Universität Regensburg
Fakultät für Wirtschaftswissenschaften
Lehrstuhl für Wirtschaftsinformatik I - Informationssysteme

Leveraging Visual Analytics for Cybersecurity

Dissertation

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft (Dr. rer. pol.)
eingereicht an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg

vorgelegt von
Fabian Konrad Böhm, M.Sc. with Honors

Berichterstatter:
Prof. Dr. Günther Pernul
Prof. Dr. Hans Reiser

Tag der Disputation: 13.05.2022

To my parents, Harald and Bärbl,
And my loving wife Lisa.

Abstract

Securing the highly complex infrastructures of modern organizations against innovative and targeted cybersecurity threats is becoming increasingly challenging. While automated security mechanisms have been significantly improved throughout recent years, they struggle to detect unknown, multi-layered attacks. Only security experts with suitable domain knowledge can surface the respective threats and incidents through in-depth analyses, which makes them an indispensable asset for effective and comprehensive cybersecurity. Therefore, it is necessary to explore how their analysis tasks can be supported and how they can be integrated into cybersecurity activities. Visual Analytics provides a well-established pathway for involving human domain knowledge in complex analytical tasks. However, the body of work applying Visual Analytics to domain problems within cybersecurity is limited and existing work often either lacks visual efficiency or falls short in involving security domain knowledge.

This dissertation sheds light on how Visual Analytics can be leveraged within the cybersecurity domain. As a foundation, the working environment of security experts, formal notions of security-relevant knowledge, and a fundamental iterative process for incident detection are defined. Subsequently, Visual Analytics approaches are leveraged to integrate domain experts into different steps of this process and to support experts' analytical tasks. First, Visual Analytics is utilized to integrate security experts into identifying imminent threats and help them interpret indicators of compromise. Second, when a threat has been detected, security experts need to be supported in the in-depth forensic analysis of the incident. In a third step, Visual Analytics enables the integration of cybersecurity analysts' knowledge into threat intelligence, which provides semantic and actionable insights about an incident.

This work provides a baseline to bridge the gap between Visual Analytics and the cybersecurity domain by highlighting how methodically sound visualization designs can support security experts with their complex analytical tasks. Thus, it indicates a way towards more efficient and comprehensive cybersecurity activities in organizations.

Acknowledgement

I have been accompanied and supported by many colleagues, friends, and family members throughout the past five years. Every one of them has helped me finish this thesis, which would have been an impossible task without all of this continuous support. Thus, I feel honored to thank all of them!

First and foremost, I owe the biggest thanks to my supervisor Prof. Dr. Günther Pernul, for guiding my journey with both professional and personal advice. Prof. Pernul was always ready to listen to the small and not-so-small challenges that arose during my doctoral studies and had an unerring sense of ways to overcome them. I am incredibly grateful for the freedom to direct my research according to my interests and to be able to continuously work on a variety of different aspects without any pressure. As a co-author of most of the research papers written during my studies, he inspired and contributed to essential ideas of this work. I would also like to thank my second supervisor, Prof. Dr. Hans Reiser, for his support.

Second, I have to thank all former and current colleagues from the IFS team. From the beginning, this team never felt like an ordinary team but like some kind of family. I owe special thanks to Manfred Vielberth for the collaboration on various topics. Thank you, Fred, for all the honest and factual discussions, but also all the fun (and the irony). Also, I have to give a special thanks to Florian Menges for his cooperation and support during some initial difficulties. Furthermore, I want to thank all my other co-authors from IFS with whom I have been honored to work: Ludwig Englbrecht, Marietheres Dietz, Benedikt Putz, Daniel Schlette, and Alexander Puchta. I also want to appreciate the next generation of IFS researchers, who have been irreplaceable colleagues ever since they joined the IFS team: Philip Empl, Magdalena Glas, and Sabrina Friedl. Last but not least, I have to thank Petra Sauer and Werner Hueber for their support in technical, emotional, and bureaucratic issues. I really enjoyed the relaxed, collaborative, and inspiring atmosphere at IFS, which we managed to keep alive even during a pandemic.

