the future.

6 Conclusion and future work

In this paper, we present an approach for acquiring and structuring incident information from human sensors to prepare it for the use within security analytics systems such as SIEM systems. Therefore, we identify the connection points of human sensors within a SIEM system (Q1) and answer the question how the reportable information can be structured (Q2). Thereby the IHaaS Incident Model and Taxonomy is deduced, which consists of the four components threat sources, threat events, entities and expected impact. The incident model builds the basis for a data format suitable for representing threat intelligence information reported by humans. An important factor while developing the data format is to maintain the compatibility with existing and well-established formats, in our case STIX. For acquiring the data from human sensors in a structured way (Q3) we propose a process where we define some constraints, which ensure that the collected data is not contradictory. Finally, the approach is evaluated from three directions. First, we prototypically implement the approach and second, an example use-case is mapped to the IHaaS Incident Model and Taxonomy to show its practicability. Finally, the use case was represented in the proposed STIX-based format.

Since the examined subject of human-as-a-sensor, especially with its focus on security, is a rather new topic, there is a lot of potential for future research. A topic marginally tackled in this paper is the connection of human-generated data with machine-generated data, which for example originates from log files. The data collected from humans may be extended by automatically or manually deriving relationships to machine data. To achieve this, different approaches such as rule-based correlation and aggregation may be used. In order to facilitate the definition of rules, it can be helpful to visualize the generated data. Therefore, existing approaches as presented by Böhm et al. (2018) could be extended to the proposed data format. Machine learning techniques also show a lot of potential regarding the correlation of data acquired by humans and machine-generated data.

Our present work considers the acquisition and structuring of information delivered by humans. However, we have not examined forensic and legal requirements. Nevertheless, considering these requirements is of great relevance especially when the collected data is supposed to be used as evidence in court afterwards. Furthermore,

human generated data may also play an important role in the incident response process and thus should be qualified as data foundation for forensic analyses. Since reports may contain personal data, the topic needs additional consideration from a legal point of view.

An additional research gap can be identified with regard to motivating employees for reporting detected incidents. On the one hand, incentives have to be created and on the other hand, barriers keeping employees from reporting have to be removed. For example, if a person reports an incident, which denigrates a colleague, it might be an unwanted result. In this context, obfuscation techniques, such as anonymization or pseudonymization, may help to solve some of these problems. Additionally, changes to the corporate culture are required, so that it is considered normal for employees to report detected incidents, as it is for example in an anti-fraud culture. In this regard the analysis and assurance of data quality is especially important due to the possibility of erroneous inputs by humans. Finally, the proposed approach is rather generic. Thus, it has to be adopted to the respective context for practical use. Especially the proposed taxonomy could be refined in order to depict more corporate information and it has to be tailored to match the corporate culture.

Acknowledgements

This research was supported by the Federal Ministry of Education and Research, Germany, as part of the BMBF DINGfest project (<https://dingfest.ur.de>). We also thank Sergey Golovanov of Kaspersky Lab for providing us with detailed information about the Dark Vishnia incident, which significantly enhanced this paper.

Funding

Not applicable.

Availability of data and materials

Source code - iHaas wizard

Project name: Client
Project home page: <http://tinyurl.com/y6kjr4q>
Archived version: 1.0
Operating system(s): Platform independent
Programming language: HTML, Typescript/JavaScript
Other requirements: Apache Webserver or similar, NPM 6.2.0 or higher
License: GNU GPL v3

Source code - sTIX server

Project name: STIX Server
Project home page: <http://tinyurl.com/y46hsvj8>
Archived version: 1.0
Operating system(s): Platform independent
Programming language: Java EE 7
Other requirements: Glassfish version 4.1.1 or higher
License: GNU GPL v3

Additional sTIX schema files

Project name: STIX-Schema
Project home page: <http://tinyurl.com/y2s3ba7k>
Archived version: 1.0
Operating system(s): Platform independent
Programming language: xml-schema
License: BSD-3-Clause
Appended as supplementary material

Received: 24 April 2019 Accepted: 29 August 2019

Published online: 22 October 2019

References

- Anti-Phishing Working Group I Report Phishing. <https://www.antiphishing.org/report-phishing/overview/>. Accessed 19.01.2019
- Appala S, Cam-Winget N, McGrew D, Verma J (2015) An Actionable Threat Intelligence system using a Publish-Subscribe communications model. Proc 2nd ACM Workshop Inf Sharing Collab Secur - WISCS '15:61–70
- Barnum S (2014) Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX). <http://stixproject.github.io/getting-started/whitepaper/>. Accessed 2019-02-21
- Bhatt S, Manadhata PK, Zomlot L (2014) The operational role of security information and event management systems. IEEE Secur Privacy 12(5):35–41
- Böhm F, Menges F, Pernul G (2018) Graph-based visual analytics for cyber threat intelligence. Cybersecurity 1(1)
- Burger EW, Goodman MD, Kampanakis P, Zhu KA (2014) Taxonomy model for cyber threat intelligence information exchange technologies. In: WISCS '14 Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security Vol. WISCS '14. pp 51–60
- Crowley C, Pescatore J (2018) Sans 2018 security operations center survey
- Dandurand L, Kaplan A, Kácha P, Kadobayashi Y, Kompanek A, Lima T, Millar T, Nazario J, Perlotto R, Young W (2015) Standards and Tools for Exchange and Processing of Actionable Information
- Fenz S, Heurix J, Neubauer T, Pechstein F (2014) Current challenges in information security risk management. Inf Manag & Comput Secur 22(5):410–430
- Fransen F, Smulders A, Kerkdijk R (2015) Cyber security information exchange to gain insight into the effects of cyber threats and incidents. Elektrotechnik & Informationstechnik 18:106–112
- Google LLC. Gmail. <https://mail.google.com/>. Accessed 19.01.2019
- Heartfield R, Loukas G, Gan D (2016) You are probably not the weakest link: Towards practical prediction of susceptibility to semantic social engineering attacks. IEEE Access 4:6910–6928
- Heartfield R, Loukas G (2018) Detecting semantic social engineering attacks with the weakest link: Implementation and empirical evaluation of a human-as-a-security-sensor framework. Comput Secur 76:101–127
- Hintzbergen J, Hintzbergen K, Smulders A, Baars H (2015) Foundations of Information Security: Based on ISO 27001 and ISO 27002. 3rd. Van Haren Publishing, Zaltbommel
- Holik F, Horalek J, Neradova S, Zitta S, Marik O (2015) The deployment of security information and event management in cloud infrastructure. In: 2015 25th International Conference Radioelektronika (RADIOELEKTRONIKA). pp 399–404
- ISO/IEC 27001: Information technology – Security techniques – Information security management systems – Requirements (2013) Technical report. Int Org Standard
- Joint Task Force Transformation Initiative (2012) Guide for Conducting Risk Assessments. National Institute of Standards and Technology, Gaithersburg, MD
- Juliadotter NV, Choo K-KR (2015) Cloud attack and risk assessment taxonomy. IEEE Cloud Comput 2(1):14–20
- Klingner S, Becker M (2012) Formal modelling of components and dependencies for configuring product-service-systems. Enterp Model Inf Syst Architectures 7(1)
- Kostakos V, Rogstadius J, Ferreira D, Hosio S, Goncalves J (2017) Human sensors. In: Participatory Sensing, Opinions and Collective Awareness. Springer, Cham. pp 69–92
- Lineberry S (2007) The human element: The weakest link in information security. J Account 204(5):44
- Marinos L (2016) ENISA Threat Taxonomy: A Tool for Structuring Threat Information
- Mello J (2017) Security Awareness Training Explosion. <https://cybersecurityventures.com/security-awareness-training-report/>. Accessed 28.02.2019
- Menges F, Pernul G (2018) A comparative analysis of incident reporting formats. Comput Secur 73:87–101
- Microsoft Corporation Deal with abuse, phishing, or spoofing in Outlook.com. <https://support.office.com/en-us/article/deal-with-abuse-phishing-or-spoofing-in-outlook-com-0d882ea5-eedc-4bed-aebc-079ffa1105a3>
- Nickerson RC, Varshney U, Muntermann J (2013) A method for taxonomy development and its application in information systems. Eur J Inf Syst 22(3):336–359

- Peffer K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. *J Manag Inf Syst* 24(3):45–77
- Rahman SS, Heartfield R, Oliff W, Loukas G, Filippopolitis A (2017) Assessing the cyber-trustworthiness of human-as-a-sensor reports from mobile devices. In: 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA). pp 387–394
- Shackelford D, SANS Institute (2015) Who's Using Cyberthreat Intelligence and How? <https://www.alienvault.com/docs/SANS-Cyber-Threat-Intelligence-Survey-2015.pdf>. Accessed 2019-02-21
- Sauerwein C, Sillaber CN, Mussmann A, Breu R (2017) Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. In: 13. Internationale Tagung Wirtschaftsinformatik, WI 2017, St. Gallen
- Golovanov S (2018) DarkVishnya: Banks attacked through direct connection to local network. <https://securelist.com/darkvishnya/89169/>
- Turnbull J (2019) The Art of Monitoring. Version 1.0.4
- Venable J, Pries-Heje J, Baskerville R (2012) A comprehensive framework for evaluation in design science research. In: International Conference on Design Science Research in Information Systems. pp 423–438
- Vielberth M, Pernul G (2018) A security information and event management pattern. In: 12th Latin American Conference on Pattern Languages of Programs (SugarLoafPLOP 2018)
- Wang D, Amin MT, Li S, Abdelzaher T, Kaplan L, Gu S, Pan C, Liu H, Aggarwal CC, Ganti R, Wang X, Mohapatra P, Szymanski B, Le H (2014) Using humans as sensors: An estimation-theoretic perspective. In: IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks. IEEE, Piscataway. pp 35–46

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

5 Improving data quality for human-as-a-security-sensor. A process driven quality improvement approach for user-provided incident information

Current status:	Published
Journal:	Information and Computer Security, Volume 29, Number 2, March 2021
Date of acceptance:	19 October 2020
Full citation:	VIELBERTH, M., ENGLBRECHT, L., AND PERNUL, G. Improving data quality for human-as-a-security-sensor. A process driven quality improvement approach for user-provided incident information. <i>Information and Computer Security</i> 29, 2 (2021), 332–349
Authors' contributions:	Manfred Vielberth 45% Ludwig Englbrecht 45% Günther Pernul 10%

Journal description: Information and Computer Security aims to cover the human aspects of security, looking at the impact of user and business behaviours when dealing with security issues. It communicates fresh ideas and cutting-edge research to academics and practitioners.

The current issue and full text archive of this journal is available on Emerald Insight at:
<https://www.emerald.com/insight/2056-4961.htm>

Improving data quality for human-as-a-security-sensor. A process driven quality improvement approach for user-provided incident information

HaaS quality improvement

Manfred Vielberth, Ludwig Englbrecht and Günther Pernul
*Chair of Information Systems, University of Regensburg,
Regensburg, Germany*

Received 26 June 2020
Revised 15 October 2020
Accepted 19 October 2020

Abstract

Purpose – In the past, people were usually seen as the weakest link in the IT security chain. However, this view has changed in recent years and people are no longer seen only as a problem, but also as part of the solution. In research, this change is reflected in the fact that people are enabled to report security incidents that they have detected. During this reporting process, however, it is important to ensure that the reports are submitted with the highest possible data quality. This paper aims to provide a process-driven quality improvement approach for human-as-a-security-sensor information.

Design/methodology/approach – This work builds upon existing approaches for structured reporting of security incidents. In the first step, relevant data quality dimensions and influencing factors are defined. Based on this, an approach for quality improvement is proposed. To demonstrate the feasibility of the approach, it is prototypically implemented and evaluated using an exemplary use case.

Findings – In this paper, a process-driven approach is proposed, which allows improving the data quality by analyzing the similarity of incidents. It is shown that this approach is feasible and leads to better data quality with real-world data.

Originality/value – The originality of the approach lies in the fact that data quality is already improved during the reporting of an incident. In addition, approaches from other areas, such as recommender systems, are applied innovatively to the area of the human-as-a-security-sensor.

Keywords Data Quality, Cybersecurity, Incident Response, Evidence Collection, Human-as-a-Security-Sensor, Security novice

Paper type Research paper

1. Introduction

In the cybersecurity domain, the answer to the question of who the weakest link in IT security is, would in most cases be the human (Lineberry, 2007; Mello, 2017). This led to constantly neglecting the human potential, even though a lot of effort is put in training employees. For example, during awareness campaigns employees are trained to distinguish between what is benign and what is conspicuous from an IT security perspective.

However, in recent years, the perspective has changed and Human-as-Problem has shifted to Human-as-Solution (Zimmermann and Renaud, 2019), with humans seen as important contributors to company-wide security. As cyberattacks become more sophisticated and target multiple points of organizations' infrastructures, although much research is still needed (Furnell and Clarke, 2012), the role of humans being solely victims changes. A human can



Information & Computer Security
© Emerald Publishing Limited
2056-4961
DOI 10.1108/ICS-06-2020-0100

ICS

contribute his conscious observation to the detection and, consequently, to the reconnaissance of an incident. For example, most advanced persistent threats in the first step solely target humans (Chen *et al.*, 2014) and thus in many cases do not leave any technical traces to detect them in early stages. Consequently, in the past few years, the human developed into an important source for security-related information and, thus, cannot be ignored in a holistic security concept, which represents a fundamental paradigm shift in the area of cybersecurity. Especially the combination of the human-provided information with technical information, such as log data and human observations, shows great potential. This development became known as the human-as-a-security-sensor paradigm (Heartfield *et al.*, 2016).

Using humans as sensors for gaining additional information about an IT security incident is a beneficial approach. However, it must also be verified whether they report false or incomplete information. A specific method must be developed and applied to ensure an acceptable quality of the delivered data from humans. In particular, appropriate mechanisms must be provided during reporting that include proper user assistance. If the user is asked about the incident after a longer period of time, important details may not be recalled anymore.

To sum up, the question arises:

RQ1. How can data quality be improved during Human-as-a-Security-Sensor reporting?

Therefore, this paper builds on existing work on the paradigm of human-as-a-security-sensor in the field of cyber threat detection, by aiming to further systematize the exploitation of user data reported in the threat detection process against a standardized set of indicators and information model criteria commonly leveraged in threat intelligence. Improvements in this area can, therefore, support reliable and automated exploitation of user-reported data through security orchestration, for faster response to credible attack vectors. Thus, improvements in data quality and structural alignment to other threat detection sources are highly attractive for more effective and autonomous integration of the human-as-a-security-sensor capabilities into emerging next-generation security architectures. Since we focus on the reporting process of a user, an ex-post improvement of the data quality of Threat Intelligence information through human-as-a-security-sensor is not a goal of this paper.

This paper is structured as follows. In Section 2, we present the applied methodology followed by an overview of the related work in the area of using information from a human-as-a-security-sensor in Section 3. A definition of human-as-a-security-sensor data quality is presented in Section 4. Section 5 provides an integrated approach for a data quality-aware human-as-a-security-sensor interview. Therein, our core contribution of a process-driven data-quality improvement is given. In Section 6, the concept is applied and evaluated to a use case. Section 7 provides a short discussion about the proposed approach. In Section 8, a summary of the concept and an outlook on future work is given.

2. Methodology

Our methodology comprises three steps. First, the foundations for the subsequent approach are laid by defining corresponding data quality dimensions and influencing factors in the context of security-related information reported by humans (1). Second, a process-driven approach to improve the provided information during the reporting of the incident by additional guidance of the user is proposed (2). Third, the feasibility of the approach is demonstrated in the evaluation by implementing the approach in a prototype and playing through a use case as an example (3).

- (1) To lay the foundation for the envisaged approach, relevant data quality dimensions are defined for the context of the paper. For this purpose, relevant data

quality dimensions and human characteristics that influence data quality are first extracted from the literature and then put into the new context. Data quality is of particular importance in this regard as there is a significant need for structured and high-quality data from humans in order to benefit from it during incident investigation in practice (Pawliński *et al.*, 2014).

HaaSS quality improvement

- (2) For the second step, an approach for utilizing human-provided incident information within enterprises is used as a baseline for the data quality improvement method. The proposed method follows a process-driven approach that is derived from approaches known from recommender systems and consists of two phases. Therein a possibility is given to acquire information from the user with minimal restrictions (Phase 1). To improve the data quality, further questioning of the user is performed by using a similarity measurement between the user-provided data and previously stored incident information (Phase 2). This concept has been influenced by recommender systems and pursues to ask the user-specific questions to differentiate the input more precisely.
- (3) To evaluate the resulting process-driven data quality improvement approach, it is prototypically implemented to show its feasibility. Further, to demonstrate that the approach achieves its goal of improving data quality it is evaluated with a set of real-world incidents and the quality improvement is demonstrated, by running through an exemplary scenario.

3. Background and related work

In-house systems, wherein employees can report non-security related information are already quite common in various forms. In this context, two of these systems can be highlighted: systems for reporting fraud and applications for reporting ideas for improvement of business or production processes Westerski *et al.* (2011). Forensic fraud management has been developed in the last few years due to increasing computer-related fraud. In Wells (2017), the implementation of adequate reporting programs has been emphasized. He states that employees who recognized suspicious fraud-related activities require a way to report these findings without the fear of being involved in the investigation. A fraud reporting hot-line is considered a useful tool for detecting fraud.

The idea of using humans for reporting security incidents is not entirely new. However, past approaches were focused on specific incident categories. For example, systems for reporting malicious emails already exist for a long time. These can be narrowed down to systems for reporting phishing emails or for flagging them as spam, which is implemented in nearly every modern email software.

Since companies are required to enable all employees to report security events or weaknesses by certain security management standards (e.g. according to control A.13.1 of the ISO 27001 (ISO/IEC, 2013)), human-to-human reporting has found its way into many organizations. Thereby, employees can for example report incidents to specific contact points such as help desks or they must fill out a form describing the incident, which in turn must be interpreted by security experts.

Heartfield and Loukas (2018) proposed a more sophisticated approach focusing on the reporting of semantic social engineering attacks. They developed a framework and show, that human sensors can outperform technical systems in many cases. Subsequently, a more general approach was introduced by Vielberth *et al.* (2019), which aims at reporting all possible kinds of security incidents that are observable by humans. Additionally, the gained data is transformed into the STIX format for structuring threat intelligence, to use it in

ICS

security analytics. This concept for a human-as-a-security-sensor incident reporting system is used as a basis for the process presented in Section 5.

Data quality assessment and improvement has been discussed intensively in the literature. Methods to improve data quality can be categorized as data-driven and process-driven strategies (Batini *et al.*, 2009). Data-driven approaches directly modify the previously collected data by using improvement techniques such as error correction or standardization. In contrast, process-driven approaches improve the processes underlying the data collection. In this paper, we focus on the process-driven strategy by improving the process for collecting data about incidents from a human-as-a-security-sensor, since it outperforms the data-driven technique in the long-term (Glowalla and Sunyaev, 2013).

Concerning human-as-a-security-sensor related data quality, Heartfield and Loukas (2018) introduced a model for predicting sensor reliability by considering, for example, the frequency of access to the attacked platform, the duration of access and the received computer security training. Rahman *et al.* (2017) describe a method for assessing the trustworthiness of human-as-a-sensor reports by analyzing automatically collected data, such as CPU usage or network traffic. Thereby, they revealed, that a small amount of recorded data can help to estimate its trustworthiness. However, these approaches do not assure data quality at the time of reporting. This is especially important since humans tend to forget certain details and thus might not be able to fix data quality issues if the reporting lies too far in the past.

As described above, there are approaches in non-security-related areas to gather and improve the information collected from humans. However, to the best of our knowledge, there is no approach that improves the process of information gathering of human-as-a-security-sensors to collect high quality data. In our framework, we propose a structured questioning approach, in combination with intelligent user assistance, that focuses on gathering as much security-related information as possible from users, while simultaneously improving its quality, which may become very relevant in case of later forensic investigations. Not only information about a specific incident is reported, but also information about a justified suspicion. This information significantly expands the scope of the observed incident.

4. Human-as-a-security-sensor data quality

An often-cited definition for data quality was introduced by Orr (1998) and describes data quality as the “measure of the agreement between the data views presented by an information system and that same data in the real world”. Thereby, one can distinguish between quality of design and quality of conformance (Helfert and Heinrich, 2003; Juran, 1999; Teboul, 1991). Heinrich *et al.* (2009) define quality of design as “the degree of correspondence between the users’ requirements and the specification” and quality of conformance as “the degree of correspondence between the specification and the existing realization”.

In the context of human-as-a-security-sensor, data quality plays an important role, as the data is generated by humans. Thus, it is important to determine whether the data quality is sufficient to be used for further analyses or even as evidence in court. In this paper, we focus on quality of conformance because it is more objective (Heinrich *et al.*, 2009). Further, we do not want to focus on the users’ requirements but rather focus on the process-driven method for quality improvement during data collection.

Since there is no clear and consistent understanding of quality dimensions related to human-as-a-security-sensor data in the literature, it is necessary to define those first. In our context, a human is acting as a sensor and, thus, combines features of a sensor and those of a

human. Therefore, we consider both areas and allocate a basic terminology for human-as-a-security-sensor data quality. Classical data quality dimensions that can be measured by analyzing human-generated data have to be considered as well. In our context, it is also important to consider the characteristics of the human sensor that influence data quality. The classical and the human characteristics are presented and defined in relation to the context hereafter. Although not all dimensions are used to measure the data quality for the proposed improvement approach, all relevant dimensions are presented in the following to give a complete view.

HaaSS quality
improvement

Data quality dimensions are extensively discussed in the literature. However, it is difficult to determine the most important dimensions in the context of human-as-a-security-sensor, because no systematic procedure and data quality criteria can be found in the literature (Helfert and Heinrich, 2003). However, three commonly cited papers (Redman, 1996; Wand and Wang, 1996; Wang and Strong, 1996) state the dimensions that are worth a closer examination. In the following, these dimensions are defined and interpreted in the context of human-as-a-security-sensor:

- **Completeness:** According to Wand and Wang (1996), “completeness is the ability of an information system to represent every meaningful state of the represented real-world system”. In our context, this refers to whether the representation of the reported incident contains all the information required for security analytics or later forensic investigations. Hence, completeness is especially dependent on whether the human sensor has observed meaningful information in its entirety and if the user has reported all of it.
- **Consistency:** Batini and Scannapieco (2016) define consistency as the “coherence of the same datum, represented in multiple copies, or different data to respect integrity constraints and rules”. Since we assume that the incident data was gathered from a human sensor in a structured way, we expect this structured approach to permit only the provision of data that adheres to all integrity constraints and rules. However, the data might not be coherent with other data related to an incident (originating from other human-as-a-security-sensors or machine-generated data).
- **Relevance:** Relevance is defined as “the extent to which data are applicable and helpful for the task at hand” by Wang and Strong (1996). In the context of human-as-a-security-sensor, the data is relevant if it provides information about the course of the incident. At the time the data was captured, it might not be possible to determine whether it is relevant or not, since it could turn out to be important later on, for example, during forensic investigations.
- **Reliability:** Wand and Wang (1996) define that “reliability indicates whether the data can be counted on to convey the right information”. With regard to the data generated by human-as-a-security-sensors, this means that the data indicates the true course of the reported incident. It is worth mentioning that the reliability declines, starting from the point in time when the incident was observed, until the actual acquisition in an information system, since humans tend to forget information. Thus, this dimension is closely related to timeliness and currency.
- **Timeliness and currency:** Heinrich *et al.* (2009) carried out a detailed comparison of definitions for timeliness and currency. A definition that fits our context comes from Wang and Strong (1996). They define currency as “the extent to which the age of the data is appropriate for the task at hand”. In our context, the task at hand is some kind of security analysis process or forensic investigation. Additionally, it is

ICS

important in a security context that the period between the observation of an incident and its actual reporting is as small as possible since the potential damage increases the longer an incident remains undetected.

In addition to the quality dimensions, the human characteristics that influence data quality can be assessed. These are not directly related to the data provided but can still help to predict whether the human sensor will provide high-quality data. In this regard, [Heartfield and Loukas \(2018\)](#) have identified several “predictors of detection efficacy”. Since our context is broader and relates to all possible incidents, we portray those predictors in a more general manner:

- **Security competencies:** Security competencies relate to knowledge about IT security. This knowledge can, for example, be acquired through security training or security awareness programs ([Heartfield and Loukas, 2018](#)). Security competencies play a crucial role since a more comprehensive knowledge of technical terminology allows a security expert to provide more complete information about an incident. Additionally, a security expert is able to assess the severity of the incident better.
- **Context-related competencies:** A security incident occurs in a certain context. This may involve the affected asset (such as the attacked information system) or other circumstances that were a component of the *incident*. The competencies of a human sensor in this particular context, play a crucial role concerning the data generated by her. For example, a human who is an expert in specific software would be more capable of delivering high-quality information about the incident (assumed it is related to the specific software) as opposed to a person who is not familiar with it.
- **General competencies:** General competencies are also expected to influence the quality of human-generated incident data. General education plays an important role in enabling people to deliver high-quality reports.
- **Trustworthiness:** A fitting definition of trustworthiness in the context of human-as-a-security-sensor comes from [Batini and Scannapieco \(2016\)](#). They state, “trustworthiness is the objective probability that the trustee performs a particular action on which the interests of the truster depend”. In our context, the interests of the truster are that the trustee delivers information about the security incident of the highest possible quality. Reputation is important here since human sensors who have provided low-quality incident information in the past can be expected to do so in the future. The predictors *mentioned* above can also influence trustworthiness significantly.

Our concept is based on the assumption that a user intentionally reports a certain suspicion of an incident or relevant information about an incident via a computer-based reporting system. The knowledge and intrinsic motives of the user are not further determined. Our concept for improving the data quality of human-as-a-security-sensors uses existing data and attempts to address the criteria of completeness and reliability. Only in a subsequent step is it possible to determine the criteria relevance based on the given information through an analysis conducted by an expert. The data quality dimensions “timeliness and currency” can be determined based on the information received but cannot be improved at the time of the reporting. For this reason, the focus of this work is on the dimensions completeness and reliability to improve the data quality of the human-as-a-security-sensor data. This can be improved by providing user assistance at the time of the reporting.

5. A data quality-aware human-as-a-security-sensor interview

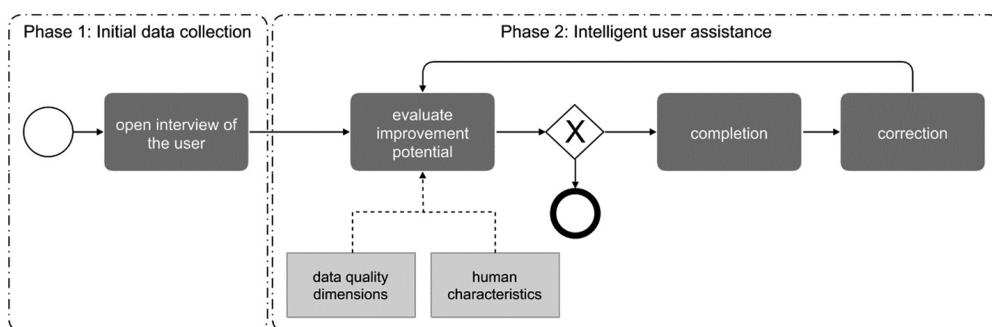
For retrieving information from human security sensors in a structured manner, we propose a generic interviewing process, as presented in Figure 1. In the first step, the human sensor can provide the information freely, without concrete suggestions, but still in a structured way based on an underlying taxonomy. This step is followed by a phase of intelligent user assistance, which should guide the user to improve the provided data quality. Therefore, the quality of the data is first measured to decide if it is already sufficient. Once the user has completed the first step, two main aspects affect data quality. Firstly, the user may have forgotten to provide certain information, and secondly, the user may have provided incorrect information. These two aspects are addressed in the subsequent steps. During the completion step, the user is guided to correct the provided data or to share additional information. Finally, during the correction, the reliability of the user input is questioned and support is provided to correct errors. The corresponding process steps are explained in more detail below.

It is worth mentioning that the process should be run as early as possible after the incident has been observed. Although, the proposed approach does not require this in general. The three steps within the intelligent user assistance phase should be carried out directly after the open interview of the user so that the user remembers the incident as well as possible.

5.1 Open interview of the user

In the first step, after the human sensor has decided to report information about a suspicion of an incident, an open interview is conducted. We argue that, at this stage, the human should be treated as freely as possible. Since the user is motivated to share information, she should not have to deal with other things that could distract or even discourage her from sharing information. However, care should be taken to ensure that the collected data is as structured as possible to facilitate further processing. An approach regarding this is described by Vielberth *et al.* (2019).

Figure 2 shows the used taxonomy for our data collection. As shown by Vielberth *et al.* (2019), most of the categories shown can be subdivided further, but for reasons of clarity, this has been omitted in the illustration. The taxonomy follows an incident model based on the Joint Task Force Transformation Initiative (2012) as well as Juliadotter and Choo (2015). It states, that an incident starts with a threat source, which initiates one or more threat events. A threat event can also initiate other threat events. These events affect entities negatively causing an adverse impact. This process is also followed when questioning the user to provide a basic structure.



HaaSS quality improvement

Figure 1.
Generic interviewing process

ICS

5.2 Evaluate improvement potential

The information provided about an incident by the users' needs to be assessed according to their usage for further IT-security analysis. A comprehensive approach for automatically measuring the quality to determine whether completion or correction is required and therefore a two-fold mechanism is applied: (1) determining the quality of the data and (2) assessing the willingness of the user to reveal more information:

- (1) To identify whether the provided information is sufficient or whether the user should be further questioned, the quality of the reported information is determined. This measurement can be carried out with the help of reliability measures derived from predefined features of the data (e.g. as described by [Heartfield and Loukas \(2018\)](#)). If the process does not reach certain minimum requirements, the steps *completion* and subsequent *correction* may be carried out repeatedly in order to improve the results until they are satisfactory. Both steps are described herein. The prediction of the degree of completeness uses the concept of calculating the similarity of the given data to pre-known incidents. This is realized with a similarity measurement and is described in detail in the process step *completion* in Chapter 5.3. If no threat intelligence sources are available, or the intelligence has no similar reference to the current attack vectors, this does not necessarily make the human sensor report less valuable. Additionally, it is hard to determine the data quality a priori, which is why the willingness of the user to report additional information might be a better indicator in many cases.
- (2) The willingness of the user to report additional information about the incident is important to guarantee that the information that he provides is useful from an IT security perspective. This means if the user transitions from a volunteer-based interview to an assisted questioning, the behavior could have changed for two reasons. Firstly, the user could have changed her mind since she has realized that the reporting could have dire consequences for herself, colleagues, or the entire company, which even could tempt to provide false answers. Second, the time the user has already spent with reporting an incident has an impact on the motivation to truthfully answer further questions. This aspect can be represented as a decreasing motivation curve of reporting incidents over time. Therefore, well-considered and targeted questions about the incident are important to obtain useful information.

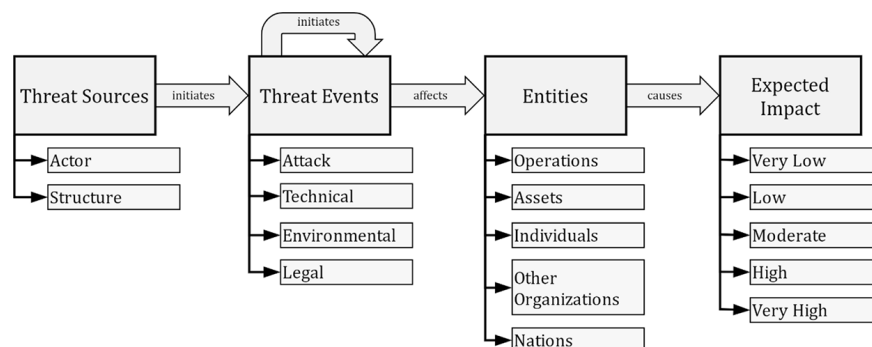


Figure 2.
Taxonomy for free
questioning of users

Source: Vielberth *et al.* (2019)

With the results of both aspects, the system can decide if a further loop could bring additional value with regard to information quality. If the possibility for improvement was identified, the process leads to the step *completion*. Since this step is integrated within the interviewing process, it can be triggered several times. To avoid misleading the user by repeating the same questions, it is necessary to track the new information retrieved after each loop once a possibility for improvement was determined and to break the loop if there was no improvement potential predicted.

HaaS quality
improvement

5.3 Completion

After all the information about the incident the user is aware of has been shared, the step “completion” guides her to provide additional data. The system selects specific attributes and asks the user to complete or correct the input based on previously provided information to improve the reporting. This is, for instance, information the user had not thought of in the first step, or simply underestimated the importance of said information and decided against reporting it.

The information already transmitted by the user at this stage is the basis for completing the information on the incident by asking further questions. Our concept proposes a comparison with pre-known incidents from a prepared database, which is an adapted technique originating from content-based recommender systems (Pazzani and Billsus, 2007). This enables the system to recognize blind spots and detect them. Thus, a response via user assistance is accomplished. Therefore, the input of the user is represented as a vector. Each value of the vector refers to a feature of a cyber-security incident and is grouped into the four areas *Threat Sources*, *Threat Events*, *Entities* and *Expected Impact*. This vector is successively compared with pre-known vectors using a similarity measure. By calculating the similarity of the input vectors and an incident vector, the similarity of the inputs to an already known event can be determined. To determine this similarity, the cosine similarity measure is used, which is explained in detail below.

Our approach incorporates the *cosine similarity* for measuring the similarity between incidents. The cosine similarity is used because it can handle missing attributes between two elements, irrespective of their magnitude. Additionally, we choose it, since it provides a more sophisticated possibility to incorporate ratings with various ranges (e.g. a scale from 1 to 10) in future applications. This enables a consideration of the certainty of the user for each item during the reporting.

The similarity of two data-sets of reported incidents, converted into two n -dimensional vectors (x and y), is defined as follows:

$$\cos(x, y) = \frac{x \cdot y}{\|x\| \cdot \|y\|} \quad (1)$$

Each attribute of a vector refers to a boolean representation of a respective incident characteristic. The calculation of the similarity narrows down the results of possible incidents. A small angle between the user input vector and the comparison vector of an incident follows in a higher cosine value and respective a similar incident description. Following this approach, we perform a selection between various known incidents, wherefore the given input vector of the user is compared with the range of known incidents.

Within the completion step, user assistance is provided by the system to ask the user about incident aspects that were not addressed until this stage. By addressing specific areas of the incident and presenting these to the user the submitted incident can be completed. An exemplary application of the cosine similarity measure is shown in [Figure 3](#). Within this

ICS

figure, (a) illustrates that the answering of feature *S5* can provide further clarification of the reported incident from the user (shown as *Vector User* in the figure) similar to *incident 4*.

5.4 Correction

In the correction step, information the human sensor has provided is scrutinized and aims to correct the user input. Only the features are questioned, which can lead to a sharper separation between the reported incident and the incidents from the incident repository. Thereby, it is verified whether the user reported the estimated truth or whether it tends toward an erroneous report because, for example, she mixed something up. Therefore, the cosine similarity as defined in equation (1) is determined again for the updated input values. This allows addressing inputs that deviate from the comparison vector. Moreover, these attributes have to be backed up with counter-questions to ensure that the user’s statement is correct.

For enabling the correction of estimated errors, previously provided statements of the user are verified by reverted questions that address the same feature. This is achieved by presenting counter-questions for the selected attributes of the user input. Thus, the user has a chance to secure the facts by answering further questions regarding the same specific feature of the incident. The approach of scientific questionnaires was used as an inspiration for our approach.

A method often used to uncover distortion effects by acquiescence is the inversion of items (Krosnick, 1999). Thereby, the items are presented in a positively formulated form, with agreement indicating a high value in the construct of interest; the same items are also presented in a negated form (“inverted”) by varying item wording. Test persons who have agreed to the positively formulated items should reject the inverted items. A renewed agreement, on the other hand, indicates acquiescence.

To sum up, the pre-provided inputs need further questioning to clarify certain details. This is illustrated in Figure 3(b). Every discrepancy detected within the correction step is documented for further evaluation (as it can for example affect the user’s trustworthiness).

6. Evaluation

The evaluation of the approach will be carried out in the following two ways. On the one hand, the approach was implemented as a prototype, which is freely accessible [1] and on the other hand, an example use-case is used to demonstrate step-by-step how our methodology works.

6.1 Prototypical implementation

The implementation builds on a prototype of a wizard from a previous work of Vielberth et al. (2019). The wizard covers the step open interview of the user and therefore enables her to report an incident in a structured way without restricting the

	Threat Sources						Threat Events						Entities						Expected Impact						Cosine similarity
	S1	S2	S3	S4	S5	S6	TE1	TE2	TE3	TE4	TE5	TE6	E1	E2	E3	E4	E5	E6	I1	I2	I3	I4	I5	I6	
Vector User	1	1	1	1	?	1	1						1	1			1		1						1,00
Incident 1	1	1					1										1								0,67
Incident 2		1	1	1			1	1					1										1		0,38
Incident 3				1	1					1	1					1	1			1				0,25	
Incident 4	1	1	1	1	1		1	1					1	1			?				1			0,78	
Incident ...																									
Incident n	1	1	1	1	1		1	1					1	1								1		0,67	

Figure 3. Exemplary application of the similarity measure

user too much. The wizard also fully covers the taxonomy shown in [Figure 2](#). Taxonomy for free questioning of users ([Vielberth et al., 2019](#)).

When the user completes the recording of the incident in the wizard, the intelligent user assistance phase continues as shown in [Figure 4](#). The user is asked questions which she can answer either with yes or no. The questions are based on information on similar incidents and are either generated automatically using the Natural Language Toolkit [2] or can be specified manually for each attribute in the taxonomy. On the one hand, the system can ask whether the user wants to complete certain information (completion) and on the other hand, whether certain information was provided incorrectly (correction).

To enable the management and creation of manually defined questions and incidents, the prototype also contains an administration tool. As shown in [Figure 4](#), the incidents reported by users must be approved by security experts to process them further, to ensure the quality of the provided data, and to prevent abuse through false reports.

HaaS quality improvement

6.2 Use-case

In the following, the functionality of the approach is described with concrete numerical values, independent of the implementation of the prototype, to enable a better reproducibility. Therefore, we applied the presented approach using an example scenario and measure relevant data-quality dimensions. For this purpose, we make certain assumptions regarding the user input data. The data for the incident repository is based on real-world cyber threat intelligence (CTI) from the IBM X-Force Exchange [3]. This is a platform for the exchange of CTI providing free access to incident information. Since the provided data mostly are incidents that are far too extensive for a vivid use case, it has been presented in a condensed form (e.g. some threat events of the course of the incidents have been omitted).

The observed incident of the user is based on a real-world incident that was recognized in mid-2020 and is a derivative of the *Emotet* malware. The procedure of the attack is as follows: First, the user receives an email with a Microsoft Word file attachment ([Figure 5](#)). Second, if the user opens the attachment the system gets infected with malware, which encrypts all data on the system and demands a ransom for decrypting the files.

[Figure 6](#) shows the user input during the free-questioning phase, which is intentionally neither complete nor correct. In addition, three example incidents from the incident repository are shown. The presented incidents are based on real-world examples:

- Incident 1: The first incident is a cyber-attack that was recorded in 2017 [4]. The attack started with a phishing attack that targeted specific individuals (so-called spear-phishing). A malicious e-mail attachment was used to gain access to the victim organizations' IT-infrastructure. This was exploited to implement a coordinated malware outbreak, which wiped all hard drives.
- Incident 2: This incident was first analyzed in January 2019 [5]. A phishing attack was also carried out. However, it did not target individuals and, thus, cannot be called a spear-phishing attack. Although login credentials had been stolen, the entities were not influenced, as the attacker did not taken advantage of the stolen data. Thus, the impact can be rated as low.
- Incident 3: The third incident is based on the so-called "Bad Rabbit" attack [6]. This is a ransomware breakout, that encrypts hard drives and demands a ransom in the form of a certain amount of Bitcoin.

ICS

Intelligent User Assistance

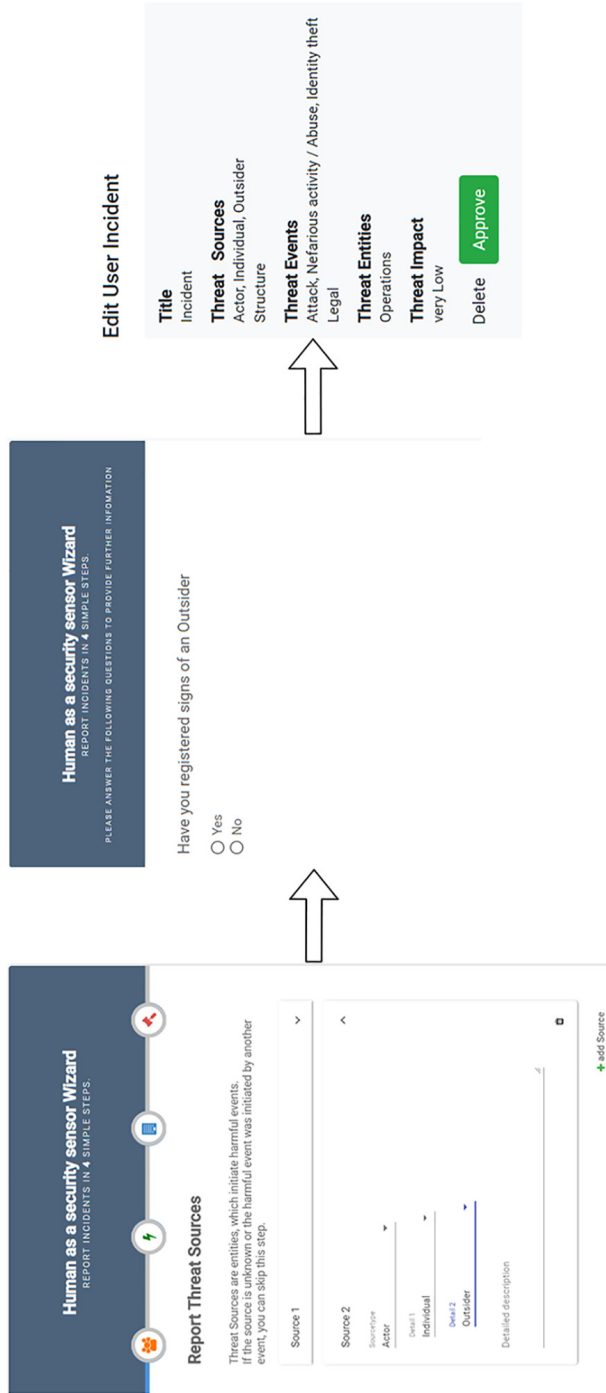


Figure 4. Screenshots and exemplary workflow of the prototypical implementation



HaaS quality improvement

Figure 5.
Phishing mail of the observed incident

	Threat Sources		Threat Events									Entities		Impact		Cosine similarity	
	Actor	Group	Social engineering	Phishing	Spearphishing	Malware	Trojans	Technical	Loss	Loss of information	compromising confidential information	Assets	Operations	Low	High		
Vector User	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1,00
Incident 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,78
Incident 2	1	1	1	1							1				1	1	0,75
Incident 3	1	1				1		1	1	1		1	1	1	1	1	0,55

Figure 6.
State of the model after the free questioning phase

6.2.1 Free questioning of user. We assume, that the user has observed and reported the described Emotet attack, which is very similar to Incident 1. However, she did not report the incident in full extent. Thus, the completeness of the report is rather low at the beginning. Here it should be noted, that the measurement of reliability and completeness differ from the measurement considered in Chapter 5.2 since we presume to know the incident that actually occurred here, which is not the case while going through the interviewing process in practice. In practice, these values have to be estimated. The initial completeness is measured based on the study of [Batini and Scannapieco \(2016\)](#) as follows:

$$\text{compl}(X, Y) = \frac{|X \cap Y|}{|Y|} = \frac{8}{13} = 0.62$$

Therefore, X is a set containing all the user inputs, and Y is the set of the incident that actually happened. This is the ratio of the number of correct user inputs and the number of elements of the actual incident. The reliability is calculated subsequently as follows:

$$\text{rel}(X, Y) = \frac{|X \cap Y|}{|X|} = \frac{8}{9} = 0.89$$

Reliability is thus the ratio between the number of correct user inputs and the total number of user inputs, containing false ones.

6.2.2 Completion. During the completion step, the similarity of the user input to the incidents in the repository is calculated. Thereby, it is noticeable, that *Incident 1* and *Incident 2* have a high cosine similarity. Thus, those two incidents are further focused

ICS

upon during the questioning of the user. To complete the data of the user, the input gets compared to the two incidents to identify and fill in gaps. In our example (comparing Figure 6), the first gap compared to *Incident 1* is that it contains a “technical” threat event, which the user did not report. In comparison to *Incident 2*, where the first gap is that the impact of the incident is low, whereby the user did not report any impact. Thus, the user gets asked whether she recognized a technical threat event and if she would estimate the impact as low (because, for instance, she might have forgotten to state it during the free-questioning step). Ideally, she can answer the first question with “yes” because she recognized a technical threat event. The second question would be answered with “no” because all the hard drives were encrypted during the incident and thus, its impact would not be estimated as low.

Figure 7 presents the results of the first completion pass. It can be noted that the cosine similarity with *Incident 1* rises, as the user’s response approaches the underlying truth. For validation, the completeness is now $9/13 = 0.69$, and the reliability $9/10 = 0.9$. As can be seen, both quality measurements are increasing.

6.2.3 Correction. The correction step also begins with the selection of the most similar incident(s). In our example, *Incident 1* is the most similar one (compare Figure 7). To improve the user data, it is searched for contradictory elements. The method finds that the user has reported that the incident contained the threat event “compromising confidential information”. However, *Incident 1* does not. Thus, the user gets asked whether this really happened. Potentially, she notices that she did not see it and for example just mixed something up. Consequently, after answering this question, the value is removed from the corresponding feature within the user vector.

As seen in Figure 8 the similarity to the first incident is further increasing. After the correction step, the completeness is $9/13 = 0.69$ and the value of the reliability is $9/9 = 1.00$; thus, it has reached the optimum level. In contrast, the completeness did not change during the correction step.

These steps get repeated, until a satisfactory degree of data quality is reached, until the user can no longer contribute anything, or until her willingness to share information is too low.

To sum it up, we have shown, that in a real-world oriented use-case, our approach step by step improves the data quality of human-as-a-security-sensor incident reports. In the real world, the user input might not be as ideal as we have assumed it. However, the user is still guided to provide data of as high quality as possible.

Figure 7. State of the model after the first completion phase

	Threat Sources		Threat Events										Entities		Impact		Cosine similarity
	Actor	Group	Social engineering	Phishing	Spearphishing	Malware	Trojans	Technical	Loss	Loss of information	compromising confidential information	Assets	Operations	Low	High		
Vector User	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00
Incident 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.82
Incident 2	1	1	1	1							1				1	1	0.71
Incident 3	1	1				1		1	1	1		1	1	1	1	1	0.61

Figure 8. State of the model after the first correction phase

	Threat Sources		Threat Events										Entities		Impact		Cosine similarity
	Actor	Group	Social engineering	Phishing	Spearphishing	Malware	Trojans	Technical	Loss	Loss of information	compromising confidential information	Assets	Operations	Low	High		
Vector User	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00
Incident 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.86
Incident 2	1	1	1	1							1				1	1	0.64
Incident 3	1	1				1		1	1	1		1	1	1	1	1	0.64

7. Discussion

To encourage users to report threats, the reporting process should be as frictionless as possible. Otherwise, a convoluted process of answering too many questions related to the report may discourage further reporting in the future. This aspect has been considered in our concept during the modeling of the two-phased interview process that asks the user for additional information during the reporting process. Compared to incident reporting procedures without immediate questions from the system, this process is slightly more time consuming for the user. Nevertheless, our concept provides the basis to find mechanisms to improve context and quality, whilst also reducing the amount of time required to information from a human.

Since the application of the similarity measure requires an existing database with incidents and corresponding features, the cold-start problem arises at the beginning of applying our concept. As shown in the evaluation, this problem was solved by using incident information from the *IBM X-Force* database. This gives the concept a sound basis for practical application. Nevertheless, the concept is strongly dependent on the stored data of incidents and requires the continuous development of the incident repository.

We found the missing ground truth data is a challenge for the concept. Although the aforementioned databases provide a source for IT security incidents (which indicate the ground truth), these data must be truthful. For implementation in a real-world environment, incident data must be enriched with the results of a profound IT forensic investigation. Only with this, the association between the characteristics of an incident and the actual course of events can be established.

The proposed concept provides additional value for the acquisition of high-quality information from a user. However, there is a risk that the requests from the system within the process step completion and correction confuse the user and he or she will revise previously made statements. This aspect has been addressed in this concept indirectly by recording the history of the statements transmitted. Providing a technical expert with this information will give a more profound picture of the development process of the reported incident.

Since we have presented an approach for gathering information from humans, remotely reminiscent of an interrogation, questions about ethical issues arise, which are cleared out in the following. The proposed interviewing process only gives hints about missing or wrong data. It does not give a general prediction if someone has intentionally retained information or provided wrong data. However, the proposed approach could pose some misuse potential if it is further developed in the wrong direction, which has to be considered in follow-up publications.

8. Conclusion and future work

To conclude, in the last years, a change of perspective is observable, from viewing the human as the weakest link in cybersecurity to perceiving her as an important source of information. Our work shows that it is important to address the data quality improvement of incident information provided by a user. To achieve that, intelligent user assistance is provided during the reporting task. A specific process-driven approach containing the steps open interview of the user, evaluate improvement potential, completion, and correction was implemented. The core contribution of our approach lies in intelligent user assistance through a targeted user interview. The user starts with the first phase “initial data collection” containing a free-questioning and within the second phase, intelligent user assistance is provided to guide her to complete or correct previous inputs. The intelligence lies in the application of methods derived from recommender systems to determine specific features of an incident to clarify the alignment and similarity of the provided information with pre-known incident entries from a prepared repository.

HaaS quality
improvement

ICS

Our approach has been evaluated by a prototypical implementation of the intelligent reporting system followed by applying the concept to a use-case. A scenario is prepared wherein data from *IBM X-Force* has been adapted, and the process of the incident reporting is played through. In the first step, the reporting process was completed without further questioning. The second step used the input data from the first phase to perform the completion and correction steps. The user was questioned about some more details regarding the incident. These questions have been determined by the proposed intelligent user assistance during the reporting process and enabled process-driven data quality improvement. This led to the answering of the research question:

RQ1. How can data quality be improved during Human-as-a-Security-Sensor reporting?

Since the time passed between the occurrence of the incident, and the reporting is extremely crucial to retrieve useful data, the process-driven approach enables this time span to be as short as possible. Hence, further clarification is performed at the point of reporting and not weeks or months later when the incident is investigated. Therefore, the system must be able to ask intelligent questions during the reporting process in order to guarantee optimal system-supported user assistance.

It is worth mentioning that the presented approach is a step in the direction that humans are part of the solution in terms of detecting security incidents. However, the approach does not address the root cause of the problem, which in many cases relates to humans (e.g. due to their vulnerability in terms of social engineering). Therefore, it is still important to address these vulnerabilities.

Future research to optimize user assistance for reporting IT security incidents should be focused on the different perceptions and classifications of incidents by different groups of people. Target-oriented handling with different levels of detail of the reported incidents has to be found. Further improvement potential for the presented approach could be identified in finding better similarity measures or utilizing approaches from the machine learning domain. Another major problem area that has hardly been considered in science is the question of how people can be motivated to report security incidents.

Notes

1. <http://go.ur.de/ihaasswizzard> (the source code can be provided upon request)
2. www.nltk.org/
3. <https://exchange.xforce.ibmcloud.com/>
4. <https://exchange.xforce.ibmcloud.com/collection/Spear-Phishing-Attacks-Preceding-Shamoon-Malware-Breakouts-eeed4eede51b9a4587f4c7c816ad6e4e>
5. <https://exchange.xforce.ibmcloud.com/collection/Phishing-Scam-Lures-Australian-Government-Contractors-Into-Disclosing-Account-Credentials-662c0fd11c387ae6b91bf4af7dc2337f>
6. <https://exchange.xforce.ibmcloud.com/collection/XFTAS-SI-2017-00001-Bad-Rabbit-51701e9c25aaaf7e02b19fa6d63ccc80>

References

- Batini, C., Cappiello, C., Francalanci, C. and Maurino, A. (2009), "Methodologies for data quality assessment and improvement", *ACM Computing Surveys*, Vol. 41 No. 3, pp. 1-52, doi: [10.1145/1541880.1541883](https://doi.org/10.1145/1541880.1541883).

- Batini, C. and Scannapieco, M. (2016), *Data and Information Quality: Dimensions, Principles and Techniques, Data-Centric Systems and Applications*, Springer International Publishing, Switzerland.
- Chen, P., Desmet, L. and Huygens, C. (2014), "A study on advanced persistent threats", in: *IFIP International Conference on Communications and Multimedia Security*, Aveiro, Portugal, pp. 63-72.
- Furnell, S. and Clarke, N. (2012), "Power to the people? The evolving recognition of human aspects of security", *Computers and Security*, Vol. 31 No. 8, pp. 983-988, doi: [10.1016/j.cose.2012.08.004](https://doi.org/10.1016/j.cose.2012.08.004).
- Glowalla, P. and Sunyaev, A. (2013), "Process-driven data quality management through integration of data quality into existing process models", *Business and Information Systems Engineering*, Vol. 5 No. 6, pp. 433-448, doi: [10.1007/s12599-013-0297-x](https://doi.org/10.1007/s12599-013-0297-x).
- Heartfield, R. and Loukas, G. (2018), "Detecting semantic social engineering attacks with the weakest link: implementation and empirical evaluation of a human-as-a-security-sensor framework", *Computers and Security*, Vol. 76, pp. 101-127, doi: [10.1016/j.cose.2018.02.020](https://doi.org/10.1016/j.cose.2018.02.020).
- Heartfield, R., Loukas, G. and Gan, D. (2016), "You are probably not the weakest link: towards practical prediction of susceptibility to semantic social engineering attacks", *IEEE Access*, Vol. 4, pp. 6910-6928, doi: [10.1109/ACCESS.2016.2616285](https://doi.org/10.1109/ACCESS.2016.2616285).
- Heinrich, B., Klier, M. and Kaiser, M. (2009), "A procedure to develop metrics for currency and its application in CRM", *Journal of Data and Information Quality*, Vol. 1 No. 1, pp. 1-28, doi: [10.1145/1515693.1515697](https://doi.org/10.1145/1515693.1515697).
- Helfert, M. and Heinrich, B. (2003), "Analyzing data quality investments in CRM: a model-based approach", in: *Eighth International Conference on Information Quality (ICIQ 2003)*, MIT, pp. 80-95.
- ISO/IEC (2013), "Information technology – security techniques – information security management systems – requirements".
- Joint Task Force Transformation Initiative (2012), *Guide for Conducting Risk Assessments*. National Institute of Standards and Technology, National Institute of Standards and Technology (NIST), Gaithersburg, MD, doi: [10.6028/NIST.SP.800-30r1](https://doi.org/10.6028/NIST.SP.800-30r1).
- Juliadotter, N.V. and Choo, K.-K.R. (2015), "Cloud attack and risk assessment taxonomy", *IEEE Cloud Computing*, Vol. 2 No. 1, pp. 14-20, doi: [10.1109/MCC.2015.2](https://doi.org/10.1109/MCC.2015.2).
- Juran, J.M. (1999), "How to think about quality", *Juran's Quality Handbook*, 5th ed., pp. 2.1-2.18.
- Krosnick, J.A. (1999), "Survey research", *Annual Review of Psychology*, Vol. 50 No. 1, pp. 537-567.
- Lineberry, S. (2007), "The human element: the weakest link in information security", *Journal of Accountancy*, Vol. 204 No. 5, pp. 44-49.
- Mello, J.P. (2017), "Security awareness training explosion", available at: <https://cybersecurityventures.com/security-awareness-training-report/> (accessed 14 October 2020).
- Orr, K. (1998), "Data quality and systems theory", *Communications of the ACM*, Vol. 41 No. 2, pp. 66-71.
- Pawliński, P., Przemysław, J., Kijewski, P., Siewierski, Ł., Jacewicz, P., Zielony, P. and Żuber, R. (2014), *Actionable Information for Security Incident Response*, European Union Agency for Network and Information Security, Heraklion, Greece.
- Pazzani, M.J. and Billsus, D. (2007), "Content-based recommendation systems", in Brusilovsky, P., Kobsa, A. and Nejdl, W. (Eds), *The Adaptive Web, Methods and Strategies of Web Personalization, Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, pp. 325-341, doi: [10.1007/978-3-540-72079-9](https://doi.org/10.1007/978-3-540-72079-9).
- Rahman, S.S., Heartfield, R., Oliff, W., Loukas, G. and Philippopolitis, A. (2017), "Assessing the cyber-trustworthiness of human-as-a-sensor reports from mobile devices", in *15th IEEE International Conference on Software Engineering Research, Management and Applications (SERA)*, IEEE, London, pp. 387-394, doi: [10.1109/SERA.2017.7965756](https://doi.org/10.1109/SERA.2017.7965756).
- Redman, T.C. (1996), *Data Quality for the Information Age*, Artech House, Boston, Mass.
- Teboul, J. (1991), *Managing Quality Dynamics*, Prentice Hall Direct.

HaaS quality improvement

ICS

- Vielberth, M., Menges, F. and Pernul, G. (2019), "Human-as-a-security-sensor for harvesting threat intelligence", *Cybersecurity*, Vol. 2 No. 1, pp. 1-15, doi: [10.1186/s42400-019-0040-0](https://doi.org/10.1186/s42400-019-0040-0).
- Wand, Y. and Wang, R.Y. (1996), "Anchoring data quality dimensions in ontological foundations", *Communications of the ACM*, Vol. 39 No. 11, pp. 86-95.
- Wang, R.Y. and Strong, D.M. (1996), "Beyond accuracy: what data quality means to data consumers", *Journal of Management Information Systems*, Vol. 12 No. 4, pp. 5-33.
- Wells, J.T. (2017), *Corporate Fraud Handbook: Prevention and Detection*, John Wiley and Sons.
- Westerski, A., Iglesias, C.A. and Nagle, T. (2011), "The road from community ideas to organisational innovation: a life cycle survey of idea management systems", *International Journal of Web Based Communities*, Vol. 7 No. 4, pp. 493-506, doi: [10.1504/IJWBC.2011.042993](https://doi.org/10.1504/IJWBC.2011.042993).
- Zimmermann, V. and Renaud, K. (2019), "Moving from a human-as-problem to a human-as-solution cybersecurity mindset", *International Journal of Human-Computer Studies*, Vol. 131, pp. 169-187, doi: [10.1016/j.ijhcs.2019.05.005](https://doi.org/10.1016/j.ijhcs.2019.05.005).

Corresponding author

Manfred Vielberth can be contacted at: manfred.vielberth@ur.de

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com

6 Security Operations Center: A Systematic Study and Open Challenges

Current status:	Published
Journal:	IEEE Access
Date of acceptance:	13 December 2020
Full citation:	VIELBERTH, M., BÖHM, F., FICHTINGER, I., AND PERNUL, G. Security Operations Center: A Systematic Study and Open Challenges. <i>IEEE Access</i> 8 (2020), 227756–227779
Authors' contributions:	Manfred Vielberth 35%
	Fabian Böhm 35%
	Ines Fichtinger 20%
	Günther Pernul 10%

Journal description: IEEE Access is a multidisciplinary, open access journal. It publishes articles that are of high interest to readers: original, technically correct, and clearly presented. The scope of this journal comprises all IEEE's fields of interest, emphasizing applications-oriented and interdisciplinary articles.

Received November 24, 2020, accepted December 13, 2020, date of publication December 17, 2020, date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3045514

Security Operations Center: A Systematic Study and Open Challenges

MANFRED VIELBERTH¹, FABIAN BÖHM¹, INES FICHTINGER¹,
AND GÜNTHER PERNUL¹, (Member, IEEE)

Chair of Information Systems, University of Regensburg, 93053 Regensburg, Germany

Corresponding author: Manfred Vielberth (manfred.vielberth@ur.de)

ABSTRACT Since the introduction of Security Operations Centers (SOCs) around 15 years ago, their importance has grown significantly, especially over the last five years. This is mainly due to the paramount necessity to prevent major cyber incidents and the resulting adoption of centralized security operations in businesses. Despite their popularity, existing academic work on the topic lacks a generally accepted view and focuses mainly on fragments rather than looking at it holistically. These shortcomings impede further innovation. In this paper, a comprehensive literature survey is conducted to collate different views. The discovered literature is then used to determine the current state-of-the-art of SOC and derive primary building blocks. Current challenges within a SOC are identified and summarized. A notable shortcoming of academic research is its focus on the human and technological aspects of a SOC while neglecting the connection of these two areas by specific processes (especially by non-technical processes). However, this area is essential for leveraging the full potential of a SOC in the future.

INDEX TERMS Security management, security operations center, security operations, SOC.

I. INTRODUCTION

According to a recent report, the average number of security breaches reported by organizations has risen by 11% from 130 in 2017 to 145 incidents in 2018 [1]. Over the last five years, this number has risen by a total of 65%. However, this report only covers detected and reported incidents, and the number of unreported incidents is probably much higher. The total annual cost of any type of cyber-attack is also growing at a steady pace [1]. Unfortunately, many attacks go undetected for a surprisingly long time. The mean time to detect an incident was 196 days in 2018, and it took another 69 days on average to contain the breach [1]. This detection time demonstrates how ineffective companies are at detecting and mitigating cyber-attacks. The reasons for this inefficiency include but are not limited to companies (1) not having an overview of their devices, systems, applications, and networks, (2) not knowing which assets to protect, (3) not knowing which tools to use and how to integrate them with the existing infrastructure, or (4) being overwhelmed by the speed technology and the ever-evolving threat landscape.

Security Operations Centers (SOCs) can provide an overarching solution for detecting and mitigating an attack if implemented correctly. They incorporate a mixture of people, processes, technologies, and governance and compliance, to effectively identify, detect, and mitigate threats, ideally before any damage occurs. However, there are a few research gaps and challenges associated with SOC. The biggest issue is the lack of a precise definition of a SOC and its components. For some researchers, a SOC is solely an entity responsible for monitoring the network. For others, it is an organizational unit encompassing all security operations, like incident management and threat intelligence. This lack of consensus hinders companies from deploying efficient SOC and researchers from further adding to the innovation of SOC. Therefore, this work's main contribution is to close this research gap by establishing a ground truth for a state-of-the-art SOC. We conduct a structured literature review to identify and subsume the current state-of-the-art.

The remainder of this paper is structured as follows. We identify related work in Section II. We describe the methodology applied to carry out this literature survey throughout Section III. Section IV is the first part of the main contribution of this work. Therein we summarize relevant work for the definition of a SOC and other more

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Huang¹.

TABLE 1. Review protocol.

Research questions	– What is the state-of-the-art of SOC as seen in research? – Which challenges need to be solved to advance the field?
Dates	January 1st, 1990 - December 31st, 2019
Databases	IEEE Xplore Digital Library ¹ , ACM Digital Library ² , SpringerLink ³ , EBSCO Host ⁴ , Wiley Online Library ⁵ , Web of Science ⁶ , Dimensions ⁷
Search criteria	English; Search keywords in Title, Abstract and Keywords
Search keywords	<i>Security Operation Center</i> OR <i>Security Operations Center</i> OR <i>Security Operations Centers</i> OR <i>Security Operation Centre</i> OR <i>Security Operations Centre</i> OR <i>Security Operations Centres</i>
Search methods	Keyword search, Backward search, Forward search
Inclusion criteria	Addresses SOC in general or part of it; Is available as a full version; Is not superseded by an included paper; Evaluates a paper included by a previous criterion

general aspects. The second main contribution is formulated in Section V, which distills the building blocks of a SOC from literature. To highlight a roadmap for future research, we identify a series of open challenges within Section VI. We conclude our work in Section VII summarizing the review.

II. RELATED WORK

A fundamental problem within a significant part of SOC literature is that it is very fragmented and widespread. Only a limited body of work has attempted to define holistic, architectural SOC frameworks so far [2]–[6]. Although researchers agree on most of the necessary capabilities, there is no clear consensus of what constitutes a SOC. Furthermore, most academic work focuses on particular characteristics of a SOC without paying much attention to the overall picture.

We identified some work partially relevant to our approach which is trying to get a more hands-on understanding of SOCs. The authors of the respective publications use semi-structured interviews [2], [7]–[11], on-site visits [2], [12], case studies [13], or ethnographic fieldwork [14]–[17]. These publications derive their definition of SOCs following a bottom-up approach leading to a limited understanding of SOCs. Interviews and on-site visits provide insight into a small fraction of specific SOC elements but do not allow conclusions upon a general state-of-the-art. We see a lack of general overview and identification of the status-quo in the field of SOC research. There is a need for a commonly agreed-upon terminology to advance the field further. We take the first step to fulfill this need.

III. METHODOLOGY

Our work aims to identify, evaluate, and synthesize relevant academic literature in the field of SOCs. Despite the real, practical significance of the topic, there is a lack of academic research, especially regarding a commonly agreed, holistic definition of SOCs. This issue makes it hard for researchers and organizations to identify relevant literature, and as a result, impedes future research and innovations in this field.

We aim to provide a guided tour through existing literature and establish a common ground truth. To conduct the review,

we follow the three stages proposed by Tranfield *et al.* [18] based on well-established guidelines [19]–[21]. The review protocol in Table 1 specifies research questions, information sources, search criteria, and relevant keywords. After the first collection of papers, we apply predefined criteria for inclusion or exclusion of papers to decrease the amount of papers and increase the quality of the literature considered for further review.

Table 1 lists the used keywords to identify relevant literature. Only publications that had the exact search term in title, abstract, or keywords are considered. Searching for “Security” AND “Operations” AND “Center” results in an immense number of papers, from which only a very small fraction is relevant to this study. Therefore, only the full term is applied to identify relevant literature. The common abbreviation “SOC” is not used to search for papers because it also abbreviates System on a Chip (SoC) and, as a result, also produces a high number of false positives. The defined keywords are used to search in the databases defined in (Table 1). We chose these databases because of their reputation within information systems, computer science, and cybersecurity. Finally, *Dimensions* is included in the list of searched databases as it provides a holistic view over a wide variety of papers reflected by the number of search results.

In total, 321 academic publications are identified using the keywords depicted in Table 2. From this set, we remove all duplicates, leaving 208 papers to analyze. Those papers are extracted, and the selection (inclusion/exclusion) criteria are applied. All available remaining papers are downloaded and their abstracts are read to decide upon their relevancy for the study, leaving a total of 158 papers.⁸ Figure 1 illustrates the publication dates of the remaining 158 papers after applying the exclusion criteria. The first paper included in the literature review was published in 2003. The number of publications about SOCs is skyrocketing since 2015, and we expect it to keep rising within the next years. Therefore, we see a strong necessity to establish a common baseline for SOC research.

⁸For transparency reasons, the full list of 321 academic publications and the filtering steps are made available via <https://go.ur.de/SOCLiterature>

TABLE 2. Search results per database.

Database	Search Criteria	Σ
IEEE Xplore	Document title, Abstract	34
ACM Digital Library	Title, Abstract, Keywords	18
SpringerLink	Title	18
EBSCO Host	AB Abstract, TI Title Only peer-reviewed	15
Wiley Online Library	Keywords, Title	4
Web of Science	Topic (Title, Abstract, Author keywords)	30
Dimensions	Title, Abstract	202
Total		321
After duplicate removal		208
After selection criteria		158

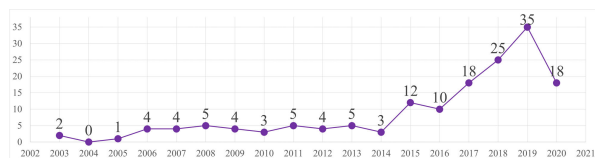


FIGURE 1. Relevant publications per year (until June 31st, 2020) identified in the structured review.

The identified literature can be categorized into two main categories *General Aspects* and *Building Blocks*. The first one summarizes the state-of-the-art regarding SOC definitions, operating models, and architectures. The second main category, *Building Blocks*, deals with the aspects which, based on literature, are comprising a SOC. Although we analyze scientific work to understand academia's current view, the topic of SOCs is highly driven by the industry as well. However, within the industry, the term *Security Operations Center* is used very ambiguously. Therefore, we only include a limited number of influential gray literature in this survey when appropriate. This literature is identified in the references used in scientific papers.

Besides the term "Security Operations Center", there is a wide variety of other, closely related terms used in the literature, e.g. Grid Security Operation Center (GSOC), Virtual Security Operation Center (VSOC), and many more. From here on, we will use the term SOC to abbreviate "Security Operations Center".

IV. GENERAL ASPECTS

This section introduces the first part of our main contribution. We subdivide this part of our work into the delimitation & definition of SOCs, their architecture, and operating models. Identified literature for these subtopics is summarized in Table 3.

A. DELIMITATION & DEFINITION

A SOC is an organizational unit operating at the heart of all security operations. It is usually not seen as a single entity or system but rather as a complex structure to manage and enhance an organization's overall security posture.

TABLE 3. Identified literature for the topic *General Aspects*.

General Aspects	References
Definition & Delimitation	[2], [3], [5], [17], [22]–[39]
Architecture	[3], [4], [6], [30], [34], [39]–[61]
Operating Models	[2], [3], [7], [25], [33], [46], [62]–[68]

Its function is to detect, analyze, and respond to cybersecurity threats and incidents employing people, processes, and technology [2], [22]–[25], [69]. Those activities can be formalized into seven dimensions or functional areas of a SOC [5], [26]. While widely accepted as utterly crucial for a company's security, SOCs are still considered a passive and reactive defense mechanism [27]–[29].

Research often describes operations within a SOC following the People, Processes, and Technologies (PPT) framework [3], [30]–[33]. This framework is used for various information technology topics like knowledge management [70] or customer relationship management [34]. Also, among SOC vendors, this framework is popular to summarize and structure their product. Although the *Governance and Compliance* aspect is often subordinated to processes, we consider it to be a category of its own due to the high importance within SOCs. It offers the framework in which people operate and according to which the processes and technologies are built. Therefore we extend the original PPT framework resulting in the People, Processes, Technology, Governance and Compliance (PPTGC) framework displayed in Figure 2.

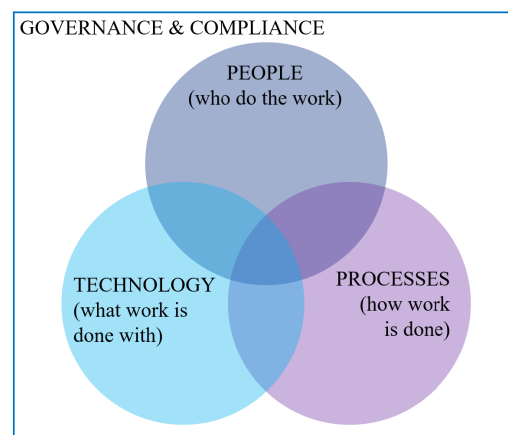


FIGURE 2. The People, processes and technology, governance & compliance (PPTGC) framework based on [70].

When implemented along with the PPTGC framework, a SOC can improve a company's security posture [36]. However, there is no clear terminology established describing a SOC. The following paragraphs delimit SOC from various other terms:

- **Computer Security Incident Response Team:** This term is often used interchangeably for a SOC although it mainly focuses on the response part once an attack has

happened. A CSIRT is an organizational unit responsible for coordinating and supporting the response to a computer security incident [71]. A CSIRT is classified either as an independent team or part of a SOC [37].

- **Network Operations Center:** A Network Operations Center (NOC) oversees identifying, investigating, prioritizing, escalating, and resolving problems [17], [38]. However, in NOCs, the addressed problems are different as the NOC focuses on incidents impacting the performance and availability of an organization's network [36], [72]. As incidents can occur on all systems not just networks, it is beneficial for organizations when the NOC and SOC teams work together.
- **Security Intelligence Center:** The term Security Intelligence Center (SIC) was first used in 2017 to describe the successor of SOCs. It aims to provide a more holistic, integrated view than a SOC and can fully visualize and manage security intelligence in one place [24]. Therefore, several technologies (e.g. Information Security (IS) knowledge management, big data processing) are combined [39].
- **Security Information and Event Management:** SIEM is an integral part of many SOCs to cover a large part of the technological requirements. It is responsible for collecting security-relevant data in a centralized manner. Thereby, it provides security analytics capabilities by correlating log events. Further functionalities enable enrichment with context data, normalizing heterogeneous data, reporting, and alerting [73]. To allow the exchange of threat information, SIEM provides a connection to cyber threat intelligence exchange platforms, and it involves human security analysts by offering visual security analytics capabilities. It includes log management capabilities by long time storage of event data.

While analyzing literature for this section, we saw the lack of a commonly agreed-upon definition for a SOC. Definitions vary widely, making it quite hard to get a grasp of what a SOC is. Additionally, a SOC takes on different responsibilities depending on the technology landscape and maturity of the organization. To ensure a clear definition of the term SOC in our work, we define our understanding of a SOC stemming from and summarizing the analyzed literature in the following paragraph:

The Security Operations Center (SOC) represents an organizational aspect of an enterprise's security strategy. It combines processes, technologies, and people to manage and enhance an organization's overall security posture. This goal can usually not be accomplished by a single entity or system but rather by a complex structure. It creates situational awareness, mitigates the exposed risks, and helps to fulfill regulatory requirements. Additionally, a SOC provides governance and compliance as a framework in which people operate and to which processes and technologies are tailored.

B. ARCHITECTURE

This section gives an overview of architectural design approaches for SOCs, which we identified within relevant SOC literature. The first part (Section IV-B1) summarizes three different general architectural approaches applied to SOC designs throughout the literature. The second part of this section (Section IV-B2) goes into more detail about specific architectures proposed throughout the years and describes the most influential ones.

1) OVERALL ARCHITECTURE

SOCs can either be structured as centralized, distributed, or decentralized entities on a high and abstract level. In the case of SOCs, a centralized architecture describes the approach where all the data is sent from different locations or subsidiaries to one central SOC for further processing [4], [34].

A distributed SOC, on the other hand, resembles one single system operating across several subsidiaries [6], [40]. It appears for users as if they are dealing with one entity. The distributed system enables all entities to retrieve, process, combine and provide security information and services to other entities [41], [42]. It allows for spreading the workload and data evenly.

The third overall architectural design for SOCs is a decentralized system, a combination of the two system designs mentioned above [39]. A decentralized SOC comprises a few SOCs with possibly limited capabilities reporting to one or more central SOCs. A shift from having one central SOC to a more decentralized architecture is observed when comparing earlier research with more recent publications. The main reason for this seems to be to avoid a single point of failure.

2) TECHNOLOGICAL ARCHITECTURES AND DESIGNS

A SOC is an organizational unit encompassing different functionalities and not just one single system. One of the first architecture models for SOCs is the SOCBox proposed by Bidou *et al.* [4], [34] and evaluated by Ganame *et al.* [43]. SOCBox defines a SOC as composed of five main modules: event generators, event collectors, message databases, analysis engines, and reaction management software.

Although the SOCBox architecture is still relevant regarding its main components, it has certain limitations as it was proposed almost 15 years ago, and technology has advanced considerably. SOCBox primarily focuses on data collection and incident management but fails to include digital forensics and reactive capabilities to prevent attacks. Moreover, the proposed architecture describes a centralized system with numerous single points of failure. Due to the complexity of modern IT landscapes and technological developments, distributed architectures are often deemed to be more appropriate [6], [41]. Therefore, the SOCBox architecture has undergone several iterations and was improved throughout the years. Its direct successor is the Distributed SOC (DSOC) proposed by the same group of authors [6].

The DSOC architecture lays the basis for the distributed Grid SOC (GSOC) architecture for critical infrastructures, which again is developed by the research teams starting the work on the original SOCBBox [40]–[42]. These three architectures highlight the shift from centralized to distributed SOC setup over time. The original SOCBBox architecture [4] was also used by Miloslavskaya [39] to design a modern SOC for big data processing.

Radu [3] states that a SOC architecture consists of a generation layer, an acquisition layer, a data manipulation layer, and an output or presentation layer. This more abstract approach to defining a SOC's technological architecture using only very few building blocks can be found in several works [30], [44]–[46]. These publications conclude that a SOC consists of similar architectural blocks: a block that summarizes the data sources, followed by a block designed to collect the data from the sources and hand it to a third block responsible for analyzing the data. The last block describes the presentation of the data analysis results. None of these blocks makes any assumptions, whether done manually or automatically.

We also identified further proposals of SOC architectures within the relevant literature, focusing on SOCs for specific use cases. Settani *et al.* [47] describe the implementation of a SOC architecture for critical infrastructure providers. Tafazzoli and Grakani propose an architecture for processing events in an OpenStack environment to detect attacks in the cloud on a very superficial level [48]. There is a wide variety of other, very specific, and domain-tailored SOC architectures [49]–[61], [74].

C. OPERATING MODELS & INFLUENTIAL FACTORS

There are numerous ways of operating a SOC. Broadly speaking, a SOC can be operated internally or externally [7], [25], [62], [63]. However, various other and more specific classifications exist. Schinagl *et al.* [2] propose clustering the different operating models based on the SOC's organizational placement and its functionality, such as an integral, a technology-driven, a partly outsourced, and a specialized SOC. A different approach to classify SOC operating models is taken by Zimmerman *et al.* [75] and adapted by Radu *et al.* [3]. They use a combination of size, authority, and the organizational model and propose to divide SOCs into five different operating models: virtual SOC, small SOC, large SOC, tiered SOC, and national SOC. Another clustering of SOC operating models applies four main categories: dedicated, virtual, outsourced, and hybrid SOC [76]. Independently of the operating model of a SOC, it has to be secured itself. A failing SOC leaves the whole rest of a company vulnerable as attacks might spread undetected. Therefore, special attention must be paid to the security of a SOC [65], [66].

Each operating model has certain advantages and disadvantages, and it is essential to come to a decision upfront. Changing the SOC structure after setting it up will require a considerable amount of time and resources [64], [77], [78].

However, the choice between SOC operating models is not a trivial task, and the implications of this choice should be thoroughly considered. The literature identifies various factors which influence this choice:

- **Company strategy:** The overall business and IT strategy should be consulted to determine which operating model fits best [76]. A SOC strategy should be defined before selecting the respective operating model [75].
- **Industry sector:** The industry sector in which a company mainly operates largely influences the scope of the SOC required [7], [76].
- **Size:** The size of a company also has an impact on the decision, since a small company might not be able to set up and run a SOC on their own [67], [68] or might not even require a rigorously defined SOC [3], [25].
- **Cost:** The costs of internally implementing and maintaining a SOC must be compared with the costs of outsourcing security operations [64]. Initially, deploying an in-house SOC might be more expensive [78], but such an option might turn out to be more cost-effective in the long term. Costs of finding, hiring, and training SOC staff constitute a significant factor, especially since they might increase due to growing skill-shortage and increasing market demand [3].
- **Time:** It takes a considerable amount of time to set up a SOC. Therefore, alignment with organizational plans and timelines is necessary. Additionally, the time to set up a SOC should be compared to the time needed for outsourcing it.
- **Regulations:** Depending on the industry sector, different regulations must be considered. Some might enforce the implementation of an operational SOC [25], others might forbid the outsourcing of SOC operations altogether, or at least to specific providers who do not comply with the respective regulations [64].
- **Privacy:** Privacy also falls under regulation and must be respected whenever dealing with personal data [3].
- **Availability:** Availability requirements should be considered [68]. Most of the time, the goal is to have a SOC operational 24/7, 365 days a year [46], [78].
- **Management support:** Management support is of crucial importance when setting up a dedicated SOC. If management is not committed and benefits of a SOC are not communicated to upper management, the team might not get the resources needed [33].
- **Integration:** The capabilities of an internal SOC need to be integrated with other IT departments [7], [63], whereas, in an external SOC, the provider needs to be integrated to get all the data needed.
- **Data loss concerns:** The SOC is most often a central place where a substantial amount of sensitive data is processed. Internal SOCs need to be highly secured, while for external SOC a trusted provider must be selected, who can ensure that the data is secured against intellectual property theft as well as accidental loss [64], [78].

TABLE 4. Identified literature for the topic *People*.

People	References
Roles & Responsibilities	[8], [14], [46], [54], [66], [79]–[81]
Recruitment & Retention	[10], [15], [32], [82]–[93]
Training & Awareness	[14], [88], [89], [93]–[96]
Collaboration & Communication	[8], [11], [12], [17], [23], [47], [97]–[99]

- **Expertise:** It takes time and money to build up expertise. The required skills for operating a SOC are not very easy to find [63], [64]. Recruitment and retention (see also Section V-A2) of personnel is a crucial factor for internal SOCs. However, the necessary skills are already present for external SOC providers. Especially in the context of SOCs, having an insight into different companies might give SOC providers a knowledge advantage [67], [68]. However, companies should be aware that outsourcing reduces in-house knowledge [3].

With this list of important factors influencing a specific SOC's operating model decision, we conclude the *General Aspects* of SOCs identified in academic literature.

V. BUILDING BLOCKS

The second part of our main contribution now focuses on the main building blocks of a SOC. We structure this part of the work following the previously described PPTGC framework. The framework translates into defining processes to optimize operations, implementing the right technology to make work more efficient, and hiring the right people with the right skills to run the processes. Therefore, the framework allows us to define a SOC and its components cohesively. We also include a dedicated section to the aspect of governance and compliance within the SOC.

A. PEOPLE

Following the PPTGC framework, we first look at the people involved in a SOC. Literature allows us to derive the various roles and responsibilities involved in running a SOC. Another important aspect discussed in related literature is the recruitment of personnel and various retention methods. Third, the importance of training and awareness programs is outlined, and fourth, collaboration and communications procedures within a SOC are identified. The relevant literature for each of these subtopics can be found in Table 4.

1) ROLES & RESPONSIBILITIES

Just like in every other organizational unit, there are several different roles and responsibilities within a SOC. Depending on scope and size, different teams are needed in different numbers. Typical core roles in a SOC are different tiers of analysts as well as dedicated managers. Based on the identified work, we derive three roles with respective responsibilities [8], [54], [66], [75], [80], [81], [100], [101]:

- **Tier 1 (Triage Specialist):** Tier 1 analysts are mainly responsible for collecting raw data as well as reviewing alarms and alerts. They need to confirm, determine, or adjust the criticality of alerts and enrich them with relevant data. For every alert, the triage specialist has to identify whether it is justified or a false positive. An additional responsibility at this level is the identification of other high-risk events and potential incidents. All these need to be prioritized according to their criticality. If occurring problems cannot be solved at this level, they are escalated to tier 2 analysts. Furthermore, triage specialists are often managing and configuring the monitoring tools.
- **Tier 2 (Incident Responder):** At tier 2 level, analysts review the more critical security incidents escalated by triage specialists and do a more in-depth assessment using threat intelligence (Indicators of Compromise, updated rules, etc.). They need to understand the scope of an attack and be aware of the affected systems. The raw attack telemetry data collected at tier 1 is transformed into actionable threat intelligence at this second tier. Incident responders are responsible for designing and implementing strategies to contain and recover from an incident. If a tier 2 analyst faces major issues with identifying or mitigating an attack, additional tier 2 analysts are consulted, or the incident is escalated to tier 3.
- **Tier 3 (Threat Hunter):** Tier 3 analysts are the most experienced workforce in a SOC. They handle major incidents escalated to them from the incident responders. They also perform or at least supervise vulnerability assessments and penetration tests to identify possible attack vectors. Their most important responsibility is to proactively identify possible threats, security gaps, and vulnerabilities that might be unknown. As they gain reasonable knowledge about a possible threat to the systems, they also should recommend ways to optimize the deployed security monitoring tools. Also, any critical security alerts, threat intelligence, and other security data provided by tier 1 and tier 2 analysts need to be reviewed at this tier.
- **SOC Manager:** SOC managers supervise the security operations team. They provide technical guidance if needed, but most importantly, they are in charge of adequately managing the team. This includes hiring, training, and evaluating team members, creating

processes, assessing incident reports, and developing as well as implementing necessary crisis communication plans. They also oversee the financial aspects of a SOC, support security audits, and report to the Chief Information Security Officer (CISO) or a respective top-level management position.

Each of these core roles is required to have a specific skill set. We summarize the identified skill sets very briefly within Figure 3. The core roles can be found in SOCs independent of their size. However, in a smaller SOC, each role's responsibilities are broader, and they are narrowed down to be more specific when the SOC grows. For example, in a small SOC with only a few analysts, everyone needs to be knowledgeable on several skills because a few employees need to cover all the arising tasks. In a bigger SOC, roles can be more specific as, for example, some analysts might be focused on network monitoring while others are experts for Windows or Linux specifics. This comes with many advantages, such as a better and faster response to threats or better separation of tasks.

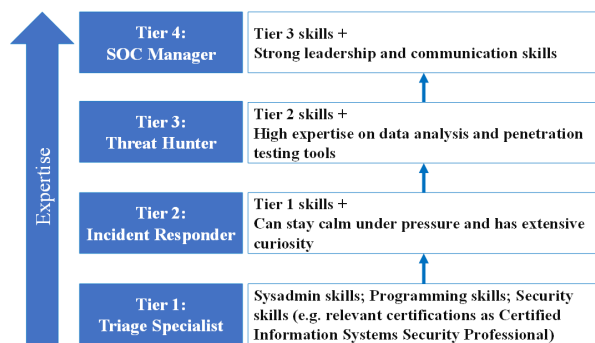


FIGURE 3. Necessary skills among SOC roles [54], [66], [75], [100], [101].

Besides the four already described essential roles, we identified additional roles that are at least to some extent involved in the daily business of a SOC [14], [46], [75], [79]. Because of the wide variety of identified roles, it is important to attempt to structure them. We have derived a list of different roles and possible interconnections between them. Figure 4 depicts those based on Olt [79]. These additional roles need to lead, work together, or cooperate with the previously described core SOC roles, which are also included in the figure. However, substantial overlap between roles and additional roles might be included in running a specific SOC. This is why we decided to group the roles into five main groups indicated through different colors in Figure 4. These groups can be adapted or expanded with additional roles when necessary:

- **Management roles:** In the context of a SOC, we identify three critical managerial roles. First of all, the *Chief Information Security Officer* defining strategies, goals, and objectives of an organization's overall security operations. A *SOC Manager* leads the SOC itself. We already described this role upfront. Inside of the SOC, the

literature includes one additional high-level management role: the *Incident Response Coordinator*, which coordinates all activities related to incident response.

- **Technical roles:** There is a wide variety of additional security specialists who need to collaborate with the SOC analysts to allow for efficient and effective SOC operations. *Malware Analysts* help with responding to sophisticated threats by performing malware reverse engineering and creating crucial results for incident response activities. To be aware of possibly ongoing attacks, *Threat Hunters* actively look for threats inside the organization, for example, by reviewing logs or outside of the organization by analyzing available TI data. This TI data is also explicitly analyzed by *Threat Intelligence Analysts* or researchers. They analyze threat intelligence from various sources and produce input for the SOC team. If parts of an attack have succeeded, *Forensic specialists* conduct detailed investigations into them. They collect and analyze forensic evidence in a legally sound manner. *Red Teams* and *Blue Teams* actively try to attack or respectively defend the organization's systems to identify vulnerabilities, and both test as well as increase the effectiveness and resilience of security mechanisms. Finally, *Vulnerability Assessment Experts* perform research to identify new, previously unknown vulnerabilities and manages known vulnerabilities with respect to business risk. These experts create detailed technical reports with their findings and support SOC analysts or incident response teams in specified vulnerability discoveries. Another vital role of this group is the *Security Engineer (SE)*. The SE develops, integrates, and maintains SOC tools. Security Engineers also define requirements for new tools. They ensure the appropriate access to tools and systems. Additional tasks are the configuration and installation of firewalls and intrusion detection/prevention systems. Furthermore, they assist in writing and updating detection rules for Security Information and Event Management (SIEM) systems.
- **Consulting roles:** The two most important roles of this group are the *Security Architect (SA)* and the *Security Consultant*. The SA plans, researches, and designs a robust security infrastructure within a company. SAs conduct regular system and vulnerability tests and implement or supervise the implementation of enhancements. They are also in charge of establishing recovery procedures. Security consultants often research security standards, security best practices, and security systems. They can provide an industry overview for an organization and compare current SOC capabilities with competitors. They can help to plan, research, and design robust security architectures.
- **External personnel:** External personnel can be included in any SOC operation, and therefore, depending on the architecture and operating model of a SOC, more or less external personnel are involved in the different SOC roles and groups.

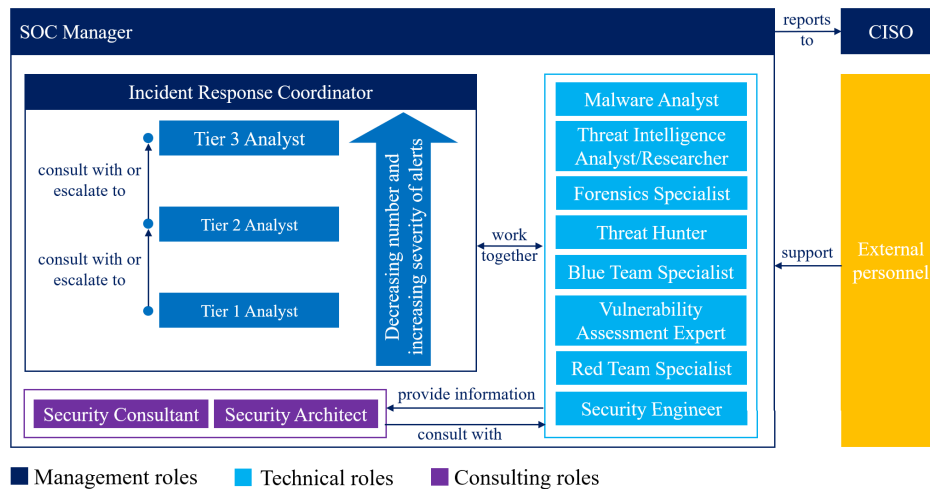


FIGURE 4. Interaction of different roles within a SOC [79].

Besides technical skills, soft skills are becoming more and more important. Desired skills include communication skills, continuous learning abilities, analytical mindset, ability to perform under stress, commitment, teamwork, curiosity, and practical organizational skills [75]. The significance of relevant soft skills grows with the level of responsibility an individual has within a SOC. Besides hard and soft skills, there is a number of useful certifications for SOC employees depending on their level, which are summarized by DeCusatis *et al.* [80].

2) RECRUITMENT & RETENTION

The people working in a SOC are the last line of defense and responsible for detecting and successfully mitigating attacks. Thus, having skilled human resources in an adequate quantity is imperative for the success of a SOC [32]. However, finding and retaining the right staff is not an easy task. The International Information System Security Certification Consortium ((ISC)²) puts the current cybersecurity workforce gap at roughly four million people on a worldwide scale, and it is still growing [102]. Therefore, recruiting new, skilled staff for SOC is getting increasingly difficult. There is little to no literature about how to specifically recruit SOC staff. Most of the relevant papers focus on retaining SOC staff and closing the skills gaps with automation.

Working in a SOC is very demanding and can be extremely stressful. Anthropological studies found that SOC analysts are often not satisfied with their job [15], [16]. They are overloaded with mundane, tedious tasks, and the currently deployed tools are not sophisticated enough to automate these tasks [82]–[84]. SOC analysts' primary responsibility, especially at tier 1, is to follow Standard Operating Procedures (SOPs), also called playbooks. This negatively impacts their creativity, growth, skills, and empowerment. Literature reveals a vicious cycle, which ultimately causes

analyst burnout in a noticeable number of cases [15], [16]. Therefore, companies should take action to increase the job satisfaction of their SOC staff. Several methods to counteract staff burnout and increase job satisfaction can be determined:

Increase Automation: Increasing automation helps decrease the amount of mundane and boring tasks [83], [84]. This can be achieved with more efficient and helpful tools deployed within the SOC. Analysts should be consulted before buying and implementing tools, and they should be engaged in the development of new tools. New possibilities for automation can be discovered by analysts themselves if they have time to reflect on their daily work [16], [85]. Technology should amplify the human capacity to be creative and apply critical thinking to solve problems. Examples are studies analyzing data triage tasks and trying to optimize the process [86]–[89].

Increase Operational Efficiency: Automating specific tasks can also help to increase operational efficiency. Additional improvements can be made by streamlining processes, ensuring that analysts have access to the data they need, and providing team communication and collaboration possibilities. An example is the preferably optimal prioritization of alerts, so analysts can focus on the most critical ones [90], or the adaptive reallocation of analysts based on the current needs [91].

Invest in Human Capital: Security professionals working in a SOC need to possess the right skills to perform their job correctly, as described above. Investing in their skills will not only contribute to their personal well-being but also benefit the company itself [92]. Skills can be enhanced by in-house or outsourced training, conference participation, observation of more senior staff, or even learning-by-doing. The more skills employees master, the more likely they are to be empowered. This empowerment enables employees to do their job efficiently

and increases their morale [16]. Gaining skills and feeling empowered, in turn, has a positive effect on the creativity of analysts. Ultimately, employees grow and increase their intellectual capacity, are empowered, and more likely to be creative. If a positive causality among the personal development factors exists, SOC staff will be gratified [16], [93]. Unfortunately, it is not always possible to exactly meet employees' expectations. Technological limitations require personnel to sometimes do tedious tasks, and budget restraints might hinder staff from going on training. Other incentives, like a competitive salary, monetary bonus, team-building or after-work activities, flexible and competitive working hours, respect, and recognition, can also play a role in keeping up the SOC staff's morale.

3) TRAINING & AWARENESS

Well-trained employees are more productive because they understand their responsibilities and tasks. Training strengthens their skills and addresses potential knowledge gaps. The quality and consistency of the work also increases [93]. Furthermore, training benefits an organization itself because employees are less likely to make mistakes. A study conducted by Accenture and the Ponemon Institute revealed that employee training could decrease the total cost of a cyber breach by about 270.000 USD [1].

For junior staff members, training is a means to equip them with the technical and soft skills required to perform well in their job. Training for juniors has a broader scope and aims to provide them with an overview of various security-related topics. For example, for a SOC tier 1 analyst, training could be given in real-time analysis, incident analysis and response, scanning and assessment, alert correlation, and many more. For more senior staff, training should be more tailored to their specific role in the SOC as employees working in a SOC are very likely specialized in specific tasks.

In general, training should consist of a mix of formal training, internal training, vendor-specific training, and on-the-job learning. Formal training is a form of structured training with predefined goals and objectives. Internal training is often taught by other team members and of a more informal nature. Thus, there is a less strict plan and internal training is more dynamic.

Vendor-specific training is used to familiarize SOC staff with deployed software (e.g. a specific SIEM system). On-the-job learning or shadowing more experienced team members is another form of acquiring the necessary skills [14]. As this type of learning is very unstructured, it is following a steep learning curve. However, it might be overwhelming for new SOC employees to deal with the flood of incoming alerts without more formal training [94]. To support them, Zhong *et al.* [88], for example, developed a system that traces and models the data triage actions of senior analysts to the present actions done in a similar context. All different training approaches have several advantages as well as disadvantages. There is only very little scientific work on

SOC-specific training methods. Further research is necessary to show how different training methods can be applied in the context of SOC's and measure their effectiveness. An interesting approach to improve on-the-job learning and training is pursued by Applebaum *et al.* [95] by developing playbooks that provide analysts with an overview of tasks and actions based on the experience of other analysts. Also, knowledge graphs representing the domain knowledge and experience of SOC analysts enable better learning and training for others [89], [95]. A relatively exotic use case is considered by Sanchez *et al.* [96]. They present particular challenges for a SOC within the space domain and emphasize employee training's unique challenges.

4) COLLABORATION & COMMUNICATION

Especially in high-pressure environments like a SOC, collaboration amongst the various team members is essential [17], [47]. A few academic resources are focusing on collaboration in SOC's. Hämornik and Krasznay [8] emphasize the need for further research about computer-supported collaborative work (CSCW) to see how computer systems can support collaborative activities. The AOH-Map developed by Zhong *et al.* [97] is a collaborative analysis report system capturing and displaying the analytical reasoning process of analysts. Afterward, analysts can look at the captured process, review past decisions, share their results with others, and divide their tasks effectively. Additionally, work between analysts needs to be divided equally depending on their skills [98]. Crémilleux *et al.* [11] propose a collaboration process to create a feedback loop between tier 1 and tier 2 SOC analysts.

An upcoming trend is the operative use of visualization platforms with collaboration features, e.g., the 3D CyberCOP platform [12], [99] distinguishes explicit collaboration through the platform and implicit collaboration through oral communication and logging every user's actions. It is imperative for the SOC team's success to have constant interaction and communication with other business units, for example, the help desk, network administrators, or even the legal team. This requires ensuring the other departments that the SOC staff is not there to watch their every move but to help [23].

B. PROCESSES

This section features academic work focusing on the processes related to a SOC. We aim for a high-level perspective, as there are different, very specific processes happening in operations. Since the goal of a SOC is to respond to or prepare for incidents, one way to structure the underlying processes is through the Incident Response Lifecycle [103], [114], [119], [120] or similar frameworks such as presented in ISO/IEC 27035:2016 [123]. According to the NIST Computer Security Incident Handling Guide [124], the Incident Response Lifecycle comprises the four steps "preparation", "detection and analysis", "containment, eradication and recovery" and "Post-incident activity", which also form the structure of the following chapter.

TABLE 5. Identified literature for the topic *Processes*.

Processes	References
Preparation	[22], [55], [67], [103]–[113]
Detection & Analysis	[4], [67], [80], [83], [114]–[118]
Containment, Eradication & Recovery	[80], [83], [97], [103], [104], [114], [117]–[122]

At this point, we would like to emphasize that, in our view, the literature only allows an incomplete picture regarding processes. For example, technical processes are treated very intensively, whereas most surrounding processes are only dealt with sporadically. These aspects are to be regarded as research gaps and are presented in the following chapter accordingly incomplete, in order to go into the gaps in more detail in chapter VI. This is especially true for “post-incident activity” since no SOC specific scientific publication deals with this topic. Therefore, it will not be considered in the following descriptions.

1) PREPARATION

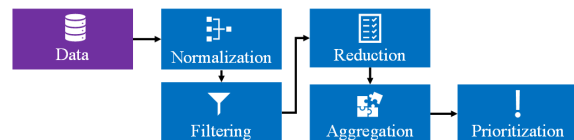
The analyzed literature mainly focuses on data collection within the topic of preparation; however, it does not give a uniform picture of which steps the data collection process is composed. However, as illustrated in Figure 5, the steps normalization with time synchronization [22], [55], [104]–[107], filtering [22], [55], [105], [106], [108], reduction [22], [109], aggregation [22], [55], [106], [109], [113] and prioritization [22], [55], [67], [103] or risk evaluation [110] were most frequently mentioned. The order of process steps is not uniform in literature, as this can vary depending on the application used. However, it is mostly described in the presented sequence. The identified process steps are explained in more detail to provide a general understanding:

Normalization: It is vital to translate the heterogeneous data formats into a uniform representation to conduct further processing. It is also essential to change all time data to one standard time zone and format [22], [77]. Synchronization helps avoid confusion in the timeline of the security events and reduces the likelihood that erroneous conclusions are made on inconsistently measured network activity. In literature, normalization is often referred to as log parsing or pre-processing.

Filtering: Since systems typically generate enormous amounts of data, it is essential to filter for data elements that are likely to contain important information from a security perspective [125].

Reduction: Reduction is like filtering, with the difference that individual, unimportant data fields are sorted out to reduce the amount of data.

Aggregation: Similar events are combined into one single data element. For example, three log entries, which indicate a log attempt to a host, could be aggregated to one single log, which states the type and number of login attempts [125].

**FIGURE 5.** The data collection process.

Prioritization: Each log data should be classified according to importance to facilitate further processing. For example, to decide how to react to events or how long the logs should be stored, it is useful to prioritize incoming data.

Considering literature about data collection specifically for SOCs, there are only two notable papers: [111] and [22]. This is probably because most SOCs deploy a software solution responsible for collecting, processing, analyzing, and displaying events and alerts [112] and thus data collection is addressed in a more technical context. Bridges *et al.* [111] conduct interviews with 13 professionals from five different SOCs to discover the current state-of-the-art and future directions for host-based data collection. They evaluate what and how host data is collected, which tools are used, and whether dynamic collection (dynamically decide how much and which data is collected depending on factors such as security posture) is used. Their major takeaway is that analysts desire a wider, less manual collection of data, but only with the right toolset to understand and work with the data. Madani *et al.* [22] propose a logging architecture for SOCs. Their architecture contains log generators, a collection server, a storage server, and a log database. The authors list SIEM vendors incorporating log management in their SIEM solution and outline their weaknesses. Normalization, filtering, reduction, rotation, time synchronization, aggregation, and integrity check are the most important functionalities. Madani *et al.* [22] underline the importance of log collection and management. However, since the paper was published in 2011, there have been no SOC specific advances in the field.

2) DETECTION AND ANALYSIS

The sheer amount of data collected in previous steps can be overwhelming, even for seasoned security practitioners and researchers. Turning this data into useful information is done through data analysis and is essentially a means to make sense of what is collected. Regarding automatic analysis and detection, the identified literature mainly focuses on specific

analysis and detection methods and technologies. However, only a few papers look at the subject area from an abstract, process-driven perspective. The following process steps were identified by merging available processes [73], [114] and by sequencing individually named steps within the stated literature. This results in a process which is comprised of the steps *Detection* [83], [114], *Analysis* [4], [115], [116], and *Alert Prioritization/Triage* [67].

- **Detection:** Incidents are detected with the help of humans or by automatic procedures. Thereby, it must be decided if the collected data indicates a security incident [114]. A more technical description of the identified detection approaches can be found in Section V-C2.
- **Analysis:** Regarding the techniques used for analysis, one can distinguish between source and target correlation, structural analysis, functional analysis, and behavior analysis [4]. Thereby, the authors describe the purpose of correlation as to enable the analysis of complex sequences by producing simple, synthesized, and accurate events.
- **Alert Prioritization/Triage:** Alert prioritization, also known as triage, can be seen as a link to containment, eradication, and recovery. It serves two primary purposes. First, to ensure that the most severe incidents are treated with priority, and second, to ensure that incidents are distributed for further processing according to available resources [67].

3) CONTAINMENT, ERADICATION, AND RECOVERY

The activities in containment, eradication, and recovery are described by Bhatt *et al.* [104] on a high level. This step aims to decide whether an incident is an unharmed event (e.g., during penetration testing), or a harmful event. In the case of a harmful incident, it is passed on to appropriate stakeholders to take further steps. In this context, Security Orchestration, Automation, and Response (SOAR) is of great importance and can be identified as a very active research area of the last two years [83], [118], [122]. According to Islam *et al.* [122] the key purpose of SOAR is the automation of processes through orchestration. The functionalities of SOAR are mainly categorized into integration, orchestration and automation. Security orchestration is a prerequisite of security automation, which is the process of automatic detection [117]. Therefore, SOAR integrates available information about security incidents (Cyber Threat Intelligence) [121] to automatically take appropriate measures to limit the damage as quickly as possible. Islam *et al.* [122] conducted a detailed survey on this topic.

A straightforward framework to tackle incidents is the Observe, Orient, Decide, Act (OODA) loop, which is a well-known analytical framework for decision-making developed by John Boyd [126]. It can be applied to incident management in the context of a SOC, as demonstrated in research [80], [97] (or similar to the Plan, Do, Check, Act loop [120]). In SOC literature [103], [114], incident management is mentioned mostly related to the incident handling

lifecycle. Thus, the Alert and Incident Management process presented in Figure 6 comprises the process steps identified by two primary standards for information security incident management [123], [124].

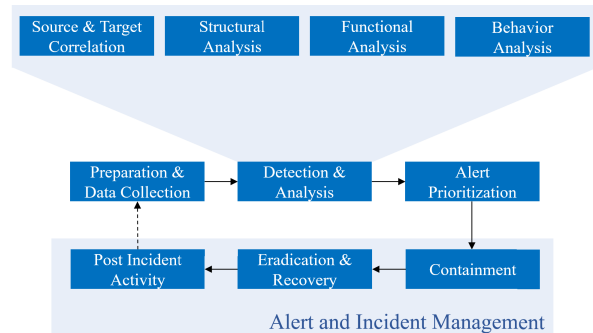


FIGURE 6. The SOC incident analysis, detection and management process.

A more detailed description of these process steps concerning SOC cannot be found in the analyzed literature, which is why the standards mentioned above must be referred to if necessary. The reason for this could be that employees know which tasks they have to carry out, but this has not been specified explicitly, which can cause problems, e.g., when staff changes. Therefore, Cho *et al.* [119] conducted a study where they show how it is possible to capture SOC staff's tacit knowledge on how they perform their tasks as processes.

C. TECHNOLOGY

This section discusses the technologies combined in a SOC. It covers the process steps from Section V-B from a technical point of view, whereby Containment, Eradication, and Recovery is not considered, as we did not find any literature dealing with SOC-specific technology covering this process step (see Table 6).

We first take a look at data collection technologies which support the preparation process mentioned in Section V-B1. Every organization should determine which devices should be monitored, what data needs to be collected, and in which format it should be stored. Moreover, depending on the data, the retention period of the data needs to be set. We then shed light on the applied methodologies and approaches to analyse data, detect threats and present the results, which can be mapped to the process detection & analysis (Section V-B2). As the interface between people and machines, the presentation of data and analysis results is of particular interest in a SOC context.

1) DATA COLLECTION

Various data collection techniques exist and can generally be classified into four categories: push/pull, distributed/centralized, real-time/historical and partial/full collection. Data can either be pulled by the data collector or pushed onto the data collector from the data source itself [77]. Furthermore, it can be collected in a centralized log collector (e.g. [171]) or

TABLE 6. Identified literature for the topic *Technology*.

Technology	References
Data Collection	[37], [47], [80], [103], [104], [107], [111], [127]–[132]
Analysis & Detection	[13], [35], [41], [43], [55], [56], [84], [133]–[157]
Presentation	[9], [12], [13], [80], [97], [99], [112], [127], [158]–[170]

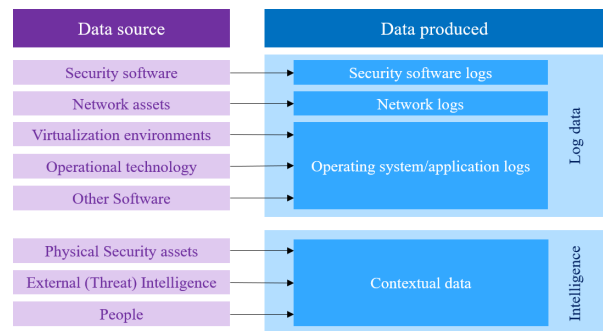
in a distributed topology (e.g. [172]) over different sub-nodes. Thereby, data can either be captured fully or partially.

Within the identified literature, data collection mainly relates to identifying data sources that capture relevant security-related information. While new data sources are continuously being created, the most common sources, its classification [127], [173], [174], and corresponding examples are:

- **Security software:** SIEM systems [80], intrusion detection/prevention systems [37], [103], [107], [128], [162], [173], [174], firewalls [37], [104], [127], [128], [174], anti-virus software [37], [111], [127], vulnerability scanners [173], identity and access management [104]
- **Network assets:** Switches [104], [173], routers [104], [128], [173], servers [104], [127], [173], hosts [104], [173], proxies [174]
- **Virtualization environments:** Hypervisor, virtual machine introspection, cloud environments [80]
- **Operational technology:** Sensors, actuators, PLCs
- **Other Software:** Open-Source Big Data Analytics [80], databases [173], identity and access management [173], mailserver [174], operating systems [111], [174]
- **Physical security assets:** Security cameras, access control
- **External (Threat) Intelligence:** Geolocation and DNS lookup [80], open source intelligence (OSINT) [47], [129], intelligence from threat sharing platforms or other organizations [130]–[132]
- **People:** Employees (Human-as-a-Security-Sensor [175]), external users.

Each of these data sources can deliver a vast amount of information, of which not all is relevant. Capturing everything may help in spotting malicious activity, but it can also negatively impact system performance. Conversely, if fewer data sources are used to collect data, an attack might go undetected. Thus, finding the right balance between capturing too much and capturing too little data is essential when designing a SOC's technological capabilities. However, as a rule of thumb, it is generally better to capture data from as many sources as possible (under performance constraints) and then rely on well established data normalization, correlation, and analysis mechanisms.

Depending on the data source, the data type collected may vary as illustrated in Figure 7. All collected data can be broadly classified into either log data or intelligence. Logs document the current state of the system and usually record all the changes occurring within the system. Logs are generally divided into operating system/application logs

**FIGURE 7.** Data sources and the type of data they produce.

and security software logs [125]. Network logs proposed by Zhiguo *et al.* [176] can be added since they have unique features and cannot be categorized perfectly into log categories. Operating systems and applications often provide data in the form of logs. These logs give the user information on system events such as the shutdown or start-up of a service, audit records, client requests and server responses, account information, usage information, etc. Security logs instead display suspicious activities, results of virus scans, etc. [125]. Intelligence provides additional context for threat analysis.

2) ANALYSIS & DETECTION

Attack detection is performed either automatically or manually. Manual detection is the detection of an incident through an internal or external person. Thereby, the detection can be performed by security experts such as analysts within the SOC or by security novices. The different roles and tasks of security experts are further discussed in Section V-A.

An example of manual detection through security novices would be if an employee receives a phishing mail and then reports it, so the security team can take appropriate measures. The concept of integrating employees into the detection process was introduced as “human-as-a-security-sensor” [175], [177] and means that employees are enabled to detect and report security incidents. Therefore, awareness training plays a crucial role as further discussed in Section V-A3. All in all, manual detection is necessary, because not all attacks can be detected through technology, especially when it comes to advanced attacks. However, automated detection cannot be neglected, because the sheer amount of data would overstrain humans. The topics of manual detection related to presentation are discussed in Section V-C3.

TABLE 7. Classification of literature with respect to applied detection methodologies and approaches.

	Detection methodologies			Detection approach classes				
	Anomaly	Signature	Specifi- cation	Statistics	Pattern	Rule	Heuristic	State
[13]	✓			✓				
[35]	✓			✓				
[41]		✓			✓	✓		
[43]	✓	✓			✓	✓		
[55]		✓			✓			
[56]	✓			✓		✓		
[84]		✓					✓	
[134]		✓			✓		✓	
[135]		✓		✓		✓		
[136]	✓	✓		✓	✓	✓	✓	
[137]		✓			✓	✓		
[138]		✓		✓	✓		✓	
[139]	✓			✓		✓		
[140]		✓		✓			✓	
[141]		✓		✓		✓	✓	
[142]		✓				✓		✓
[143]	✓			✓		✓		
[144]		✓			✓	✓		✓
[145]		✓		✓				✓
[146]	✓				✓			
[147]		✓					✓	
[148]	✓					✓	✓	
[149]	✓	✓				✓		✓
[150]		✓		✓		✓		
[151]	✓	✓		✓	✓			
[152]	✓					✓		
[153]	✓	✓		✓				
[154]		✓				✓		
[155]		✓		✓			✓	
[156]	✓			✓			✓	
Σ	14	21	0	16	10	16	10	4

Regarding automatic analysis and detection, the identified literature mainly focuses on specific analysis and detection methods and technologies. To show the state-of-the-art analytical methods, those mentioned in the literature are classified in Table 7. Therefore, a well-accepted classification scheme of Liao *et al.* [178] was used. It distinguishes between detection methodologies and detection approaches.

Anomaly-based or behavior-based methodologies use the system's normal behavior as a foundation and try to detect deviations. *Signature-based* or also knowledge-based methods use accumulated knowledge of attacks and is very useful to detect known attacks or exploitation of known system vulnerabilities. Therefore, it is important to regularly update the knowledge base. *Specification-based* methodologies focus on detecting incidents based on predefined profiles or protocols. Hybrid methodologies use a mixture of the three described detection methodologies.

Concerning detection approaches, *statistics-based* detection is one of the oldest methods used for intrusion detection and uses statistical properties and statistical tests like mean, median or variance, to detect deviation between the normal

behavior and observed behavior. Threshold metrics, hidden Markov models and multivariate models are examples of statistical based detection approaches. *Pattern-based* and *Rule-based* approaches use either predefined patterns, learned patterns or rules for detection. An example for rule-based detection are support vector machines. *Heuristic-based* approaches are inspired by biological concepts as for example artificial neural networks. *State-based* approaches try to infer the behavior of attacks within the network for example by utilizing finite state machines.

Table 7 shows, that all used detection methodologies are either anomaly- or signature-based. In none of the analyzed papers, the potential of specification-based incident detection was leveraged. In contrast, each detection approach class can be assigned an approach described in the literature, whereby a focus on statistics- and rule-based approaches is recognizable. To enhance detection independent of the utilized approach Karaçay *et al.* [133] propose a principle that allows intrusion detection even when end-to-end encryption was used and Smith [157] suggests that user behaviour analytics (UBA) should be used more intensively, since misused credentials are a great threat.

TABLE 8. Identified literature for the topic *Governance & Compliance*.

Governance & Compliance	References
Standards & Guidelines	[3], [30], [36], [60], [179]
Security Audits & Maturity Assessments	[2], [5], [63]
Metrics	[23], [30], [46], [57], [68], [81], [85], [163], [180]–[186]

3) PRESENTATION

From a technological view, most identified publications focus on specific visualization tackling problems related to SOCs. They are briefly outlined in the following. DeCusatis [80] describes an attack visualization based on force diagrams and hive plots. Settani *et al.* [158] shows how a map and dashboard-based visualization of incidents and a mobile visualization enables on-site personnel to make qualified decisions. Besides, Erola *et al.* [159] present an approach that combines machine learning and information from business processes with visual analytics to guide SOC employees through the decision-making process. Similarly, Sopan *et al.* [9] aim at visually supporting SOC analysts by automating decision-making using a machine learning model. However, they also present the model visually to enable the machine learning model's decisions to be understood. The Situ platform [13] has the goal to visualize the context of an incident for leveraging the experience of security experts. In contrast to the approaches described above, the CyberCOP [12], [99], [160] platform relies on three-dimensional visualization. The VISNU project [112], [161], [162] takes a similar approach, which improves the collaboration of multiple SOCs in different organizations by displaying network data in three dimensions. Thereby, they aim at the collaboration of multiple analysts in one environment by providing different views on the same incident. The concept of mind maps is leveraged by the AOH-Map [97] software, which visualizes all the identified traces of an attack to exchange it with collaborating analysts. Hassell *et al.* [163] combine network simulation with its visualization for optimizing its resilience against threats. Payer *et al.* [164] rely on Virtual Reality (VR) to analyze threats, allowing new types of interactions. To enhance tactical situational awareness within a SOC Mullins *et al.* [170] describe three suitable visualizations.

Starting 2018, increasing interest in sonification and its potential for SOCs can be identified [165] as it was implemented within the SIEM system of a SOC [166]. This showed that humans can detect attacks by listening to network traffic [127], [167] in specific contexts [168].

A fairly new approach to SOC is data presentation using storytelling presented by Afzaliseresht *et al.* [169]. This involves translating the analysis results into a narrative story containing more or less details depending on the users' level of knowledge. In a SOC setting within a research institution, this approach is advantageous in terms of cognitive load.

D. GOVERNANCE AND COMPLIANCE

The following section discusses the governance and compliance aspect of a SOC (see Table 8). IT governance is responsible for ensuring the effective and efficient use of IT systems by providing a strategic direction, developing standards, policies and procedures, and implementing them. Compliance ensures that companies adhere to external rules, for example standards and regulations and internal rules, for example policies and procedures. Additionally, compliance is essentially the feedback loop of security governance, because it shows how governance rules are applied in practice. The following section will look at three aspects of governance and compliance: how security audits are performed, current metrics in a SOC and standards and guidelines related to SOCs. It should be noted that metrics play a major role in maturity assessment, so the two sections partly overlap.

1) STANDARDS & GUIDELINES

Today, many organizations are struggling to decide whether they need a SOC, which kind of SOC they need, and what components their SOC should have. There are no renowned holistic SOC standards or industry specific guidelines to help companies with their decisions [3]. However, a SOC can help to ensure that certain compliance regulations are met [30], [179] and many of the standards focus on one domain or task within a SOC. We provide a list of these standards in Table 9.

Another noteworthy standard is provided by the European Telecommunications Standards Institute (ETSI) [187] providing guidelines for building and operating a secured SOC. It mainly focuses on requirements to be met by the service provider operating a SOC for the telecommunication industry. Some private organizations have started to provide companies with best practices and recommendations, for example by conducting a survey [188]. There is only very little work on establishing best practices for a SOC [36], [60].

2) SECURITY AUDITS & MATURITY ASSESSMENTS

A SOC can help companies in conducting internal and external IT (security) audits. In an IT audit, the IT infrastructure, policies, and procedures are examined and evaluated. Independent and unbiased parties usually perform external audits. An example would be a typical year-end audit in the banking sector, which assesses the compliance of its IT capabilities against relevant standards. Depending on the type and scope of the audit, different IT capabilities are assessed. Because a SOC collects valuable log data from almost all

TABLE 9. Standards related to SOC domains or tasks.

Domain or task	Standards
Cyber Security in general	ISO/IEC 27001 and 27002, IEC 62443, ANSI/ISA 62443, NIST Cybersecurity Framework, NIST Special publication 800-12, NIST Special publication 800-14, NIST Special publication 800-26
Data Logging	DCID, FFIEC, ISO 17799, DISA, NIST SP 800-92, NIST SP 800-53, PCI DSS, FDA GXP
Incident Management	SANS Incident Handler's Handbook, ISO/IEC 27035:2016, NIST Special publication 800-83, NIST Special publication 800-61, ITIL
Business Continuity Management	ISO 22301:2012, ISO 22313:2012, ISO/FDIS 22313, BSI-Standard 100-4
Digital Forensics	ISO/IEC 27037:2012, ISO/IEC JTC 1 SC 27, ISO/IEC 27041:2015, ISO/IEC 27042:2015, NIST SP 800-86
IT Governance	COBIT, ITIL, Information Security Assurance - Capability Maturity Model (ISA-CMM)
Vulnerability Management	SANS Implementing a Vulnerability Management Process, NIST SP 800-40, ISACA Vulnerability Management
Privacy	EU-GDPR

systems, and hosts some relevant capabilities itself, it is an invaluable source of data for IT auditors. Advanced SIEM tools aggregate security information from across the company and generate reports for compliance audits. This information can be used to prove compliance with laws and regulations. Additionally, the SOC team can help determine the IT risks for the company.

Of course, the SOC itself should have controls in place, which should be audited regularly. An example for an internal SOC audit and its findings is given by NASA [189]. Due to the lack of widely accepted standards and guidelines, external assessments are not offered by independent parties. However, there is literature proposing methods to assess the current maturity of the SOC capabilities as well as the overall effectiveness of the SOC [63]. Common maturity models are compared and summarized into five capability maturity stages: non-existent, initial, repeatable, defined process, reviewed and updated, and continuously optimized [63]). In practice a similar maturity assessment approach is presented in an industry guideline from IBM [190]. Schinagl *et al.* [2] assess the effectiveness of a SOC by identifying the degree to which identified building blocks have been implemented. These approaches enable SOC owners to uniformly assess the maturity of their capabilities and

to spot the areas which still need to be improved. It also allows various companies to compare their SOC operations and benchmark against each other, if the data is made available, enabling the collaboration between SOCs. To locate collaboration areas of SOCs, a questionnaire-based approach is proposed by Kowtha *et al.* [5]. The authors describe a model for characterizing SOCs by the seven dimensions of scope, activities, organizational dynamics, facilities, process management and external interactions.

3) METRICS

Metrics are quantifiable measures used to track and assess the status of a process or system. Metrics are mainly used to support strategic decisions, to assure the quality, or to gain tactical oversight [191]. A considerable body of literature exists in the field of security metrics [192], [193], and many of those metrics can be directly applied to a SOC. However, there is very little scientific literature on how those security metrics can be used in a SOC, let alone metrics specifically covering SOCs. Ganame and Bougeois [180] propose metrics to assess the security level of different sites in a multi-site network in real-time. Their goal is to see whether threats are occurring in a network or not. Aiming to improve the resiliency of networks, Hassell *et al.* [163] test their simulation software using resiliency metrics. They criticize the lack of standardized metrics to evaluate resiliency techniques. Ganesan *et al.* [181], [194] propose an optimization model to dynamically schedule analysts and dynamically assign them to sensors to decrease total time for alert investigation and increase the Level of Operational Effectiveness (LOE). Some literature, however, comes from SOC vendors [188], [195]. Typical metrics used in a SOC include:

- **General SOC metrics:**

- **Coverage [188]:** A SOC can only monitor a limited amount of assets due to resource constraints, which raises the question of how many of them are covered. *Examples:* Number of monitored assets, coverage (number of monitored assets vs. number of assets)

- **Performance metrics:** Measurement of the performance is crucial for managing and improving a SOC. Historical performance metrics enable comparability between work-shifts or longer time periods [68]. Agyepong *et al.* [85] conducted an extensive survey about performance metrics for SOCs and proposed a consecutive framework [186]. *Examples:* False positive rate [30], [68], average analysis time [68], readiness level [81], [181], Mean Time to Detect [185]

- **People metrics:** To improve the performance of security analysts inside a SOC it is necessary to measure human activities and workflows [68]. *Examples:* Security analyst performance [68], number of incidents closed in one shift [188], workload [195]

- **Technical metrics:**

- **Threat metrics:** A threat is the potential damage posed by vulnerabilities. Thus, these metrics are closely related and, in most cases, based on

vulnerability and threat metrics. *Examples:* Security level [180], threat actor attribution [188]

- **Vulnerability metrics:** In general, vulnerabilities can be exploited by attackers or can cause a security incident. Thus, it is particularly important for SOCs to be aware of possible weak spots. *Examples:* Vulnerability exposure [182], time-to-vulnerability remediation [182], vulnerability severity [182], incidents due to known vs. unknown vulnerabilities [188]
- **Risk metrics:** Risks are in most cases assessed in real time, which is also summarized under the term situational awareness [46]. The evaluation of risks is especially important, when it comes to choosing appropriate security measures. *Examples:* Risk posture [23], [46], [183], [184], [188], risk per system [81], [180], key risks [195]
- **Alert metrics:** Alerts are in most cases generated automatically by technologies such as SIEM systems or intrusion detection systems, based on the analysis of sensor data [181]. Each alert should go through an alert analysis process [194] in order to decide upon possible measures. *Examples:* Time per alert investigation [181], alert generation rate [181], number of alerts that remain un-analyzed [81], criticality of an alert [180]
- **Incident metrics:** An incident is an occurrence, that causes harm to an organization and a SOC aims at averting incidents or reducing the caused harm. As incidents are a very central element of SOCs, appropriate metrics are essential. *Examples:* Incident priority [23], number of incidents [68], [183], [188], number of successful attacks [163], recovery time [181], costs per incident [188], mitigation success [195]
- **Resiliency metrics:** Cyber resilience is crucial, if an environment is compromised in order to continue operations with as little damage as possible [163]. *Examples:* Time spent per attack [163], defensive efficiency [163], attack noise [163], number or time of disruptions [163], [188].
- **Governance and Compliance metrics:**
 - **Compliance metrics:** Since compliance to all regulatory guidelines and standards is hardly possible, it is useful to define compliance goals and accordingly appropriate metrics. Additionally, it can be of value to provide measures for compliance audits. *Examples:* Number of policy violations [30], [57], percentage of systems with tested security controls
 - **Maturity metrics:** Usually refers to the level of maturity as described in Section V-D2

The classification is not always strict and lines are blurry. For example, some people metrics might be classified as governance and compliance metrics.

To overcome the many problems with current security metrics, a few things should be considered. It is important to clearly define what the objectives of the metrics

are and how their success/failure can be measured. Some SOC vendors use the S.M.A.R.T. management objectives framework developed by Doran [196], as a guide to develop metrics [195], [197].

VI. CHALLENGES

Throughout Sections IV and V, we focused on our first research question in terms of the state-of-the-art of a SOC. We already mentioned a series of challenges that impose the development and improvement of SOCs. Within the following paragraphs, we now briefly describe these challenges in response to our second research questions regarding the challenges needing to be solved to advance the field of SOC research. Every SOC naturally faces different challenges depending on its operating model, architecture, scope, or size. However, we derive several challenges applicable to most SOCs. Although many of the challenges are somewhat related, we try to describe them as independently as possible and along with the PPTGC framework, which we followed throughout this work. Figure 8 gives an overview of these challenges and highlights some relevant dependencies between them.

A. PEOPLE

1) MONOTONOUS AND DEMOTIVATING TASKS

As mentioned earlier, there is a vast number of alerts coming into the SOC every second. Even though tools are trying to display only true positive alerts, the number of false positives is still very high. Every incoming alert needs to be manually investigated by an analyst, most of the time at tier 1 level. The analysts need to open the alert and determine whether it is a false positive or not. Sometimes it takes seconds to come to a decision, sometimes minutes or even hours. Performing this task over and over again is very repetitive and monotonous as several works have shown previously [8], [11], [16], [32]. Additionally, this task is very demanding on a security analysts' capability of information processing and analytical reasoning due to the vast amount of data [94]. Although doing a very monotonous task, the analysts are working under high pressure and have high responsibility. Any incorrect decision can lead to unpredictable consequences for the company if an incident unfolds. This issue, combined with time pressure faced in a SOC and the lack of creativity needed to solve the tasks causes analyst boredom, which finally could lead to burnout [8], [16]. Additionally, the non-challenging nature of tasks and the fact that most analysts need to follow predefined procedures all the time limits their ability to react to new and innovative threats in the future [11]. An exciting direction for retaining SOC analysts' motivation might be the inclusion of gamification aspects into the SOC operations. When the tasks become too mundane and frustrating for the SOC employees, it is tough to retain skilled staff [30], [32]. This amplifies the next challenge in the context of people within SOCs.

2) LACK OF SKILLED STAFF AND DIFFICULT RETENTION

A very severe challenge companies will continue to face is the lack of skilled security staff [3], [8], [80]. In addition

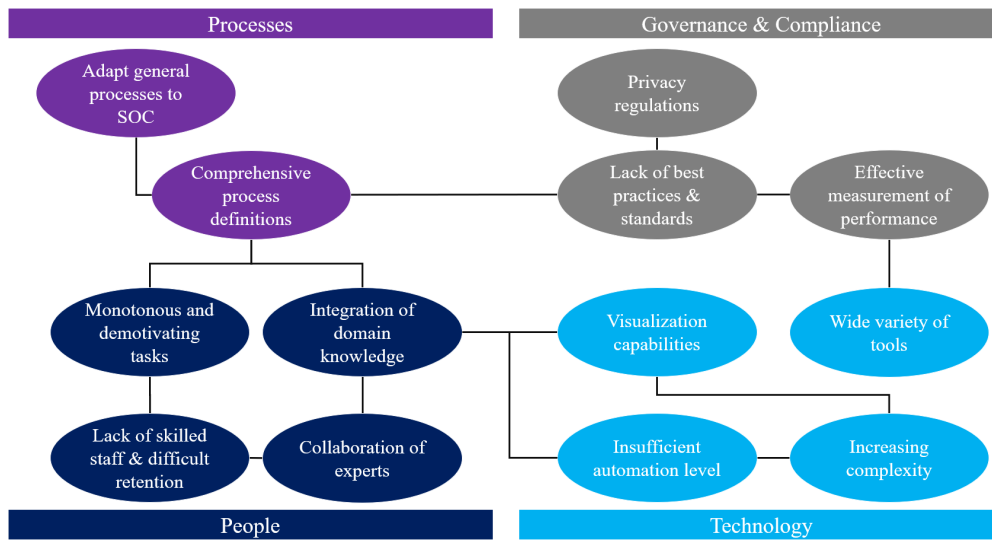


FIGURE 8. Challenges for SOC research.

to that, the nature of the work as highlighted in the previous chapter leads to a high turnover rate of personnel. This means companies have to spend many resources on training new staff, unless they are willing to spend their resources on retaining the staff. We identified some options in literature to retain staff like training or after-work activities (Section V-A2). However, the lack of job-related security training is still apparent [6], [32]. Practical experience is required to perform data triage, but it is considered hard to get the practical training and experience in the first place [98]. Tier 1 analysts are not always empowered to perform more challenging tasks to improve their knowledge and experience. A lack of feedback from senior analysts intensifies the challenge and can cause frustration [11]. Some technological solutions are trying to overcome the problem by capturing past activities and decisions from experienced staff so the more junior can profit and learn from this data. However, capturing the tacit knowledge involved in the decision-making is a challenging task [98]. Despite this fact, some approaches, especially from Human-Computer Interface (HCI) and respective communities, have been trying to capture the reasoning behind analytical decisions for quite some time [198]. These aspects can help to improve SOC's working conditions.

3) COLLABORATION OF EXPERTS

Collaboration between analysts is still rare, and analysts usually work on a problem independently [12]. This challenge might either stem from the time pressure the staff is facing or the lack of appropriate collaboration platforms. The same applies to communication, which is mostly carried out directly between analysts. This type of communication is necessary but also time-consuming and inefficient [97]. Once again, the absence of an appropriate communication platform

for SOC-specific requirements reduces the staff's interactions overall. Only with the appropriate means to collaborate and communicate SOC analysts from any tier can learn from each other and, therefore, improve their efficiency and motivation.

4) INTEGRATION OF DOMAIN KNOWLEDGE

Identifying threats and incidents gets increasingly harder as IT infrastructures grow and expand from the cyberspace into the physical world, for example through the use of cyber-physical systems [83]. Current automated threat detection tools work pretty well for detecting well-known attacks, as they operate based on signatures and attack patterns [13], [159]. Therefore, unknown situations remain undetected as no rule is defined for them yet. To detect unknown attacks, it is inevitable to include domain knowledge of security experts and even non-security experts. Security experts are valuable as they have a deep understanding of security routines, requirements and have already taken countermeasures. However, non-security experts (e.g. engineers) become more and more indispensable as they have the knowledge which is often necessary to decide whether an alert or the reported behavior is malicious or benign, especially in the context of cyber-physical systems.

Additionally, it is necessary to communicate knowledge of automated analyzes like machine learning models to the SOC staff to understand and comprehend what their analyzes algorithms learned. Tying human experts and machines closer together and providing them processes and technologies to transfer knowledge in either direction is a crucial challenge for SOC's. Only when we succeed in leveraging both domain knowledge from humans and explicit knowledge from machines, we face the next generation of cyber threats.

B. PROCESSES

1) COMPREHENSIVE PROCESS DEFINITIONS

The review showed that there is only very little literature on the processes within a SOC. As these processes are the core of understanding SOCs and deploying them effectively, the lack of precisely defined processes hinders academia from entirely comprehending what organizations are doing within a SOC. Thus, room for small improvements, let alone innovations, are very hard to identify on an abstract level. This might be the reason for the imbalanced results regarding processes and technology. As there is no abstract, high-level understanding of a SOC's processes, many researchers focus on trying to improve technologies that might be useful with no clear understanding of which specific process or task of a SOC needs improvement. Also, having a clear understanding of a SOC's processes, tasks, and interfaces requires the integration with other business processes. This blind spot needs to be closed by academia to understand the processes running in SOCs. Only then will it be possible to advance the current proliferation that is imminent in SOCs in a sustainable manner. Especially "post-incident activity" is barely mentioned in SOC literature, although it is of great importance as it mainly deals with learning and iterative improvement.

2) ADAPT GENERAL PROCESSES TO SOC

Several security standards, regulations, and frameworks [123], [124] define general security-related processes that give rise to the assumption that these can be related at least partially to SOC. These can therefore serve as a basis for a SOC specific process landscape. However, our analysis has not identified any academic literature dealing with how these processes can be related to SOCs. Further research should aim to identify the aspects that apply to SOCs, adapt those to SOC, and extend them by SOC specifics. This could lead simply to a more comprehensive definition and understanding of the processes.

C. TECHNOLOGY

1) INCREASING COMPLEXITY

We see three major challenges for SOCs resulting from the increased complexity of the IT and OT environment in a company: First, the infrastructure is becoming more complicated and intertwined, making it difficult to maintain situational awareness and a cohesive overview. Managers and analysts have poor visibility into the network because they cannot keep track of all the devices in the network [7]. Second, the data captured from the infrastructure is as heterogeneous as its sources [22], [32], [94], making it hard to process, analyze, understand, and link. It also impedes the discovery of whether an event is part of a bigger attack [11]. Third, having more data sources increases the overall number of events and, in many cases, the number of false-positive alerts. It is often mentioned that there is too much (useless) data in general [22], and too many (false positive) alerts [9], [25], [32], [159], [164]. Analysts are overloaded with a high vol-

ume of such alerts and face a typical "needle in a haystack" problem when trying to filter the noise [12], [159]. There is not much discussion about the negative impact of false positives on SOCs, although there are controversial opinions like Kokulu *et al.* [7].

2) WIDE VARIETY OF TOOLS

In many SOCs, the previous problem is approached by implementing and deploying various SOC tools, for example, a SIEM system. However, deploying a variety of tools does not solve the overall problem, at least not immediately. Tools need to be configured and maintained, which is a time- and resource-consuming process [159]. If tools are not maintained properly, they increase the amount of data and false positives to be dealt with for the analysts. Different tools are necessary because most of them only offer a solution to a specific problem. Therefore, a variety of tools is needed to cover all capabilities within a SOC. Integrating them so that they can run smoothly together poses a further challenge [4], [23]. For example, tools typically only cover the standard IT technologies and have no visibility into operational technology. Some tools also suffer from poor usability and regular malfunctioning [7]. This makes the job for analysts much more complicated than it should be and has a negative effect on the detection rate of a SOC. Lastly, tools might be chosen for compliance or budget reasons, not because they are helpful or practical [15].

3) VISUALIZATION CAPABILITIES

Having the right visualization capabilities is another challenge. Generally, there is too much data to be able to visualize it properly [173]. Visualizations need to be simple and easily accessible, as well as precise and informative [12]. However, there is no perfect solution, and a trade-off between these two requirements is necessary. Selecting the right visualization technique is rigid and very dependent on the context and tasks that should be solved with the visualization.

Nonetheless, appropriate visualizations are crucial for an efficient and effective SOC team. Additionally, visualizations are a great deal to support the transfer of knowledge between humans and machines. They can serve as an intermediary allowing analysts to understand machine learning models and improve automated analyses by implicit human input and domain knowledge [199].

4) INSUFFICIENT LEVEL OF AUTOMATION

There is also an insufficient level of automation of SOC components [7]. Many of the tasks carried out in a SOC, e.g. threat hunting, scanning alerts, or responding to incidents, still require a significant portion of manual work in a context where human resources are scarce. The insufficient level of automation is caused by the fact that analysts' tasks are hard to automate. However, automation is needed to reduce the manual and repetitive tasks many SOC analysts have to perform today. There is already a considerable body of literature focusing on the applicability of machine learning

techniques to automate the detection of attacks. Unfortunately, many techniques prove to only be successful under certain conditions or for specific types of attacks. These techniques and their comprehensiveness and effectiveness in detecting attacks need to be compared. More user studies should be conducted to evaluate their usability. Additionally, machine learning approaches produce a high number of false positives. Determining whether an alert is real requires further investigation by the analysts based on tacit knowledge.

D. GOVERNANCE AND COMPLIANCE

1) EFFECTIVE MEASUREMENT OF SOC PERFORMANCE

Even though measuring a SOC's performance and effectiveness is one of the most important governance tasks, many of the currently established metrics are considered inefficient [7], [171]. Additionally, if the metrics are too focused on performance, analysts might be incentivized to work for general statistics [16], [200], as described in Section V-D3. This fuels the need for uniform metrics proving the value of a SOC to management.

2) LACK OF BEST PRACTICES & STANDARDS

Some SOC capabilities, like incident management, are already very advanced. Consequently, many standards and industry best practices can be implemented for these specific capabilities. They can then be audited to see whether they adhere to the standard. Other capabilities are less advanced and have no universal standard. Unfortunately, there is no holistic SOC standard or framework, making it hard to audit a cohesive and complex SOC. The lack of best practices also means that there is no actual decision support for organizations. Decision-makers struggle to choose the right operating model, the right scope, the right capabilities, and even the right tools to support the capabilities. Best practices, either from academia or industry, are needed to enable companies to set up SOCs fitted to their needs. Currently, many guidelines on SOCs are written by security vendors [77], [190]. Despite their valuable contributions to the development of SOCs, they are biased to a certain extent, which further highlights the need for independent standards and impartial industry guidelines. Researchers alone cannot solve this problem. They need to collaborate with regulators, standardization entities, and industry expertise.

3) PRIVACY REGULATIONS

Existing privacy standards and regulations leave many questions regarding collecting and analyzing data unanswered. The company needs to determine if they capture sensitive information, if they could avoid it, and how they can anonymize or at least pseudonymize the data without losing their value. However, there is not much work providing guidelines to decide whether data contains sensitive information or not and even less work giving practical advice on the anonymization of data and still detecting incidents using the

anonymized data. Another challenge on the rise is to define the right policies and procedures.

VII. CONCLUSION

The main objective of this work is to identify and compile the current state-of-the-art of SOCs. To thoroughly achieve this goal, we needed to explore the frontiers of academic literature on the topic. This work's central part consists of a comprehensive literature review on SOCs from a pure research viewpoint. Its objective is to take a close look at SOCs in general but also include their components. The survey is conducted systematically to avoid the exclusion of any relevant information. We planned the review, meaning that the used search terms included various keywords and terms relevant to SOCs. This work includes as many aspects of SOCs as possible. Using the PPTGC framework, various components of a SOC are generally classified into either people, processes, technology, or governance and compliance. We describe these SOC components as currently defined in the literature.

We use the relevant literature and the defined state-of-the-art to identify major challenges that hinder further development and innovation for SOCs. The challenges can also serve as a guideline for future research aiming to improve SOCs. Regarding the people working in a SOC, we see a major challenge in recruiting and retaining staff. Training and Awareness play an essential role in addressing this challenge while also helping to increase the company's overall security posture. When looking at the various processes in a SOC, it is imperative to integrate them with other processes across the whole organization. Analyzing processes regarding SOCs, we can also see that academia and practice lack a thorough and comprehensive definition of the specific processes included in a SOC and their interactions. Without a proper definition of processes, it might not be possible to advance the current state-of-the-art. Technologies promise relief from many repetitive tasks in a SOC; however, most of them are not advanced enough to deliver on the expectations and hype they have created. To maximize the potential of deployed technological solutions, they need to be aligned with and integrated with the rest of an organization's technological infrastructure. Lastly, an immaturity of SOC governance and compliance aspects has been identified. Compared to people or technological components of a SOC, comprehensive standards and industry-specific guidelines are lacking. This kind of immaturity generally impedes security audits and overall SOC assessments. The lack of standards also prevents various SOC components from advancing since a common baseline of the status-quo has not yet been agreed upon. As we have mainly analyzed academic literature, to provide a more comprehensive picture we aim to include a more practical view by considering information such as case studies in future research.

Concluding, SOCs surely help companies to be prepared for cyber-attacks. However, they need to be planned thoroughly, implemented, and integrated very carefully, assessed

regularly, and improved continually to unveil their full potential. If done correctly, they improve companies' ability to prevent hacks, financial losses, and personal data breaches.

REFERENCES

- [1] *The Cost of Cybercrime*, Accenture and Ponemon Institute, New York, NY, USA, 2018.
- [2] S. Schinagl, K. Schoon, and R. Paans, "A framework for designing a security operations centre (SOC)," in *Proc. 48th Hawaii Int. Conf. Syst. Sci.*, Kauai, HI, USA, Jan. 2015, pp. 2253–2262.
- [3] S. Radu, "Comparative analysis of security operations centre architectures; Proposals and architectural considerations for frameworks and operating models," in *Innovative Security Solutions for Information Technology and Communications* (Lecture Notes in Computer Science), vol. 10006. Cham, Switzerland: Springer, 2016, pp. 248–260.
- [4] R. Bidou, J. Bourgeois, and F. Spies, "Towards a global security architecture for intrusion detection and reaction management," in *Information Security Applications* (Lecture Notes in Computer Science), vol. 2908. Berlin, Germany: Springer, 2004, pp. 111–123.
- [5] S. Kowtha, L. A. Nolan, and R. A. Daley, "Cyber security operations center characterization model and analysis," in *Proc. IEEE Conf. Technol. Homeland Secur. (HST)*, Waltham, MA, USA, Nov. 2012, pp. 470–475.
- [6] A. Karim Ganame, J. Bourgeois, R. Bidou, and F. Spies, "A global security architecture for intrusion detection on computer networks," *Comput. Secur.*, vol. 27, no. 1–2, pp. 30–47, Mar. 2008.
- [7] F. B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupe, and G.-J. Ahn, "Matched and mismatched SOCs," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, New York, NY, USA, Nov. 2019, pp. 1955–1970.
- [8] B. Hámornik and C. Krasznay, "A team-level perspective of human factors in cyber security: Security operations centers," in *Advances in Human Factors in Cybersecurity*, vol. 593 D. Nicholson, Ed. Cham, Switzerland: Springer, 2018, pp. 224–236.
- [9] A. Sopan, M. Berninger, M. Mulakaluri, and R. Katakam, "Building a machine learning model for the SOC, by the input from the SOC, and analyzing it for the SOC," in *Proc. IEEE Symp. Visualizat. Cyber Secur. (VizSec)*, Berlin, Germany, Oct. 2018, pp. 1–8.
- [10] V. Rooney and S. Foley, "What you can change and what you can't: Human experience in computer network defenses," in *Secure IT Systems* (Lecture Notes in Computer Science), vol. 11252, N. Gruschka, Ed. Cham, Switzerland: Springer, 2018, pp. 219–235.
- [11] D. Crémilleux, C. Bidan, F. Majorczyk, and N. Prigent, "Enhancing collaboration between security analysts in security operations centers," in *Risks and Security of Internet and Systems*, vol. 11391. Cham, Switzerland: Springer, 2019, pp. 136–142.
- [12] A. Kabil, T. Duval, N. Cuppens, G. Le Comte, Y. Halgand, and C. Ponchel, "3D cybercop: A collaborative platform for cybersecurity data analysis and training," in *Cooperative Design, Visualization, and Engineering* (Lecture Notes in Computer Science), vol. 11151, Y. Luo, Ed. pp. 176–183. Cham, Switzerland: Springer, 2018, pp. 176–183.
- [13] J. R. Goodall, E. D. Ragan, C. A. Steed, J. W. Reed, G. D. Richardson, K. M. T. Huffer, R. A. Bridges, and J. A. Laska, "Situ: Identifying and explaining suspicious behavior in networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 204–214, Jan. 2019.
- [14] S. C. Sundaramurthy, J. Case, T. Truong, L. Zomlot, and M. Hoffmann, "A tale of three security operation centers," in *Proc. ACM Workshop Secur. Inf. Workers*, New York, NY, USA, 2014, pp. 43–50.
- [15] S. C. Sundaramurthy, M. Wesch, X. Ou, J. McHugh, S. R. Rajagopalan, and A. G. Bardas, "Humans are dynamic—our tools should be too," *IEEE Internet Comput.*, vol. 21, no. 3, pp. 40–46, May 2017.
- [16] S. Sundaramurthy, "An anthropological study of security operations centers to improve operational efficiency," Ph.D. dissertation, Dept. Comput. Sci. Eng., Univ. South Florida, Tampa, FL, USA, 2017.
- [17] J. M. Brown, S. Greenspan, and R. Biddle, "Incident response teams in IT operations centers: The T-TOCs model of team functionality," *Cognition, Technol. Work*, vol. 18, no. 4, pp. 695–716, Nov. 2016.
- [18] D. Tranfield, D. Denyer, and P. Smart, "Towards a methodology for developing evidence-informed management knowledge by means of systematic review," *Brit. J. Manage.*, vol. 14, no. 3, pp. 207–222, Sep. 2003.
- [19] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quart.*, vol. 26, no. 2, pp. 13–23, 2002.
- [20] Y. Levy and T. J. Ellis, "A systems approach to conduct an effective literature review in support of information systems research," *Inf. Sci., Int. J. Emerg. Transdiscipline*, vol. 9, pp. 181–212, Dec. 2006.
- [21] C. Okoli, "A guide to conducting a standalone systematic literature review," *Commun. Assoc. Inf. Syst.*, vol. 37, pp. 879–910, May 2015.
- [22] A. Madani, S. Rezayi, and H. Gharaee, "Log management comprehensive architecture in security operation center (SOC)," in *Proc. Int. Conf. Comput. Aspects Social Netw. (CASoN)*, Salamanca, Spain, Oct. 2011, pp. 284–289.
- [23] M. Mutemwa, J. Mtsweni, and L. Zimba, "Integrating a security operations centre with an Organization's existing procedures, policies and information technology systems," in *Proc. Int. Conf. Intell. Innov. Comput. Appl. (ICONIC)*, Plaine Magnien, Mauritius, Dec. 2018, pp. 1–6.
- [24] N. Miloslavskaya, "Analysis of SIEM systems and their usage in security operations and security intelligence centers," in *Biologically Inspired Cognitive Architectures (BICA) for Young Scientists*, vol. 636. Cham, Switzerland: Springer, 2018, pp. 282–288.
- [25] N. Miloslavskaya, A. Tolstoy, and S. Zapechnikov, "Taxonomy for unsecure big data processing in security operations centers," in *Proc. IEEE 4th Int. Conf. Future Internet Things Cloud Workshops (FiCloudW)*, Vienna, Austria, Aug. 2016, pp. 154–159.
- [26] C.-H. Han, S.-T. Park, and S.-J. Lee, "The enhanced security control model for critical infrastructures with the blocking prioritization process to cyber threats in power system," *Int. J. Crit. Infrastruct. Protection*, vol. 26, Sep. 2019, Art. no. 100312.
- [27] J. Kaplan, T. Bailey, C. Rezek, D. O'Halloran, and A. Marcus, "Engage attackers with active defense," in *Beyond Cybersecurity*. Hoboken, NJ, USA: Wiley, 2015, pp. 123–139.
- [28] G. Wang, Z. Yan, and J. Chen, "A method for software trusted update on network security equipment," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 569, Jul. 2019, Art. no. 052086.
- [29] A. Shah, K. A. Farris, R. Ganesan, and S. Jajodia, "Vulnerability selection for remediation: An empirical analysis," *J. Defense Model. Simul., Appl., Methodol., Technol.*, vol. 21, no. 4, Sep. 2019, Art. no. 154851291987412.
- [30] C. Onwubiko, "Cyber security operations centre: Security monitoring for protecting business and supporting cyber defense strategy," in *Proc. Int. Conf. Cyber Situational Awareness, Data Analytics Assessment (CyberSA)*, London, U.K., Jun. 2015, pp. 1–10.
- [31] C. Onwubiko and K. Ouazzane, "Cyber onboarding is Broken," in *Proc. Int. Conf. Cyber Secur. Protection Digit. Services*, Oxford, U.K., Jun. 2019, pp. 1–13.
- [32] S. Mansfield-Devine, "Creating security operations centres that work," *Netw. Secur.*, vol. 2016, no. 5, pp. 15–18, May 2016.
- [33] M. Majid and K. Ariffi, "Success factors for cyber security operation center (SOC) establishment," in *Proc. 1st Int. Conf. Informat., Eng., Sci. Technol.*, Bandung, IN, USA, May 2019, pp. 1–11.
- [34] J. Bourgeois, A. Ganame, I. Kutenko, and A. Ulanov, "Software environment for simulation and evaluation of a security operation center," in *Information Fusion and Geographic Information Systems* (Lecture Notes in Geoinformation and Cartography). Berlin, Germany: Springer, 2007, pp. 111–127.
- [35] A. Bialas, M. Michalak, and B. Flisiuk, "Anomaly detection in network traffic security assurance," in *Engineering in Dependability of Computer Systems and Networks*, vol. 987. Cham, Switzerland: Springer, 2020, pp. 46–56.
- [36] D. Kelley and R. Moritz, "Best practices for building a security operations center," *Inf. Syst. Secur.*, vol. 14, no. 6, pp. 27–32, Jan. 2006.
- [37] L. Ajaz, B. Aslam, and U. Khalid, "Security operations center—A need for an academic environment," in *Proc. World Symp. Comput. Netw. Inf. Secur. (WSCNIS)*, Hammamet, Tunisia, Sep. 2015, pp. 1–7.
- [38] O. Podzins and A. Romanovs, "Why siem is irreplaceable in a secure it environment?" in *Proc. Open Conf. Electr., Electron. Inf. Sci.*, Vilnius, Lithuania, May 2019, pp. 1–5.
- [39] N. Miloslavskaya, "Security intelligence centers for big data processing," in *Proc. 5th Int. Conf. Future Internet Things Cloud Workshops (FiCloudW)*, Prague, Czech Republic, Aug. 2017, pp. 7–13.
- [40] J. Bourgeois and R. Syed, "Managing security of grid architecture with a grid security operation center," in *Proc. Int. Conf. Secur. Cryptogr.*, Milan, Italy, 2009, pp. 403–408.
- [41] R. H. Syed, J. Pazardzievska, and J. Bourgeois, "Fast attack detection using correlation and summarizing of security alerts in grid computing networks," *J. Supercomput.*, vol. 62, no. 2, pp. 804–827, Nov. 2012.

- [42] R. H. Syed, M. Syrame, and J. Bourgeois, "Protecting grids from cross-domain attacks using security alert sharing mechanisms," *Future Gener. Comput. Syst.*, vol. 29, no. 2, pp. 536–547, Feb. 2013.
- [43] A. Ganame, J. Bourgeois, R. Bidou, and F. Spies, "Evaluation of the intrusion detection capabilities and performance of a security operation center," in *Proc. Int. Conf. Secur. Cryptogr.*, 2006, pp. 48–55.
- [44] X. Hu and C. Xie, "Security operation center design based on D-S evidence theory," in *Proc. Int. Conf. Mechatronics Autom.*, Luoyang, China, Jun. 2006, pp. 2302–2306.
- [45] S. Yuan and C. Zou, "The security operations center based on correlation analysis," in *Proc. IEEE 3rd Int. Conf. Commun. Softw. Netw.*, Xi'an, China, May 2011, pp. 334–337.
- [46] E. G. Amoroso, "Cyber attacks: Awareness," *Netw. Secur.*, vol. 2011, no. 1, pp. 10–16, Jan. 2011.
- [47] G. Settanni, F. Skopik, Y. Shovgenya, R. Fiedler, M. Carolan, D. Conroy, K. Boettinger, M. Gall, G. Brost, C. Ponchel, M. Haustein, H. Kaufmann, K. Theuerkauf, and P. Olli, "A collaborative cyber incident management system for European interconnected critical infrastructures," *J. Inf. Secur. Appl.*, vol. 34, pp. 166–182, Jun. 2017.
- [48] T. Tafazzoli and H. Gharaee Garakani, "Security operation center implementation on OpenStack," in *Proc. 8th Int. Symp. Telecommun. (IST)*, Tehran, Iran, Sep. 2016, pp. 766–770.
- [49] J.-S. Li, C.-J. Hsieh, and H.-Y. Lin, "A hierarchical mobile-agent-based security operation center," *Int. J. Commun. Syst.*, vol. 26, no. 12, pp. 1503–1519, Dec. 2013.
- [50] J.-S. Li and C.-J. Hsieh, "Implementation of the distributed hierarchical security operation center using mobile agent group," in *Proc. Int. Symp. Comput., Commun., Control Autom. (3CA)*, Tainan, Taiwan, May 2010, pp. 79–82.
- [51] G. Chamiekar, M. Cooray, L. Wickramasinghe, Y. Koshila, K. Abeywardhana, and A. Senarathna, "Autosoc: A low budget flexible security operations platform for enterprises and organizations," in *Proc. Nat. Inf. Technol. Conf. (NITC)*, Colombo, Sri Lanka, 2017, pp. 100–105.
- [52] E. Falk, S. Repcek, B. Fiz, S. Hommes, R. State, and R. Sasnauskas, "VSOC—A virtual security operating center," in *Proc. IEEE Global Commun. Conf.*, Singapore, Dec. 2017, pp. 1–8.
- [53] U. Glasser, P. Jackson, A. Araghi, and H. Shahir, "Intelligent decision support for marine safety and security operations," in *Proc. IEEE Int. Conf. Intell. Secur. Inform.*, Vancouver, BC, Canada, May 2010, pp. 101–107.
- [54] B. AlSabbagh and S. Kowalski, "A framework and prototype for a socio-technical security information and event management system (STSIEM)," in *Proc. Eur. Intell. Secur. Informat. Conf. (EISIC)*, Uppsala, Sweden, Aug. 2016, pp. 192–195.
- [55] F. Sailhan and J. Bourgeois, "Log-based distributed intrusion detection for hybrid networks," in *Proc. 4th Annu. workshop Cyber Secur. informaiton Intell. Res.*, New York, NY, USA, 2008, pp. 1–6.
- [56] P. Bienias, G. Kolaczek, and A. Warzynski, "Architecture of anomaly detection module for the security operations center," in *Proc. IEEE 28th Int. Conf. Enabling Technologies: Infrastruct. Collaborative Enterprises (WETICE)*, Naples, Italy, Jun. 2019, pp. 126–131.
- [57] A. Chowdhary, D. Huang, G.-J. Ahn, M. Kang, A. Kim, and A. Velazquez, "SDNSOC: Object oriented SDN framework," in *Proc. ACM Int. Workshop Secur. Softw. Defined Netw. Function Virtualization*, New York, NY, USA, 2019, pp. 7–12.
- [58] D. Crooks and L. Valsan, "Wlsg security operations centre working group," *Proc. Sci.*, vol. 1, no. 1, pp. 1–25, 2017.
- [59] D. Crooks, L. Valsan, K. Mohammad, S. McKee, P. Clark, A. Boutcher, A. Padée, M. Wójcik, H. Giemza, and B. Kreukniet, "Operational security, threat intelligence & distributed computing: The wlsg security operations center working group," *EPJ Web Conferences*, vol. 214, p. 15, May 2019.
- [60] D. Crooks and L. Valsan, "Building a minimum viable security operations centre for the modern grid environment," in *Proc. Int. Symp. Grids Clouds*, Trieste, Italy, Nov. 2019, p. 10.
- [61] P. Danquah, "Security operations center: A framework for automated triage, containment and escalation," *J. Inf. Secur.*, vol. 11, no. 4, pp. 225–240, 2020.
- [62] D. Forte, "An inside look at security operation centres," *Netw. Secur.*, vol. 2003, no. 5, pp. 11–12, 2003.
- [63] P. Jacobs, A. Arnab, and B. Irwin, "Classification of security operation centers," in *Proc. Inf. Secur. South Afr.*, Johannesburg, South Africa, Aug. 2013, pp. 1–7.
- [64] D. Forte, "State of the art security management," *Comput. Fraud Secur.*, vol. 2009, no. 10, pp. 17–18, Oct. 2009.
- [65] N. Miloslavskaya, "Security operations centers for information security incident management," in *Proc. IEEE 4th Int. Conf. Future Internet Things Cloud (FiCloud)*, Vienna, Austria, Aug. 2016, pp. 131–136.
- [66] F. David Janos and N. Huu Phuoc Dai, "Security concerns towards security operations centers," in *Proc. IEEE 12th Int. Symp. Appl. Comput. Intell. Informat. (SACI)*, Timisoara, Romania, May 2018, pp. 000273–000278.
- [67] A. Shah, R. Ganesan, and S. Jajodia, "A methodology for ensuring fair allocation of CSOC effort for alert investigation," *Int. J. Inf. Secur.*, vol. 18, no. 2, pp. 199–218, Apr. 2019.
- [68] M. Khalili, M. Zhang, D. Borbor, L. Wang, N. Scarabeo, and M.-A. Zamor, "Monitoring and improving managed security services inside a security operation center," *ICST Trans. Secur. Saf.*, vol. 5, no. 18, Apr. 2019, Art. no. 157413.
- [69] C. Crowley and J. Pescatore, "Sans 2018 security operations center survey," SANS Inst., Swansea, U.K., Tech. Rep., 2018.
- [70] G. D. Bhatt, "Knowledge management in organizations: Examining the interaction between technologies, techniques, and people," *J. Knowl. Manage.*, vol. 5, no. 1, pp. 68–75, Mar. 2001.
- [71] R. Ruefle, "Defining computer security incident response teams," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep., 2007.
- [72] D. Robb, "How to manage a security operations center," eSecurity Planet, Nashville, TN, USA, Tech. Rep., 2019.
- [73] M. Vielberth and G. Pernul, "A security information and event management pattern," in *Proc. 12th Latin Amer. Conf. Pattern Lang. Prog. (SLPLoP)*, 2018, pp. 1–5.
- [74] F. Alruwaili and T. Gulliver, "SocaaS: Security operations center as a service for cloud computing environments," *Int. J. Cloud Comput. Services Sci.*, vol. 3, no. 2, pp. 87–96, 2014.
- [75] C. Zimmerman, "Ten strategies of a world-class cybersecurity operations center," MITRE Corp., Bedford, MA, USA, Tech. Rep., 2014.
- [76] H. Security, "Choosing a soc service model: The key considerations," Huntsman Secur., London, U.K., Tech. Rep., 2018.
- [77] J. Muniz, G. McIntyre, and N. AlFardan, *Security operations center: Building, operating, and maintaining your SOC*. Indianapolis, IN, USA: Cisco Press, 2015.
- [78] *Outsourced Soc Vs. Internal Soc: How to Choose*, Linkbynet, Montreal, QC, Canada, 2018.
- [79] C. Olt, "Establishing security operation centers for connected cars," *ATZelectronics worldwide*, vol. 14, no. 5, pp. 40–43, May 2019.
- [80] C. DeCusatis, R. Cannistra, A. Labouseur, and M. Johnson, "Design and implementation of a research and education cybersecurity operations center," in *Cybersecurity and Secure Information Systems* (Advanced Sciences and Technologies for Security Applications), vol. 33. Cham, Switzerland: Springer, 2019, pp. 287–310.
- [81] R. Ganesan, A. Shah, S. Jajodia, and H. Cam, "Optimizing alert data management processes at a cyber security operations center," in *Adversarial and Uncertain Reasoning for Adaptive Cyber Defense* (Lecture Notes in Computer Science), vol. 11830. Cham, Switzerland: Springer, 2019, pp. 206–231.
- [82] C. Zhong, J. Yen, P. Liu, and R. F. Erbacher, "Learning from Experts' experience: Toward automated cyber security data triage," *IEEE Syst. J.*, vol. 13, no. 1, pp. 603–614, Mar. 2019.
- [83] C. Islam, M. Babar, and S. Nepal, "Automated interpretation and integration of security tools using semantic knowledge," in *Advanced Information Systems Engineering* (Lecture Notes in Computer Science), vol. 11483. Cham, Switzerland: Springer, 2019, pp. 513–528.
- [84] Y. Kanemoto, K. Aoki, M. Iwamura, J. Miyoshi, D. Kotani, H. Takakura, and Y. Okabe, "Detecting successful attacks from IDS alerts based on emulation of remote shellcodes," in *Proc. IEEE 43rd Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, Milwaukee, WA, USA, Jul. 2019, pp. 471–476.
- [85] E. Agyepong, Y. Cherdantseva, P. Reinecke, and P. Burnap, "Challenges and performance metrics for security operations center analysts: A systematic review," *J. Cyber Secur. Technol.*, vol. 76, no. 3, pp. 1–28, 2019.
- [86] C. Zhong, J. Yen, P. Liu, R. Erbacher, C. Garneau, and B. Chen, "Studying analysts' data triage operations in cyber defense situational analysis," in *Theory Models for Cyber Situation Awareness* (Lecture Notes in Computer Science), vol. 10030. Cham, Switzerland: Springer, 2017, pp. 128–169.

- [87] C. Zhong, J. Yen, P. Liu, and R. F. Erbacher, "Automate cybersecurity data triage by leveraging human Analysts' cognitive process," in *Proc. IEEE IEEE 2nd Int. Conf. Big Data Secur. Cloud*, Apr. 2016, pp. 357–363.
- [88] C. Zhong, T. Lin, P. Liu, J. Yen, and K. Chen, "A cyber security data triage operation retrieval system," *Comput. Secur.*, vol. 76, pp. 12–31, Jul. 2018.
- [89] A. Pingle, A. Piplai, S. Mittal, A. Joshi, J. Holt, and R. Zak, "Relext: Relation extraction using deep learning approaches for cybersecurity knowledge graph improvement," in *Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Mining*, 2019, pp. 879–886.
- [90] A. Shah, R. Ganesan, S. Jajodia, and H. Cam, "Adaptive reallocation of cybersecurity analysts to sensors for balancing risk between sensors," *Service Oriented Comput. Appl.*, vol. 12, no. 2, pp. 123–135, Jun. 2018.
- [91] A. Shah, R. Ganesan, S. Jajodia, and H. Cam, "A two-step approach to optimal selection of alerts for investigation in a CSOC," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1857–1870, Jul. 2019.
- [92] S. Sundaramurthy, A. Bardas, J. Case, X. Ou, M. Wesch, J. McHugh, and R. Rajagopalan, "A human capital model for mitigating security analyst burnout," in *Proc. 11th Symp. Usable Privacy Secur.*, Ontario, ON, Canada, 2015 pp. 347–359.
- [93] T. Sander and J. Hailpern, "UX aspects of threat information sharing platforms," in *Proc. 2nd ACM Workshop Inf. Sharing Collaborative Secur.*, New York, NY, USA, 2015, pp. 51–59.
- [94] T. Lin, C. Zhong, J. Yen, and P. Liu, "Retrieval of relevant historical data triage operations in security operation centers," in *From Database to Cyber Security* (Lecture Notes in Computer Science), vol. 11170. Cham, Switzerland: Springer, 2018, pp. 227–243.
- [95] A. Applebaum, S. Johnson, M. Limiero, and M. Smith, "Playbook oriented cyber response," in *Proc. Nat. Cyber Summit (NCS)*, Huntsville, Alabama, Jun. 2018, pp. 8–15.
- [96] S. Sanchez, R. Mazzolin, I. Kechaoglou, D. Wiemer, W. Mees, and J. Muylaert, "Cybersecurity space operation center: Countering cyber threats in the space domain," in *Handbook Space Security*, K.-U. Schroggl, Ed. Cham, Switzerland: Springer, 2020, pp. 921–939.
- [97] C. Zhong, A. Alnusair, B. Sayer, A. Troxell, and J. Yao, "AOH-map: A mind mapping system for supporting collaborative cyber security analysis," in *Proc. IEEE Conf. Cognit. Comput. Aspects Situation Manage. (CogSIMA)*, Las Vegas, NV, USA, Apr. 2019, pp. 74–80.
- [98] A. Shah, R. Ganesan, S. Jajodia, and H. Cam, "Optimal assignment of sensors to analysts in a cybersecurity operations center," *IEEE Syst. J.*, vol. 13, no. 1, pp. 1060–1071, Mar. 2019.
- [99] A. Kabil, T. Duval, N. Cuppens, G. Le Comte, Y. Halgand, and C. Ponchel, "From cyber security activities to collaborative virtual environments practices through the 3D cybercop platform," in *Information Systems Security* (Lecture Notes in Computer Science), vol. 11281. Cham, Switzerland: Springer, 2018, pp. 272–287.
- [100] A. Vault, "How to build a security operations center," Alien Vault, San Mateo, CA, USA, Tech. Rep., 2017.
- [101] O. Cassetto, "Security operations center roles and responsibilities," Exabeam, Foster City, CA, USA, Tech. Rep., 2019.
- [102] *Strategies for Building and Growing Strong Cybersecurity Teams: Cybersecurity Workforce Study*, International Information System Security Certification Consortium, Clearwater, FL, USA, 2019.
- [103] A. Chin-Ching Lin, H.-K. Wong, and T.-C. Wu, "Enhancing interoperability of security operation center to heterogeneous intrusion detection systems," in *Proc. 39th Annu. Int. Carnahan Conf. Secur. Technol.*, Las Palmas, Spain, 2005, pp. 216–221.
- [104] S. Bhatt, P. K. Manadhata, and L. Zomlot, "The operational role of security information and event management systems," *IEEE Secur. Privacy*, vol. 12, no. 5, pp. 35–41, Sep. 2014.
- [105] D. Zhang and D. Zhang, "The analysis of event correlation in security operations center," in *Proc. 4th Int. Conf. Intell. Comput. Technol. Autom.*, Guangdong, Shenzhen, Mar. 2011, pp. 1214–1216.
- [106] Z. Qu and L. Wang, "The design of a correlation analysis engine model based on Carma_VE algorithm," in *Proc. IEEE Int. Symp. Med. Edu.*, Jinan, China, Aug. 2009, pp. 1267–1270.
- [107] B. Bösch, "Approach to enhance the efficiency of security operation centers to heterogeneous ids landscapes," in *Critical Information Infrastructures Security* (Lecture Notes in Computer Science), vol. 7722. Berlin, Germany: Springer, 2013, pp. 1–9.
- [108] F. Sallhan, J. Bourgeois, and V. Issarny, "A security supervision system for hybrid networks," in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (Studies in Computational Intelligence), vol. 149, R. Lee, Ed. Berlin, Germany: Springer, 2008, pp. 137–149.
- [109] M. E. Verma and R. A. Bridges, "Defining a metric space of host logs and operational use cases," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Seattle, WA, USA, Dec. 2018, pp. 5068–5077.
- [110] M. Alam, S.-U.-R. Malik, Q. Javed, A. Khan, S. B. Khan, A. Anjum, N. Javed, A. Akhuzada, and M. K. Khan, "Formal modeling and verification of security controls for multimedia systems in the cloud," *Multimedia Tools Appl.*, vol. 76, no. 21, pp. 22845–22870, Nov. 2017.
- [111] R. Bridges, M. Iannacone, J. Goodall, and J. Beaver, "How do information security workers use host data? A summary of interviews with security analysts," 2018, *arXiv:1812.02867v1*. [Online]. Available: <https://arxiv.org/abs/1812.02867>
- [112] B. Song, J. Choi, S.-S. Choi, and J. Song, "Visualization of security event logs across multiple networks and its application to a CSOC," *Cluster Comput.*, vol. 22, no. S1, pp. 1861–1872, Jan. 2019.
- [113] D. Weissman and A. Jayasumana, "Integrating IoT monitoring for security operation center," in *Proc. Global Internet Things Summit (GIoTS)*, Dublin, Ireland, Jun. 2020, pp. 1–6.
- [114] M. Nabil, S. Soukainat, A. Lakkabi, and O. Ghizlane, "SIEM selection criteria for an efficient contextual security," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, Marrakech, Morocco, May 2017, pp. 1–6.
- [115] Y.-C. Cheng, C.-H. Chen, C.-C. Chiang, J.-W. Wang, and C.-S. Lai, "Generating attack scenarios with causal relationship," in *Proc. IEEE Int. Conf. Granular Comput.*, Fremont, CA, USA, Nov. 2007, p. 368.
- [116] G. Gonzalez Granadillo, M. El-Barbori, and H. Debar, "New types of alert correlation for security information and event management systems," in *Proc. 8th IFIP Int. Conf. New Technol., Mobility Secur. (NTMS)*, Larnaca, Cyprus, Nov. 2016, pp. 1–7.
- [117] C. Islam, M. A. Babar, and S. Nepal, "A multi-vocal review of security orchestration," *ACM Comput. Surv.*, vol. 52, no. 2, pp. 1–45, May 2019.
- [118] K. Hughes, K. McLaughlin, and S. Sezer, "Dynamic countermeasure knowledge for intrusion response systems," in *Proc. 31st Irish Signals Syst. Conf. (ISSC)*, Letterkenny, Ireland, Jun. 2020, pp. 1–6.
- [119] S. Y. Cho, J. Happa, and S. Creese, "Capturing tacit knowledge in security operation centers," *IEEE Access*, vol. 8, pp. 42021–42041, 2020.
- [120] M. H. Khyavi, "Isms role in the improvement of digital forensics related process in soc's," 2015, *arXiv:2006.08255*. [Online]. Available: <https://arxiv.org/abs/2006.08255>
- [121] W. Yang and K.-Y. Lam, "Automated cyber threat intelligence reports classification for early warning of cyber attacks in next generation soc," in *Information and Communications Security*, vol. 11999, J. Zhou, X. Luo, Q. Shen, and Z. Xu, Eds. Cham, Switzerland: Springer, 2020, pp. 145–164.
- [122] C. Islam, M. A. Babar, and S. Nepal, "Architecture-centric support for integrating security tools in a security orchestration platform," in *Software Architecture* (Lecture Notes in Computer Science), vol. 12292, A. Jansen, I. Malavolta, H. Muccini, I. Ozkaya, O. Zimmermann, Eds. Cham, Switzerland: Springer, 2020, pp. 165–181.
- [123] *Information Technology - Security Techniques—Information Security Incident Management—Part 1: Principles of Incident Management*, Standard Iso/iec 27035-1:2016, 2016.
- [124] P. Cichonski, T. Millar, T. Grance, and K. Scarfone, "Computer security incident handling guide: Special publication 800-61 revision 2," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. 800-61, 2012.
- [125] K. Kent and M. Souppaya, "Guide to computer security log management: Recommendations of the national institute of standards and technology," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. 800-92, 2006.
- [126] F. Osinga, *Science, Strategy and War: The Strategic Theory of John Boyd*. London, U.K.: Routledge, 2007.
- [127] C. Falk and J. Dykstra, "Sonification with music for cybersecurity situational awareness," in *Proc. 25th Int. Conf. Auditory Display (ICAD)*, Jun. 2019, pp. 50–55.
- [128] D. Ambawade, P. Kedar, and J. Bakal, "A comprehensive architecture for correlation analysis to improve the performance of security operation center," in *Innovations in Computer Science and Engineering* (Lecture Notes in Networks and Systems), vol. 8. Singapore: Springer, 2017, pp. 205–216.
- [129] M. Almukaynizi, E. Marin, E. Nunes, P. Shakarian, G. I. Simari, D. Kapoor, and T. Siedlecki, "DARKMENTION: A deployed system to predict enterprise-targeted external cyberattacks," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Miami, FL, USA, Nov. 2018, pp. 31–36.

- [130] R. Graf and R. King, "Neural network and blockchain based technique for cyber threat intelligence and situational awareness," in *Proc. 10th Int. Conf. Cyber Conflict (CyCon)*, Tallinn, Estonia, May 2018, pp. 409–426.
- [131] R. Graf and R. King, "Secured transactions technique based on smart contracts for situational awareness tools," in *Proc. 12th Int. Conf. Internet Technol. Secured Trans. (ICITST)*, Cambridge, U.K., Dec. 2017, pp. 81–86.
- [132] D.-R. Tsai, W.-C. Chen, Y.-C. Lu, and C.-W. Wu, "A trusted security information sharing mechanism," in *Proc. 43rd Annu. Int. Carnahan Conf. Secur. Technol.*, Zurich, Switzerland, Oct. 2009, pp. 257–260.
- [133] L. Karaçay, E. Savaa, and H. Alptekin, "Intrusion detection over encrypted network data," *Comput. J.*, vol. 63, no. 4, pp. 604–619, Apr. 2020.
- [134] M. M. Baskaran, T. Henretty, J. Ezick, R. Lethin, and D. Bruns-Smith, "Enhancing network visibility and security through tensor analysis," *Future Gener. Comput. Syst.*, vol. 96, pp. 207–215, Jul. 2019.
- [135] K. Berlin, D. Slater, and J. Saxe, "Malicious behavior detection using windows audit logs," in *Proc. 8th ACM Workshop Artif. Intell. Secur.*, New York, NY, USA, 2015, pp. 35–44.
- [136] P. Burnap, R. French, F. Turner, and K. Jones, "Malware classification using self organising feature maps and machine activity data," *Comput. Secur.*, vol. 73, pp. 399–410, Mar. 2018.
- [137] Q. Chen, R. Islam, H. Haswell, and R. Bridges, "Automated ransomware behavior analysis: Pattern extraction and early detection," in *Proc. Int. Conf. Sci. Cyber Secur.*, 2019, pp. 199–214.
- [138] K. Demertzis, N. Tziritas, P. Kikiras, S. L. Sanchez, and L. Iliadis, "The next generation cognitive security operations center: Adaptive analytic lambda architecture for efficient defense against adversarial attacks," *Big Data Cognit. Comput.*, vol. 3, no. 1, p. 6, Jan. 2019.
- [139] H. M. Farooq and N. M. Otaibi, "Optimal machine learning algorithms for cyber threat detection," in *Proc. 20th Int. Conf. Comput. Model. Simul. (UKSim)*, Cambridge, U.K., Mar. 2018, pp. 32–37.
- [140] C. Feng, S. Wu, and N. Liu, "A user-centric machine learning framework for cyber security operations center," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Beijing, China, Jul. 2017, pp. 173–175.
- [141] W. Feng, S. Wu, X. Li, and K. Kunkle, "A deep belief network based machine learning system for risky host detection," 2017, *arXiv:1801.00025*. [Online]. Available: <https://arxiv.org/abs/1801.00025>
- [142] J. D. Hernandez Guillen, A. Martin del Rey, and R. Casado-Vara, "Security countermeasures of a SCIRAS model for advanced malware propagation," *IEEE Access*, vol. 7, pp. 135472–135478, 2019.
- [143] S. Hiruta, S. Ikeda, S. Shima, and H. Takakura, "Ids alert priority determination based on traffic behavior," in *Advances in Information and Computer Security* (Lecture Notes in Computer Science), vol. 11689. Cham, Switzerland: Springer, 2019, pp. 189–206.
- [144] K.-F. Hong, C.-C. Chen, Y.-T. Chiu, and K.-S. Chou, "Ctracer: Uncover C&C in advanced persistent threats based on scalable framework for enterprise log data," in *Proc. IEEE Int. Congr. Big Data*, New York, NY, USA, Jun. 2015, pp. 551–558.
- [145] C. Mao, H. Pao, C. Faloutsos, and H. Lee, "Sbad: Sequence based attack detection via sequence comparison," in *Privacy and Security Issues in Data Mining and Machine Learning* (Lecture Notes in Computer Science), vol. 6549. Berlin, Germany: Springer, 2011, pp. 78–91.
- [146] H. Mao, C. Wu, E. Papalexakis, C. Faloutsos, K. Lee, and T. Kao, "Malspot: Multi2 malicious network behavior patterns analysis," in *Advances in Knowledge Discovery and Data Mining* (Lecture Notes in Computer Science), vol. 8443. Cham, Switzerland: Springer, 2014, pp. 1–14.
- [147] Y. Niu and Y. C. Peng, "Application of radial function neural network in network security," in *Proc. Int. Conf. Comput. Intell. Secur.*, Suzhou, China, Dec. 2008, pp. 458–463.
- [148] Y. Niu, Q. Zhang, Q. Zheng, and H. Peng, "Security operation center based on immune system," in *Proc. Int. Conf. Comput. Intell. Secur. Workshops (CISW)*, Heilongjiang, China, Dec. 2007, pp. 97–103.
- [149] A. Oprea, Z. Li, T.-F. Yen, S. H. Chin, and S. Alrwais, "Detection of early-stage enterprise infection by mining large-scale log data," in *Proc. 45th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw.*, Rio de Janeiro, Brazil, Jun. 2015, pp. 45–56.
- [150] A. Oprea, Z. Li, R. Norris, and K. Bowers, "MADE: Security analytics for enterprise threat detection," in *Proc. 34th Annu. Comput. Secur. Appl. Conf.*, New York, NY, USA, Dec. 2018, pp. 124–136.
- [151] H.-K. Pao, C.-H. Mao, H.-M. Lee, C.-D. Chen, and C. Faloutsos, "An intrinsic graphical analysis based on alert correlation analysis for intrusion detection," in *Proc. Int. Conf. Technol. Appl. Artif. Intell.*, Hsinchu, Taiwan, Nov. 2010, pp. 102–109.
- [152] R. Vaidyanathan, A. Ghosh, Y.-H. Cheng, A. Yamada, and Y. Miyake, "On the use of BGP AS numbers to detect spoofing," in *Proc. IEEE Globecom Workshops*, Miami, FL, USA, Dec. 2010, pp. 1606–1610.
- [153] S. Wu, J. Fulton, N. Liu, C. Feng, and L. Zhang, "Risky host detection with bias reduced semi-supervised learning," in *Proc. Int. Conf. Artif. Intell. Comput. Sci.*, New York, NY, USA, Jul. 2019, pp. 34–40.
- [154] T.-F. Yen, A. Oprea, K. Onarlioglu, T. Leetham, W. Robertson, A. Juels, and E. Kirda, "Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks," in *Proc. 29th Annu. Comput. Secur. Appl. Conf.*, New York, NY, USA, 2013, pp. 199–208.
- [155] N. Yi, Z. Qi-Lun, and P. Hong, "Network security management based on data fusion technology," in *Proc. 7th Int. Conf. Comput.-Aided Ind. Des. Conceptual Des.*, Hangzhou, China, May 2006, pp. 889–892.
- [156] P. Dymora and M. Mazurek, "An innovative approach to anomaly detection in communication networks using multifractal analysis," *Appl. Sci.*, vol. 10, no. 9, p. 3277, May 2020.
- [157] M. Smith, "The soc is dead, long live the soc!" *Inow*, vol. 62, no. 1, pp. 34–35, 2020.
- [158] G. Settanni, Y. Shovgenya, F. Skopik, R. Graf, M. Wurzenberger, and R. Fiedler, "Acquiring cyber threat intelligence through security information correlation," in *Proc. 3rd IEEE Int. Conf. Cybern. (CYBCONF)*, Exeter, U.K., Jun. 2017, pp. 1–7.
- [159] A. Erola, I. Agrafiotis, J. Happa, M. Goldsmith, S. Creese, and P. Legg, "Richerpicture: Semi-automated cyber defence using context-aware data analytics," in *Proc. Int. Conf. On Cyber Situational Awareness, Data Anal. Assessment*, London, U.K., Aug. 2017, pp. 1–8.
- [160] A. Kabil, T. Duval, N. Cuppens, G. L. Comte, Y. Halgand, and C. Ponchel, "Why should we use 3D collaborative virtual environments for cyber security?" in *Proc. IEEE 4th VR Int. Workshop Collaborative Virtual Environ. (3DCVE)*, Reutlingen, Germany, Mar. 2018, pp. 1–2.
- [161] T. Kwon, J.-S. Song, S. Choi, Y. Lee, and J. Park, "VISNU: A novel visualization methodology of security events optimized for a centralized SOC," in *Proc. 13th Asia Joint Conf. Inf. Secur. (AsiaJCS)*, Guilin, China, Aug. 2018, pp. 1–7.
- [162] B. Song, S. Choi, J. Choi, and J. Song, "Visualization of intrusion detection alarms collected from multiple networks," in *Information Security* (Lecture Notes in Computer Science), vol. 10599. Cham, Switzerland: Springer, 2017, pp. 437–454.
- [163] S. Hassell, P. Beraud, A. Cruz, G. Ganga, S. Martin, J. Toennies, P. Vazquez, G. Wright, D. Gomez, F. Pietryka, N. Srivastava, T. Hester, D. Hyde, and B. Mastropietro, "Evaluating network cyber resiliency methods using cyber threat, vulnerability and defense modeling and simulation," in *Proc. IEEE Mil. Commun. Conf.*, Orlando, FL, USA, Oct. 2012, pp. 1–6.
- [164] G. Payer and L. Trossbach, "The application of virtual reality for cyber information visualization and investigation," in *Evolution of Cyber Technologies and Operations*, vol. 63, M. Blowers, Ed. Cham, Switzerland: 2015, pp. 71–90.
- [165] L. Axon, B. Alahmadi, J. Nurse, M. Goldsmith, and S. Creese, "Sonification in security operations centres: What do security practitioners think?" in *Proc. Workshop Usable Secur.*, Reston, VA, USA, 2018, pp. 1–12.
- [166] L. Axon, J. Happa, A. van Janse Rensburg, M. Goldsmith, and S. Creese, "Sonification to support the monitoring tasks of security operations centres," *IEEE Trans. Dependable Secure Comput.*, early access, Jul. 29, 2019, doi: [10.1109/TDSC.2019.2931557](https://doi.org/10.1109/TDSC.2019.2931557).
- [167] L. Axon, J. Happa, M. Goldsmith, and S. Creese, "Hearing attacks in network data: An effectiveness study," *Comput. Secur.*, vol. 83, pp. 367–388, Jun. 2019.
- [168] L. Axon, B. A. Alahmadi, J. R. C. Nurse, M. Goldsmith, and S. Creese, "Data presentation in security operations centres: Exploring the potential for sonification to enhance existing practice," *J. Cybersecurity*, vol. 6, no. 1, Jan. 2020, Art. no. tyaa004.
- [169] N. Afzaliseresht, Y. Miao, S. Michalska, Q. Liu, and H. Wang, "From logs to stories: human-centred data mining for cyber threat intelligence," *IEEE Access*, vol. 8, pp. 19089–19099, 2020.
- [170] R. Mullins, B. Nargi, and A. Fouse, "Understanding and enabling tactical situational awareness in a security operations center," in *Advances in Human Factors in Cybersecurity*, vol. 1219, I. Corradini, E. Nardelli, and T. Ahram, Eds. Cham, Switzerland: Springer, 2020, pp. 75–82.
- [171] Z. Wang and Y. Zhu, "A centralized HIDS framework for private cloud," in *Proc. 18th IEEE/ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput. (SNPD)*, Kanazawa, Japan, Jun. 2017, pp. 115–120.
- [172] R. Gad, M. Kappes, and I. Medina-Bulo, "Monitoring traffic in computer networks with dynamic distributed remoting packet capturing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 5759–5764.

- [173] H. Shiravi, A. Shiravi, and A. A. Ghorbani, "A survey of visualization systems for network security," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 8, pp. 1313–1329, Aug. 2012.
- [174] R. Marty, *Applied Security Visualization*. Boston, MA, USA: Addison-Wesley, 2009.
- [175] M. Vielberth, F. Menges, and G. Pernul, "Human-as-a-security-sensor for harvesting threat intelligence," *Cybersecurity*, vol. 2, no. 1, p. 35, Dec. 2019.
- [176] G. Zhiguo, X. Luo, J. Chen, F. L. Wang, and J. Lei, Eds., *Emerging Research in Web Information Systems and Mining* (Communications in Computer and Information Science). Berlin, Germany: Springer, 2011.
- [177] R. Heartfield, G. Loukas, and D. Gan, "You are probably not the weakest link: Towards practical prediction of susceptibility to semantic social engineering attacks," *IEEE Access*, vol. 4, pp. 6910–6928, 2016.
- [178] H. Liao, C. Richard Lin, Y. Lin, and K. Tung, "Intrusion detection system: A comprehensive review," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 16–24, 2013.
- [179] N. Miloslavskaya, "SOC-and SIC-based information security monitoring," in *Recent Advances in Information Systems and Technologies* (Advances in Intelligent Systems and Computing), vol. 570. Cham, Switzerland: Springer, 2017, pp. 364–374.
- [180] A. K. Ganame and J. Bourgeois, "Defining a simple metric for real-time security level evaluation of multi-sites networks," in *Proc. IEEE Int. Symp. Parallel Distrib. Process.*, Miami, FL, USA, Apr. 2008, pp. 1–8.
- [181] R. Ganesan and A. Shah, "A strategy for effective alert analysis at a cyber security operations center," in *From Database to Cyber Security* (Lecture Notes in Computer Science), vol. 11170. Cham, Switzerland: Springer, 2018, pp. 206–226.
- [182] K. A. Farris, A. Shah, G. Cybenko, R. Ganesan, and S. Jajodia, "VULCON: A system for vulnerability prioritization, mitigation, and management," *ACM Trans. Privacy Secur.*, vol. 21, no. 4, pp. 1–28, Oct. 2018.
- [183] T. Sadamatsu, Y. Yoneyama, and K. Yajima, "Practice within Fujitsu of security operations center: Operation and security dashboard," *Fujitsu Sci. Tech. J.*, vol. 52, no. 3, pp. 52–58, 2016.
- [184] L. Allodi and F. Massacci, "Security events and vulnerability data for cybersecurity risk estimation," *Risk Anal.*, vol. 37, no. 8, pp. 1606–1627, Aug. 2017.
- [185] C. Onwubiko and K. Ouazzane, "SOTER: A playbook for cybersecurity incident management," *IEEE Trans. Eng. Manag.*, early access, May 6, 2020, doi: [10.1109/TEM.2020.2979832](https://doi.org/10.1109/TEM.2020.2979832).
- [186] E. Agyepong, Y. Cherdantseva, P. Reinecke, and P. Burnap, "Towards a framework for measuring the performance of a security operations center analyst," in *Proc. Int. Conf. Cyber Secur. Protection Digit. Services (Cyber Secur.)*, Dublin, Republic of Ireland, Jun. 2020, pp. 1–8.
- [187] G. Gaudin, H. Debar, A. Fillette, J. deMeer, A. Rennoch, P. Saadé, and J. Saugeot, *Guidelines for Building and Operating a Secured Security Operations Center (SOC)*, document ETSI GS ISI 007, 2018.
- [188] C. Crowley and J. Pescatore, "Common and best practices for security operations centers: Results of the 2019 SOC survey," SANS, Bethesda, MD, USA, Tech. Rep., 2019.
- [189] "Audit of NASA's security operations center," Nat. Aeronaut. Space Admin., Washington, DC, USA, Tech. Rep. ig-18-020, 2018.
- [190] *Strategy Considerations for Building a Security Operations Center: Optimize Your Security Intelligence to Better Safeguard Your Business From Threats*, IBM, Armonk, NY, USA, 2013.
- [191] W. Jansen, *Directions in Security Metrics Research*. Gaithersburg, MD, USA: Diane Publishing, 2010.
- [192] R. M. Savola, "Towards a taxonomy for information security metrics," in *Proc. ACM Workshop Qual. Protection (QoP)*, 2007, pp. 28–30.
- [193] P. Black, K. Scarfone, and M. Souppaya, "Cyber security metrics and measures," in *Wiley Handbook of Science and Technology for Homeland Security*. Hoboken, NJ, USA: Wiley, 2008, pp. 1–15.
- [194] R. Ganesan, S. Jajodia, and H. Cam, "Optimal scheduling of cybersecurity analysts for minimizing risk," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 4, pp. 1–32, Jul. 2017.
- [195] J. Moran, "Key performance indicators (KPIS) for security operations and incident response: Identifying which KPIS should be set, monitored and measured," DFLABS, Milano, IT, USA, Tech. Rep., 2019.
- [196] G. Doran, "There's a SMART way to write management's goals and objectives," *Manage. Rev.*, vol. 70, no. 11, pp. 35–36, 1981.
- [197] R. Cambra, "Metrics for operational security control," SANS Inst., Swansea, U.K., Tech. Rep., 2004.
- [198] K. Xu, S. Attfield, T. J. Jankun-Kelly, A. Wheat, P. H. Nguyen, and N. Selvaraj, "Analytic provenance for sensemaking: A research agenda," *IEEE Comput. Graph. Appl.*, vol. 35, no. 3, pp. 56–64, May 2015.
- [199] M. Wagner, A. Rind, N. Thür, and W. Aigner, "A knowledge-assisted visual malware analysis system: Design, validation, and reflection of KAMAS," *Comput. Secur.*, vol. 67, pp. 1–15, Jun. 2017.
- [200] M. Hummer, S. Groll, M. Kunz, L. Fuchs, and G. Pernul, "Measuring identity and access management Performance—An expert survey on possible performance indicators," in *Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy*, Funchal, Portugal, 2018, pp. 233–240.



MANFRED VIELBERTH received the bachelor's and master's degrees in management information systems with a specialization in cyber security from the University of Regensburg, Germany. He is currently pursuing the Ph.D. degree with the Chair of Information Systems, University of Regensburg. Since February 2017, he has been a Research Assistant with the Chair of Information Systems, University of Regensburg. His research interest includes human aspects in the security analytics domain. On the expert side, this mainly comprises improving processes for better integrating security analysts within a Security Operations Center. In terms of security novices, this primarily covers capturing reports about security incidents in the context of the Human-as-a-Security-Sensor paradigm.



FABIAN BÖHM received the master's degree (Hons.) in management information systems from the Elite Program, University of Regensburg, and the Polytechnic University of Catalonia, Barcelona, in 2016. He is currently pursuing the Ph.D. degree with the Chair of Information Systems, University of Regensburg. Since 2017, he has been a Research Assistant with the Chair of Information Systems, University of Regensburg. His research interest includes the application of

Visual Analytics for cybersecurity. His primary focus within this topic is to leverage Visual Analytics approaches to integrate human domain knowledge into different cybersecurity areas. The core research results show the possibilities offered by Visual Analytics in crucial security domains as Cyber Threat Intelligence, Identity and Access Management, Security Analytics, and Digital Forensics.



INES FICHTINGER received her B.Sc. and M.Sc. degrees in management information systems with a specialization in cyber security from the University of Regensburg, Germany. She is currently working at Deloitte Belgium as a Senior Cyber Security Consultant. Her research interests include security operations and SOC-as-a-service, as well as evaluating the current cyber security posture of companies and helping them design a strategy to reach their desired security posture.



GÜNTHER PERNUL (Member, IEEE) received the diploma and Ph.D. degrees (Hons.) in business informatics from the University of Vienna, Austria. He is currently a Professor with the Department of Information Systems, University of Regensburg, Germany. Previously, he held positions at the University of Duisburg-Essen, Germany; the University of Vienna; the University of Florida, Gainesville; and the College of Computing, Georgia Institute of Technology, Atlanta.

His research interests include data and information-security aspects, data protection and privacy, data analytics, and advanced datacentric applications.

...

7 CTI-SOC2M2 – The quest for mature, intelligence-driven security operations and incident response capabilities

Current status:	Published
Journal:	Computers & Security
Date of acceptance:	15 September 2021
Full citation:	SCHLETTE, D., VIELBERTH, M., AND PERNUL, G. CTI-SOC2M2 – The quest for mature, intelligence-driven security operations and incident response capabilities. <i>Computers & Security</i> 111, 102482 (2021), 1–20
Authors' contributions:	Daniel Schlette 45%
	Manfred Vielberth 45%
	Günther Pernul 10%

Journal description: Computers & Security is the most respected technical journal in the IT security field. With its high-profile editorial board and informative regular features and columns, the journal is essential reading for IT security professionals around the world.



Available online at www.sciencedirect.com

ScienceDirect

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

Computers
&
Security

TC 11 Briefing Papers



CTI-SOC2M2 – The quest for mature, intelligence-driven security operations and incident response capabilities



Daniel Schlette^{1,*}, Manfred Vielberth, Günther Pernul

University of Regensburg, Universitätsstraße 31, Regensburg 93053, Germany

ARTICLE INFO

Article history:

Received 30 April 2021

Revised 7 September 2021

Accepted 15 September 2021

Available online 21 September 2021

Keywords:

Maturity model

Cyber threat intelligence (CTI)

Security operations center (SOC)

Incident response

Security orchestration

Automation and response (SOAR)

Cybersecurity

ABSTRACT

Threats, cyber attacks, and security incidents pertain to organizations of all types. Everyday information security is essentially defined by the maturity of security operations and incident response capabilities. However, focusing on internal information only has proven insufficient in an ever-changing threat landscape. Cyber threat intelligence (CTI) and its sharing are deemed necessary to cope with advanced threats and strongly influence security capabilities. Therefore, in this work, we develop CTI-SOC2M2, a capability maturity model that uses the degree of CTI integration as a proxy for SOC service maturity. In the course, we examine existing maturity models in the domains of Security Operations Centers (SOCs), incident response, and CTI. In search of adequate maturity assessment, we show threat intelligence dependencies through applicable data formats. As the systematic development of maturity models demands, our mixed methodology approach contributes a new in-depth analysis of intelligence-driven security operations. The resulting CTI-SOC2M2 model contains CTI formats, SOC services and is complemented with an evaluation through expert interviews. A prototypical, tool-based implementation is aimed to document steps towards the model's practical application.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Despite the proliferation of advanced systems for cyber defense, it remains a major challenge to build, assess and improve security operations and incident response capabilities within an organization (Ahmad et al., 2021). This situation is paired with sophisticated threat actors in constant search for unprepared and insecure organizations. Besides, a cybercrime economy is monetizing victims' information and vulnerabili-

ties. Consequently, organizations are forced to implement adequate security operations. As different attackers exchange information about publicly known vulnerabilities, exploits, and successful tactics, this is a call to action for information security defenders.

In coping with attackers, the sharing of Cyber Threat Intelligence (CTI) has emerged as an essential measure (Brown et al., 2015). The benefits of collaboration and sharing of contextualized security information about threats, cyber attacks, and security incidents are additional external insights

* Corresponding author.

E-mail addresses: daniel.schlette@ur.de (D. Schlette), manfred.vielberth@ur.de (M. Vielberth), guenther.pernul@ur.de (G. Pernul).

¹ www.go.ur.de/ifs.

<https://doi.org/10.1016/j.cose.2021.102482>

0167-4048/© 2021 Elsevier Ltd. All rights reserved.

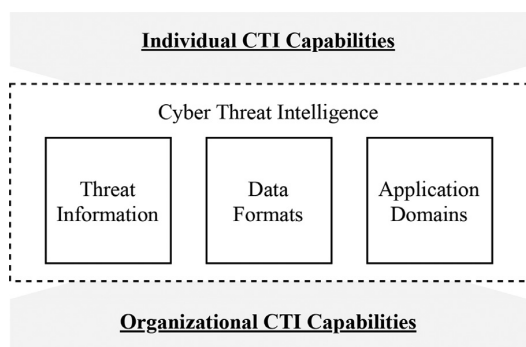


Fig. 1 – Cyber Threat Intelligence concept and capabilities.

and faster, intelligence-driven response. In essence, one organization's security incident description is another organization's threat intelligence. However, using CTI demands a common representation realized with structured data formats and frameworks. Therefore, to leverage CTI formats' full potential, these formats must be sufficiently integrated into organizational processes and tools. CTI is defined as being both actionable threat information (i.e., CTI artifacts) and a comprehensive concept that relates to individuals (i.e., security experts) (Shin and Lowry, 2020) and organizations with security services (see Fig. 1 and Section 3.2).

Security Operations Centers (SOCs) constitute a crucial element bridging CTI and organizational integration (Zimmerman, 2014). By definition, a SOC bundles organizational security roles, essential security services, and tools. Thus, in general, a SOC is responsible for CTI but sometimes dedicated organizational CTI units exist (Brown and Lee, 2021). CTI and its formats enable a well-functioning and effective SOC. While interacting primarily with SOC services, CTI formats also build the foundation of various technologies, such as Security Information and Event Management (SIEM) systems, Intrusion Detection and Prevention Systems (IDS/IPS), or Threat Intelligence Sharing Platforms (TISP).

In order to avoid the proverbial search for the needle in the haystack, SOCs depend on (external) support. We argue that this is possible primarily through a more data-driven approach to detecting and responding to security incidents. Therefore, mature security operations and incident response capabilities in modern organizations rely on data sources integrated via CTI formats (Kokulu et al., 2019). Current studies indicate a need for systematic integration with organizations, technology, and individuals (Lakshmi et al., 2021).

To improve an existing SOC, assessing its current state and finding deficiencies are pivotal objectives. These objectives can be achieved using a Capability Maturity Model (CMM), which fits the determined purpose and scope. Typically, CMMs combine a defined, rigorous academic methodology for development and a practical use case for which the model aims to assess and improve organizational capabilities (de Bruin et al., 2005).

In this work, we seek to better integrate Cyber Threat Intelligence and Security Operations Center. The challenge of assessing and improving intelligence-driven SOC maturity is addressed with an integrated approach. We thereby target stakeholders such as SOC Managers and Chief Information Security Officers (CISO) responsible for an organization's information security management and tactical security operation. With our proposed capability maturity model CTI-SOC2M2 we contribute a comprehensive method towards CTI-based SOC maturity focused on SOC services.

The iterative development methodology we apply leads to the following elementary steps:

- We analyze existing maturity models in the domains of SOC, CTI, and incident response.
- We develop SOC services based on a literature corpus.
- We map CTI formats and SOC services and build an integrated capability maturity model.

The remainder of this paper is structured as follows. Section 2 shows a motivating example and outlines the need for mature SOC services. In Section 3 we introduce SOC and CTI details. Capability maturity model development, including the methodology of this paper, is presented in Section 4. Then, Section 5 covers related maturity models and introduces our integrated three-tiered CTI-SOC2M2 architecture based on CTI formats mapped to SOC services. Maturity assessment with the proposed model and a prototypical implementation of the self-assessment tool build Section 6. Evaluating relevance and applicability form Section 7. Section 8 discusses contributions and limitations, while Section 9 concludes this paper.

2. Motivating example

In early 2021 unknown vulnerabilities in Microsoft Exchange Servers were detected and exploited by various threat actors. In the following section, we use this real-world attack to illustrate the necessity of adequate SOC services and the effective use of CTI and CTI formats. The example emphasizes beneficial aspects of threat intelligence for security operations and the difference between mature and immature SOC services.

Investigations of the Microsoft Exchange Server hack revealed both the timeline of events (Krebs, 2021) and the elementary steps of the attack (Microsoft Threat Intelligence Center (MSTIC), 2021). Based on the detection of 4 vulnerabilities and at first absent and later delayed patching by affected organizations, threat actors (e.g., Hafnium) performed the following actions:

1. **Scan** – The attacker first performs network scans for on-premises Microsoft Exchange Servers with versions susceptible to vulnerability CVE-2021-26855.
2. **Authentication bypass** – The attacker then uses server-side request forgery (SSRF) to bypass authentication and access the server.
3. **Remote Code Execution** – The attacker then uses additional vulnerabilities (e.g., CVE-2021-26857 and CVE-2021-26858) to run code and write files.

4. **Post-exploitation** – The attacker finally installs web shells used for command-and-control communication, exfiltrates information, escalates privileges, drops ransomware, and performs lateral movement.

The compromise of Microsoft Exchange Servers became possible due to threat actors exchanging information about vulnerabilities and exploits before SOCs had access to information on detecting and mitigating the attack. Besides, a lack of vulnerability management and security monitoring supported the widespread exploitation of organizations worldwide.

As it is recommended to patch information systems as fast as possible, situations such as the Microsoft Exchange Server hack point to additional mandatory security operations and the use of CTI and CTI formats. CTI provides the means to detect and mitigate any security compromise swiftly. CVE-IDs have been published on March 2nd 2021² and can be used by SOCs to be aware of the attacker's scan and authentication bypass actions. However, numerous organizations still were breached due to missing processes incorporating CTI. On AlienVault's Open Threat Exchange, CVE-IDs and various indicators of related adversary activity were quickly gathered³. This aggregated CTI can be retrieved using CTI formats such as OpenIOC 1.1 or STIX2.1. Likewise, other CTI sharing platforms list CTI on precise exploits and support export with TAXII⁴. Besides, courses of action to mitigate a compromised Microsoft Exchange Server and connected organizational networks are helpful to pursue incident response⁵. Therefore, it is in the best interest of any SOC to incorporate external CTI and have a thorough understanding of its formats.

Consequently, mature SOC services driven by threat intelligence allow organizations to better defend and mitigate post-exploitation actions. In contrast, SOCs missing external CTI face the complex task of detecting abnormal behavior solely from logs and events. As numerous breaches show, various organizations still struggled to cope with the situation long after CTI could be used. Again, this fact documents the need for mature SOC services.

3. Background

A thorough understanding of information security operations requires consideration of associated organizational concepts. The state-of-the-art of Security Operations Centers (SOCs) and incident response are detailed in the following. Next to these organizational aspects and processes, Cyber Threat Intelligence (CTI) is described. The foundations of threat intelligence are data, data formats, and data sources and relate to security operations.

² <https://nvd.nist.gov/vuln/detail/CVE-2021-26855>.

³ <https://otx.alienvault.com/pulse/6079bf21c21b824801b7a2a5>.

⁴ <https://exchange.xforce.ibmcloud.com/collection/In-the-Wild-Exploits-Seen-Targeting-MS-Exchange-8ec52986bb85fd000a3cf396677fbc1c>.

⁵ <https://github.com/microsoft/CSS-Exchange/tree/main/Security>.

3.1. Security operations center and incident response

To protect IT assets, today's organizations use SOCs. Vielberth et al. (2020) define a SOC as an organizational aspect consisting of four building blocks: people, processes, technology, and governance and compliance. Thereby, governance and compliance provide an encompassing framework. The primary goal of a SOC is to manage and enhance an organization's overall security posture, which usually cannot be achieved by a single entity or system. Instead, it requires a more complex structure. One cause of complexity is the plethora of SOC activities. A SOC creates situational awareness, mitigates security-related risks, and helps to fulfill regulatory requirements. Besides, SOC roles such as SOC analyst or SOC manager are required. The SOC roles are connected to the use of appropriate tools. For instance, SOC analysts typically use SIEM systems to analyze events and identify potential security incidents (Onwubiko, 2015; Zimmerman, 2014).

Of particular interest for both researchers and practitioners is incident response and its relation to SOC. It remains an open question whether incident response is part of a SOC. Two points of view are represented in literature with justifying arguments.

First, if one considers the 24/7 nature of a SOC, security analysts work in shifts around the clock. These analysts mainly analyze events in order to identify possible security-relevant events. Here, incident response and the actions performed are not part of a SOC as they follow incident detection. It can be argued that a dedicated Computer Security Incident Response Team (CSIRT) becomes active only in the case of an incident. Thus, in practice, CSIRT employees usually pursue main activities differently than responding to incidents (Ahmad et al., 2012). In contrast, SOC analysts pursue the analysis of security events full-time, depending on the company's size. This separation between SOC and incident response is based on two components: time and roles.

Second, if one considers the capabilities and activities combined within a SOC, there is an overlap between SOC and incident response, and the clear distinction becomes difficult. Primary SOC tasks such as detection and analysis of incidents and targeted incident response depend upon each other and include feedback loops. Thus, it is necessary to strive for a strong interconnection, if not integration, of SOC and CSIRT. To achieve a strong integration of those two sub-areas, it is of central importance to exchange and manage relevant threat intelligence effectively and efficiently (Onwubiko and Ouazane, 2020).

Finally, the inter-connectedness of SOC and incident response is documented within the incident response life cycle. In essence, this common concept to describe incident response includes several steps iterated in the process (Ab Rahman and Choo, 2015). In the case of the Incident Response Life Cycle developed by the National Institute of Standards and Technology (NIST) (Cichonski et al., 2012) there are four elementary steps. Based on Preparation, incident Detection & Analysis are conducted. These steps are followed by Containment, Eradication & Recovery. Then, the last step covers Post-Incident Activity. For all steps, feedback to previous steps is envisioned. Due to the elements outlined above, we de-

side to take on an integrated SOC and incident response perspective.

3.2. Cyber threat intelligence

Security information and threat reports build the basis of Cyber Threat Intelligence (CTI). The various types of threat intelligence range from *Indicators of Compromise (IoCs)* to *Tactics, Techniques and Procedures (TTPs)* and mitigating *Courses of Action (CoAs)* (Mavroeidis and Bromander, 2017). In addition, security information concerning vulnerabilities, exploit targets, risks and attack attribution are also considered CTI. While the most prominent examples of CTI are the more technical IoCs such as malicious IP addresses, domain names, and malware hashes, CTI includes the means to describe essentially any actionable information related to cyber attacks and security incidents (Tounsi and Rais, 2018).

Besides the threat information itself, the concept of CTI refers to processes to derive relevant knowledge from observed data. Initially, CTI is generated through detailed analysis and contextualization. Then, CTI sharing, including collaboration and dedicated platforms, builds another elementary part of CTI (Skopik et al., 2016). Ultimately, CTI is aimed to contribute to cyber defense by fostering security assessment and improving defensive security measures. Its use, therefore, targets decision-making processes and security operations.

CTI sharing is an essential application domain within the CTI concept and connects it with SOC. CTI sharing involves at least two parties: A producer and a consumer. While a CTI producer (e.g., a security analyst of a manufacturing company or a security vendor) creates CTI based on evidence, a CTI consumer retrieves external threat intelligence for further use. Besides informal and bilateral CTI sharing, dedicated platforms and communities can be involved. For example, an Information Sharing and Analysis Center (ISAC) can collect and distribute CTI to its member organizations via a Threat Intelligence Sharing Platform (TISP). In the context of CTI sharing, employees of SOCs produce and consume CTI on behalf of their organization.

CTI sharing and other application domains demand standardization. On a more granular level, data formats ensure standardization and support certain aspects of data quality assessment (Schlette et al., 2021). They are used to structure the different types of security-relevant information and the threat reports themselves. The role of different CTI formats has been explored by research highlighting their importance as a driving force for specific CTI use cases (Dandurand et al., 2014; Menges and Pernul, 2018).

Finally, associated with the CTI concept, two levels of capabilities can be identified – individual CTI capabilities and organizational CTI capabilities (see Fig. 1). Shin and Lowry (2020) capture the individual-level CTI capabilities and define three distinct CTI capability dimensions required for CTI practitioners to handle CTI artifacts effectively. Based on the Triarchic Theory of Intelligence (TTI), analytical or component intelligence, practical or contextual intelligence, and experiential intelligence represent practitioners' skills of the CTI capability comprising them. In contrast, we propose a capability maturity model that is centered on organizational-level capabilities. Our model has a different, more technical

Table 1 – Two perspectives on progress: the capability and maturity levels (adopted from the CMMI Product Team (2010)).

Level	Capability	Maturity
Level 0	Incomplete	
Level 1	Performed	Initial
Level 2	Managed	Managed
Level 3	Defined	Defined
Level 4		Qualitatively Managed
Level 5		Optimizing

focus integrating CTI and SOC. Nevertheless, its CTI formats, SOC services, and application domains relate to the analytical capabilities of the individual and thus link the two CTI capability levels.

4. Capability maturity model development

Although different maturity models exist, the primary objectives remain the same. As such, a maturity model serves four purposes (Ahern et al., 2004; Becker et al., 2009; de Bruin et al., 2005; CMMI Product Team, 2010): it provides a framework for assessing the current state of an object of interest in terms of its capabilities and other indicators of maturity; it allows the measurement of progress along a path of maturity stages, thus indicating a way for improvement; lastly, it allows benchmarking, though this requires an elaborated maturity model as well as widespread use. Since a SOC can generally be understood as a service provider (Zimmerman, 2014), a closer look at CMMI-SVC's structure and content is warranted (CMMI Product Team, 2010). The model framework consists of various process areas (PAs) or services divided into several thematic service categories. A *service* is defined by its purpose, specific and generic goals, as well as specific and generic practices to achieve those goals. Broadly speaking, whereas specific goals and practices denote the particular features of a service, the generic goals and practices apply across all services and denote the service's level of institutionalization. Institutionalization, in turn, is understood as the level of integration of the process in the overall organization and its consistent performance.

For assessing the current state, the CMMI provides two perspectives: whereas the first provides a view on the capability level of individual services, the second provides a view of the maturity level of multiple services across the organization. As such, an individual service's capability can progress from incomplete to performed, managed, and finally to defined - corresponding to levels 0–3, respectively (cf. Table 1). On the other hand, the organization's maturity can progress from initial to managed, defined, quantitatively managed, and finally to optimizing - corresponding to levels 1–5, respectively.

4.1. Methodology

The proliferation and development of maturity models have considerably increased since the inception of the most popular maturity model: the Capability Maturity Model (CMM) go-

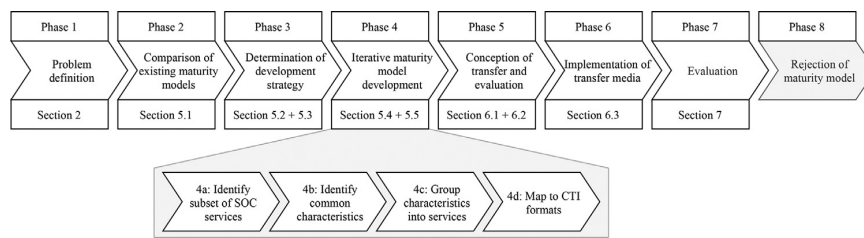


Fig. 2 – Methodology.

ing back to [Humphrey \(1988\)](#). However, as early as 2005, research has pointed out the lack of methodological rigor in maturity model development ([Becker et al., 2009](#); [de Bruin et al., 2005](#); [Mettler, 2009](#)). As [Mettler \(2009\)](#) points out, this is not just a concern for academics. Like [de Bruin et al. \(2005\)](#) before, he argues that the absence of a theoretical basis not just limits the generalizability but also strongly questions the validity of actual appraisals undertaken on the basis of such maturity models. Consequently, practitioners may face results that fail to deliver the promised benefits: acquiring a comprehensive picture of the current state of the object of interest and a path forward detailing where progress is needed to achieve higher stages of maturity.

We address the concerns raised by existing literature and use the methodology shown in [Fig. 2](#) for developing our integrated CTI-SOC2M2. Also, the structure of our paper follows the methodology. It is largely based on the widely-used methodology for maturity model development proposed by [Becker et al. \(2009\)](#). However, since the CMM methodology does not deal with the delimitation of services - which is an essential contribution of CTI-SOC2M2 - in detail, we have supplemented it with [Nickerson et al. \(2013\)](#)'s methodology for taxonomy development. The resulting eight phases of the methodology are briefly described in the following:

Phase 1, problem definition, prescribes that the target domain and the audience of the maturity model be made explicit. In addition, the problem relevance must be justified. In other words, the scope should be clearly defined ([de Bruin et al., 2005](#)). All of these requirements are met in the introductory section and emphasized with the motivating example.

Phase 2, comparison of existing maturity models, serves two purposes. First, it makes sure that research efforts are not wasted on problems that have already been solved. Second, it provides a sound basis for the following phase of determining the development strategy.

Phase 3, determination of development strategy, is entered once it is clear that no existing model suffices. Consequently, [Becker et al. \(2009\)](#) discern between four general development strategies: 1. the design of a completely new model; 2. the enhancement of existing models; 3. the combination of several models into a new one; and finally, 4. the transfer of structures or content from existing models to new domains. We find the principle of cumulative knowledge development applicable. Therefore, there is no need to design a maturity model from scratch, as proposed by strategy one. Further, plenty of literature has used the basic structure and generic terminology pro-

vided in prominent maturity models like CMMI ([CMMI Product Team, 2010](#)). The most sensible strategy for the present case is deemed to be strategy four.

Phase 4, iterative maturity model development, is divided into four steps: selecting the design level, selecting the approach, designing the level, and finally testing the result ([Becker et al., 2009](#)). As an approach to delimiting the SOC services, we have followed the methodology of [Nickerson et al. \(2013\)](#). Thereby the "empirical-to-conceptual approach" applies because of a well-established body of knowledge in the area of SOC. Then, to grasp this knowledge, a structured literature review is conducted. The adapted methodology of [Nickerson et al. \(2013\)](#) comprises the three steps to identify the subset of SOC services, identify common characteristics, and group characteristics into services. To meet the requirements of the intended model, a fourth step includes mapping services and CTI formats.

Those four steps are explained in more detail in the following: (a) *Identify subset of SOC services*, was conducted with the help of literature analysis. Thereby, all publications dealing with SOC capabilities can be considered. In the methodology of [Nickerson et al. \(2013\)](#), capabilities are referred to as objects (as they are not SOC specific). Here, no classification is performed, so the capabilities are available in different degrees of abstraction and can overlap in some cases. (b) *Identify common characteristics*, aims to find properties of the identified SOC capabilities. Thereby, the characteristics that differ across the capabilities and thus enable classification are essential. This step is, to some extent, carried out intuitively, as an objective identification of the properties is not always possible. Based on this, (c) *Group characteristics into services*, is the next operation. The groups form the SOC services, whereby an umbrella term must be identified for each group. Finally, (d) *Map to CTI formats*, connects services by comparing their respective goals and areas of application. For CTI formats, multiple assignments are possible since individual CTI formats can affect several SOC services.

Phase 5, conception of transfer and evaluation, is conducted by adapting standard procedures to our specific requirements. For this, some possibilities are already given by [Becker et al. \(2009\)](#), which can be used as established means. To enable a targeted approach, we first define the requirements for transfer and evaluation and, building on this, carry out phases 6 and 7.

Phase 6, implementation of transfer media, describes the development of a prototypical tool that an employee of an organization can use to perform a maturity assessment of their

SOC. Thereby, particular attention is paid to intuitive applicability and an appealing visual presentation of capability and maturity levels.

Phase 7, evaluation, investigates the relevance and applicability of the proposed maturity model. Evaluation is achieved in two ways: 1) a user study is conducted to illustrate the relevance of the defined problem, and 2) in-depth expert interviews demonstrate the model's practical applicability and relevance.

Phase 8, rejection of maturity model, is about constantly assessing if the maturity model still fits. It re-confirms the results of the evaluation. If the results do not meet the requirements, a new design and development iteration is carried out. However, eventually, a model must be rejected as ever-changing requirements make it impossible to adapt. We list this phase for completeness. Please note that neither reassessment nor rejection due to missing adaptation applies to the iteratively developed model within our work.

5. Integrated SOC capability maturity model

Within an integrated SOC capability maturity model, the SOC concept is accompanied by inter-connected domains. We opt for this approach with an additional focus on CTI due to previously described overlaps between the organizational concepts SOC, CSIRT, and the use of CTI for security operations. In the following, we compare related maturity models specifically targeting SOC, CSIRT, Incident Response (IR), and CTI. We then derive a three-tier capability maturity model architecture. From highest tier to lowest tier, maturity levels, SOC services, and CTI formats specify the integrated model.

5.1. Related maturity models

Organizations seek guidance on how to build, assess and improve capabilities bundled within a SOC. Guidelines and recommendations published in recent years aim to provide details on the many aspects of consideration (Taurins, 2020; Zimmerman, 2014). Also, SOC capability and maturity models have been developed. Towards an integrated and CTI-focused SOC capability maturity model, we address the first research question about the absence or existence of specific maturity models by searching and examining related maturity models. Our initial search was conducted in early 2021 and includes peer-reviewed academic literature and gray literature on maturity models. Existing maturity models can broadly be classified into two groups. The first group includes models proposed by academia, both peer-reviewed and non-peer-reviewed. The second group includes maturity models proposed by organizations and special interest groups in information security. We indicate the origin of a given maturity model according to the two groups. As SOC and incident response typically comprise aspects of CTI, we cover these models before exclusive CTI-related models.

Initiated in 2016, SOC-CMM by Van Os (2016) is a SOC-centered capability maturity model based on a scholarly study. Besides the current version SOC-CMM 2.1 and its MS-Excel self-assessment tool, there exists a separate model adapted to the needs of incident response teams. SOC-CMM covers

the domains of business, people, process, technology, and services in detail. In contrast to SOC-CMM, we argue that CTI is not only a separate service but builds the basis for a variety of SOC services. Therefore, we emphasize SOC services, CTI dependencies, and the intelligence-driven underlying of security operations and incident response. Another extension to SOC-CMM is introduced within the academic work of Acartürk et al. (2020). The authors conclude that generic continuous improvement methods from the field of quality management can be supportive elements.

Based on the results of a comparative study on generic cybersecurity capability maturity models (Rea-Guaman et al., 2017), we take the *Cybersecurity Capability Maturity Model (C2M2)* into account. As C2M2 is cybersecurity-oriented and targeted at an organizational environment, it relates to core SOC characteristics. C2M2 is a collaborative product of the US Department of Energy and Carnegie Mellon University and includes ten domains (e.g., threat and vulnerability management) for which objectives are defined (Christopher et al., 2014). When we define SOC services, these domains are assessed and either aligned or excluded based on their granularity and fit.

A prominent maturity model for CSIRT is the *Security Incident Management Maturity Model (SIM3)* introduced by Stikvoort (2015). SIM3 is a maturity model and online self-assessment tool currently provided by the non-profit Open CSIRT Foundation (OCF). Organizations within the incident response community use SIM3 to certify organizational incident management. While covering essential capabilities in the domains of tools and processes, we identify a gap within SIM3 regarding the systematic use of CTI for organizational processes.

Maturity assessment of incident response with a lesser focus on organizational integration is addressed by the non-profit CREST (The Council for Registered Ethical Security Testers). The *Cyber Security Incident Response Assessment Tool (CSIR-MAT)* covers basic assessment for three phases: prepare, respond, and follow-up. However, CREST also provides the *Cyber Threat Intelligence Maturity Assessment Tool (CTI-MAT)*. Here, threat intelligence is structured according to a life cycle from direction to review. We find that the processing and dissemination operations in CTI-MAT include CTI formats.

Other maturity models – CTI-CMM (Lourenco, 2018) and CTIM (Luchs and Doerr, 2020) – are outlined by ENISA, and by researchers at Hasso Plattner Institut and TU Delft respectively. Both introduce a CTI life cycle focusing on data and its categorization. We identify maturity levels and indicators valuable for maturity assessment.

Finally, an academic proposal towards a CTI maturity model is introduced by Sillaber et al. (2018). The high-level maturity model is derived from previously conducted expert interviews and focused on inter-organizational CTI sharing. Application scenarios for CTI play a role for more mature organizations. Thus, we consider the use of CTI formats a reasonable condition for maturity improvement.

Foundations. All related maturity models adhere to the standard multi-tier structure with different maturity levels, different capability levels, or both. Besides, maturity models for SOC, CSIRT, or incident response typically cover people, processes, and technology. We conclude that these foundations provide both guidance and flexibility for develop-

ing an integrated and CTI-focused SOC capability maturity model.

5.2. Design decisions

Capability maturity models consist of standard components and have specific characteristics. Therefore, the development of a CMM is accompanied by multiple design decisions. In the following, we outline the design decisions of our model, which will be detailed in the subsequent sections of this paper. Please note that the CMMI, the methodology by Becker et al. (2009), and the arguments below guide our design decisions.

Objective. CMMs fulfill one or more specified objectives. The objective of our model is the development of a mature, intelligence-driven SOC. We reason that as data and its analysis are key to successful business operations (business intelligence), the same holds for security operations (threat intelligence).

Scope. CMMs are defined by their scope. The scope of our model is the operationalization of CTI in SOC. We focus on organizational CTI capabilities required for SOC services. Consequently, we build on widely used CTI formats as these are crucial for understanding and using CTI.

Users. CMMs have a specified target audience. Our model is aimed at SOC and information security personnel. Most CMMs address managerial positions with executive powers. We envision SOC managers, SOC consultants, and CISOs to apply our model within organizations.

Tiers. CMMs consist of hierarchically structured elements for which we use the term *tiers*. From top to bottom, maturity levels, capabilities and capability levels, and indicators represent the three standard tiers in CMMI. We define four maturity levels to capture SOC maturity concerning CTI. Therefore, we adapt the CMMI naming convention. Capabilities in our model are represented by six SOC services identified in the literature. The decision for SOC services as capabilities is influenced by the model's scope. Our model further includes six capability levels to show the implementation of a given SOC service. Capability levels are closely linked to indicators. CTI formats serve as indicators and, mapped to the SOC services, determine capability levels. The design decision for CTI formats is based on the assumption that organizational use of CTI involves formats. Thus, indicative questions based on CTI focal points address the degree of CTI format coverage and represent the capability levels.

Mapping. CMMs must deal with mapping the individual tiers. At the center of our model, we tie CTI formats and SOC services together and thus realize mapping indicators to capabilities. We leverage the well-known NIST incident response life cycle to map capabilities and their levels to maturity levels. We argue that SOCs aim for complete life cycle coverage. However, as resources are scarce, SOCs will improve CTI maturity step-by-step and start with preventive measures. Also, for the degree of CTI format consideration, we assume a successive approach with source, quality, and integration of CTI formats elementary for organizational use and backed by CTI literature.

Approach in a nutshell. In short, our model is based on how well organizations handle CTI formats. Mapping CTI for-

ats to SOC services allows determining a capability level for these services. Then, all SOC services combined determine the intelligence-driven maturity of the SOC (see Fig. 3).

5.3. Architecture

Having discussed related maturity models, we determine a maturity model development strategy according to the procedure put forward by Becker et al. (2009). As existing maturity models already provide an initial setup, we dismiss the option to develop an entirely new capability maturity model. Instead, we integrate and refine existing elements and direct focus on the area of CTI and CTI formats. The decision for CTI formats is influenced by the significance of the CTI concept for operational cybersecurity (Brown and Lee, 2021; Shin and Lowry, 2020) and formats required as its core. We further reject the option to use the industrial Software Process Improvement and Capability Determination (SPICE) methodology (Dorling, 1993). This decision in favor of the CMMI approach is influenced by the scientific prevalence of CMMI and its more general direction than SPICE's engineering aspects.

CTI-SOC2M2. The architecture of our CTI-focused model, referred to as **CTI-SOC2M2**, is visualized in Fig. 3. From left to right, the architecture consists of three tiers – CTI formats, SOC services and the associated capability levels, and maturity levels for the intelligence-driven SOC.

On the lower tier of CTI-SOC2M2, threat intelligence focus is realized via *CTI formats*. The CTI formats in this tier function as an indicator and directive for capability fulfillment, eventually leading to maturity assessment. CTI formats being part of the CTI concept represent organizational CTI capabilities as they link CTI artifacts and application domains. In the context of the CTI-SOC2M2, the CTI formats are assigned to SOC services.

On the central tier, we organize *SOC services* representing procedural elements. CTI-SOC2M2 is a capability maturity model centered exclusively on services. We opt for this approach because SOC services are integral to the operationalization of CTI. Nevertheless, it is worth mentioning that complementary maturity models and frameworks (e.g., ISO 27001) covering governance and people exist. As SOC services use technologies, where applicable, we point to relevant dependencies.

Both the central tier and the lower tier document the decision for an integrated maturity model. As it can be observed that maturity models in highly-specific domains exist (e.g., Digital Forensic Readiness (Engbrecht et al., 2020)), we aim to combine detailed elements of CTI and SOC. Therefore, we map CTI formats to SOC services allowing capability assessment for SOC services via the CTI formats. CTI, in general, has the benefit of introducing external insights into threats not (directly) visible within an organization. Besides, we also aim to provide enough differentiation from generic information security maturity models by integrating an extensive SOC study's research results.

On the upper tier, *maturity levels* indicate the current state of SOC maturity with regard to CTI. These maturity levels are dependent on the capability levels reached by individual SOC services and guide step-wise improvement.

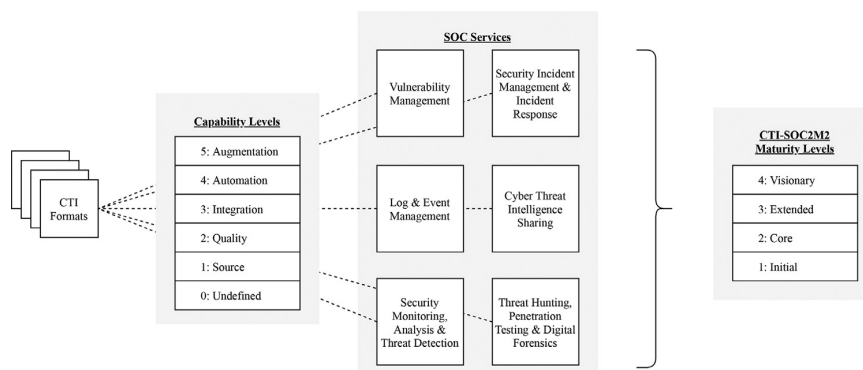


Fig. 3 – CTI-SOC2M2 Architecture.

Table 2 – Related maturity models.

Model	Documentation	Category
SOC-CMM	Van Os (2016)	SOC
SOC-CMM*	Acartürk et al. (2020)	SOC
C2M2	Christopher et al. (2014)	SOC
SIM3	Stikvoort (2015)	CSIRT
CSIR-MAT	CREST (2014)	IR
CTI-MAT	CREST (2016)	CTI
CTI-CMM	Lourenco (2018)	CTI
CTIM	Luchs and Doerr (2020)	CTI
CTI-MM	Sillaber et al. (2018)	CTI

Table 3 – CTI formats and frameworks.

CTI Category	CTI Format
Threat Report	IODEF
	MISP
	STIX
	TAXII
	VERIS
	CACAO
Course of Action (CoA)	OpenC2
	RE&CT
	CAPEC
Tactic, Technique, Procedure (TTP)	Cyber Kill Chain
	Diamond model
	MITRE ATT&CK
	CVSS
Scoring System	CPE
	CWE
Security Enumeration	CVE
	GENE (logs)
Indicator of Compromise (IoC)	OpenIOC
	Sigma (logs)
	Snort (traffic)
	YARA (files)
	Zeek (traffic)

From a functional perspective, CTI-SOC2M2 is based on qualitative and quantitative measures. Qualitative assessment of CTI formats leads to a capability level for the mapped SOC service. It is worth mentioning that organizations can deem CTI formats not applicable for their context and omit these from qualitative capability assessment. Thus, high capability levels can be achieved in our model by considering only a few selected CTI formats. In contrast, the overall intelligence-driven SOC maturity is based on a quantitative approach. Therefore, capability levels of different SOC services are considered and assessed (later discussed in Section 6.2).

5.4. CTI formats

Before mapping CTI formats to SOC services, an over-view of relevant formats and frameworks is displayed in Table 3. We base our selection of these formats and frameworks on the influential work by Dandurand et al. (2014), consider promising new format developments, and ignore deprecated formats. Format categorization is based on generic CTI elements (Mavroeidis and Bromander, 2017).

IoC formats are often closely related to defensive security systems such as SIEM tools, IDS, or IPS. Their primary purpose is the identification of rules or patterns within logs, network traffic, or files. Security enumerations are used to identify relevant artifacts (e.g., IT assets or vulnerabilities). A systematic approach to assist decision-making is described by scoring systems. While TTP formats cover mainly attacker behavior

and the offensive side, CoA formats describe procedural countermeasures conducted as a method of defense. Finally, encompassing threat report formats handle aggregated CTI and include aspects of other CTI formats.

5.5. SOC services

SOC services are the foundation for SOC operation. We use primary data from a previous SOC study (Vielberth et al., 2020) for the definition and iterative development of SOC services. The data used is the result of a structured literature review. The literature corpus covers 158 publications on SOC in general⁶. We only consider parts of the literature helpful and thus explicitly cite only highly relevant works. However, we want to

⁶ <https://ieeexplore.ieee.org/ielx7/6287639/8948470/9296846/supp1-3045514.xlsx?arnumber=9296846>.

Table 4 – SOC field studies and identified SOC services.

Reference	SOC Services
Kowtha et al. (2012)	Log & Event Management Analysis Security Incident Management Incident Response CTI Sharing Threat Hunting
Jacobs et al. (2013)	Vulnerability Management Log & Event Management Security Monitoring Analysis Incident Response Threat Hunting Penetration Testing
Onwubiko (2015)	Vulnerability Management Log & Event Management Analysis Threat Detection Incident Response CTI Sharing Digital Forensics
Settanni et al. (2017)	CTI Sharing CTI Usage

point to details and supplementary material of the literature review, which we use as starting point.

SOC elements of varying granularity can be categorized into topic areas. This categorization is of particular interest for the clustering of SOC services and CTI-SOC2M2 development. We perform an in-depth examination of publications that cover the two topic areas capability or maturity. Outlined SOC processes (e.g., Kowtha et al. (2012); Schinagl et al. (2015)) provide further guidance for the development of SOC services.

This knowledge is then combined with the methodology by Nickerson et al. (2013) to form clusters which we name SOC services. The approach by Nickerson et al. (2013) is particularly suitable as our classification is essentially a single-layer taxonomy. Thereby, it pays special attention to mutual exclusiveness, which, according to de Bruin et al. (2005), is a major requirement for the defined services. Besides, we use existing and sufficiently documented SOC maturity models (e.g., SOC-CMM (Van Os, 2016)) to iteratively validate our SOC services. Mapping of SOC services and CTI formats is then assisted by core aspects defining each SOC service and work on CTI format categorizations (Hernandez-Ardieta et al., 2013).

Even though the majority of the papers in the literature corpus deal with SOC services, SOC is usually not considered holistically. Thus, often only selected sub-areas or sub-services are analyzed. In particular, there is a strong focus on log & event management and analysis. Table 4 lists the most important field studies that take a holistic view of a SOC, as these are particularly important for identifying SOC services. Within the table the identified SOC services within these studies are listed.

5.5.1. Vulnerability management

Existing vulnerabilities define threats to information systems and IT infrastructure. Vulnerability management is part of

different maturity models, industry standards, and SOC research (Farris et al., 2018). However, as vulnerability management deals with an adequate handling of known vulnerabilities, it is a SOC service influenced by CTI (Chismon and Ruks, 2015). Therefore, maturity assessment must consider relevant CTI formats and sources such as exploit and vulnerability databases. Applicable CTI formats to vulnerability management cover both security enumerations and scoring systems. They provide a common understanding, reference, and assessment of vulnerability severity to guide decision-making. Additionally, vulnerabilities relate to IT assets.

CTI formats:

- CPE – Common Platform Enumeration allows reference and identification of classes of IT assets. In its current version 2.3., CPE is maintained by the National Institute of Standards and Technology (NIST). Each IT asset's characteristics are described via string-based format (Cheikes et al., 2011).
- CVE – Common Vulnerabilities and Exposures is a security enumeration to refer to vulnerabilities in IT assets uniquely. It is maintained by MITRE and a community. Its data constitutes the basis for the US National Vulnerability Database (NVD) (Baker et al., 1999).
- CVSS – Common Vulnerability Scoring System version 3 provides a formal procedure to specify the severity of a vulnerability ranging from 1 to 10. It is maintained by FIRST and can be applied in different modes, including a primary assessment and consideration of organizational and environmental factors (Forum of Incident Response and Security Teams (FIRST), 2019).

Referring to the motivational example, the latest version of the Microsoft Exchange Server affected is specified as `cpe:2.3:a:microsoft:exchange_server:2019:cumul._update_8` and CVE-2021-26855 is CVSS-rated 9.8.

5.5.2. Log and event management

Logs and events capture information about system processes and system states. As a consequence, log and event management is concerned with internal data required for security analysis. Being part of various industry standards, IT operations, and SIEM tools, this service is essential for SOC (Madani et al., 2011). But, to conduct effective log and event management, external CTI can foster security assessment and alignment to security goals. It is thus necessary to consider data formats describing attacker behavior documented by logs and system events. These data formats go beyond the essential log formats such as Syslog (Gerhards et al., 2009), NCSA (Apache HTTP Server Project, 1995), EVTX (Microsoft, 2018), or Common Event Format (CEF) (ArcSight, 2010) and describe threat detection patterns.

CTI formats:

- GENE - Go Evtx sigNature Engine and rule format version 1.6 aims to provide signatures for Windows event logs. The open-source format proposed in 2018 by RawSec company centers on JSON-described rules (RawSec - Quentin Jerome, 2018).

- Sigma – Generic Signature Format for SIEM Systems is an open-source format aimed at log files and log events. The project driven by [Roth and Patzke \(2017\)](#) includes a YAML-based format specification to describe threat identifiers and allow detection.

Referring to the motivational example, a Microsoft Exchange Server captures its logs and events in .evtx-files. Using GENE might be a feasible approach to determine anomalies.

5.5.3. Security monitoring, analysis & threat detection

Security monitoring is a continuous approach to ensure an organization's security goals. At the center of a SOC, security monitoring copes with an aggregate view of IT assets and their security ([Onwubiko, 2015](#)). In conjunction with security monitoring, security analysis and threat detection can yield additional insights into specific security aspects and identify threats. While it is possible to conduct security monitoring, analysis, and threat detection without threat intelligence, the general threat landscape can provide essential clues. Contrasted with CTI on current malware, command and control servers, and ongoing cyber attacks, variations witnessed in network traffic and system behavior allow organizations to initiate appropriate follow-up steps. CTI formats applicable for this SOC service mainly include information on IoC. As an example for additional CTI, we also list the MITRE ATT&CK framework and the comprehensive STIX format.

CTI formats:

- OpenIOC – The OpenIOC format allows description of different types of IoCs. Developed by Fireeye (formerly Mandiant) in 2013 the current schema version 1.1 is XML-based, open-source, and has a criteria section to match specified values against, for example, files or processes ([Ross et al., 2013](#)).
- Snort – The Snort format provides rules for detecting network traffic threats in combination with the open-source IDS/IPS tool. Snort is maintained by Cisco Talos and a community. Its rules are based on a custom schema and describe actions and detection parameters ([Snort Team, 2021](#)).
- Zeek – The Zeek signature format for network traffic supports matching threat patterns. The format is part of the open-source Zeek (formerly Bro) network security monitoring tool, which contains additional components. Maintenance is realized by a community ([The Zeek Project, 2021](#)).
- MITRE ATT&CK – The ATT&CK framework categorizes and details adversary behavior with tactics, techniques, and mitigation procedures. Maintained by MITRE, the framework evolved and comprises both information for IT and OT environments ([Strom et al., 2018](#)).
- STIX – In version 2.1, Structured Threat Information Expression (STIX) is a comprehensive CTI format capturing low-level cyber-observables and information on TTPs, CoAs and their dependencies. Initiated in 2012, STIX is maintained by OASIS and centers on JSON-based threat reports ([OASIS Cyber Threat Intelligence \(CTI\) Technical Committee, 2020a](#)).
- see also CPE, CVE, CVSS, and Sigma.

Referring to the motivational example, available OpenIOC-based indicators for malware assist detection of a Microsoft Exchange Server compromise. Other malicious IP addresses enable alerts in IDS and are grouped in STIX2.1 threat reports.

5.5.4. Threat hunting, penetration testing & digital forensics

Threat hunting, penetration testing, and digital forensics are all concerned with detailed investigations. In-depth analyses aggregated in this SOC service go one step further than security monitoring and aim to find evidence of ongoing attacks, malware, existing vulnerabilities, and procedural deficiencies. As it is common practice to conduct threat hunting, penetration testing, and digital forensics to test actively and identify incidents ([Hámornik and Krasznay, 2018](#)), these activities rely on information. While it is necessary to resort to internal information, this is often not sufficient. However, the use of external CTI can integrate typical attack patterns and other identifying elements. Therefore, CTI formats describing TTPs, software weaknesses, and IoCs are of relevance. Together with appropriate technology, comprehensive CTI formats such as MISP can further assist specific actions.

CTI formats:

- YARA – The YARA rule format allows to describe patterns to match against files. Developed at VirusTotal, the open-source YARA tool and format support detection of malicious files. Community driven open-source rule repositories exist ([VirusTotal - Victor Alvarez, 2014](#)).
- CWE – Common Weakness Enumeration is focused on software flaws. It is maintained by MIRE and lists software weaknesses by three categories (i.e., software development, hardware design and research concepts) ([MITRE, 2020](#)).
- CAPEC – Common Attack Pattern Enumeration and Classification is used to refer to attack patterns. Maintained by MITRE, CAPEC is focused on common application weaknesses and categorizes patterns by mechanisms and domains of attack ([CAPEC Team, 2020](#)).
- Cyber Kill Chain – Various cyber kill chains exist. For example, the Lockheed Martin cyber kill chain describes various stages of an attack and thereby assists detection and defense ([Hutchins et al., 2011](#)).
- Diamond model – The diamond model supports intrusion analysis with four adjacent categories: adversary, capability, victim, and infrastructure. These define core characteristics of an adversary and its campaign ([Caltagirone et al., 2013](#)).
- MISP – Open Source Threat Intelligence Platform and format centers on events, attributes and tags to comprehensively describe threat intelligence. The open-source project supported by Computer Incident Response Center Luxembourg (CIRCL) and the European Union allows CTI collection and sharing ([Wagner et al., 2016](#)).
- see also CVE, OpenIOC, Snort, and MITRE ATT&CK.

Referring to the motivational example, analysis of other attack campaigns by threat actors exploiting the Microsoft Exchange Server vulnerability is relevant. Analysis and threat hunting can start with threat actor Hafnium and its use of web shells.

5.5.5. Security incident management & incident response

A security incident may have various causes potentially leading to harm for an organization. As a security incident is a type of event that violates security policies, it is essential to manage security incidents and respond with appropriate measures. One aspect of managing is incident triage leading to a prioritization of actions based, for example, on impact or available resources (Shah et al., 2019). The greater concept of security incident management and incident response is currently gaining momentum. A variety of dedicated Security Orchestration, Automation and Response (SOAR) systems (Neiva et al., 2020) and the underlying incident response standardization aim to establish a more efficient approach. Besides, for incident response training, cyber ranges are discussed. Thus, combined with existing ticketing systems, it becomes necessary to consider CTI formats centering on CoAs to support these use cases.

CTI formats:

- CACAO – Collaborative Automated Course of Action Operations for Cyber Security format version 1.0 is centered on incident response workflows. CACAO is maintained by OASIS. It is capturing information about procedural logic and actions in JSON-based playbooks and supports sharing (OASIS, 2021).
- OpenC2 – Open Command and Control format version 1.0 represents commands for incident response machine-to-machine communication. Maintained by OASIS, granular OpenC2 actions are JSON-based and transferred to defensive systems for execution (OASIS, 2020).
- RE&CT – The RE&CT framework includes a matrix representation of incident response stages and actions. Besides, the ATC project behind RE&CT introduced YAML-based playbooks (ATC Project, 2020).
- see also CVE, OpenIOC, STIX, and MISP.

Referring to the motivational example, integrating the PowerShell script into a CACAO playbook is beneficial for the incident response workflow. IP addresses belonging to Command-and-Control infrastructure can be blocked using a firewall and initiating an outbound traffic re-direct with OpenC2 message.

5.5.6. Cyber threat intelligence sharing

Cyber Threat Intelligence is not only part of other SOC services but also a SOC service of itself. Based on incident reporting, the Cyber Threat Intelligence sharing service copes with comprehensive threat reports. Ensuring adequate gathering of external information and internal dissemination is at the center of this SOC service. Nevertheless, using CTI demands comprehensive and structured CTI formats which encapsulate relevant threat information. Therefore, threat report formats are included to build a knowledge base about cyber attacks, threats, and security incidents. The sharing aspect of CTI is also incorporated in threat report formats focusing on data transfer.

CTI formats:

- IODEF – Incident Object Description Exchange Format version 2 supports the representation and the exchange of security incident reports and indicators. The IETF standard

is based on XML and includes objects for different types of CTI (Danyliw, 2016).

- TAXII – Trusted Automated eXchange of Indicator Information format version 2.1 is used for transferring and collecting STIX-based threat reports. Associated with STIX and maintained by OASIS, the framework includes services, HTTPS transport, and client-server architecture to facilitate the sharing of CTI (OASIS Cyber Threat Intelligence (CTI) Technical Committee, 2020b).
- VERIS – Vocabulary for Event Recording and Incident Sharing version 1.3.2 enables incident description based on the categories actors, actions, assets and attributes. Developed by Verizon, it is open-source and uses JSON to represent CTI. A VERIS Community Database (VCDB) includes public security incidents (VERIS Community, 2021).
- see also Sigma, YARA, STIX, and MISP.

Referring to the motivational example, using a TAXII server to query CTI provides fast access and supports following SOC services. Also, IODEF and VERIS can document security incidents and build a historical database with both high- and low-level incident descriptions.

6. Maturity assessment with CTI-SOC2M2

The use of CTI-SOC2M2 for maturity assessment is based on CTI formats that serve as indicators. These indicators and indicative questions allow organizations to self-assess their current capability and maturity level and show steps towards improvement.

6.1. CTI formats and capability levels

Typically, capability maturity models use indicative questions to determine capability fulfillment and the associated capability level. We follow a generic and qualitative approach applicable to all individual CTI formats and SOC services. As the indicative questions for capability levels pertain directly to CTI and CTI formats, similarities with maturity models designed solely for CTI exist. It is also worth mentioning that capability levels are built upon each other. Lower levels, e.g., *Source* and *Quality*, are necessary requirements to reach the next capability level (e.g., *Integration*). After selecting applicable CTI formats according to the specific organizational setting, the following categories and indicative questions must be answered to assess any given CTI-based SOC service.

Capability Levels:

- 0: Undefined** – CTI and CTI formats have not yet been considered.
- 1: Source** – Have you determined and assessed the source of CTI with the mentioned CTI format(s)?
- 2: Quality** – Have you applied appropriate measures to assess the quality of the CTI structured with the mentioned CTI format(s)?
- 3: Integration** – Have you integrated CTI and the mentioned CTI format(s) into your organizational processes and technology architecture?

- 4: **Automation** – Have you automated retrieval, use and dissemination of CTI based on the mentioned CTI format(s)?
- 5: **Augmentation** – Have you set-up a monitoring mechanism to cope with new developments within CTI and new CTI format(s)?

We determined the capability levels and indicative questions by considering focal points of the CTI concept. Previous studies emphasized the importance of CTI sharing platforms and the source of CTI (Bauer et al., 2020; Bouwman et al., 2020). Besides, the quality of CTI and the expressiveness of CTI formats are highly relevant (Li et al., 2019; Schaberreiter et al., 2019; Schlette et al., 2021). SOC services are specified by more granular processes, which must handle the integration of CTI, its formats, and technology. Integration builds a prerequisite to fully achieve effectiveness via automation. Towards the ultimate goal of security orchestration, automation, and incident response, CTI is one essential element (Islam et al., 2019). However, the current developments show that the state-of-the-art of CTI is constantly shifting (Brown and Lee, 2019). Therefore, it is necessary to monitor and continuously extend the organizational understanding of CTI formats and associated concepts.

6.2. SOC services and maturity levels

SOC services are assessed based on the CTI formats and indicative questions mentioned in Section 6.1. Transitioning from SOC services and capability levels to overall CTI-SOC maturity levels demands a methodology outlined below.

Maturity Levels:

- 1: **Initial** – Capability level 2 is reached for Log & Event Management, Security Monitoring, Analysis & Threat Detection and Vulnerability Management.
- 2: **Core** – Capability level 2 is reached for Security Incident Management & Incident Response and Cyber Threat Intelligence Sharing. All previous services reached capability level 3.
- 3: **Extended** – Capability level 2 is reached for Threat Hunting, Penetration Testing & Digital Forensics. All previous services reached capability level 3.
- 4: **Visionary** – Capability level 4 is reached by all SOC services.

We first define four maturity levels: *Initial*, *Core*, *Extended*, and *Visionary*. The naming of these maturity levels indicates SOC functionalities addressed by CTI. Improvement of CTI-SOC maturity within an organization depends on the individual fulfillment levels for the SOC services. As CTI-SOC2M2 adheres to a step-wise approach, maturity levels are downgraded if underlying SOC service capabilities cease to exist.

Our methodology for CTI-SOC2M2 is inspired by the NIST Incident Response Life Cycle (Cichonski et al., 2012). Reaching higher maturity levels is equivalent to addressing more aspects of the Incident Response Life Cycle more thoroughly (see Fig. 4). Whereas organizations aim to implement all the individual aspects of the incident response life cycle, we envision organizations with limited resources and new to the concept of CTI and CTI formats to approach CTI-driven SOC maturity

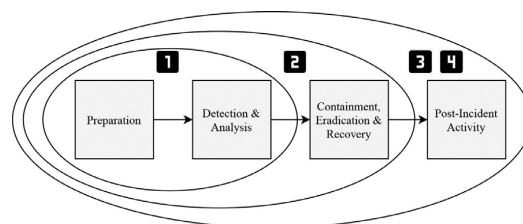


Fig. 4 – Maturity level scope compared to NIST Incident Response Life Cycle (Cichonski et al., 2012).

step by step. Consequently, an organization might not cover all its aspects concerning CTI formats but can still have some generic measures in place. Also, organizations might decide to remain on a specific maturity level due to the delegation of certain SOC services.

Therefore, with *Log and Event Management*, *Security Monitoring*, *Analysis & Threat Detection* and *Vulnerability Management* on capability level 2: *Quality* an initial maturity level is reached and likewise aspects of preparation, detection and analysis of the incident response life cycle covered by CTI formats. Progress towards core maturity is possible when the aforementioned SOC services reach level 3: *Integration* and additionally *Security Incident Management & Incident Response* and *Cyber Threat Intelligence Sharing* reach level 2: *Quality*. This is accompanied by covering containment, eradication & recovery of the incident response life cycle with CTI formats. Including *Threat Hunting*, *Penetration Testing & Digital Forensics* on level 2: *Quality* as well as progressing the other SOC services towards 3: *Integration* leads to an extended CTI-SOC. This further implies covering post-incident activity of the incident response life cycle in detail. Finally, starting with at least capability level 4: *Automation* for all SOC services the visionary maturity level is reached. Aspects of the incident response life cycle are advanced and additional progress with 5: *Augmentation* for SOC services is still possible.

Motivational Example. We want to document the CTI-focus of an illustrative SOC which reached the *extended* maturity level and emphasize aspects of the motivational example introduced in Section 2. While an acceptable maturity is already reached with the *core* maturity level, this illustrative SOC covers all SOC services with at least integrated CTI formats. Concerning the Microsoft Exchange Server breach, the organization witnessed a compromise but has a sufficiently integrated log and event management where the public Sigma rule⁷ has been analyzed and is part of the organizational CTI process for SIEM systems. Security monitoring, analysis, and threat detection integrate SIEM systems, IDS, and firewalls with CTI formats leading to the detection of network traffic to IP 218.103.234.[.]104 also listed in a queried STIX2.1 threat report. Vulnerability management covers processes considering newly published CVE-IDs. This is aimed to avoid missing other related vulnerabilities (e.g., CVE-2021-27065). NVD is actively and regularly searched, compared to CVEs in threat re-

⁷ https://github.com/SigmaHQ/sigma/blob/master/rules/web/web_exchange_exploitation_hafnium.yml.

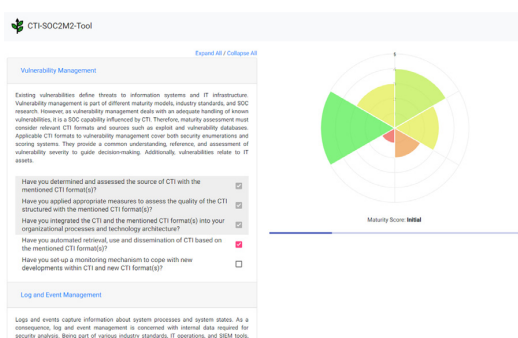


Fig. 5 – CTI-SOC2M2 self-assessment tool with radar chart for SOC services.

ports, and complemented with other vulnerability databases (e.g., OSV - Open Source Vulnerabilities⁸). For CTI sharing, the organization hosts its own MISP instance to share security information between the headquarter and branches. The security incident management and incident response SOC service is triggered by the high severity indicated by the CVSS score and focused on web shell removal. A CACAO playbook is used to define and structure the granular actions. A suspicious file `shell.aspx` was submitted to OTX, deleted, and the Microsoft Exchange Server patched. Threat hunting, penetration testing, and digital forensics were then conducted for in-depth analysis. Based on the different types of web shells the SHA-1 hash `eb8d39ce08b32a07b7d847f6c29f4471cd8264f2` was found and its actions analyzed. As a result, a YARA rule was adapted.

With SOC services this mature, the illustrative organization avoided the exfiltration of large amounts of data and a widespread lateral movement by attackers on the organizational networks. In this particular case, patches became available only after the first successful exploits by threat actors had been conducted. As a result, even a mature CTI-driven SOC could not prevent the initial attack. However, as response measures were applied promptly, further negative consequences could be limited. In addition, risk management was always aware of the ongoing operations and could advise decisions.

6.3. Prototypical implementation

The presented maturity model is based on the assumption that employees within an organization record the capability level for each SOC service. In this scenario, self-assessment is best supported with a suitable tool. Essentially, the tool provides two functions. On the one hand, it can be used to record capability levels. On the other hand, it will calculate and display the overall SOC maturity based on the methodology outline in Section 6.2.

Fig. 5 depicts the structure of the prototypical CTI-SOC2M2 self-assessment tool. A demo version of the implementation can be accessed online (<https://antumin.github.io/>

⁸ <https://osv.dev/>.

[CTI-SOC2M2/](#)). The tool layout is divided into two parts. On the left-hand side, the CTI-based capability levels for each SOC service can be recorded. For this purpose, a description is intended to help the user understand the SOC service characteristics. A drop-down menu then allows the user to select a capability level referring to CTI and CTI formats. On the right-hand side, the maturity level is displayed. The maturity level of the assigned to the overall SOC is stated, and a radar chart visualizes the capability breakdown for the SOC services. This visualization enables users to immediately identify deficient SOC services and improve CTI efforts to progress towards a more mature SOC.

From a technical perspective, the tool is implemented as a web app. This decision allows for platform-independent use independent from other commercial software such as Microsoft Excel, typically used for maturity models. The web app was developed using the Angular⁹ framework, based on HTML, Javascript, and CSS. The source code is open-source and published on GitHub¹⁰ enabling further development and future research.

7. Evaluation

This section concludes the methodology of capability maturity model development outlined earlier. We use a mixed-method approach, combining a quantitative user study with a qualitative evaluation based on expert interviews. With the two components, we aim to document relevance and applicability.

User study

To show the relevance of SOC maturity and its threat intelligence focus, we conducted an international user study.

Design & Procedure: The impact of a security analyst's CTI-based skills on the ability to detect attacks is explored with three phases:

1. *Assessment of pre-knowledge and attack detection skills:* During the first phase, participants are asked questions using a questionnaire that measures their pre-knowledge of CTI formats. In addition, the participants are asked how accurately they can detect attacks and how extensive their attack detection knowledge is.
2. *CTI-based SOC training:* The second phase forms a training session. Participants learn to understand a CTI format for describing detection rules and indicators of compromise using dSIEM¹¹, an open source SIEM system based on Elasticsearch¹². For this purpose, introductory videos and texts are provided. The SIEM system is made available and data from real attacks is inserted.
3. *Assessment of post-knowledge and attack detection skills:* In order to verify the effect of the training on the participants

⁹ <https://angular.io/>.

¹⁰ <https://github.com/antumin/CTI-SOC2M2>.

¹¹ <https://www.dsiem.org/>.

¹² <https://www.elastic.co/>.

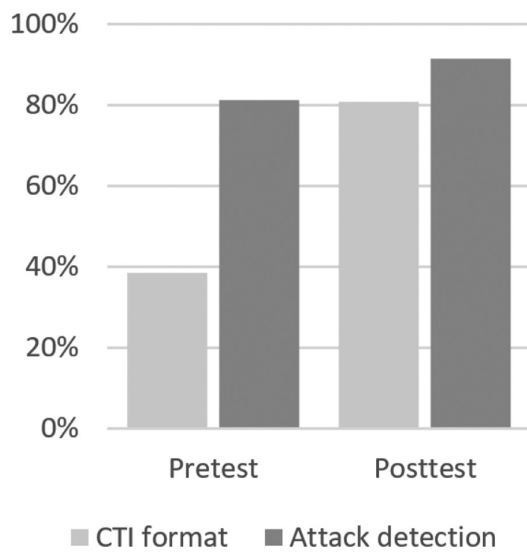


Fig. 6 – Participants' knowledge on CTI formats and attack detection before and after the user study.

skills to detect attacks, the assessment with the questionnaire from phase 1 is conducted a second time.

The user study was conducted at a German and a Greek university with $n = 44$ participants. All of them were students with an IT security background. 22 participants were Greek and 22 German, whereby 12 female and 32 male students participated. 20 students were undergraduate, and 24 postgraduate. The complete data set of the user study can be found as public data on GitHub (<https://github.com/DigitalTwinSocCyberberrange/userStudy>).

Results: During phase 1 and phase 3, the participants were asked eight questions to assess their knowledge about CTI formats and their skills in detecting attacks. In Fig. 6 the percentage of correct answers during phase 1 (pretest) and phase 3 (posttest) are shown. In order to statistically show the increase in knowledge, a t-test was conducted. The user study showed that the mean of correctly answered questions about CTI formats significantly increased by 42.05% ($t = -3.448$, $SD = 0.331368$, $p < .001$). Additionally, a significant mean increase of 10.23% ($t = -3.448$, $SD = 0.196763$, $p = .0013$) considering attack detection can be observed. The t-test shows that the training had a significant positive effect on both variables. We conclude that training on CTI formats can have a positive effect on the overall attack detection knowledge. This fact is particularly interesting as we base our approach on a qualitative assessment of CTI formats, leading to SOC service capability levels.

Expert interviews

To validate the conceptual development and relevance of CTI-SOC2M2 and its structure, we conducted a number of expert interviews. The security experts selected for the interviews work in different industries and have different degrees

of knowledge about maturity models and SOC. While the role of two interviewees can be understood as senior SOC manager or analyst, the third interviewee has previously worked with maturity models. None of the participants is currently aware of a dedicated CTI-based SOC maturity model.

Design & Procedure: Using a semi-structured approach, the interview design and procedure includes the following four phases (Lazar et al., 2010):

1. **Introduction:** We start the interview to determine the interviewees understanding of common terminology and prior experiences with SOC services, CTI formats and maturity models. Afterwards, we introduce the proposed CTI-SOC2M2 and outline basic elements. We actively encourage interviewees to directly voice criticism and mention issues throughout the interview.
2. **CTI-SOC2M2:** In this phase we aim to get additional feedback on the individual SOC services and CTI formats. Although the result of an iterative process, we further aim to validate the capability maturity model with focus on its development and structure. We discuss methodological decisions with the interviewees to identify whether they support the relevance and applicability of the results. We also ask the participants to name aspects of CTI and SOC they see missing and which might require a more detailed explanation.
3. **Maturity Assessment:** This phase is focused on using the CTI-SOC2M2 and issues faced when doing so. To enable the interviewees to work with the proposed capability maturity model we explain aspects such as indicative question about CTI formats for capability levels and the foundations of maturity levels. During the interviews, the participants can also access the CTI-SOC2M2 tool. The primary goal in this interview phase is to identify whether the CTI-SOC2M2 approach is comprehensible, applicable and the self-assessment tool provides adequate functionality. We ask the interviewees whether there are further aspects they think would enhance the understanding and use of CTI-SOC2M2.
4. **Wrap-Up:** Last, we conclude the interviews with a summarizing discussion. We discuss with participants whether limitations to CTI-SOC2M2 are methodological, conceptual or based on implementation. This phase also includes collecting participants' ideas of features deemed useful and extensions to improve our approach.

Results: The interviews were scheduled for 60 minutes. The following results are first divided according to the four interview phases:

1. **Introduction:** Table 5 summarizes background information about the interviewed experts.
2. **CTI-SOC2M2:** Reflected by their knowledge the interviewees focused on CTI or SOC elements of the proposed CTI-SOC2M2. Above all, the interviewees unanimously stated the importance of an intelligence-driven SOC. The approach to assess SOC maturity with CTI and CTI formats was perceived as innovative. Paired with scientific methodology on CMM development the proposed model was seen as coherent. While the CTI formats were only

Table 5 – General information on the interview participants.

	Position	Business Branch	Organization's Size	SOC Knowledge	CMM Knowledge
#1	IT-Security Manager	Consulting	ca. 500.000	high	medium
#2	Senior Security Architect	Automotive	ca. 40.000	medium	medium
#3	Security Expert	Education	ca. 5.000	medium	high

partially known to the interviewees, the SOC services provided enough differentiation to cluster the essential SOC activities. As it was pointed out that SOC services are not independent, the participants referred to other organizational IT services. We see this as an important aspect and envision this within capability level *Integration* (e.g., organizational risk management). The completeness concern voiced by one interviewee had been addressed with a broad literature corpus used for taxonomy development. Concerning relevance and practical necessity to combine CTI and SOC, the participants all strongly agreed with a more data-centric approach in organizations.

3. *Maturity Assessment*: The participants' answers concerning the maturity assessment covered the step-wise approach to improving the maturity and further explanations about the capability levels. The mapping to the NIST incident response life cycle was seen as a helpful structuring element. One interviewee pointed out that the highest maturity level should emphasize the boldness of the CTI-focused SOC. Thus we re-considered our naming convention and opt for *visionary* as the highest maturity level. When explaining capability levels to the interviewees, it became apparent to direct future work to a more fine granular specification of data quality and possible metrics. The self-assessment tool was considered an essential element to the adoption of the CTI-SOC2M2, documenting the need for visualization.
4. *Wrap-Up*: The final phase revealed different perceptions on SOC and its services. Red teaming and threat hunting were topics of discussion as they apply only to sophisticated organizations with a strong focus on information security. In the same direction, the inclusion of active defense services beyond the use of honey pots was mentioned. We acknowledge that this is a possible SOC service. However, jurisdiction and existing laws in various regions prohibit its use. Therefore, we do not include this SOC service specifically.

We also apply the thematic synthesis by [Cruzes and Dybå \(2011\)](#) to the results of our expert interviews. Whereas the authors' approach is typically applied to academic literature and coding the content of primary studies, we use the interviews instead. Our starting point to thematic synthesis is the question: How do experts perceive CTI-SOC2M2? After the coding of data, the approach involves translating codes into themes. Within the expert interviews, we identified several codes. Due to the limited information available from the interviews, we identified a comparatively small number of codes. These codes were then directly transformed into four higher-order themes (see [Table 6](#)). Please note, each code constitutes a component of the respective theme. Also, we do not weigh and order the individual codes. Relevance, applicability, compre-

hensibility, and limitations represent the higher-order themes and build the model to answer the specified question.

Thematic synthesis of the expert interviews resulted in four higher-order themes – *relevance*, *applicability*, *comprehensibility*, and *limitations*. The following excerpts address some of the codes in [Table 6](#). Overall, the interviewees' feedback includes various aspects relating to the relevance of our proposed maturity model. As one interviewee stated, SOC services and CTI formats cannot be implemented separately but must be integrated (codes: SOC services, CTI concept). Further, a valuable contribution to practice stems from the CMM and the self-assessment tool. Here, interviewee perceptions include the flexibility of the model regarding specific CTI formats (code: flexibility). Comprehensibility is centered on the scientific methodology. The interviewees point to the conciseness of CTI-SOC2M2 based on CMM components (code: naming conventions). At last, the innovative approach faces limitations. This higher-order theme mainly concerns the granularity of the model (code: CTI data quality).

To subsume, experts perceive CTI-SOC2M2 as a relevant capability maturity model and see a valuable contribution. Applied to IT management, the self-assessment tool supports its actual use. In addition, the scientific methodology and the well-known reference framework foster comprehensibility. Nevertheless, some limitations must be accepted or addressed by the target audience itself.

[Cruzes and Dybå \(2011\)](#) conclude the thematic synthesis with an assessment of its trustworthiness. Trustworthiness is specified by the concepts of credibility, confirmability, dependability, and transferability. In the context of our thematic synthesis, credibility is addressed by the selection of the interviewees. We selected three interview participants that have sufficient experience in information security and are familiar with CMMs (see [Table 5](#)). Concerning confirmability, we opt for separating coding the interview results between different researchers and aggregating the outcome. Doing so allows us to avoid potential individual biases. Dependability as the stability of data is partially applicable to our synthesis. We assume the interview data to be stable. Finally, transferability refers to valuable insights beyond the scope of CTI formats, SOC and CMMs. The interview guide adapted for our purpose and the applied synthesis method document transferability.

8. Discussion

Novel aspects of this work are the integration of CTI and SOC services within CTI-SOC2M2. The capability maturity model provides an adequate foundation for assessing a SOC based on the CTI used. However, similar to other research efforts, this model has limitations, which are worth discussing. It should

Table 6 – Thematic synthesis of expert interviews: themes and codes.

Themes	Codes
Relevance	CMMs, SOC services, CTI concept, data-driven, CTI feeds, tactical level, strategic decisions, threat information, target audience
Applicability	Self-assessment, tool support, IT management, flexibility, existing CMMs, defined goal, SOC service selection, completeness
Comprehensibility	Naming conventions, scientific methodology, NIST incident response life cycle, stakeholders
Limitations	Active defense, CTI data quality, SOC service exclusion, metrics

be noted that the proposed model can only provide an indicator of overall SOC maturity, as the focus is exclusively on the integration of CTI and SOC services. Scoring a high maturity level in our CTI-focused model does not necessarily mean a high overall SOC maturity. Thus, a combination with more holistic models covering governance and roles might be recommended depending on the specific use case.

Furthermore, there is a challenge that applies to most maturity models, especially those developed in research. Both the development and the application always contain a certain degree of subjectivity, which can hardly be eliminated. In the development phase of the model, the degree of subjectivity can be controlled through methodology, but it cannot be avoided entirely. Also, when determining the capability levels using the model, a certain degree of subjectivity on the users' self-assessment cannot be avoided. In future work, the methodological procedure for developing the model could be supplemented by other methods. For example, conducting a Delphi study is a frequently chosen approach. However, from a research perspective, we decided to capture the current state-of-the-art presented in academic literature using a literature review and performing validation with expert interviews.

With the help of the two-stage evaluation, it could be shown that the problem definition is relevant and the proposed model appreciated by experts. However, many ways of maturity model evaluation could supplement the procedure described in this paper. We chose the most frequently used option conducting expert interviews, which is also the most useful for the present case.

Finally, it should be mentioned that for maturity models, completeness and perfection are strived for but not realized. Instead, a maturity model must be understood as a living model that evolves and is adapted to new requirements. The same applies to the present CTI-SOC2M2. Additional aspects such as CTI quality offer the opportunity to specify individual metrics but go beyond the scope of this paper. We are aware that weaknesses will emerge during practical use, which must then be addressed. In addition, the requirements for a SOC and the possibilities of CTI will change in the future, which is why the model must be expanded accordingly.

Implications for literature

As seen in Fig. 1, the CTI concept comprises application domains where CTI artifacts structured with CTI formats are used. This organizational CTI focus has not been examined by existing literature. Complementing existing research on individual CTI capabilities (Shin and Lowry, 2020), with our CTI-SOC2M2, we address the organizational CTI capabilities level, which is realized by mapping CTI formats to SOC services.

A second implication for literature is the aggregation of SOC literature concerning services and capabilities. While the literature corpus is based on previous work, we derive relevant information to structure SOC services for our model. In conjunction with CTI formats, two currently separate research areas are connected.

With CTI-SOC2M2, we build a first basis for the quest for mature, intelligence-driven security operations and incident response capabilities. However, for future work, we see the necessity to examine further 1) CTI quality and CTI automation, 2) incident response and CTI, and conduct 3) field studies based on CTI-SOC2M2.

Implications for industry

Maturity models evolved from best practices and have become popular in the industry. They are of particular interest at a higher management level. As some examples show, SOC-related models often provide a holistic view of SOC (see Table 2 and Section 5.1).

The popularity and use of CMMs in the industry are due to several reasons. One reason is that otherwise abstract factors, such as management success, can be measured with them. Regarding a SOC, it is possible to measure how it compares to other SOCs without consulting otherwise problematic metrics (e.g., the number of successful attacks). Another reason is that CMMs link theory and practice. More specifically, CMMs allow checking how close one's SOC comes to a theoretically and, in some cases, scientifically complete SOC.

In the previous presentation of the maturity model, we have taken a more academic and theoretical view. However, since a maturity model should be seen as a bridge between theory and practice, the implications of the model for the industry are presented subsequently.

In most cases, the current view of SOCs is people-driven and bases actions and decisions on the knowledge and intuition of analysts and other staff. Contrary, CTI-SOC2M2 aims at a more data-driven view, where the procedures within a SOC are based on available CTI – ideally resulting in a (partial) automation of incident handling processes. The use of CTI formats and SOC services in CTI-SOC2M2 contributes to security operations by guiding practitioners towards a more effective SOC.

Both SOC and CTI have only recently found their way into practice. However, different forms of SOC services and CTI formats predate their concepts. It is, therefore, necessary to align SOC services and CTI formats and adapt the latter to current needs and future requirements (e.g., Digital Twin based security operations (Dietz et al., 2020)). We contribute by emphasizing the importance of considering CTI and SOC together.

The resulting implications for industry can be subsumed as leveraging information about the external threat landscape. CTI formats lead to the operationalization of CTI in organizations and possibly improve proactive and reactive security measures. Organizations using CTI-SOC2M2 have a structured yet flexible model at hand to assess and improve their CTI-driven SOC maturity.

9. Conclusion

This paper presents CTI-SOC2M2, a capability maturity model that aims at assessing the maturity of SOCs based on the use of CTI. Special attention is paid to a structured and methodical approach. In addition to the maturity model itself, the contribution is divided into three parts: First, existing maturity models in the area of SOC, CTI, and incident response are collected and analyzed. Second, as a basis for the new maturity model, the activities in a SOC are clustered by services with the help of a structured literature review. Third, to finally develop the model, the most common CTI formats are mapped to the SOC services for which they are relevant. A mixed-method approach was performed to evaluate the relevance and applicability of the model. For this purpose, a quantitative user study and expert interviews were combined. The results show that the problem addressed by CTI-SOC2M2 is relevant and that the developed model is considered useful by experts. Implications resulting from CTI-SOC2M2 for literature center on the extension of existing research. The combined consideration of CTI and SOC leads to more mature security operations and incident response capabilities. Currently fragmented research is aggregated by our model. Implications for industry settle on the operationalization of CTI. Therefore, (semi)-structured CTI formats are leveraged to achieve implementation of the CTI concept in organizations. The proposed model is an essential step to assess the current state of a SOC and its ability to cope with external threat information. Its actual use within several organizations will eventually determine the models' success.

Credit Author Statement

DS and MV carried out the capability maturity model development and the underlying clustering of SOC services. DS contributed threat intelligence formats, mapping and assessment methodology while MV provided SOC content, mapping and implemented the prototype. GP participated in the formal definition of CTI-SOC2M2 supporting selection and structure of its elements. DS, MV and GP also helped to draft the manuscript revising it critically for important intellectual content. All authors read and approved the final manuscript.

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.

Acknowledgements

This research was supported by the Federal Ministry of Education and Research (BMBF), Germany, as part of the DEVISE project and by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (BayStMWi), as part of the IN-SIST project.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.cose.2021.102482](https://doi.org/10.1016/j.cose.2021.102482)

REFERENCES

- Ab Rahman NH, Choo K-KR. A survey of information security incident handling in the cloud. *computers & security* 2015;49:45–69.
- Acartürk C, Ulubay M, Erdur E. Continuous improvement on maturity and capability of security operation centres. *IET Inf. Secur.* 2020.
- Ahern DM, Clouse A, Turner R. *CMMI Distilled: A practical introduction to integrated process improvement*. Addison-Wesley Professional; 2004.
- Ahmad A, Hadgkiss J, Ruighaver AB. Incident response teams—challenges in supporting the organisational security function. *Computers & Security* 2012;31(5):643–52.
- Ahmad A, Maynard SB, Desouza KC, Kotsias J, Whitty MT, Baskerville RL. How can organizations develop situation awareness for incident response: a case study of management practice. *Computers & Security* 2021;101:102–22.
- Apache HTTP Server Project, 1995. NCSA Common Log Format. Last accessed 2021-07-01, <https://httpd.apache.org/docs/trunk/logs.html#common>.
- ArcSight, 2010. Common Event Format.
- ATC Project, 2020. RE&CT framework documentation. Last accessed 2021-02-01, <https://atc-project.github.io/atc-react/>.
- Baker DW, Christey SM, Hill WH, Mann DE. The development of a common enumeration of vulnerabilities and exposures, Vol. 7; 1999. p. 9.
- Bauer S, Fischer D, Sauerwein C, Latzel S, Stelzer D, Breu R. Towards an evaluation framework for threat intelligence sharing platforms. In: *Proceedings of the 53rd Hawaii International Conference on System Sciences*; 2020. p. 1–10.
- Becker J, Knackstedt R, Pöppelbuß J. Developing maturity models for it management. *Business & Information Systems Engineering* 2009;1(3):213–22.
- Bouwman X, Griffioen H, Egbers J, Doerr C, Klievink B, van Eeten M. A different cup of TI? the added value of commercial threat intelligence. In: *29th USENIX Security Symposium (USENIX Security 20)*; 2020. p. 433–50.
- Brown R, Lee RM. The evolution of cyber threat intelligence (cti): 2019 sans cti survey. SANS Institute 2019.
- Brown R, Lee RM. 2021 Sans cyber threat intelligence (cti) survey. SANS Institute 2021.
- Brown S, Gommers J, Serrano O. From cyber security information sharing to threat management. In: *Proceedings of the 2nd ACM workshop on information sharing and collaborative security*; 2015. p. 43–9.
- de Bruin T, Michael Rosemann, Ronald Freeze, Uday Kulkarni. Understanding the main phases of developing a maturity assessment model. *ACIS 2005 Proceedings - 16th Australasian Conference on Information Systems*, 2005.

- Caltagirone S, Pendergast A, Betz C. In: Technical Report. The diamond model of intrusion analysis. Center For Cyber Intelligence Analysis and Threat Research, Hanover Md; 2013.
- CAPEC Team, 2020. Schema documentation - schema version 3.4. Last accessed 2021-04-01, <https://capec.mitre.org/documents/schema/index.html>.
- Cheikes BA, Waltermire D, Scarfone K. In: Technical Report. Common Platform Enumeration: Naming Specification Version 2.3. Maryland, USA: National Institute of Standards and Technology; 2011. NIST Interagency Report 7695
- Chismon D, Ruks M. In: Technical Report. Threat intelligence: Collecting, analysing, evaluating. MWR InfoSecurity, CERT-UK; 2015.
- Christopher JD, Gonzalez D, White DW, Stevens J, Grundman J, Mehravari N, Dolan T. In: Technical Report. Cybersecurity Capability Maturity Model (C2M2). US Department of Energy (DOE); 2014.
- Cichonski P, Millar T, Grance T, Scarfone K. Computer security incident handling guide. NIST Special Publication 2012;800(61):1-147.
- CMMI Product Team, 2010. Cmmi for services, version 1.3: Improving processes for providing better services. https://resources.sei.cmu.edu/asset_files/TechnicalReport/2010_005_001_15290.pdf.
- CREST, 2014. Cyber Security Incident Response Maturity Assessment Tool (CSIR-MAT). <https://www.crest-approved.org/2018/07/20/cyber-security-incident-response-maturity-assessment/index.html>.
- CREST, 2016. Cyber Threat Intelligence Maturity Assessment Tool (CTI-MAT). <https://www.crest-approved.org/2020/01/10/cyber-threat-intelligence-maturity-assessment-tool/index.html>.
- Cruzes DS, Dybå T. Recommended steps for thematic synthesis in software engineering. In: 2011 international symposium on empirical software engineering and measurement. IEEE; 2011. p. 275-84.
- Dandurand L, Kaplan A, Kácha P, Kadobayashi Y, Kompanek A, Lima T, Millar T, Nazario J, Perlotto R, Young W. In: Technical Report. Standards and tools for exchange and processing of actionable information. European Union Agency for Network and Information Security (ENISA); 2014.
- Danyliw R. In: Technical Report. The Incident Object Description Exchange Format Version 2. Internet Engineering Task Force (IETF); 2016. <https://tools.ietf.org/html/rfc7970>.
- Dietz M, Vielberth M, Pernul G. Integrating digital twin security simulations in the security operations center. In: Proceedings of the 15th International Conference on Availability, Reliability and Security; 2020. p. 1-9.
- Dorling A. Spice: software process improvement and capability determination. *Software Quality Journal* 1993;2(4):209-24.
- Englbrecht L, Meier S, Pernul G. Towards a capability maturity model for digital forensic readiness. *Wireless Networks* 2020;26(7):4895-907.
- Farris KA, Shah A, Cybenko G, Ganesan R, Jajodia S. Vulcon: a system for vulnerability prioritization, mitigation, and management. *ACM Transactions on Privacy and Security* 2018;21(4):1-28. doi:10.1145/3196884.
- Forum of Incident Response and Security Teams (FIRST), 2019. Common Vulnerability Scoring System version 3.1: Specification document - revision 1. Last accessed 2021-02-01, <https://www.first.org/cvss/specification-document>.
- Gerhards R, et al. In: Technical Report. The syslog protocol. RFC 5424, March; 2009.
- Hámorník BP, Krasznay C. A Team-level Perspective of Human Factors in Cyber Security: Security Operations Centers. In: *Advances in Intelligent Systems and Computing*, Vol 593. Cham: Springer International Publishing; 2018. p. 224-36. doi:10.1007/978-3-319-60585-2_21.
- Hernandez-Ardieta JL, Tapiador JE, Suarez-Tangil G. Information sharing models for cooperative cyber defence. In: 2013 5th International Conference on Cyber Conflict (CYCON 2013). IEEE; 2013. p. 1-28.
- Humphrey WS. Characterizing the software process: a maturity framework. *IEEE Software* 1988;5(2):73-9. doi:10.1109/52.2014.
- Hutchins EM, Cloppert MJ, Amin RM. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research* 2011;1(1):80.
- Islam C, Babar MA, Nepal S. A multi-vocal review of security orchestration. *ACM Computing Surveys (CSUR)* 2019;52(2):1-45.
- Jacobs P, Arnab A, Irwin B. In: 2013 Information Security for South Africa. Classification of security operation centers. IEEE; 2013. doi:10.1109/ISSA.2013.6641054.
- Kokulu FB, Soneji A, Bao T, Shoshitaishvili Y, Zhao Z, Doupé A, Ahn G-J. 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; 2019. p. 1955-70.
- Kowtha S, Nolan LA, Daley RA. Cyber security operations center characterization model and analysis. In: 2012 IEEE Conference on Technologies for Homeland Security (HST). IEEE; 2012. p. 470-5.
- Krebs, B., 2021. A basic timeline of the exchange mass-hack. <https://krebsonsecurity.com/2021/03/a-basic-timeline-of-the-exchange-mass-hack/>.
- Lakshmi, R, Naseer, H, Maynard, S, Ahmad, A. Sensemaking in cybersecurity incident response: The interplay of organizations, technology and individuals. arXiv preprint arXiv:2107.02941 2021.
- Lazar J, Feng JH, Hochheiser H. Research methods in human-Computer interaction. Burlington: Morgan Kaufmann; 2010.
- Li VG, Dunn M, Pearce P, McCoy D, Voelker GM, Savage S. Reading the tea leaves: A comparative analysis of threat intelligence. In: 28th USENIX Security Symposium (USENIX Security 19); 2019. p. 851-67.
- Lourenco M. In: Technical Report. CTI Capability Maturity Model. European Union Agency for Network and Information Security (ENISA); 2018.
- Luchs M, Doerr C. In: Technical Report. Measuring your Cyber Threat Intelligence Maturity. Hasso Plattner Institut and TU Delft; 2020.
- Madani A, Rezayi S, Gharaee H. Log management comprehensive architecture in security operation center (soc). In: 2011 International Conference on Computational Aspects of Social Networks (CASoN). IEEE; 2011. p. 284-9. doi:10.1109/CASON.2011.6085959.
- Mavroeidis V, Bromander S. Cyber threat intelligence model: an evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence. In: 2017 European Intelligence and Security Informatics Conference (EISIC). IEEE; 2017. p. 91-8.
- Menges F, Pernul G. A comparative analysis of incident reporting formats. *Computers & Security* 2018;73:87-101. doi:10.1016/j.cose.2017.10.009.
- Mettler, T., 2009. A design science research perspective on maturity models in information systems.
- Microsoft, 2018. Windows Event Log. Last accessed 2021-07-01, <https://docs.microsoft.com/en-us/windows/win32/wes/windows-event-log>.
- Microsoft Threat Intelligence Center (MSTIC). In: Technical Report. HAFNIUM targeting Exchange Servers with 0-day

- exploits. Microsoft; 2021. <https://www.microsoft.com/security/blog/2021/03/02/hafnium-targeting-exchange-servers/>.
- MITRE, 2020. Common Weakness Enumeration - a community-developed list of software & hardware weakness types. Last accessed 2021-02-01, <https://cwe.mitre.org/index.html>.
- Neiva C, Lawson C, Bussa T, Sadowski G. In: Technical Report. 2020 Market Guide for Security Orchestration, Automation and Response Solutions. Gartner; 2020.
- Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. *European Journal of Information Systems* 2013;22(3):336–59. doi:10.1057/ejis.2012.26.
- OASIS. Open command and control (openc2) language specification version 1.0 - Committee specification 02. OASIS; 2020. Last accessed 2020-11-15, <https://docs.oasis-open.org/openc2/oc2ls/v1.0/cs02/oc2ls-v1.0-cs02.html>.
- OASIS. CACAO Security playbooks version 1.0 - Committee specification 01. OASIS; 2021. Last accessed 2021-01-15, <https://docs.oasis-open.org/cacao/security-playbooks/v1.0/security-playbooks-v1.0.html>.
- OASIS Cyber Threat Intelligence (CTI) Technical Committee. STIX™ Version 2.1: Committee specification 01. OASIS; 2020. Last accessed 2021-01-01, <https://docs.oasis-open.org/cti/stix/v2.1/stix-v2.1.html>.
- OASIS Cyber Threat Intelligence (CTI) Technical Committee. TAXII™ Version 2.1: Committee specification 01. OASIS; 2020. Last accessed 2020-10-20, <https://docs.oasis-open.org/cti/taxii/v2.1/taxii-v2.1.html>.
- Onwubiko C. 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). IEEE; 2015. p. 1–10. doi:10.1109/CyberSA.2015.7166125.
- Onwubiko C, Ouazzane K. Soter: a playbook for cybersecurity incident management. *IEEE Trans. Eng. Manage.* 2020;1–21. doi:10.1109/TEM.2020.2979832.
- RawSec - Quentin Jerome, 2018. Go EvtX Signature Engine. Last accessed 2021-04-01, <https://rawsec.lu/doc/gene/1.6/writerules.html>.
- Rea-Guaman AM, San Feliu T, Calvo-Manzano JA, Sanchez-Garcia ID. Comparative study of cybersecurity capability maturity models. In: *International Conference on Software Process Improvement and Capability Determination*. Springer; 2017. p. 100–13.
- Ross, D., Shiffer, J., Dell, T., Gibb, W., Wilson, D., 2013. OpenIOC 1.1 Schema. Last accessed 2021-04-01, https://github.com/mandiant/OpenIOC_1.1.
- Roth, F., Patzke, T., 2017. Sigma - Generic Signature Format for SIEM Systems. Last accessed 2021-04-01, <https://github.com/SigmaHQ/sigma/wiki/Specification>.
- Schaberreiter T, Kupfersberger V, Rantos K, Spyros A, Papanikolaou A, Ilioudis C, Quirchmayr G. A quantitative evaluation of trust in the quality of cyber threat intelligence sources. In: *Proceedings of the 14th International Conference on Availability, Reliability and Security*; 2019. p. 1–10.
- Schinagl S, Schoon K, Paans R. A framework for designing a security operations centre (soc). In: *2015 48th Hawaii International Conference on System Sciences*. IEEE; 2015. p. 2253–62.
- Schlette D, Böhm F, Caselli M, Pernul G. Measuring and visualizing cyber threat intelligence quality. *Int. J. Inf. Secur.* 2021;20(1):21–38.
- Settanni G, Shovgenya Y, Skopik F, Graf R, Wurzenberger M, Fiedler R. Acquiring cyber threat intelligence through security information correlation. In: *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*. IEEE; 2017. p. 1–7. doi:10.1109/CYBCONF.2017.7985754.
- Shah A, Ganesan R, Jajodia S. A methodology for ensuring fair allocation of csoc effort for alert investigation. *Int. J. Inf. Secur.* 2019;18(2):199–218. doi:10.1007/s10207-018-0407-3.
- Shin B, Lowry PB. A review and theoretical explanation of the 'cyberthreat-intelligence (cti) capability' that needs to be fostered in information security practitioners and how this can be accomplished. *Computers & Security* 2020;92:101761.
- Sillaber C, Sauerwein C, Mussmann A, Breu R. Towards a maturity model for inter-organizational cyber threat intelligence sharing: a case study of stakeholders' expectations and willingness to share. *Proceedings of Multikonferenz Wirtschaftsinformatik (MKWI 2018)* 2018:1409–20.
- Skopik F, Settanni G, Fiedler R. A problem shared is a problem halved: a survey on the dimensions of collective cyber defense through security information sharing. *Computers & Security* 2016;60:154–76. doi:10.1016/j.cose.2016.04.003.
- Snort Team, 2021. Writing Snort Rules. Last accessed 2021-04-01, <https://www.snort.org/documents>.
- Stikvoort D. In: Technical Report. SIM3: Security Incident Management Maturity Model. OCF, S-CURE and PRESECURE; 2015.
- Strom BE, Applebaum A, Miller DP, Nickels KC, Pennington AG, Thomas CB. In: Technical Report. MITRE ATT&CK: Design and philosophy. The MITRE Corporation; 2018.
- Taurins E. In: Technical Report. How to set up CSIRT and SOC - Good Practice Guide. European Union Agency for Network and Information Security (ENISA); 2020.
- The Zeek Project, 2021. Signature Framework. Last accessed 2021-04-01, <https://docs.zeek.org/en/current/frameworks/signatures.html>.
- Tounsi W, Rais H. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & security* 2018;72:212–33.
- Van Os, R., 2016. SOC-CMM: Designing and evaluating a tool for measurement of capability maturity in security operations centers.
- VERIS Community, 2021. Veris - the vocabulary for event recording and incident sharing. Last accessed 2021-04-01, <http://veriscommunity.net/index.html>.
- Vielberth M, Böhm F, Fichtinger I, Pernul G. Security operations center: a systematic study and open challenges. *IEEE Access* 2020;8:227756–79.
- VirusTotal - Victor Alvarez, 2014. Signature Framework. Last accessed 2021-04-01, <https://yara.readthedocs.io/en/stable/>.
- Wagner C, Dulaunoy A, Wagener G, Iklody A. MISP - the design and implementation of a collaborative threat intelligence sharing platform. In: Katzenbeisser S, Weippl E, Blass E-O, Kerschbaum F, editors. In: *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security - WISCS'16*. ACM; 2016. p. 49–56. doi:10.1145/2994539.2994542.
- Zimmerman C. In: Technical Report. Cybersecurity Operations Center. The MITRE Corporation; 2014.
- Daniel Schlette** received the master's degree (Hons.) in management information systems from the Elite Graduate Program, University of Regensburg, in 2019. He is currently pursuing the Ph.D. degree with the Chair of Information Systems, University of Regensburg. Since 2019, he has been a Research Assistant with the Chair of Information Systems, University of Regensburg. His research interests include the field of Cyber Threat Intelligence. His primary focus within this topic is to leverage structured data formats, explore aspects of Security Orchestration, Automation and Response (SOAR), and data quality.
- Manfred Vielberth** received the bachelor's and master's degrees in management information systems with a specialization in cyber security from the University of Regensburg, Germany. Since February 2017, he is pursuing the Ph.D. degree with the Chair of Information Systems, University of Regensburg. His research interest

includes human aspects in the security analytics domain. On the expert side, this mainly comprises improving processes for better integrating security analysts within a Security Operations Center. In terms of security novices, this primarily covers capturing reports about security incidents in the context of the Human-as-a-Security-Sensor paradigm.

Günther Pernul (Member, IEEE) received the diploma and Ph.D. degrees (Hons.) in business informatics from the University of Vienna, Austria. He is currently a Professor with the Department

of Information Systems, University of Regensburg, Germany. Previously, he held positions at the University of Duisburg-Essen, Germany; the University of Vienna; the University of Florida, Gainesville; and the College of Computing, Georgia Institute of Technology, Atlanta. His research interests include data and information-security aspects, data protection and privacy, data analytics, and advanced datacentric applications.

8 Integrating Digital Twin Security Simulations in the Security Operations Center

Current status:	Published
Conference:	15th International Conference on Availability, Reliability and Security, ARES 2020, Dublin, Ireland, August 25 - 28, 2020
Date of acceptance:	08 June 2020
Full citation:	DIETZ, M., VIELBERTH, M., AND PERNUL, G. Integrating Digital Twin Security Simulations in the Security Operations Center. In <i>Proceedings of the 15th International Conference on Availability, Reliability and Security (2020)</i> , ACM, pp. 1–9
Authors' contributions:	Marietheres Dietz 45% Manfred Vielberth 45% Günther Pernul 10%

Conference description: The International Conference on Availability, Reliability and Security brings together researchers and practitioners in the area of dependability since 2006. ARES highlights the various aspects of security – with special focus on the crucial linkage between availability, reliability and security. ARES aims at a full and detailed discussion of the research issues of security as an integrative concept that covers among others availability, safety, confidentiality, integrity, maintainability and security in the different fields of applications.

Integrating Digital Twin Security Simulations in the Security Operations Center

Marietheres Dietz
marietheres.dietz@ur.de
Department of Information Systems
University of Regensburg
Germany

Manfred Vielberth
manfred.vielberth@ur.de
Department of Information Systems
University of Regensburg
Germany

Günther Pernul
guenther.pernul@ur.de
Department of Information Systems
University of Regensburg
Germany

ABSTRACT

While industrial environments are increasingly equipped with sensors and integrated to enterprise networks, current security strategies are generally not prepared for the growing attack surface that resides from the convergence of their IT infrastructure with the industrial systems. As a result, the organizations responsible for corporate security, the Security Operations Center (SOC), are overwhelmed with the integration of the industrial systems.

To facilitate monitoring the industrial assets, digital twins represent a helpful novel concept. They are the virtual counterparts of such assets and provide valuable insights through collecting asset-centric data, analytic capabilities and simulations. Moreover, digital twins can assist enterprise security by simulating attacks and analyzing the effect on the virtual counterpart. However, the integration of digital twin security simulations into enterprise security strategies, that are mainly controlled by the SOC, is currently neglected.

To close this research gap, this work develops a process-based security framework to incorporate digital twin security simulations in the SOC. In the course of this work, a use case along with a digital twin-based security simulation provides proof of concept. It is demonstrated how a man-in-the-middle attack can be performed in a simulated industry setting and how it affects the systems. Moreover, we show how the resulting system logs can support the SOC by building technical rules to implement in Security Information and Event Management (SIEM) systems.

CCS CONCEPTS

• **Computer systems organization** → **Embedded and cyber-physical systems**; • **Security and privacy** → **Virtualization and security**.

KEYWORDS

digital twin, security operations center, security information and event management, security framework

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ARES 2020, August 25–28, 2020, Virtual Event, Ireland

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-8833-7/20/08...\$15.00

<https://doi.org/10.1145/3407023.3407039>

ACM Reference Format:

Marietheres Dietz, Manfred Vielberth, and Günther Pernul. 2020. Integrating Digital Twin Security Simulations in the Security Operations Center. In *The 15th International Conference on Availability, Reliability and Security (ARES 2020)*, August 25–28, 2020, Virtual Event, Ireland. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3407023.3407039>

1 INTRODUCTION

The increasing integration of industrial assets in corporate networks leads to a convergence of operational technology (OT) and information technology (IT). While this phenomena entails various benefits, e.g. enhanced monitoring and predictive maintenance, the attack surface eventually increases.

In current enterprises, the Security Operations Center (SOC) upholds corporate security. On the technical side, the SOC is commonly supported by Security Information and Event Management (SIEM) systems. The capabilities of such systems range from managing security-relevant data, over security analyses to deduce security rules or patterns as well as to check the adherence of security rules. However, contemporary SIEM systems mainly monitor the IT infrastructure, while the incorporation of industrial systems is pressing.

A novel paradigm, currently considered an essential milestone in companies, especially in those pursuing Industry 4.0, is the digital twin. It refers to a virtual representation of an enterprise asset – at most an industrial one. It relies, amongst others, on simulation technology to analyse potential outcomes of physical processes and to determine machine fatigue. Moreover, some digital twins even employ prescriptive maintenance that provides maintenance solutions on top. Next to optimization of production, digital twins may contribute to corporate security. For instance, simulating an attack on an industrial asset with digital twins might provide information about system weaknesses and behavior under attack. Thus, it provides potential assistance for SOC and SIEM systems by delivering novel security insights on industrial systems that are currently neglected.

This paper tackles this issue by proposing an effective integration of the digital twin paradigm to SOC and SIEM systems that provides novel potentials for security. The main contributions can be summarized as follows:

- development of a process-based security framework and its formal requirements
- proposition of a use case for demonstration
- evaluation by prototypical implementation

The remainder of this work is organized as follows. Section 2 provides the foundations for our research, outlines related works and

ARES 2020, August 25–28, 2020, Virtual Event, Ireland

Dietz et al.

the research gap. Subsequently, Section 3 proposes a process-based security framework to integrate digital twin security operations with SIEM and SOCs and states formal requirements to achieve this integration. In Section 4, we evaluate our framework by a use case and implemented prototype. Finally, a conclusion of our work is drawn and future work is stated in Section 5.

2 BACKGROUND

The following sections lay the foundation of this work and introduce the respective related work. The first two focus on the SOC and SIEM concept. The next section presents the digital twin paradigm, while the subsequent section addresses works employing simulations in the security field and points out the addressed research gap.

2.1 Security Operations Center

The Security Operations Center (SOC) represents an *organizational aspect* of a security strategy in an enterprise by providing *procedures, technologies and people* [17, 25]. It is usually not seen as a single entity or system, but rather as a complex structure to manage and enhance the overall security posture of an organization, whereby the core purpose of a SOC is the protection of the organization's system infrastructure. Thereto, it integrates, monitors and analyses all security-relevant systems and events in a central point. In general, the activities within a SOC can be classified as reactive and proactive, although these cannot always be clearly separated. An integral task of the SOC is to handle alerts and take countermeasures to protect data and applications. Furthermore, it provides governance and compliance as a framework, in which people operate and to which processes and technologies are tailored. When installed and operated correctly, a SOC improves an organization's security posture, creates situational awareness, mitigates the exposed risks, and helps to fulfill regulatory requirements [14]. Since people play an essential role in the security of companies, this is also an important part of a SOC. From the SOC manager to the analyst, a variety of roles can be defined, whereby a SOC must take care of staffing and recruitment. Furthermore, the security awareness of employees can also be assigned to a SOC. To realize the technical side for security operations, SOCs commonly employ, amongst others, SIEM systems as central tools.

2.2 Security Information and Event Management

A key aspect of today's SIEM systems is that it provides a holistic and centralized view on all security relevant systems of an organization, whereas other systems (such as Network Intrusion Systems) only take a limited perspective on selected systems or functions.

A pattern describing SIEM abstractly was proposed in [28]. Compared to the anatomy of SIEM introduced by [18], it describes SIEM in more detail by considering relevant interfaces and a broader breakdown of relevant components. In general, a SIEM provides means for collecting data (such as log data or network flows) from various heterogeneous sources and reports about incidents by humans [27]. To improve its value for further processing, the collected data gets enriched with context data and is normalized into a uniform format. The core of a SIEM is the correlation and analysis

module, which interconnects the gathered data and deduces possible security incidents or abnormalities. Most SIEM systems provide interfaces for sharing the gained threat intelligence externally with other systems on the one hand, and interfaces for human or expert interaction on the other hand. In case of an incident, measures for incident response can be taken either automatically or supported by staff that gets informed by alerts.

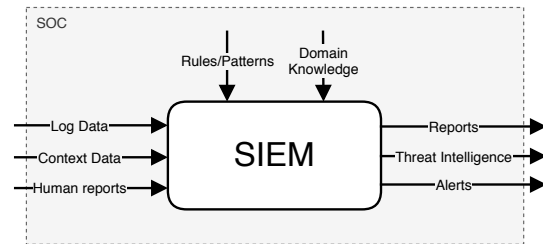


Figure 1: Black box view of SIEM embedded in a SOC.

Moreover, recent works propose to deploy SIEM systems to handle the industrial security by incorporating OT systems next to IT systems [23]. In the course of this paper the in- and outputs of a SIEM are of special importance, as these provide interaction options with digital twins. In Figure 1, these are illustrated by a black box view of a SIEM.

2.3 Digital Twin

Generally, a digital representation of any real-world counterpart such as a system, product, process or other enterprise asset over its lifecycle can be referred to as digital twin [2]. The digital twin comprises asset-specific data and typically adds context to the data by semantic technologies. Based on the semantically linked data, analyses such as predictive maintenance can be conducted [5]. Moreover, the data allows to model the real-world counterpart virtually in order to conduct simulations [12]. Overall, simulations play a decisive role in digital twin research [20]. Dependent on the specific use case, digital twins are capable of asset management ranging from simple monitoring to autonomy. Figure 2 illustrates this paradigm. It highlights the simulation aspect of the paradigm which presents the focus of this work in respect to digital twin research.

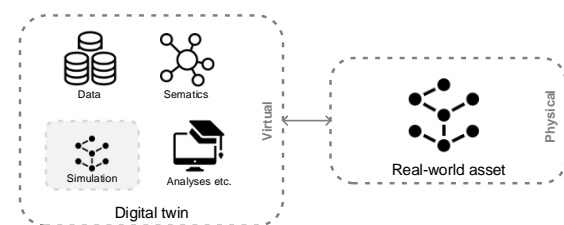


Figure 2: The Digital Twin paradigm.

Various authors mention the importance of security in the digital twin concept (e.g. [10], [24] and [26]), which can also be considered from two different angles: At first, security mechanisms may support the digital twin such as presented in [7]. The second perspective proposes the application of the digital twin concept to enhance security, e.g. [8] and [9]. For the latter, the digital twin provides versatile potentials to enhance current security analyses [6], such as analyses on an asset's historical data, emulated asset-environments to simulate cyber-attacks and transmission of an asset's current state (e.g. by sensor data). Our work focuses on how the simulation of security-relevant incidents within a digital twin can provide helpful insights for SOCs and support the enhancement of SIEM systems.

2.4 Simulations and their Use in Cyber Security

In contrast to traditional intrusion detection mechanisms, simulations of cyber security incidents (e.g. attack simulation) can lead to completely new ways for security monitoring. Simulations differ from simple analyses as they are based on a model, e.g. of a real-world asset. Moreover, simulations depend on user-specified settings and parameters. Consequently, security can benefit from cyber security simulations in the following ways [6]: Simulations enable repeatability and offer to compress the time interval. Additionally, simulations can show a system's behavior under a broad range of specified configurations like during a security incident, which supports the comprehension of emergent as well as prospective behavior. Most importantly, simulations run in a standalone-virtual environment and therefore do not affect the physical environment. Moreover, whenever historical data is missing, resulting data of simulations might deliver important input. Despite all these benefits, simulations are mostly neglected in current security approaches. However, during recent years, some authors have approached this area as follows:

Testbeds, cyber ranges and honey pots. The generation of testbeds is commonly employed for critical infrastructure testing and represents an approach that usually combines virtual and physical components. For instance, the Opnet module [19] is applied in numerous testbeds (e.g. [13], [3]) to integrate physical network devices with the virtual part of the testbeds. Testbeds are often used to test planned infrastructures in terms of functionality, but also to assess the level of security e.g. by spotting vulnerabilities. Cyber ranges are virtual environments to develop IT systems or infrastructures [11]. Their main purpose is to provide a training environment with tools that help improve the security as well as stability and performance of IT infrastructures and systems [22]. Honey pots are commonly employed to attract attackers by emulating real-world systems and simulating their behavior. Their usage mainly aims at extracting attack methodologies including the attackers' tactics, techniques and procedures (TTP). To create a more realistic system for the attacker than virtualized machine environments, honey pots perform better when automation hardware is integrated [23].

Our approach extends work of [9], where a digital twin of a cyber-physical system (CPS) is generated to run a man-in-the-middle (MITM) attack simulation on it. However, the security functionalities only lie within their digital twin solution, while our approach crosses borders beyond the digital twin through the inclusion of

SOCs and SIEM systems. To the best of our knowledge, a holistic approach to employ simulations to enhance corporate security management is missing. Therefore, we suggest applying simulations of security incidents (e.g. threats, attacks) in the virtual counterparts of assets (digital twins) and to combine the gathered information with SOC and SIEM systems that protect these assets.

3 PROCESS-BASED SECURITY FRAMEWORK

This chapter introduces the proposed approach to holistically address security by combining the digital twin concept, SOC and SIEM systems. The resulting process-based framework is illustrated in Figure 3, which applies the Business Process Model and Notation (BPMN) modelling technique to describe the process, its corresponding actors, systems and artifacts. Each process activity is explained in the individual sections that refer to the respective BPMN swim lane. The first part addresses the SOC activities, the next part explains the process steps within the digital twin and afterwards the activities conducted by the security analytics tools are presented. Finally, the formal requirements of this framework are summarized.

3.1 The Role of the Security Operations Center

In order to detect possible security breaches or system security weaknesses, a large amount of data often must be collected and processed – usually handled by SIEM systems. However, SIEM systems also require domain knowledge about threats etc. and thus, highly depend on cognitive processes [4]. For instance, although SIEM systems allow setting and monitoring security rules, these rules must be created in advance by experts, who we assume to be organized within a SOC.

Determination of simulation settings. This first activity of the process is conducted by the SOC. Within this activity, the security experts decide the purpose of the subsequent simulation in the digital twin. This includes deciding which security incident to be simulated as well as which parameters and settings to use. For instance, they could decide to run an attack on the communication between a host to another host using an MITM attack. Another security simulation purpose could be to test if specified security rules apply – preventing the exploitation of vulnerabilities. The output of this activity is the *Simulation settings*-artifact.

Incident Analysis. Subsequent to the security simulation by the digital twin, the SOC studies the simulation output. Thereby, it analyzes the *Incident data* produced by the digital twin simulation. The outcome of this analysis is the *Incident information*-artifact. This information is further used in the subsequent incident detection and handling of the test SIEM to verify the derived SIEM logic/patterns in order to catch the security incidents.

3.2 Security Simulation with the Digital Twin

To ensure the security of the system, attack detection mechanisms and rules require constant review and development. The Digital Twin can serve this purpose. Thus, a central part of our framework lies in simulating security incidents within a digital twin.

In contrast to the related work introduced in section 2.4, the digital twin provides a *sheer digital approach* with *strong focus on the asset* instead of the attack [6]. Moreover, it allows to model modular

ARES 2020, August 25–28, 2020, Virtual Event, Ireland

Dietz et al.

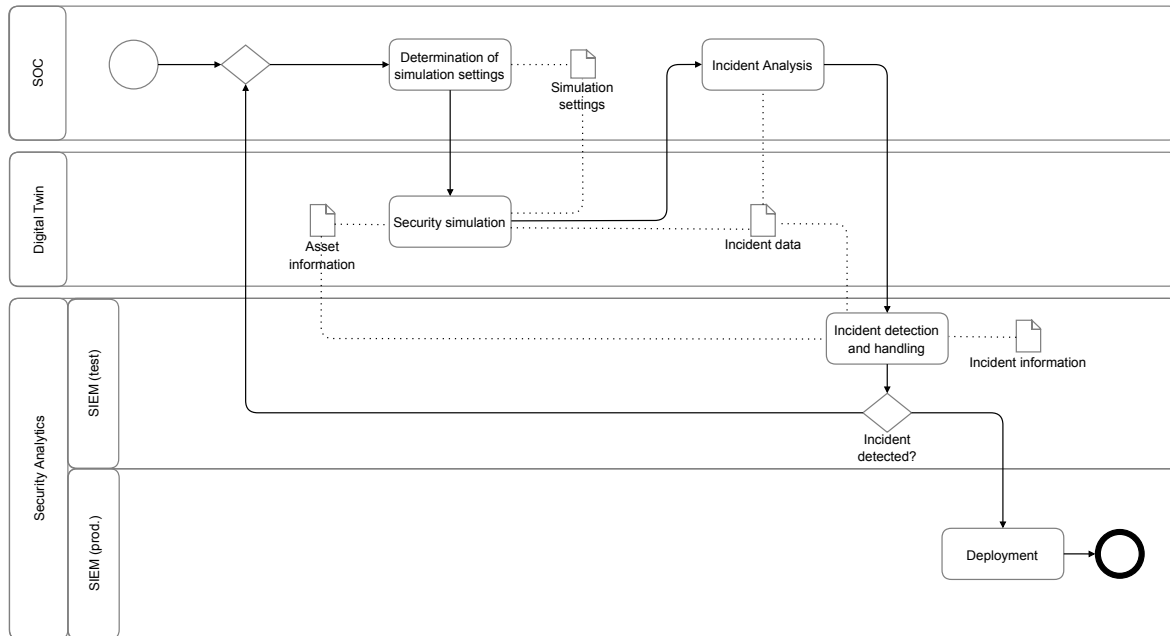


Figure 3: Process-based Security Management Framework integrating Digital Twin security simulations.

components of an asset as well as their network and interplay. Another advantage presents the *bi-directional connection* of the real-world asset with its digital twin. Therefore, subsequent to the evaluation of the simulation outcome, the real-world counterpart can be adapted to the desired setting by transmitting the respective commands and parameters. Possible benefits of this approach are damage limitation of the physical asset, lower prototyping costs, reduction of asset downtime and certainly the promotion of a high security level.

Security simulation. At first, the digital twin virtually represents the real-world asset with the help of the *Asset information*-artifact. The previous activity of the SOC sets the purpose of the simulation and its related conditions. Thereby, the purpose of the simulation can be distinguished between:

- **System security testing:** The simulation investigates security configurations of the emulated systems to investigate potential misconfigurations or vulnerabilities. It further provides feedback whether the weakness is present. For instance, in order to reveal security weaknesses, specified security rules of the system might be tested.
- **Pentesting:** While system security testing involves the assessment and reporting of security weaknesses, pentesting attempts to exploit these vulnerabilities in order to verify the feasibility of malicious activity. In short, it assesses the system's sensitivity to security incidents.
- **Attack simulation:** The simulation of an attack reveals the behavior of the system under attack. Its subsequent analysis

can derive patterns. These can be signature-based, anomaly-based or specification-based [16].

Note that the presented simulation types are not absolutely distinct from each other and might exist additional as well as hybrid forms. By setting the system's conditions and details (*Simulation settings*-artifact), the simulation can be created. Each security simulation produces certain data, e.g. logs reporting occurred events. The produced *Incident data* of the digital twin security simulation is analyzed in the succeeding activity by the SOC.

3.3 Security Analytics with SIEM

The benefit of a SIEM is its holistic view on multiple IT systems. In order to detect security incidents, it correlates events and analyses them. Thereby, various techniques for attack detection are applied. The most common method is rule based detection [18] but is expanded by more advanced methods in modern SIEMs. To avoid influencing the production system negatively during testing a test SIEM can be set up in the test environment, where it can be linked to the digital twin simulation.

Incident detection and handling. Based on the *Incident information*-artifact, the SIEM ideally detects the incident. The incident should be detected in the *Incident data* produced by the digital twin security simulation. In case the incident detection was not successful, either the simulation settings or the analysis of the incident by the SOC were erroneous or incomplete. Thus, these steps must be reevaluated and repeated. *Incident information* can for example be detection rules, which were deduced manually by SOC analysts or

automatically derived patterns by machine learning methods such as anomaly detection. To determine the risk of the incident, *asset information* such as the asset value is needed.

Deployment. If the incident was detected successfully in the test environment, the security rules or security settings, such as a trained detection model, are transferred to the production SIEM so that the incident can be detected upon occurrence in the production environment. Therefore, the utilized technology in the production SIEM should be the same as in the test SIEM in order to avoid side effects during deployment.

3.4 Formal requirements

On the basis of the explanations of the framework's activities, the requirements to be met are formally stated in the following.

REQUIREMENT 1 (DETERMINATION OF SIMULATION SETTINGS). A set of simulation settings $S = \{s_1, s_2, \dots, s_n\}, n \in \mathbb{N} \setminus \{0\}$ is determined.

REQUIREMENT 2 (SECURITY SIMULATION). Asset information I_{Asset} linked with simulation settings S are required to build the simulation SIM , which produces incident data D_{Inc} :

$$I_{Asset} \bullet S \rightarrow SIM \rightarrow D_{Inc}$$

REQUIREMENT 3 (INCIDENT ANALYSIS). Incident information I_{Inc} is deduced from data D_{Inc} :

$$D_{Inc} \models I_{Inc}$$

$I_{Inc} = \{I_{Det}, I_{Rel}, I_{Pri}\}$, whereby I_{Det} is information to incident detection (such as rules), I_{Rel} is the detection reliability and I_{Pri} the priority or severity of the incident.

REQUIREMENT 4 (INCIDENT DETECTION + DEPLOYMENT). Incident information I_{Inc} implemented in a functional SIEM that detects the incident:

$$SIEM(I_{Inc}, D_{Inc}) \rightarrow Inc$$

Determination of risk I_{Ris} , based on I_{Asset} , I_{Rel} and I_{Pri} :

$$I_{Ris} = I_{Rel} \bullet I_{Pri} \bullet I_{Asset}$$

Each of these requirements corresponds to at least one activity of the process-based framework (Figure 3). Moreover, all artifacts of the framework, which serve as inputs resp. outputs of the activities, can be found formally stated:

- *Asset information:* I_{Asset}
- *Incident data:* D_{Inc}
- *Incident information:* I_{Inc}

In regard to the swimlanes, the simulation carried out in the *Digital Twin* and the *SIEM* system are formalized into *SIM* and *SIEM*.

4 EVALUATION

To evaluate the proposed approach a security simulation for a digital twin of an industrial filling plant is implemented. Furthermore, it is demonstrated how this output can serve the SOC to build SIEM rules. This is tested by implementing a SIEM tool that directly receives the logs produced by the digital twin simulation. In the following, the use case (an industrial filling plant) and the tools used for the prototypical implementation are introduced. Afterwards, we concentrate on the conceptual setting of the attack. Afterwards, the obtained results are summarized. We review each activity of

our process-based security framework (Figure 3) in our use case, and show how the formal requirements can be met. The evaluation is concluded by a discussion of the use case and the proposed approach.

4.1 Use Case and Tools

The physical process and the relevant components of our use case are illustrated in Figure 4. The industrial filling plant consists of a tank that contains some liquid and an actuator, e.g. a motor-driven valve (MV), which controls the outflow of the tank. The liquid flows through the pipe into a bottle. The tank, the bottle and the pipe each have a sensor that reads the liquid level (LL) resp. the flow level (FL). Three programmable logic controllers (PLCs) monitor the sensors and the actuator. In the use case setting, PLC2 controls the sensor measuring the flow level in the pipes (Sensor2-FL) and PLC3 controls the liquid level of the bottle that is to be filled (Sensor3-LL-bottle). Meanwhile, PLC1 gets hold of the sensor value of the tank's liquid level (Sensor1-LL-tank) and follows a control strategy for the motor-driven valve (Actuator1-MV). To accomplish the control of the actuator, PLC1 receives the other sensor values controlled by PLC2 and PLC3. The network communication is realized over Ethernet/IP (ENIP) and organized by ENIP tags that store the sensor values in the respective PLCs. For instance, the "Sensor3-LL-bottle"-tag is stored in PLC3, and can be requested and received by PLC1.

Figure 6 shows the use case's network infrastructure. Next to the PLCs, a Human-Machine Interface (HMI) allows direct control the actuator (open/close). Together, the Figures 4 and 6 can be referred to as *Asset information-artifact* as presented in our process-based framework.

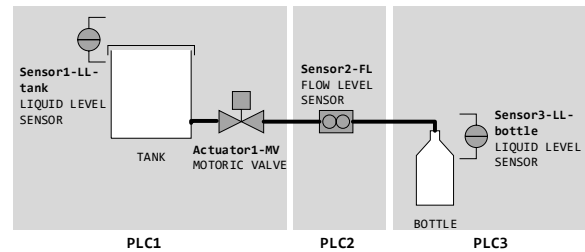


Figure 4: Use case – An industrial filling plant.

The following tools are utilized for prototypical implementation of the filling plant simulation within the digital twin. Commonly, Mininet¹ is employed for simulations with a digital twin [8, 9]. To fit our use case, we chose a technology relying on Mininet, namely MiniCPS², which enables the simulation of industrial assets and originates from research [1]. MiniCPS is tailored for industrial settings: It simulates of traditional industrial systems like PLCs, HMIs and common industrial network communication over ENIP or Modbus. Moreover, an underlying database supports the simulation of the physical process by storing the current states, e.g. the liquid

¹<http://mininet.org>

²<https://github.com/scy-phy/minicps>

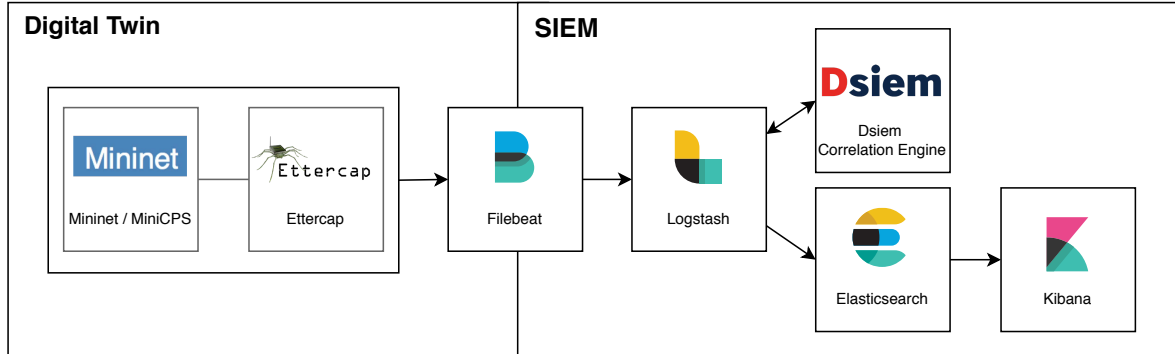


Figure 5: Implemented micro-service architecture for digital twin and SIEM deployment – managed by the SOC.

levels, from which the simulated sensors get their values. Moreover, each PLC in MiniCPS can be implemented with a own ENIP server that can store tags [1]. For instance, in the use case setting, PLC2 stores the value of Sensor2 as ENIP tag ("SENSOR2-FL") in its ENIP server. Additionally, we added a logging strategy to capture the respective system's events (system logs) to each of the MiniCPS PLCs main function. Currently, they are stored in the logs folder of the project, from where they can be directly transferred to the SIEM system.

To simulate the attackers behavior, the attacker node in our setting currently makes use of tools like Ettercap³ to capture the network traffic and start a MITM attack. The respective source code with installation details, further description of the use case implementation and the resulting logs are available for the public and can be found at Github⁴.

The technical side of the incident analysis within the SOC is supported by the SIEM tool Dsiem⁵, which builds upon Elasticsearch, Logstash and Kibana⁶. These tools are under open source licence to a large extent, and are commonly favored in research for their in-depth comprehensibility.

The overall architecture of the tools and their interaction is shown in Figure 5. From the SIEM system side, Filebeat is responsible for the collection of the log data that is transmitted by the digital twin simulation into the files of the logs folder. Within Filebeat, the log files are monitored and in case new lines are added, these are transmitted to Logstash. Logstash normalizes the data by parsing the logs line by line and transforming them into semi-structured JSON-documents in order to enable and facilitate further processing and readability. Dsiem is a correlation engine, which detects incidents based on rules and offers the possibility to trigger an alarm. If an alarm is triggered, it is pushed back to logstash to pass it along the pipeline. Elasticsearch provides the data storage and query execution. Finally, Kibana visualizes the data, displays alarms and offers analytics capabilities, which enables experts to manually analyze incidents.

³<https://www.ettercap-project.org/>

⁴<https://github.com/FrauThes/DigitalTwin-SIEM-integration>

⁵<https://www.dsiem.org/>

⁶<https://www.elastic.co/>

The total environment is realized in the form of a micro-service architecture, where each component is deployed within a docker container. This facilitates the later transfer into the production system and simplifies the use of the framework for research in order to build upon or extend it. Furthermore, individual components are easily replaceable if more suitable ones for the respective environment emerge.

4.2 Attack Setting

SOC: Determination of simulation settings. The first activity of process-based framework (see Figure 3), defines the attack scenario. In our use case, the SOC determines the attacker to be present in the network of the filling plant as shown in Figure 6.

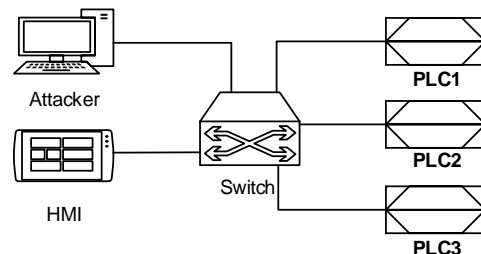


Figure 6: The attacker inside the network topology of the industrial filling plant.

Moreover, the SOC lets the attacker start an ARP spoofing MITM attack between PLC1 that monitors the actuator (open/close motor-driven valve) and PLC3, which sends the liquid level of the bottle to PLC1. More precisely, the attacker only shortly sniffs the network traffic between PLC1 and PLC3 and stops the attack. This is repeated about every two minutes. The goal is to not completely stop the network traffic for a longer time period, so that the physical process seems to be running regular and is not stopped at all. This attack strategy proceeds similarly to the spying phase of Advanced Persistent Threats (APTs). With this procedure, the SOC tests if

their log management strategy is sufficient to detect this simple attack.

In terms of requirements, our use case produces the set of settings $S = \{\text{attacker in network, attack, targets, repeat, attack duration}\}$, whereby $\text{attacker in network} = \text{true}$, $\text{attack} = \text{MITM ARP spoofing}$, $\text{targets} = \{\text{PLC1, PLC3}\}$, $\text{repeat} = \{\text{true, 2 min}\}$, $\text{attack duration} = 15 \text{ s}$ that fulfills REQUIREMENT 1.

4.3 Results

Digital Twin: Security simulation. The purpose of the performed simulation as dictated by the SOC, can be summarized as a hybrid of attack simulation and pentesting as it reveals how the attack affects the systems and their response. To build the simulation SIM , the general information about the asset I_{Asset} , i.e. the information about the network topology and the physical process, is linked with the simulation settings S (REQUIREMENT 2). To start the attack, the attacker node of the simulation executes the shell commands for the MITM attack.

```
INFO 03/16/2020 13:31:59 10.0.0.1 main_loop
Liquid level (SENSOR 3) under
BOTTLE_M['UpperBound']: 0.84 < 0.90
-> open mv (ACTUATOR 1).
INFO 03/16/2020 13:32:01 10.0.0.1 main_loop
Flow level (SENSOR 2) under
SENSOR2_THRESH: 2.45 < 3.00
-> leave mv status (ACTUATOR 1).
WARNING 03/16/2020 13:32:06 10.0.0.1 main_loop
Liquid level (SENSOR 3) is not received.
Program is unable to proceed properly
INFO 03/16/2020 13:32:08 10.0.0.1 main_loop
Flow level (SENSOR 2) under
SENSOR2_THRESH: 2.45 < 3.00
-> leave mv status (ACTUATOR 1).
```

Listing 1: System logs of PLC1

The output of the security simulation SIM is the incident data D_{Inc} (REQUIREMENT 2), which, in this case, are several log files, i.e. the system logs of PLC1, PLC2 and PLC3. This security simulation output of the digital twin as referenced in Figure 3 (*Incident data-artifact*), can be found at GitHub⁷. Listing 1 shows an extract of the logged system events of PLC1 (10.0.0.1). As can be seen from the logs, the control strategy of PLC1 works fine until the value of Sensor3 that is managed by PLC3 cannot be received any more. This causes the program to produce error messages and leads to the actuator being left at the state as it was at last. In the case of Listing 1, it remains open. The physical result would be a bottle overflow, which would only be stopped when the tank reaches its lower-bound threshold or after the attacker stops the MITM attack.

SOC: Incident Analysis. The analysis of simulation output D_{Inc} by the SOC suggest that the ARP poisoning MITM attack allows to place the attacker successfully between PLC1 and PLC3 and results in a denial-of-service (DOS) of network communication between PLC1 and PLC3. Thereby, the pattern of the ARP MITM attack can be deduced from D_{Inc} , and the *Incident information-artifact* I_{Inc} is created (REQUIREMENT 3): The artifact mainly consists of multiple correlation rules I_{Det} to detect the attack. Therefore, it can

⁷<https://github.com/FrauThes/DigitalTwin-SIEM-integration/tree/master/example-logs>

be detected in two stages. The first stage detects, that the sensor data of PLC3 was not received. If the condition of the first stage is fulfilled, the second stage waits for log data within a specific time window, that shows, that PLC3 is operating normally and that there must be a communication problem between PLC1 and PLC3, which in turn indicates a potential MITM attack. Figure 7 exemplary shows the generalized detection rule in the form of a decision tree. The actually JSON-formatted and implemented correlation rules can be found at Github⁸.

Furthermore, parameters for risk calculation are preset during this step. First, the priority of the incident I_{Pri} is set. This parameter determines the severity of the incident. In the case of Dsiem the priority is represented as a value between 1 and 5. Since the severity of the attack is estimated as medium, $I_{Pri} = 3$ is assumed. Second, the reliability I_{Rel} of the detection is set to a value between 1 and 10. This is based on the two-stage rules as explained above. If it was recognized that no more data is received from PLC3, $I_{Rel} = 3$ is predefined. The reliability of the first stage is rather low, as the warning might also indicate a general connection problem of PLC3. However, if the second stage rule (PLC3 is operating normally) applies in addition, $I_{Rel} = 8$ is predefined, since an attack is more likely, but there is still the possibility of a false positive.

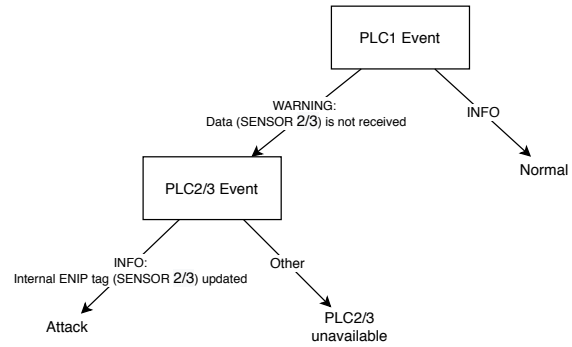


Figure 7: Deduced detection rules for MITM Attack as decision tree

SIEM: Incident detection and handling. To enable incident investigation and the derivation of detection rules or patterns a SIEM system is utilized. To instantly detect the attack within the produced incident data D_{Inc} and generate an alert, the security rule is implemented: $SIEM(I_{Inc}, D_{Inc})$. The physical result of Listing 1 would be a slight overflow of the bottle. The SIEM alert can inform the SOC that an incident Inc is detected. The SOC, in turn, can react as quickly as possible or even install new mechanisms that completely prevent the overflow scenario. During this step, the risk of the incident I_{Ris} is calculated based on I_{Rel} , I_{Pri} , and I_{Asset} (more specifically the value of the asset). In the case of Dsiem the formula $I_{Ris} = \frac{I_{Rel} \cdot I_{Pri} \cdot I_{Asset}}{25}$ is applied. This results in a risk value $I_{Ris} = \frac{8 \cdot 3 \cdot 2}{25} = 1,92$ which triggers a low risk alarm. The calculated risk I_{Ris} and the detected incident Inc fulfill REQUIREMENT 4.

⁸https://github.com/FrauThes/DigitalTwin-SIEM-integration/blob/master/deployments/docker/conf/dsiem/configs/directives_dsiem-digital_twin.json

ARES 2020, August 25–28, 2020, Virtual Event, Ireland

Dietz et al.

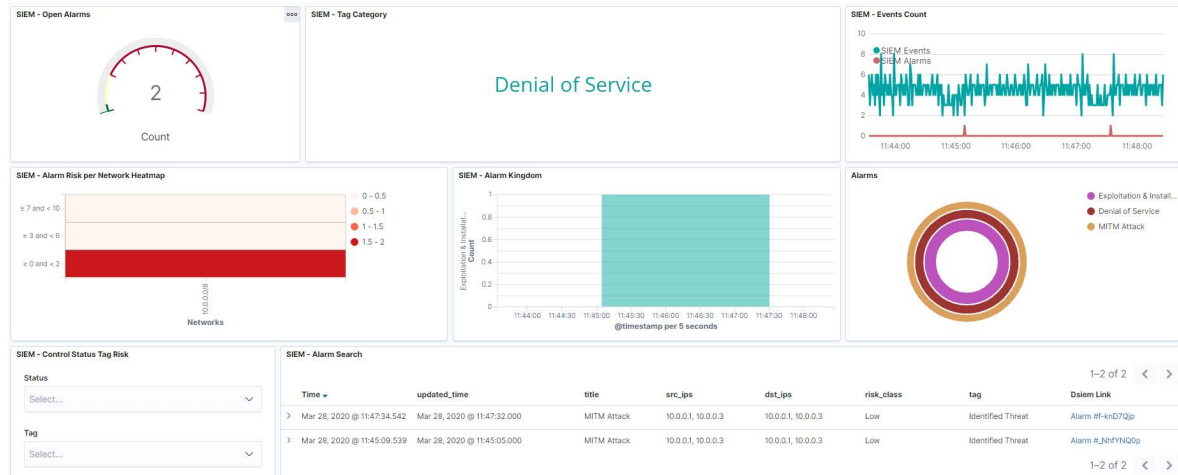


Figure 8: Screenshot of the Dsiem incident detection.

Figure 8 shows the visualization of the prototypical recognition of the incident by Dsiem. In the upper right corner a line chart shows the log data stream (turquoise) and the detected attacks (red peaks). The heatmap on the left side shows the risk level of the detected incidents and allows a quick overview of the threat level. If the incident is successfully detected, one can proceed to deploy the implementation in the production SIEM.

SIEM: Deployment. As in the use case, the asset is in the Planning & Design phase, no actual physical asset exists. Therefore, the deployment is currently accomplished in the virtual environment (cf. further explanation in the following section) by carrying out the deduced security rule.

4.4 Discussion

In order to reach significant results, the use case and its simulation requires to be realistic. To achieve a setup as close as possible to reality, the filling plant use case is oriented on the input of a security expert of an industrial company⁹ that plans, develops and manufactures machines and complete lines for the fields of process, filling and packaging technology. Additional statistics support the use case settings. For instance, a recent study shows that the communication's over Industrial Ethernet annual growth is at 20%, whereby ENIP is applied the most [21]. Moreover, regarding the current states of OT solutions, software patching remains a slow process [15], representing easy entry points for the attackers. So, the attacker already being in the network topology as illustrated in Figure 6 is a realistic setting indeed. While the ARP spoofing represents also a realistic attack against ICS, future attack simulations could cover more advanced and more affecting scenarios (e.g. ransomware attacks).

Regarding the process-based framework, it might be questionable to what extent other simulation technologies for cyber security can be applied instead of the digital twin. While the concept might be

⁹<https://www.krones.com/en/>

realizable with technologies such as testbeds, the digital twin is however the most fitting solution for the following reasons: Digital twins cover the whole lifecycle of their physical counterparts. This means that they can be created in the Planning & Design phase, where no physical counterpart might exist yet, but data can already be gathered. In terms of security, this enables creating an asset by following security-by-design paradigm as the digital twin might already simulate a planned industrial asset and security-related vulnerabilities can be found before the real-world assets exists and operates [6]. This goes hand in hand with the digital twins being mere virtual technologies, and, esp. in comparison to testbeds, no physical assets are mandatory. The main argument for employing digital twins in our framework is that digital twins contain by far more data and analytic capacities than simulations alone (see Figure 2), such as asset-centric data (e.g. context data, domain- and expert-knowledge). This data presents additional important input for SOC and SIEM systems (cf. Figure 1).

5 CONCLUSION AND FUTURE WORK

In this work, a process-based framework for integrating digital twin security simulations in SOCs is developed. By prototypical implementation and use case demonstration, it is shown that the suggested approach can be accomplished in practice. As the use case's digital twin security simulation repeats the MITM attack only about every two minutes and sniffs the occurring traffic for only a few seconds, the attack wouldn't be noticed without logs. This is reminiscent of the espionage phase of APTs, which often target industrial environments. However, in industrial systems currently many proprietary, but no standardized, log management solutions are pursued. In this work, we highlight how the digital twin security simulation can support the SOC in their security strategies (e.g. log management and monitoring).

The presented approach is the first of its kind to combine the research areas digital twin and SOC and thus, requires demonstration

of feasibility first. In the scope of this work, an evaluation in terms of efficiency or performance is not carried out yet. However, future work will address the performance-based evaluation as well as the creation of more and different attacks to simulate, and the subsequent processing of the simulation output in the SOC and SIEM systems. As addressed above, future work could further extend the framework, e.g. by passing on further security-relevant data from the digital twin to the SOC.

Moreover, future research should tackle to integrate even further – towards cyber threat intelligence (CTI). For instance, the structured description of attacks (e.g. STIX format) could be applied to the simulation data. This includes the report of found vulnerabilities, e.g. via lists like the Common Vulnerabilities and Exposures (CVE). Seeing the even bigger picture, the presented approach could work together with further attack knowledge to simulate the attacks as close to reality as possible, e.g. by integrating knowledge from honeypots to get current methods of the attackers.

Regarding SIEM systems, one major research challenge lies in the complexity of creating detection and correlation rules. Since this requires the definition of multiple parameters and since this syntax for rule definition is designed for SIEM experts, SOC analysts struggle with creating rules. Future work should address this problem by designing approaches, which are targeted at a broader security audience and offer a higher usability – while still providing the ability to define rules for detection of complex incidents. Lowcode approaches, tailored to the special demands within a SOC, could present promising solutions to this problem.

ACKNOWLEDGMENTS

This work is partly performed under the ZIM SISSeC project (<https://www.it-logistik-bayern.de/produktionslogistik/projekt-sissec>), which is supported under contract by the German Federal Ministry for Economic Affairs and Energy (16KN085725).

We further thank Andreas Reisser from the Krones Group (<https://www.krones.com/en/>) for the helpful remarks to the industrial use case and the professional support from the industry.

REFERENCES

- [1] Daniele Antonioli and Nils Ole Tippenhauer. 2015. MiniCPS: A Toolkit for Security Research on CPS Networks. In *Proceedings of the First ACM Workshop on Cyber-Physical Systems Security and/or Privacy (CPS-SPC '15)*. ACM, New York, NY, USA, 91–100.
- [2] Stefan Boschert, Christoph Heinrich, and Roland Rosen. 2018. Next Generation Digital Twin. In *Proceedings of the 12th International Symposium on Tools and Methods of Competitive Engineering (TMCE 2018)*, 209–217.
- [3] B. Chen, N. Pattanaik, A. Goulart, K. L. Butler-purpy, and D. Kundur. 2015. Implementing attacks for modbus/TCP protocol in a real-time cyber physical system test bed. In *2015 IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR)*, 1–6. <https://doi.org/10.1109/CQR.2015.7129084>
- [4] Marcello Cinque, Domenico Cotroneo, and Antonio Pecchia. 2018. Challenges and Directions in Security Information and Event Management (SIEM). In *2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. IEEE, 95–99. <https://doi.org/10.1109/ISSREW.2018.00-24>
- [5] Marietheres Dietz and Günther Pernul. 2020. Digital Twin: Empowering Enterprises Towards a System-of-Systems Approach. *Business & Information Systems Engineering* 62, 2 (2020), 179–184. <https://doi.org/10.1007/s12599-019-00624-0>
- [6] Marietheres Dietz and Günther Pernul. 2020. Unleashing the Digital Twin's Potential for ICS Security. *IEEE Security Privacy* (2020). <https://doi.org/10.1109/MSEC.2019.2961650>
- [7] Marietheres Dietz, Benedikt Putz, and Günther Pernul. 2019. A Distributed Ledger Approach to Digital Twin Secure Data Sharing. In *Data and Applications Security and Privacy XXXIII*, Simon N. Foley (Ed.), Springer International Publishing, Cham, 281–300. https://doi.org/10.1007/978-3-030-22479-0_15
- [8] Matthias Eckhart and Andreas Ekelhart. 2018. A Specification-Based State Replication Approach for Digital Twins. In *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and Privacy (CPS-SPC '18)*. ACM, New York, NY, USA, 36–47. <https://doi.org/10.1145/3264888.3264892>
- [9] Matthias Eckhart and Andreas Ekelhart. 2018. Towards Security-Aware Virtual Environments for Digital Twins. In *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security (CPSS '18)*, 61–72. <https://doi.org/10.1145/3198458.3198464>
- [10] Matthias Eckhart and Andreas Ekelhart. 2019. *Digital Twins for Cyber-Physical Systems Security: State of the Art and Outlook*. Springer International Publishing, Cham, 383–412. https://doi.org/10.1007/978-3-030-25312-7_14
- [11] B. Ferguson, A. Tall, and D. Olsen. 2014. National Cyber Range Overview. In *2014 IEEE Military Communications Conference*, 123–128. <https://doi.org/10.1109/MILCOM.2014.27>
- [12] Michael Grieves and John Vickers. 2017. *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*. Springer International Publishing, Cham, 85–113. https://doi.org/10.1007/978-3-319-38756-7_4
- [13] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu. 2013. Cyber-Physical Security Testbeds: Architecture, Application, and Evaluation for Smart Grid. *IEEE Transactions on Smart Grid* 4, 2 (2013), 847–855. <https://doi.org/10.1109/TSG.2012.2226919>
- [14] Diana Kelley and Ron Moritz. 2006. Best Practices for Building a Security Operations Center. *Information Systems Security* 14, 6 (2006), 27–32. <https://doi.org/10.1201/1086.1065898X/45782.14.6.20060101/91856.6>
- [15] Peter Kieseberg and Edgar Weippl. 2018. Security Challenges in Cyber-Physical Production Systems. In *Software Quality: Methods and Tools for Better Software and Systems*, Dietmar Winkler, Stefan Biffl, and Johannes Bergsmann (Eds.), Springer International Publishing, Cham, 3–16. https://doi.org/10.1007/978-3-319-71440-0_1
- [16] Hung-Jen Liao, Chun-Hung [Richard] Lin, Ying-Chih Lin, and Kuang-Yuan Tung. 2013. Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications* 36, 1 (2013), 16 – 24. <https://doi.org/10.1016/j.jnca.2012.09.004>
- [17] Afsaneh Madani, Saeed Rezayi, and Hossein Gharraee. 2011. Log management comprehensive architecture in Security Operation Center (SOC). In *2011 International Conference on Computational Aspects of Social Networks (CASoN)*. IEEE, 284–289. <https://doi.org/10.1109/CASON.2011.6085959>
- [18] David Miller, Shon Harris, Allen Harper, Stephen VanDyke, and Chris Blask. 2011. *Security information and event management (SIEM) implementation*. McGraw-Hill, New York, NY.
- [19] Sparsh Mittal. 2014. OPNET: An Integrated Design Paradigm for Simulations.
- [20] Elisa Negri, Luca Fumagalli, and Marco Macchi. 2017. A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manufacturing* 11 (2017), 939–948. <https://doi.org/10.1016/j.promfg.2017.07.198>
- [21] Joakim Nideborn. 2019. Industrial network market shares 2019 according to HMS. <https://www.hms-networks.com/news-and-insights/news-from-hms/2019/05/07/industrial-network-market-shares-2019-according-to-hms>. [Online; accessed 19-Mar-2020].
- [22] Cuong Pham, Dat Tang, Ken-ichi Chinen, and Razvan Beuran. 2016. CyRIS: A Cyber Range Instantiation System for Facilitating Security Training. In *Proceedings of the Seventh Symposium on Information and Communication Technology (SoICT '16)*. ACM, New York, NY, USA, 251–258. <https://doi.org/10.1145/3011077.3011087>
- [23] R. Piggan and I. Buffey. 2016. Active defence using an operational technology honeypot. In *11th International Conference on System Safety and Cyber-Security (SSCS 2016)*, 1–6. <https://doi.org/10.1049/cp.2016.0860>
- [24] Juan E. Rubio, Rodrigo Roman, and Javier Lopez. 2018. Analysis of Cybersecurity Threats in Industry 4.0: The Case of Intrusion Detection. In *Critical Information Infrastructures Security*, Gregorio D'Agostino and Antonio Scala (Eds.), Springer International Publishing, Cham, 119–130. https://doi.org/10.1007/978-3-319-99843-5_11
- [25] Stef Schinagl, Keith Schoon, and Ronald Paans. 2015. A Framework for Designing a Security Operations Centre (SOC). In *2015 48th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2253–2262. <https://doi.org/10.1109/HICSS.2015.270>
- [26] Thomas H.J. Uhlemann, Christian Lehmann, and Rolf Steinhilper. 2017. The Digital Twin: Realizing the Cyber-Physical Production System for Industry 4.0. In *Procedia CIRP*, Vol. 61. Elsevier B.V., 335–340. <https://doi.org/10.1016/j.procir.2016.11.152>
- [27] Manfred Vielberth, Florian Menges, and Günther Pernul. 2019. Human-as-a-security-sensor for harvesting threat intelligence. *Cybersecurity* 2, 23 (2019). <https://doi.org/10.1186/s42400-019-0040-0>
- [28] Manfred Vielberth and Günther Pernul. 2018. A Security Information and Event Management Pattern. In *12th Latin American Conference on Pattern Languages of Programs (SugarLoafPLoP)*. The Hillside Group.



9 A Digital Twin-Based Cyber Range for SOC Analysts

Current status:	Published
Conference:	35th Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy, DBSec 2021, Calgary, Canada, July 19 - 20
Date of acceptance:	27 Mai 2021
Full citation:	VIELBERTH, M., GLAS, M., DIETZ, M., KARAGIANNIS, S., MAGKOS, E., AND PERNUL, G. A Digital Twin-Based Cyber Range for SOC Analysts. In <i>Data and Applications Security and Privacy XXXV</i> , vol. 12840 of <i>Lecture Notes in Computer Science</i> . Springer, Cham, 2021, pp. 293–311
Authors' contributions:	Manfred Vielberth 30% Magdalena Glas 20% Marietheres Dietz 15% Stylios Karagiannis 15% Emmanouil Magkos 10% Günther Pernul 10%

Conference description: DBSec is an annual international conference covering research in data and applications security and privacy. The conference seeks submissions from academia, industry, and government presenting novel research on all theoretical and practical aspects of data protection, privacy, and applications security.



A Digital Twin-Based Cyber Range for SOC Analysts

Manfred Vielberth¹ , Magdalena Glas¹ , Marietheres Dietz¹ ,
Stylianos Karagiannis² , Emmanouil Magkos² , and Günther Pernul¹ 

¹ Chair of Information Systems, University of Regensburg, Regensburg, Germany
{manfred.vielberth,magdalena.glas,marietheres.dietz,
guenther.pernul}@ur.de

² Department of Informatics, Ionian University, Corfu, Greece
{skaragiannis,emagos}@ionio.gr

Abstract. Security Operations Centers (SOCs) provide a holistic view of a company's security operations. While aiming to harness this potential, companies are lacking sufficiently skilled cybersecurity analysts. One approach to meet this demand is to create a cyber range to equip potential analysts with the skills required. The digital twin paradigm offers great benefit by providing a realistic virtual environment to create a cyber range. However, to the best of our knowledge, tapping this potential to train SOC analysts has not been attempted yet. To address this research gap, a concept of a digital twin-based cyber range for SOC analysts is proposed and implemented. As part of the virtual training environment, several attacks against an industrial system are simulated. Being provided with a SIEM system that displays the real-time log data, the trainees solve increasingly complex tasks in which they have to detect the attacks performed against the system. Thereby, they learn how to interact with a SIEM system and create rules that correlate events aiming to detect security incidents. To evaluate the implemented cyber range, a comprehensive user study demonstrates a significant increase of knowledge within SIEM-related topics among the participants. Additionally, it indicates that the cyber range was subjectively perceived as a positive learning experience by the participants.

Keywords: Cyber range · Security operations center · Digital twin

1 Introduction

As cyber-attacks become increasingly sophisticated and use more and more points of attack, it is essential to establish a holistic view of organizations' security. As a recently published report [2] indicates, organizations are becoming better at detecting and mitigating direct attacks. However, more advanced attacks are on the rise, targeting the victim indirectly through weak spots in the business ecosystem or the supply chain. Over the recent years, Security Operations Centers (SOCs) have emerged to address this problem by providing a holistic

© IFIP International Federation for Information Processing 2021

Published by Springer Nature Switzerland AG 2021

K. Barker and K. Ghazinour (Eds.): DBSec 2021, LNCS 12840, pp. 293–311, 2021.

https://doi.org/10.1007/978-3-030-81242-3_17

view of organizations' cybersecurity. However, this has increased the demand for security personnel, making it difficult to find enough well-trained analysts for SOCs. This is worsened by the so-called "alert burnout", since an analyst's daily work can be quite tedious and tiring. According to a SANS survey [23], the key to low attrition rates is to invest more in analysts' training. Therefore, it is crucial to create a means to train analysts as quickly and effectively as possible, considering that the requirements can vary from company to company. To create a suitable training environment, cyber ranges can be used to train analysts by simulating realistic scenarios without disrupting business operations. To be as close as possible to the specifics of the company, the integration of a digital twin is a promising option. Thereby, the relevant section of the company infrastructure for which the experts are to be trained can be mirrored, creating a training environment that barely differs from the company's real environment.

The contribution of this paper is twofold. First, we examine which components of a digital twin can be used for cyber ranges. Based on this, a cyber range for SOC analysts is designed and prototypically implemented. To show that the proposed concept offers advantages for the training of security analysts, it is evaluated through an extensive empirical user study.

The remainder of this paper is structured as follows. Section 2 provides the foundation of the conducted research. In Sect. 3, the digital twin's potential for cyber ranges is outlined along with the current research gap. Based on that, Sect. 4 proposes a concept for a digital twin-based cyber range, including a scenario and learning concept and concludes with a description of the prototypical implementation of the concept. Section 5 covers the evaluation of the concept in the form of a comprehensive user study by presenting the methodology and the results of the evaluation. Finally, the work is concluded in Sect. 6.

2 Background and Related Work

2.1 Cyber Range

As conventional training methods that only focus on transferring theoretical knowledge do not meet the demand for practical knowledge and skills within the cybersecurity domain, cyber ranges have gained attention over the past years [32]. Generally, cyber ranges are virtual environments, which are used for cybersecurity training [28]. As the name indicates, the expression is derived from shooting range, as both provide an environment in which people can be trained without harming or interfering with the environment for which they are educated. Application areas range from public settings such as military defense and intelligence, academic and educational, to commercial purposes driven by the industry [29].

The idea of using cyber ranges to train specialists in attack detection and in cybersecurity in general is not entirely new. For example, the Austrian Institute of Technology recently introduced a cyber range of industrial control systems [20], not only targeting education, but also serving as a platform for conducting research and development by testing new approaches and methods. This

is only one example in this context. For a deeper insight into related approaches, we would like to refer to two extensive literature reviews which provide a good overview of preliminary work [29, 32]. Although some works in this area exist, to the best of our knowledge, no approach combines the digital twin's potential with the concept of cyber ranges for educating SOC analysts to date. Additionally, the effect of the approaches on the obtained knowledge has not gained sufficient attention in previous works.

According to Yamin et al. [32], a cyber range can be described by following a taxonomy with six domains. However, the description of a cyber range does not necessarily have to consider all domains, but instead, can focus on selected ones. As this paper applies the taxonomy for describing the developed cyber range, the six domains are elaborated briefly in the following:

Scenario: A scenario defines the storyline and context of a training exercise performed in a cyber range. It supports the purpose of the training, such as education, experimenting, or testing. Thereby, it is allocated to a domain (e.g., networking, critical infrastructure, or IoT). Additionally, a scenario can either be static or dynamic. A dynamic scenario means that changes are made during the exercise, for example, by simulating infrastructure components.

Environment: The environment presents the topology in which the scenario is executed. This includes the underlying technology used to build a system model (simulation, emulation, hardware, or hybrid).

Teaming: Teaming describes which teams are part of the scenario. The most important teams are a red team with the goal to exploit vulnerabilities of the system, and a blue team with the task to defend the system against attacks. Teams can also be autonomous if specific technologies automate them.

Learning: The learning domain covers explanatory elements of a scenario such as texts, images, or video clips used for initial knowledge transfer.

Monitoring: Participants' actions can be monitored in real-time during an exercise by using appropriate tools.

Management: This domain covers how management tasks, such as role and resource allocation, are performed. It also comprises interfaces for controlling the scenario or the environment during the exercise.

Furthermore, it is worth mentioning in this context that the term can be narrowed down further. Kavallieratos et al. [17] define a cyber-physical range as a testbed that enables the testing of the security posture of cyber-physical systems. The cyber range presented in this paper can be assigned to this class.

2.2 Security Operations Center (SOC)

The term Security Operations Center has been around in research for more than a decade. However, attention has significantly increased in the last three to five years as SOC's have emerged as a central pivotal point for security operations in practice [30]. The SOC represents an organizational aspect of an enterprise's security strategy. It combines processes, technologies, and people [21, 27] to manage and enhance an organization's overall security posture. This goal can usually

296 M. Vielberth et al.

not be accomplished by a single entity or system, but rather by a complex structure. It creates situational awareness, mitigates the exposed risks, and helps to fulfill regulatory requirements [19]. Additionally, a SOC provides governance and compliance as a framework in which people operate and to which processes and technologies are tailored. A central role within a SOC is taken by security analysts. Using appropriate tools, they can attempt to detect security incidents, then analyze them and react appropriately. Therefore, the success of a SOC depends to a large extent on the skills and training of the analysts. Within a SOC, a SIEM system is usually used as the central tool [31]. A SIEM aims to collect security-relevant data (usually log data) in a central location and analyze it in a correlated manner to detect security incidents. For this purpose, SIEM systems use detection rules that are usually created by analysts, in most cases in JSON or XML format. These fulfill the purpose of triggering an alert if defined conditions within the log data apply.

2.3 Digital Twin

The digital twin refers to a concept that differs in meaning depending on its application area [22]. In general, a digital twin can be defined as a virtual representation of any real-world asset (e.g., system or process). The digital twin accompanies its real-world asset's lifecycle, which may range from phases like idea/planning over operation to decommissioning [6]. The digital twin gathers data about its real-world twin during these phases and enriches the data with semantics [3]. This way, the twin is able to represent its counterpart in-depth and provides a solid basis for simulations and further analytical measures.

Especially in cybersecurity, the digital twin holds several benefits [26]. It can support lifecycle security [11], including the security-by-design paradigm by offering simulations and system testing, in which the security level of the asset can be assessed. Moreover, digital forensics may profit from the vast data and documentary capabilities of a digital twin [7].

3 Investigating the Potential of the Digital Twin for Building Cyber Ranges

In order to extract what digital twins offer for cyber ranges, we must first regard the foundation of digital twin deployment in cybersecurity. According to [11], the digital twin is required to provide sufficient fidelity for security measures that rely on its data. A digital twin offering this characteristic can then be successfully implemented for cybersecurity. This definition presents the prerequisite for combining digital twins and cyber ranges. Currently, one work conceptually proposes to utilize a digital twin as a cyber range [4]. However, an implementation has not been realized to date. In their approach, the digital twin is merely applied as cyber range with the purpose of security training, while other purposes are not considered. However, the digital twin originally serves completely

different purposes, such as monitoring and controlling its counterparts' operation [6]. Thus, in this paper, we propose to use the digital twin as a valuable input to create a cyber range rather than turning it into one. In this matter, we investigate which digital twin characteristics can provide valuable input for cyber ranges in the following. The core parts included in a digital twin represent (a) *data, enhanced with semantic technologies*, (b) *analysis, simulation and other intelligent services* as well as (c) *access control and interfaces* [6].

Data of the digital twin's real-world counterpart is produced along its lifecycle, stored in the digital twin and given context by adding *semantics* [3]. This data supports high-fidelity modeling of the counterpart to virtually represent the real-world system. Added semantics offer better comprehension and modeling of the connection and the context of the system's components. This can prove to be an essential input for creating cyber ranges as well. To maximize the training potential of cyber ranges, the virtually represented system and related security incidents should resemble reality as close as possible. This way, security analysts can be trained in a highly realistic environment. However, not all data held in a digital twin may be relevant for building cyber ranges. The virtual system, used to build a cyber range might represent only a part of a complex real-world system, e.g., by focusing on the network level. In this case, the physics-related data of the system might not be of interest. Moreover, the resulting data of digital twin analyses (like predictive maintenance) typically are not relevant. In general, only a subset of digital twin data is required for creating the cyber range – depending on the complexity level, granularity, and the part of the system being represented.

Analysis, simulation, and other intelligent services represent operation modes of a digital twin. According to [7], three modes can be used for security purposes as well: analysis, simulation, and replication. Table 1 summarizes these modes and their potential benefits for building cyber ranges. Each operation mode relies on digital twin data and has already been tackled in terms of security in some works (see Table 1). **Analysis** usually takes historical/state data of the physical counterpart into account to apply analytical measures such as anomaly detection, pattern recognition, etc. For cyber ranges, this data has to be virtually reproduced (moderate effort). However, there is no virtual system

Table 1. Digital twin security operation modes and their potential for cyber ranges.

Operation mode	Required data	Related work	Benefit	Effort
Analysis	Historical/state data	[24]	Low	Moderate
Simulation	Specification data (for emulation)	[8, 10]	High	Moderate
Replication	Specification data (for emulation), historical/state data (stimuli)	[9, 13]	Moderate	High

298 M. Vielberth et al.

that can be explored by security analysts (low benefit). **Simulation**, in contrast, requires only specification data to build the emulation. On top of the emulation, different (security) scenarios can be applied to a virtual system to create a simulation, where the security analyst in training can not only see produced data of the virtual system but also interact with the system (high benefit). Moreover, the simulations can be taken from digital twins and directly used in or tailored to the cyber range (moderate effort). **Replication**, on the other hand, requires high effort to be used for cyber ranges as it relies on integrating not only specification data to build the emulation, but also on current state data of the physical counterpart to defer the stimuli changing the systems state. However, it only provides moderate benefit as the system is always in synchronization with its real-world counterpart and alternative scenarios (e.g., security incidents and countermeasures) cannot be tested.

Other important parts of digital twins are *access control and interfaces* (e.g., implemented in [25]). Although control mechanisms in digital twins for accessing their data and analytic capabilities represent no relevant input for cyber ranges, interfaces might be used to transmit data from the digital twin.

To conclude, some parts of the digital twin offer benefits for building cyber ranges. Especially the operation mode simulation can be used to create a virtual environment close to reality. Such a system simulation model can be directly transferred from the digital twin into the cyber range and – if necessary – customized to meet the cyber range’s needs. The interfaces part of the digital twin might help to transfer the model, while additional data might help to create simulation scenarios or to get an overview of the system that is virtually represented. Overall, the simulation capability of the digital twin presents a valuable input for cyber ranges and will be concentrated on in the following.

4 A Cyber Range for SOC Analysts

To create the cyber range, it is first necessary to define the *learning objectives*, the *target group*, and the *requirements*. In the case of our cyber range, the analyst in training - hereafter referred to as the trainee - should be introduced to the tasks of a SOC analyst and learn how to work with a SIEM system. In the process, he or she should acquire the following skills and know-how:

- S1: Knowledge of how selected incidents or attacks on the industrial system work.
- S2: Manual detection of anomalies or incidents by analyzing log data with a SIEM system.
- S3: Create both syntactically correct and semantically appropriate rules to detect the incidents.

The target group are individuals who want to achieve skills in security analytics within a SOC – for example, because they want to work as analysts in a SOC in the future. They are assumed to have basic cybersecurity skills but have never worked with a SIEM system or in a SOC. Even though incident response often

lies within an analyst's responsibilities, this will not be considered in this cyber range as, in our opinion, it is too complex to start with and would create too steep a learning curve. However, this could be addressed in future work.

One requirement for the cyber range is to run it entirely virtual in order for the trainees to take part in the cyber range remotely without physical presence. This allows the trainees to take part without too much effort. Additionally, it facilitates the evaluation with an international user study. Furthermore, since the user study is to take place in times when – due to COVID-19 restrictions – face-to-face contact should be kept to a minimum, conducting it in a classroom setting is not an option.

In the following, first the general concept of the cyber range is presented. Based on this, the scenario is described with the help of which the user should acquire the skills outlined above. Subsequently, the prototypical implementation of the cyber range is elaborated upon, with a brief description of the technologies used.

4.1 Cyber Range Concept

Our cyber range consists of five main building blocks (compare Fig. 1): A *virtual environment*, a *SOC*, a *management and monitoring unit*, a *learning management system*, and the *digital twin*, which lies outside of the cyber range. Thus, it represents a security analytics service [12] combined with cyber range specific components. In the following, these building blocks are explained in more detail.

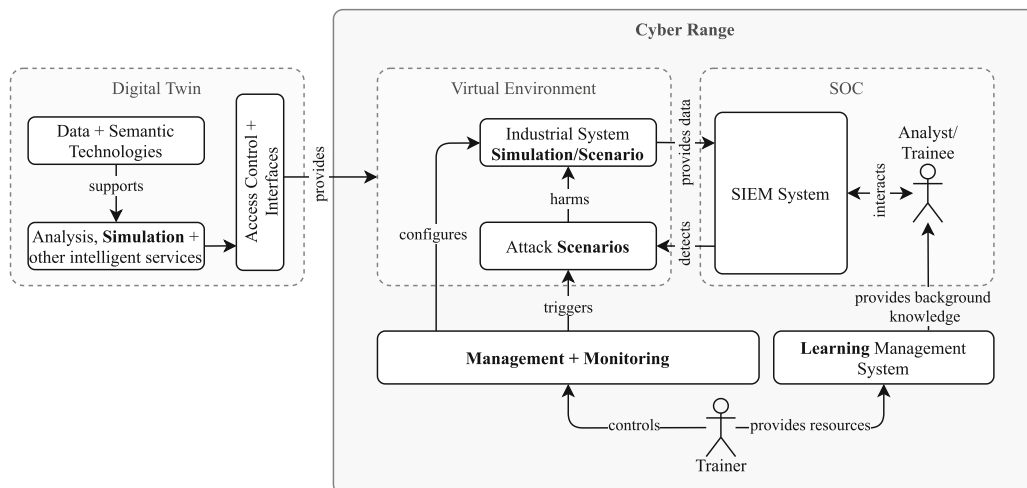


Fig. 1. Basic concept of the digital twin-based cyber range for SOC analysts.

As investigated in Sect. 3, the **digital twin** provides a system *simulation* model used to create the *virtual environment*. The simulation model is supported by specification data, enabling a realistic simulation of the physical counterpart with which the trainee can dynamically interact, like with the real system within

300 M. Vielberth et al.

the organization. The data for the simulation is provided through respective *interfaces* and protected by *access control* capabilities.

The **virtual environment** implements and reflects the scenario of the cyber range through the simulation. For this purpose, an *industrial system* is simulated on the one hand and simulated *attacks* harming the industrial system are carried out on the other. Thereby, the planned training *scenario* is reproduced, guiding the trainee through several training units similar to a playbook, elaborated in more detail in Sect. 4.2. In the process, the simulated industrial system produces log data documenting its operation and providing traces pointing to the attack scenarios.

Within the **SOC** building block, a *SIEM system* is provided, which provides the actual point of interaction with the trainee. The SIEM represents the system for which an analyst is trained, and ideally is also a system in practical use in the trainee's organization. This ensures that the trainee learns to work with a system that is as close to the real SIEM as possible or even identical to it. The log data of the industrial system is fed into the SIEM. In the first step, the trainee interacts with the SIEM to analyze and manually detect the simulated attacks based on the available data. In the next step, the trainee can use this to create correlation rules in the SIEM, which detect attacks automatically.

The **learning management system (LMS)** provides additional learning material for the trainee and introduces the scenario. This information can be presented in various forms, such as videos or simple textual descriptions. In our case, an introduction to the functioning of SIEM systems and the structure of SIEM rules is provided. In addition, hints on the attacks are given to make it easier to get started using the SIEM. These materials are prepared by the trainer and are included in the LMS so that they can be accessed during the procedure. A more detailed description of the prepared media is given in Sect. 4.2.

With the help of the **management and monitoring** building block, the trainer can oversee the trainees' progress during training. Additionally, it configures the simulation of the industrial system and automatically triggers attack simulations depending on the progress of the training.

4.2 Scenario and Learning Concept

The scenario represented by the cyber range is an Industrial Control System (ICS)-based setting of a filling plant. Thereto, the simulation from the digital twin is used, which enables a realistic representation of the industrial filling plant. Figure 2 illustrates the setting in a simplified way for better understanding. The filling plant consists of a tank containing liquid that is to be filled into bottles. The tank is equipped with a sensor measuring the liquid level at regular intervals. To control how much liquid is bottled, the system includes a motoric valve that can be opened and closed. The flow-level sensor is being used to check how much liquid flows through the pipe towards the bottle at any given time. The level of the bottle itself is monitored with another sensor. Each sensor and the actuator is controlled by one of the three Programmable Logic Controllers (PLCs) connected through a switch via Ethernet, which store the sensor data and communicate

A Digital Twin-Based Cyber Range for SOC Analysts 301

via Ethernet/IP. The interface between the employees and the industrial plant is realized with the help of a Human-Machine Interface (HMI). This allows an employee to read the measured and logged sensor values and intervene in the plant's operation. Within the scenario, it is assumed that an attacker has gained direct access to the network of the industrial plant. This allows him or her to carry out various attacks, which can then be detected in the SIEM.

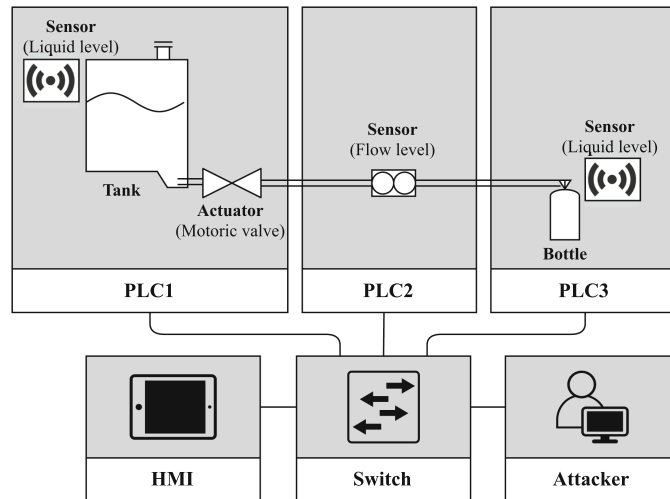


Fig. 2. ICS Scenario of the cyber range.

As shown in Fig. 3 the trainee is guided through the scenario by several learning materials provided by the LMS. Each step within the scenario is accompanied by a task that the trainee must complete.

The scenario is designed to slowly introduce to rule creation by requiring the trainee to solve increasingly elaborate tasks. It starts with a general introduction, where only simple questions about the events captured by the SIEM have to be answered. Once the first step is complete, increasingly complex attacks are simulated one after another, which the trainee must first detect manually (S2). Then he or she is required to create rules (S3) that automatically detect these attacks. The rules to be created also increase in complexity. In order not to overtax the trainee, large parts of the rule are initially given, and the trainee only has to add certain parts. Then, starting with the scenario step “log file manipulation”, the trainee has to create the whole rules themselves. The complexity of the rules to be learned can be divided into three difficulty levels: Starting with very simple rules for which only one condition must be met, to multi-stage rules that build on each other and for which several conditions must be met, to rules that also query an IP address range.

The LMS provides various media to support the trainee's learning between each scenario step. These are either explanatory texts or videos that convey knowledge for the subsequent step in the scenario. In each case, the simulated

302 M. Vielberth et al.

attack is briefly presented from the attacker’s point of view (S1) to provide guidance on what the trainee must look for in the SIEM. It also explains how to use the SIEM and how rules are structured. Gamification elements are used to motivate the trainee during the training session. The trainee receives points for each task he or she solves and can use them to move up levels. If a task is answered incorrectly, the trainee can correct the answer, but points are lost to prevent solutions from simply being guessed. If the trainee is stuck, hints can be bought with earned points, which guide towards the solution.

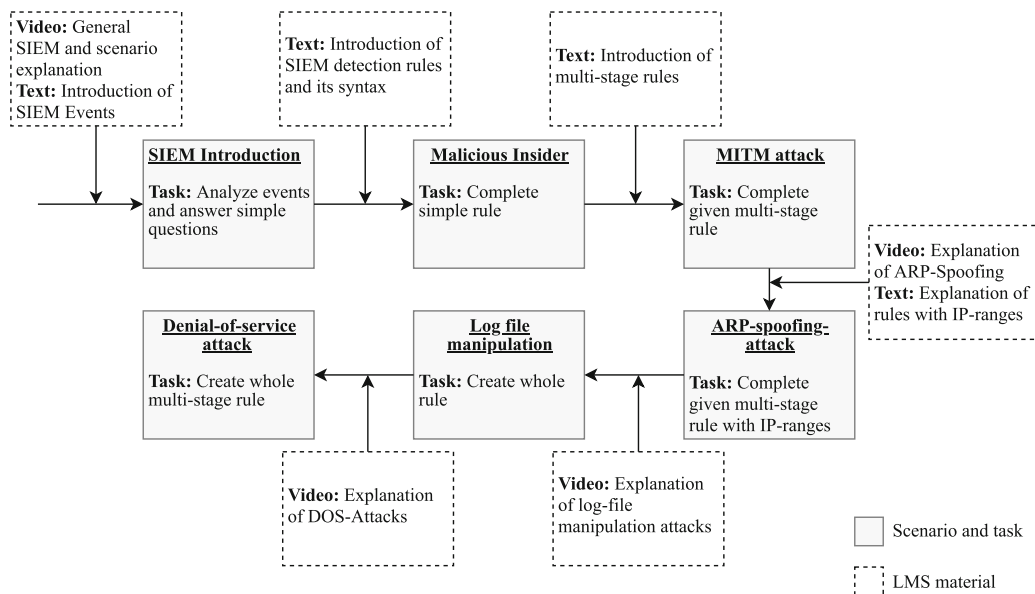


Fig. 3. Learning concept for the cyber range.

4.3 Prototypical Implementation

The overall architecture of the cyber range is shown in Fig. 4. To simulate the industrial system, the digital twin’s simulation component is transferred to the cyber range to create a realistic virtual environment. The simulation is realized with MiniCPS¹, an academic framework for simulating cyber-physical systems which builds upon Mininet². To monitor the network traffic, a firewall captures the TCP-traffic within the network and detects certain abnormalities such as ambiguous responses to ARP-requests. The firewall functionalities are implemented with scapy³. The PLCs and the HMI produce system logs on the main functions of the filling process and the firewall monitoring, which are stored as log files in a common logs directory.

¹ <https://github.com/scy-phy/minicps>.

² <http://mininet.org/>.

³ <https://scapy.net/>.

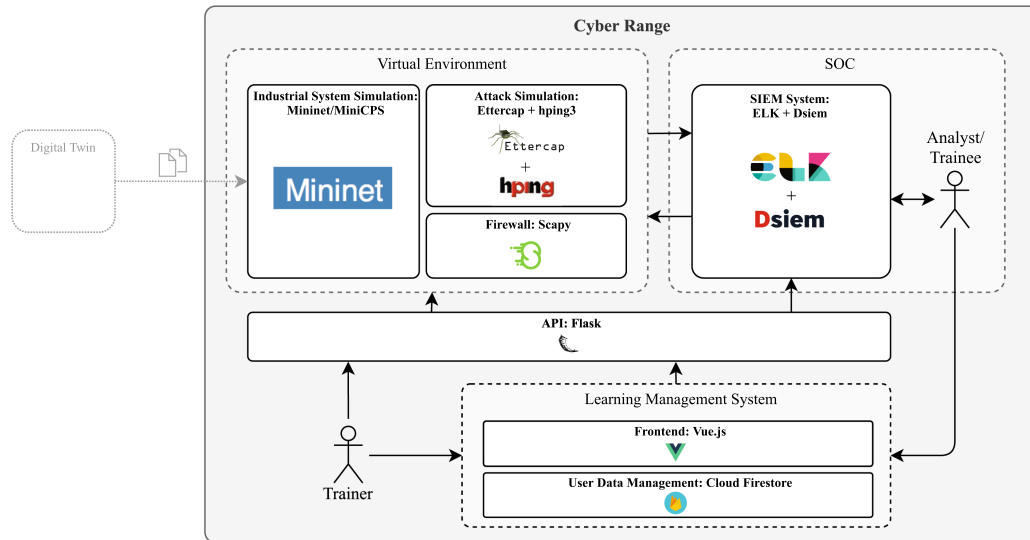


Fig. 4. Architecture of the prototypical implementation.

As described in Sect. 4.2, the attacker performs various attacks against the network components. To implement the attacks, the network tools Ettercap⁴ (for the ARP-Spoofing/Man-In-The-Middle-Attack) and hping3⁵ (for the Denial-Of-Service-Attack) are used. The Log-File-Manipulation-Attack is performed by simply deleting the log file in which the system logs of PLC1 are stored. For the filling plant simulation to produce consistent system logs over the cyber range's lifetime, the attacks are automated and repeated periodically. The open-source tool Dsiem⁶ is the implemented SIEM system of the cyber range. It builds upon Elasticsearch, Logstash, Filebeat, and Kibana. With Logstash and Filebeat, the aforementioned log data is parsed and normalized as so-called SIEM events, which are then forwarded to Dsiem. Dsiem correlates SIEM events with predefined rules to generate SIEM alarms. Finally, these SIEM events and alarms are transferred to Elasticsearch and visualized in Kibana. The virtual environment and the SIEM system are realized as a microservice-infrastructure separated from the LMS and with each component being deployed in a docker container. This modular architecture facilitates reusing the infrastructure for future work and enables its extension as well as the replacement of one or more of the components.

The LMS is realized with the JavaScript framework Vue.js⁷. A screenshot of the user interface of the cyber range is presented in the Appendix (Fig. 6). One section of the LMS displays a Kibana-based SIEM dashboard for Dsiem. It visualizes the SIEM events produced by the digital twin-based simulation and the

⁴ <https://www.ettercap-project.org/>.

⁵ <http://www.hping.org/>.

⁶ <https://www.dsiem.org/>.

⁷ <https://vuejs.org/>.

304 M. Vielberth et al.

SIEM alarms triggered by the Dsiem rules and enables the trainees to interact with the SIEM system in real-time. The other section of the LMS consists of the provided learning material and the tasks the trainees need to complete. The trainee's current score, and the scores of the other trainees taking part in the training at the same time, are displayed on a scoreboard. This functionality is implemented by storing each trainee's current score in a Realtime Firestore⁸. Additionally, a timestamp is saved whenever a trainee completes a task. This enables the trainer to monitor the trainees' progress while the cyber range is being conducted.

The SIEM and the LMS are connected via a REST-API implemented with Flask⁹. Every time a trainee creates a detection rule by completing one of the tasks, an API request is set off to activate the respective rule in Dsiem. Dsiem then starts triggering alarms based on the new rule which are visualized on the SIEM dashboard inside the LMS. The LMS, therefore, enables the trainees to interact directly with the SIEM system and see the impact of detection rules without having to gain a deeper understanding of the project structure of the SIEM system beforehand. Furthermore, the Flask API provides functions for the trainer to interact with the microservice architecture of the digital twin-based simulation and the SIEM system. These functions can be used to start and stop the infrastructure and reset single components in case any technical issues occur while the cyber range training is being conducted. The source code of the project, together with further documentation, is available on GitHub¹⁰.

5 User Study Evaluation

5.1 Method

To measure the effectiveness of the cyber range, it is necessary to evaluate whether it leads to an improvement of the participants' knowledge or skill level. Since a cyber range in our case is similar to a serious game according to the definition of Girard et al. [15], methods from this context can be applied to measure the effectiveness. Besides qualitative methods [16], it is possible to quantitatively evaluate this by measuring the participants' skills and knowledge before and after the training [15]. In the present case, to the best of our knowledge, a comparable system targeting the training of analysts within a SOC does not exist. Therefore it is not possible to evaluate the increase of performance of participants of the cyber range training against participants of a control group in order to compare it to a similar training concept. Instead, it is more suitable to use a one group pre-test/post-test design proposed by Hauge et al. [16] to show whether or not an increase in knowledge has been achieved. Therefore, two assessment questionnaires are constructed consisting of 13 multiple-choice questions (Q1–Q13) for evaluating the learning outcomes of the cyber range. These aim at testing the

⁸ <https://firebase.google.com/docs/firestore>.

⁹ <https://flask.palletsprojects.com/en/1.1.x/>.

¹⁰ <https://github.com/DigitalTwinSocCyberrange>.

knowledge of the participants, whereby four answer options are given for each question. These questionnaires are disseminated before and after the training to measure the improvement of the participant's knowledge.

As the cyber range concept should not only lead to an increase of knowledge but also provide a positive learning experience, the training aims to attract the participant's attention and provide a high level of engagement. Metrics for measuring the engagement levels of the participants are provided by Keller's ARCS model of motivational design [18] which has been used in the past to evaluate security and privacy educational approaches before [14]. It focuses on the intrinsic attributes enhancing motivation, and includes metrics that relate to Attention, Relevance, Confidence, and Satisfaction. The ARCS model can be extended by an extra metric for perceived learning, which measures the subjective impression of whether learning has occurred [1, 5]. This part of the evaluation was implemented by constructing a feedback questionnaire based on the ARCS model, extended by the perceived learning condition. Thereto, the participants can indicate the degree of agreement to 16 statements, with a Likert scale ranging from 1 to 5 ("completely disagree" to "fully agree") after the training.

Participants. Participants were recruited in cybersecurity-related courses at both the University of Regensburg (Germany) and Ionian University (Greece). This ensures that all participants have at least a basic knowledge of cybersecurity, reflecting the target group of the cyber range. In total $n = 44$ test persons participated in the study: 22 German students and 22 Greek students, whereby 12 were female and 32 male. 24 students were undergraduate and 20 were post-graduate students.

Procedure. The study was conducted entirely online over several video conferencing sessions. For each session, 10 virtual machines with one cyber range each were available, limiting the simultaneous number of participants to 10. The user study was divided into three phases. After a short welcome and introduction to the cyber range at the start of the session, the participants were asked to complete the first questionnaire to record their previous knowledge. In the second phase, they were asked to open the cyber range and complete the training contained within. Participation was not time-limited, but most of the participants completed all tasks after a maximum of 2 h. After having completed the second phase, the test persons were asked to fill in the two remaining questionnaires in the third phase, which tested their knowledge afterwards and assessed their motivation during the training. During the execution of the cyber range, we ensured that the trainer intervened as little as possible in the test persons' performance of the tasks in order to avoid influencing them and their results.

5.2 Results

To show that the participants of the study achieved a learning effect, the results from the assessment of the pre-, and post-knowledge are analyzed in the following. The study's questions can be divided into three classes: General knowledge

about cybersecurity attacks, general knowledge about SIEM, and specific knowledge about the structure and functionality of SIEM detection rules. Figure 5 shows the results of the pre-, and post-test. Thereto, the mean percentage of correctly answered questions in both test runs is visualized. The dashed lines indicate the mean in the respective knowledge classes.

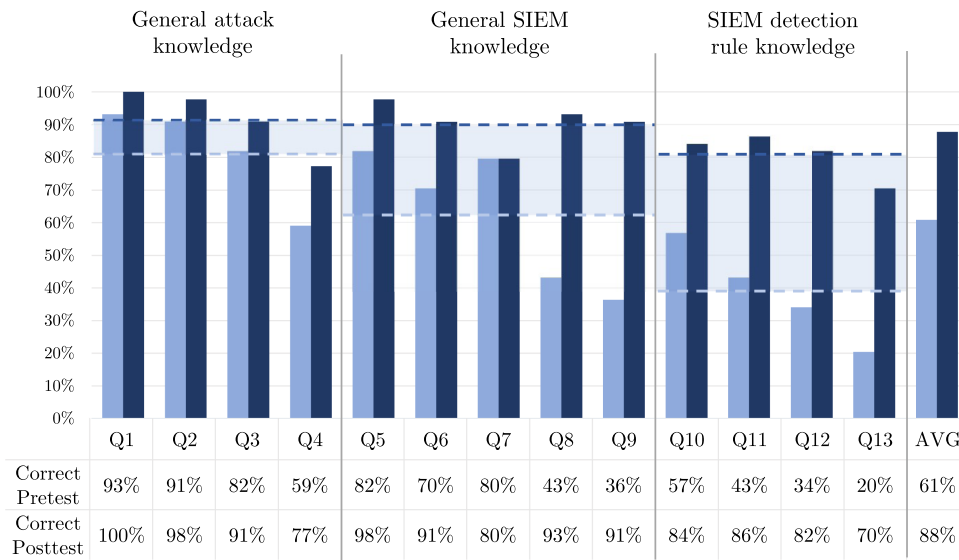


Fig. 5. Comparison of test persons’ knowledge (measured by percentage of correct answers for questions Q1 to Q13) before and after participation in the cyber range. Grouped by knowledge classes, with the dashed lines visualizing the mean of each class.

A paired t-test was conducted to examine the increase in knowledge overall and across the individual knowledge classes¹¹. It shows that the mean of correctly answered questions significantly increased by 26.92% ($t = -12.472, SD = 0.143191, p < 0.001$). In the first class, “general attack knowledge” the mean increase is smaller (10.23%) and less significant ($t = -3.448, SD = 0.196763, p = 0.0013$). This is, however, expected because the test persons possess a certain level of pre-knowledge in cybersecurity and therefore about simple attacks. Thus, the increase from an already high level is smaller. Within the class “general SIEM knowledge”, an increase of 28.18% is observed ($t = -7.398, SD = 0.252681, p < 0.001$). Based on the pre-test, it could be determined that some pre-knowledge was already present within this class. However, a significant increase could still be achieved. Within the “SIEM detection rule knowledge” class, a significant increase of 42.05% is indicated ($t = -8.417, SD = 0.331368, p < 0.001$).

Since an increase in knowledge does not necessarily show that the cyber range was a positive experience for the participants, it is necessary to evaluate the results from the feedback survey. The aggregated results can be found in

¹¹ The SPSS output of the t-test can be found in Fig. 7 in the appendix.

Table 2. The results indicate that the cyber range was, in general, received quite well by the test persons. Both the mean and the median are at least 4 for all conditions on a scale of 1 to 5 (where a higher value indicates the participants' agreement).

Table 2. Results of the feedback questionnaire.

Condition	Mean	Median	Standard deviation
Attention	4.395	5	0.753
Relevance	4.352	4	0.724
Confidence	4.090	4	0.778
Satisfaction	4.284	4	0.738
Perceived learning	4.460	5	0.602

To ensure a high standard of reproducibility and reusability, the anonymized data of all the results and the used questionnaires are available as a public data set¹².

5.3 Discussion

Overall, the results of the user study reveal that an increase in knowledge could be achieved among the participants. Although the increase in general knowledge about attacks (S1) was quite small, a significant increase in knowledge about attack detection using a SIEM system (S2 and S3) is shown – leading to the conclusion that the previously defined goals are achieved. Taking into consideration the results of the evaluation, in the following, we discuss some details we found to be particularly noteworthy.

Within the cyber range, the participants were able to score points by solving the tasks provided as described in Sect. 4.2. The score of a participant thereby indicates to what extent he or she was able to solve the tasks without requiring many attempts to provide the correct solution. While this score was not explicitly used for evaluating the effectiveness of the cyber range, we find it worth examining - especially for participants with particularly high or low increase in knowledge. Five participants showed a notably large increase in knowledge in the assessment questionnaire from 50% or less to more than 90% after participating in the cyber range. The score results of these participants vary from 43 to 100 out of 101 possible points. This shows that though initially failing some tasks of the cyber range, a participant can still gain a large increase of knowledge. In contrast, three participants did not present any improvement in the pre-, and post-assessment. These participants achieved comparably low scores ranging

¹² <https://github.com/DigitalTwinSocCyberRange/userStudy>.

308 M. Vielberth et al.

from 28 to 33 points. This indicates difficulties in engaging with the overall approach. However, it is noteworthy that these participants still provided positive feedback on the cyber range.

Considering the results of the feedback survey, a noticeable aspect is a somewhat lower result for Confidence compared to the other values. This is also confirmed by some participants' oral feedback, who told us that they were somewhat overstrained at the beginning. In our estimation, this was mainly due to an information overload, as they were confronted with both the SIEM and the LMS. In the future, the cyber range could be adapted so that trainees are not shown all information from the start, but only selected content that is then gradually expanded. The value for perceived learning also sticks out, indicating whether the participants themselves assess whether they learned something during the procedure. With a value of 4.460, it is slightly higher than the others. This confirms the result from comparing the pre-, and post-test, as the participants themselves also have the impression of having gained knowledge.

6 Conclusions

This work demonstrates how cyber ranges can be utilized for training security analysts in a SOC. It shows that cyber ranges are suitable for the acquisition of general knowledge about SIEM as well as for specific training on how to create SIEM rules. The provided cyber range concept builds upon the simulation component of a digital twin of an industrial filling plant. This ensures that the analysts are trained based on a realistic scenario. To show the increase in knowledge and the perceived learning experience, the concept is implemented and evaluated in an international study among both Greek and German participants. To the best of our knowledge, this is the first cyber range to utilize the potential of a digital twin, specifically targeting the training of SOC analysts.

Like any other research effort, this paper contains limitations. Since, to our knowledge, no approach with the same objective exists, it was not possible to compare the knowledge gains. However, we were able to show that a cyber range is, in general, suitable for imparting knowledge. Nonetheless, we did not concentrate on an evaluation comparing our cyber range to other concepts.

In summary, this work provides a new approach to train SOC analysts. By proposing security training, it addresses the current problem of the increasing demand for security analysts personnel, which will continue to grow. Furthermore, the attack detection training of SOC analysts is only one of many possible applications of the presented cyber range. Among many other possibilities, it could also be used for penetration testing of industrial plants or incident response exercises in future research.

Appendix

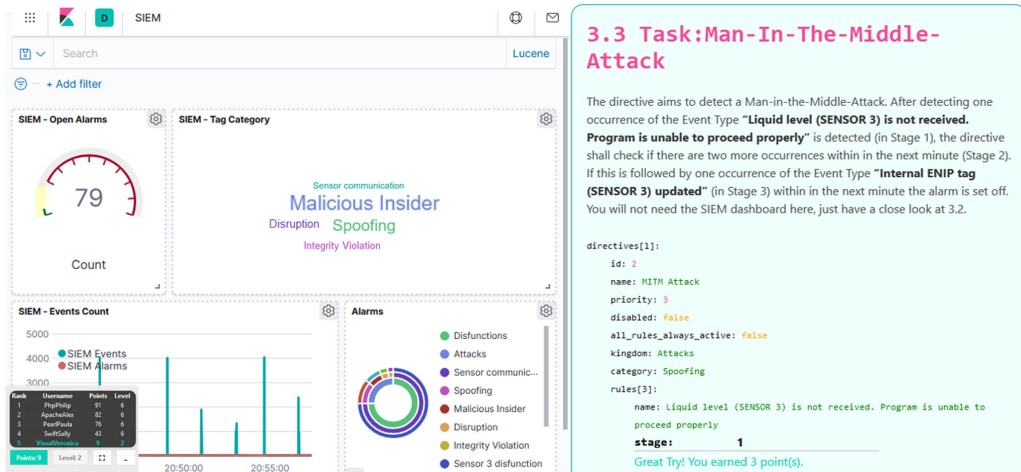


Fig. 6. Screenshot of the cyber range interface: SIEM dashboard and LMS

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Err. Mean
Pair 1	Correct_pre	,60839	44	,139608	,021047
	Correct_post	,87762	44	,148677	,022414
Pair 2	Correct_pre_att	,81250	44	,187742	,028303
	Correct_post_att	,91477	44	,142069	,021418
Pair 3	Correct_pre_SIEM	,62273	44	,280252	,042250
	Correct_post_SIEM	,90455	44	,146199	,022040
Pair 4	Correct_pre_rule	,38636	44	,255489	,038516
	Correct_post_rule	,80682	44	,240298	,036226

Paired Samples Test									
Paired Differences									
		Mean	Std. Deviation	Std. Err. Mean	95% Conf. Interval of the Diff.		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Correct_pre - Correct_post	-,269231	,143191	,021587	-,312765	-,225697	-12,472	43	,000
Pair 2	Correct_pre_att - Correct_post_att	-,102273	,196763	,029663	-,162094	-,042451	-3,448	43	,001
Pair 3	Correct_pre_SIEM - Correct_post_SIEM	-,281818	,252681	,038093	-,358640	-,204996	-7,398	43	,000
Pair 4	Correct_pre_rule - Correct_post_rule	-,420455	,331368	,049956	-,521199	-,319710	-8,417	43	,000

Fig. 7. SPSS output of the t-test

References

1. Barzilai, S., Blau, I.: Scaffolding game-based learning: impact on learning achievements, perceived learning, and game experiences. *Comput. Educ.* **70**, 65–79 (2014)
2. Bissel, K., Lasalle, R., Dal Cin, P.: Third annual state of cyber resilience report. Accenture (2020)
3. Boschert, S., Heinrich, C., Rosen, R.: Next generation digital twin. In: Proceedings of the 12th International Symposium on Tools and Methods of Competitive Engineering, TMCE 2018, pp. 209–217 (2018)

310 M. Vielberth et al.

4. Bécue, A., et al.: CyberFactory1 – securing the industry 4.0 with cyber-ranges and digital twins. In: 2018 14th IEEE International Workshop on Factory Communication Systems (WFCS), pp. 1–4 (2018)
5. Caspi, A., Blau, I.: Social presence in online discussion groups: testing three conceptions and their relations to perceived learning. *Soc. Psychol. Educ.* **11**(3), 323–346 (2008). <https://doi.org/10.1007/s11218-008-9054-2>
6. Dietz, M., Pernul, G.: Digital twin: empowering enterprises towards a system-of-systems approach. *Bus. Inf. Syst. Eng.* **62**(2), 179–184 (2019). <https://doi.org/10.1007/s12599-019-00624-0>
7. Dietz, M., Pernul, G.: Unleashing the digital twin’s potential for ICS security. *IEEE Secur. Priv.* **18**(4), 20–27 (2020)
8. Dietz, M., Vielberth, M., Pernul, G.: Integrating digital twin security simulations in the security operations center. In: Proceedings of the 15th International Conference on Availability, Reliability and Security, ARES 2020. ACM, New York (2020)
9. Eckhart, M., Ekelhart, A.: A specification-based state replication approach for digital twins. In: Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy, CPS-SPC 2018, pp. 36–47. ACM, New York (2018)
10. Eckhart, M., Ekelhart, A.: Towards security-aware virtual environments for digital twins. In: Proceedings of the 4th ACM Workshop on Cyber-Physical System Security (CPSS 2018), pp. 61–72 (2018)
11. Eckhart, M., Ekelhart, A.: Digital twins for cyber-physical systems security: state of the art and outlook. In: Security and Quality in Cyber-Physical Systems Engineering, pp. 383–412. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25312-7_14
12. Empl, P., Pernul, G.: A flexible security analytics service for the industrial IoT. In: Proceedings of the 2021 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems, pp. 23–32. ACM, New York (2021)
13. Gehrman, C., Gunnarsson, M.: A digital twin based industrial automation and control system security architecture. *IEEE Trans. Ind. Inf.* **16**, 669–680 (2020)
14. Giannakas, F., Papasalouros, A., Kambourakis, G., Gritzalis, S.: A comprehensive cybersecurity learning platform for elementary education. *Inf. Secur. J.* **28**(3), 81–106 (2019)
15. Girard, C., Ecalle, J., Magnan, A.: Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *J. Comput. Assist. Learn.* **29**(3), 207–219 (2013)
16. Hauge, J.B., et al.: Study design and data gathering guide for serious games’ evaluation. In: Tennyson, R., Connolly, T.M., Hainey, T., Boyle, E., Baxter, G., Moreno-Ger, P. (eds.) *Psychology, Pedagogy, and Assessment in Serious Games. Advances in Game-Based Learning*, pp. 394–419. IGI Global (2014)
17. Kavallieratos, G., Katsikas, S.K., Gkioulos, V.: Towards a cyber-physical range. In: Proceedings of the 5th on Cyber-Physical System Security Workshop - CPSS 2019, pp. 25–34. ACM Press, New York (2019)
18. Keller, J.M.: Development and use of the ARCS model of instructional design. *J. Instr. Dev.* **10**(3), 2–10 (1987). <https://doi.org/10.1007/BF02905780>
19. Kelley, D., Moritz, R.: Best practices for building a security operations center. *Inf. Syst. Secur.* **14**(6), 27–32 (2006)
20. Leitner, M., et al.: AIT cyber range: flexible cyber security environment for exercises, training and research. In: Proceedings of the European Interdisciplinary Cybersecurity Conference, pp. 1–6 (2020)

A Digital Twin-Based Cyber Range for SOC Analysts 311

21. Madani, A., Rezayi, S., Gharaee, H.: Log management comprehensive architecture in Security Operation Center (SOC). In: 2011 International Conference on Computational Aspects of Social Networks (CASoN), pp. 284–289. IEEE (2011)
22. Negri, E., Fumagalli, L., Macchi, M.: A review of the roles of digital twin in CPS-based production systems. *Procedia Manuf.* **11**, 939–948 (2017)
23. Pescatore, J., Filkins, B.: Closing the critical skills gap for modern and effective security operations centers (SOCs). SANS Institute (2020)
24. Pokhrel, A., Katta, V., Colomo-Palacios, R.: Digital twin for cybersecurity incident prediction: a multivocal literature review. In: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, ICSEW 2020, pp. 671–678. ACM, New York (2020)
25. Putz, B., Dietz, M., Empl, P., Pernul, G.: EtherTwin: blockchain-based secure digital twin information management. *Inf. Process. Manag.* **58**(1), 102425 (2021)
26. Rubio, J.E., Roman, R., Lopez, J.: Analysis of cybersecurity threats in industry 4.0: the case of intrusion detection. In: D’Agostino, G., Scala, A. (eds.) *Critical Information Infrastructures Security*. LNCS, vol. 10707, pp. 119–130. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99843-5_11
27. Schinagl, S., Schoon, K., Paans, R.: A framework for designing a security operations centre (SOC). In: 2015 48th Hawaii International Conference on System Sciences, pp. 2253–2262. IEEE (2015)
28. Tian, Z., et al.: A real-time correlation of host-level events in cyber range service for smart campus. *IEEE Access* **6**, 35355–35364 (2018)
29. Ukwandu, E., et al.: A review of cyber-ranges and test-beds: current and future trends. *Sensors* **20**(24), 7148 (2020)
30. Vielberth, M., Bohm, F., Fichtinger, I., Pernul, G.: Security operations center: a systematic study and open challenges. *IEEE Access* **8**, 227756–227779 (2020)
31. Vielberth, M., Pernul, G.: A security information and event management pattern. In: 12th Latin American Conference on Pattern Languages of Programs (SugarLoafPLoP 2018), pp. 1–12. The Hillside Group (2018)
32. Yamin, M.M., Katt, B., Gkioulos, V.: Cyber ranges and security testbeds: scenarios, functions, tools and architecture. *Comput. Secur.* **88**, 101636 (2020)

Bibliography

- [1] ACCENTURE, AND PONEMON INSTITUTE LLC. The cost of cybercrime, 2018.
- [2] BASKERVILLE, R., BAIYERE, A., GERGOR, S., HEVNER, A., AND ROSSI, M. Design science research contributions: Finding a balance between artifact and theory. *Journal of the Association for Information Systems* 19, 5 (2018), 358–376.
- [3] BECKER, J., KNACKSTEDT, R., AND PÖPPELBUSS, J. Developing maturity models for it management. *Business & Information Systems Engineering* 1, 3 (2009), 213–222.
- [4] BENBASAT, AND ZMUD. The identity crisis within the is discipline: Defining and communicating the discipline’s core properties. *MIS Quarterly* 27, 2 (2003), 183.
- [5] CHEN, P., DESMET, L., AND HUYGENS, C. A study on advanced persistent threats. In *IFIP International Conference on Communications and Multimedia Security* (2014), pp. 63–72.
- [6] CICHONSKI, P., MILLAR, T., GRANCE, T., AND SCARFONE, K. Computer security incident handling guide. *NIST Special Publication 800*, 61 (2012), 1–147.
- [7] CROWLEY, C., AND PESCATORE, J. Common and best practices for security operations centers: Results of the 2019 soc survey, 2019.
- [8] ELSAYED, M. A., AND ZULKERNINE, M. Predictdeep: Security analytics as a service for anomaly detection and prediction. *IEEE Access* 8 (2020), 45184–45197.
- [9] EMPL, P., AND PERNUL, G. A flexible security analytics service for the industrial iot. In *Proceedings of the 2021 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems* (New York, NY, USA, 04282021), ACM, pp. 23–32.
- [10] EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* (2016), 1–88.
- [11] EUROPEAN UNION AGENCY FOR CYBERSECURITY. ENISA threat landscape for supply chain attacks, 2021.

- [12] FEHLING, C., BARZEN, J., BREITENBÜCHER, U., AND LEYMAN, F. A process for pattern identification, authoring, and application. In *Proceedings of the 19th European Conference on Pattern Languages of Programs - EuroPLoP '14* (New York, USA, 2014), ACM Press, pp. 1–9.
- [13] HEARTFIELD, R., AND LOUKAS, G. Detecting semantic social engineering attacks with the weakest link: Implementation and empirical evaluation of a human-as-a-security-sensor framework. *Computers & Security* 76 (2018), 101–127.
- [14] HEVNER, A., VOM BROCKE, J., AND MAEDCHE, A. Roles of digital innovation in design science research. *Business & Information Systems Engineering* 61, 1 (2019), 3–8.
- [15] HEVNER, A. R., MARCH, S. T., PARK, J., AND RAM, S. Design science in information systems research. *MIS Quarterly* 28, 1 (2004), 75–105.
- [16] JOINT TASK FORCE TRANSFORMATION INITIATIVE. *Guide for conducting risk assessments*. National Institute of Standards and Technology, Gaithersburg, MD, 2012.
- [17] KAVANAGH, K. M., AND BUSSA, T. Magic quadrant for security information and event management. *Technical Report - Gartner Inc.* (2017).
- [18] KAVANAGH, K. M., BUSSA, T., AND COLLINS, J. Magic quadrant for security information and event management. *Technical Report - Gartner Inc.* (2021).
- [19] KOCHER, P., HORN, J., FOGH, A., GENKIN, D., GRUSS, D., HAAS, W., HAMBURG, M., LIPP, M., MANGARD, S., PRESCHER, T., SCHWARZ, M., AND YAROM, Y. Spectre attacks: Exploiting speculative execution. In *2019 IEEE Symposium on Security and Privacy (SP)* (2019), IEEE, pp. 1–19.
- [20] LINEBERRY, S. The human element: The weakest link in information security. *Journal of Accountancy* 204, 5 (2007), 44.
- [21] MAHMOOD, T., AND AFZAL, U. Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools. In *2013 2nd National Conference on Information Assurance (NCIA)* (2013), IEEE, pp. 129–134.
- [22] MELLO, J. P. Security awareness training explosion. <https://cybersecurityventures.com/security-awareness-training-report/>, 2017.
- [23] MENGES, F., AND PERNUL, G. A comparative analysis of incident reporting formats. *Computers & Security* 73 (2018), 87–101.
- [24] MOHURLE, S., AND PATIL, M. M. A brief study of wannacry threat: Ransomware attack 2017. *International Journal of Advanced Research in Computer Science* 8 (2017), 1938–1940.

- [25] NICKERSON, R. C., VARSHNEY, U., AND MUNTERMANN, J. A method for taxonomy development and its application in information systems. *European Journal of Information Systems* 22, 3 (2013), 336–359.
- [26] NONAKA, I., AND VON KROGH, G. Perspective—tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory. *Organization Science* 20, 3 (2009), 635–652.
- [27] ÖSTERLE, H., BECKER, J., FRANK, U., HESS, T., KARAGIANNIS, D., KRCDMAR, H., LOOS, P., MERTENS, P., OBERWEIS, A., AND SINZ, E. J. Memorandum on design-oriented information systems research. *European Journal of Information Systems* 20, 1 (2011), 7–10.
- [28] PESCATORE, J., AND FILKINS, B. Closing the Critical Skills Gap for Modern and Effective Security Operations Centers (SOCs). SANS Institute, 2020.
- [29] SCHLETTE, D., CASELLI, M., AND PERNUL, G. A comparative study on cyber threat intelligence: The security incident response perspective. *IEEE Communications Surveys & Tutorials* (2021), 1.
- [30] SCHNEIER, B. *Secrets and Lies: Digital Security in a Networked World*, 15 ed. John Wiley & Sons, 2015.
- [31] TRANFIELD, D., DENYER, D., AND SMART, P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management* 14, 3 (2003), 207–222.
- [32] ZIMMERMANN, V., AND RENAUD, K. Moving from a ‘human-as-problem’ to a ‘human-as-solution’ cybersecurity mindset. *International Journal of Human-Computer Studies* 131 (2019), 169–187.

Appendix

Curriculum Vitae

Manfred Vielberth, M.Sc.

Chair of Information Systems (IFS)
Faculty of Business, Economics, Management Information Systems
University of Regensburg

ACADEMIC EXPERIENCE

2017 - present	Research Assistant <i>Chair of Information Systems, University of Regensburg</i>
2014 - 2016	Graduate Student Research Assistant <i>Faculty of Business, Economics, Management Information Systems, University of Regensburg</i>
2011 - 2014	Student Research Assistant <i>Faculty of Business, Economics, Management Information Systems, University of Regensburg</i>

RESEARCH PROJECT INVOLVEMENT

2021 - present	INSIST (StMWi) - Industrial IoT Security Operations Center <i>Bavarian Ministry of Economic Affairs, Regional Development and Energy</i>
2021 - present	DEVISE (BMBF) - Data Quality Management for Improving In- formation Security <i>Federal Ministry of Education and Research</i>
2016 - 2020	DINGfest (BMBF) - Detection, Visualization, and Forensic Anal- ysis of Security Incidents <i>Federal Ministry of Education and Research</i>

EDUCATION

2017 - present	PhD Student <i>University of Regensburg</i>
2014 - 2017	Master of Science <i>University of Regensburg</i> Thesis title: <i>Classifying images by privacy using machine learning technologies: theoretical review and implementation (translated from German)</i>
2011 - 2014	Bachelor of Science <i>University of Regensburg</i> Thesis title: <i>Implementation and evaluation of trust-based privacy alerts for Facebook (translated from German)</i>
2002 - 2011	High School Graduation <i>Willibald Gymnasium Eichstätt</i>

INDUSTRY EXPERIENCE

2012 - 2013 | **Intern (6 months)**, Information Security Management
Krones AG, Neutraubling

2010 | **Working student (2 months)**
Osram AG, Eichstätt

TEACHING

2017 - 2021 | **Co-Lecturer** - Security of data-intensive Applications
Graduate lecture at University of Regensburg

2017 - 2021 | **Tutor** - Informationsystems - Developments and Trends
Graduate lecture at University of Regensburg

2017 - 2021 | **Tutor** - Corporate Databases
Undergraduate lecture at University of Regensburg

2017 - 2021 | **Coordinator** - Student Seminars
Theoretical and practical seminars at University of Regensburg

2011 - 2016 | **Student Tutor** - Fundamentals of Management Information Systems
Undergraduate lecture at University of Regensburg

2011 - 2016 | **Student Tutor** - Operational Information Processing
Undergraduate lecture at University of Regensburg

SERVICE TO THE RESEARCH COMMUNITY

2021 | **Journal Reviewer**
Computers in Human Behavior Reports

2017 - present | **External Reviewer**
ARES 2021, CCGRID 2021, ESORICS 2021, SECRIPT 2021, ARES 2020, ESORICS 2020, ICCCN 2020, NSS 2020, SECRIPT 2020, TrustCom 2020, DBSec 2019, ESORICS 2019, IFIPTM 2019, ISPEC 2019, SECRIPT 2019, CAiSE 2019, ICISSP 2019, ARES 2018, ESORICS 2018, ISPEC 2018, SECRIPT 2018, STM 2018, SugarLoafPLoP 2018, CAiSE 2018, ICISSP 2018, e-Democracy 2017, ESORICS 2017, NSS 2017, SECRIPT 2017, TrustBus 2017

PUBLICATIONS

- [1] VIELBERTH, M., AND PERNUL, G. A security information and event management pattern. In *12th Latin American Conference on Pattern Languages of Programs (Sugar-LoafPLoP)*. The Hillside Group, 2018, pp. 1–12
- [2] MENGES, F., BÖHM, F., VIELBERTH, M., PUCHTA, A., TAUBMANN, B., RAKOTONDRAVONY, N., AND LATZO, T. Introducing DINGfest: An architecture for next generation SIEM systems. In *Sicherheit*. 2018, pp. 257–260
- [3] VIELBERTH, M., MENGES, F., AND PERNUL, G. Human-as-a-security-sensor for harvesting threat intelligence. *Cybersecurity* 2, 23 (2019), 1–15
- [4] VIELBERTH, M. Security Information and Event Management (SIEM). *Encyclopedia of Cryptography, Security and Privacy*, Springer Berlin Heidelberg, (2019), 1–3
- [5] VIELBERTH, M. Security Operations Center (SOC). *Encyclopedia of Cryptography, Security and Privacy*, Springer Berlin Heidelberg, (2019), 1–3
- [6] DIETZ, M., VIELBERTH, M., AND PERNUL, G. Integrating Digital Twin Security Simulations in the Security Operations Center. In *Proceedings of the 15th International Conference on Availability, Reliability and Security* (2020), ACM, pp. 1–9
- [7] VIELBERTH, M., BÖHM, F., FICHTINGER, I., AND PERNUL, G. Security operations center: A systematic study and open challenges. *IEEE Access* 8 (2020), 227756–227779
- [8] BÖHM, F., VIELBERTH, M., AND PERNUL, G. Bridging Knowledge Gaps in Security Analytics. In *Proceedings of the 7th International Conference on Information Systems Security and Privacy* (Online Streaming, 2021), SCITEPRESS, pp. 98–108
- [9] MENGES, F., LATZO, T., VIELBERTH, M., SOBOLA, S., PÖHLS, H. C., TAUBMANN, B., KÖSTLER, J., PUCHTA, A., FREILING, F. C., REISER, H. P., AND PERNUL, G. Towards GDPR-compliant data processing in modern SIEM systems. *Computers & Security* 103, 102165 (2021), 1–19
- [10] VIELBERTH, M., ENGLBRECHT, L., AND PERNUL, G. Improving data quality for human-as-a-security-sensor. A process driven quality improvement approach for user-provided incident information. *Information and Computer Security* 29, 2 (2021), 332–349
- [11] VIELBERTH, M., GLAS, M., DIETZ, M., KARAGIANNIS, S., MAGKOS, E., AND PERNUL, G. A Digital Twin-Based Cyber Range for SOC Analysts. In *Data and Applications Security and Privacy XXXV*, vol. 12840 of *Lecture Notes in Computer Science*. Springer, Cham, 2021, pp. 293–311
- [12] SCHLETTE, D., VIELBERTH, M., AND PERNUL, G. CTI-SOC2M2 – the quest for mature, intelligence-driven security operations and incident response capabilities. *Computers & Security* 111, 102482 (2021), 1–20
- [13] BÖHM F., VIELBERTH, M., AND PERNUL, G. Formalizing and Integrating User Knowledge into Security Analytics. Submitted to *SN Computer Science*, (2021), 1–28