Third, I need to thank my friends and family. I am especially thankful to my parents for their support and patience throughout the entire course of my academic studies, starting more than ten years ago.

Finally, I would like to take this opportunity to thank my wife Lisa for her tireless support, unsparing honesty, and most importantly, her unconditional love.

Contents

List of Tables

List of Figures

List of Acronyms

CPS	Cyber-Physical System
CTI	Cyber Threat Intelligence
DF	Digital Forensics
DLT	Distributed Ledger Technology
FM	Fileless Malware
HyperSec	Hyperledger Security Explorer
IAM	Identity and Access Management
ICT	Information and Communications Technologies
IDL	Incident Detection Lifecycle
IoC	Indicators of Compromise
IoT	Internet of Things
IT	Information Technology
IV	Information Visualization
KAVAS	Knowledge-assisted Visual Analytics for STIX
LDF	Live Digital Forensics
NBGM	Nested Blocks and Guidelines Model
RQ	Research Question
SA	Security Analytics
SIEM	Security Information and Event Management
SLR	Structured Literature Review
SOC	Security Operations Center
STIX2	Structured Threat Information Expression Version 2
VA	Visual Analytics
VSA	Visual Security Analytics

Part I

Outline of the Dissertation

1 Introduction

Information and Communications Technologies (ICT) permeate almost every aspect of economic, private, and public life by now. These technologies have evolved from a supporting factor to the central component of any modern civilization in the last two decades. The ever-increasing utilization of Information Technology (IT) and the digitization of our society are in every respect driving forces and enablers of growth and prosperity. Digitization plays a decisive role in the competitiveness of companies and institutions. Thus, as we increasingly rely on IT and digital infrastructures, our society has become entirely dependent on them in most areas. This leads to our society and economy becoming more efficient but also provides an increased surface for attacks, where unauthorized actors try to steal critical information from ICT systems, disrupt business operations, or damage an organization's assets [11].

Theft from, disruptions, or failures of digital infrastructures can have dramatic consequences for companies, institutions, and private persons. Thus, organizations are trying to protect themselves against threats in our highly interconnected world. For these actors, direct and indirect consequences of a successful attack can be substantial [19, 20]. For this reason, tremendous sums have been invested for years in the implementation of security concepts, and investments are not going to stop here [12, 13]. While organizations improve their defenses, researchers drive the state-of-the-art of information security forward [8]. However, despite the amount of resources invested, defenders seem to be losing the battle, as implied by an increasing quantity of successful attacks [1, 21]. The risks of being attacked and associated costs are at an all-time high and continually rising [16]. From this perspective, the investments made do not seem to have the desired effect. As a result, the threat of cyberattacks is one of the major risks to economic success and growth [40]. The increase of interconnected devices in digital infrastructures and, thus, the increased attack surface is just one of the reasons for this. Furthermore, the intensity and, above all, professionalism of attacks steadily accelerates, which means that companies are constantly confronted with novel, unknown, and highly complex attack patterns.

However, a closer and more differentiated look reveals that the situation can be improved. Progress is being made and research helps to improve protection, detection, containment, and response mechanisms. In recent years, a great deal of emphasis has been placed on automating a wide range of cybersecurity tasks. While automation and the application of advanced analysis methods are undoubtedly important, the involvement of human domain knowledge is still essential. This becomes clear when taking a look at the central challenges that information security is facing. On the one hand, the volume of monitored and analyzed data grows with increasing digitization. In particular, current developments such as the implementation of the Internet of Things (IoT) and the prevalence of Cyber-Physical System (CPS) are increasing the quantity of relevant data. On the other hand, these developments lead to complex and heterogeneous infrastructures that change dynamically. This makes protecting them a complicated task. A third

significant challenge are the threats themselves. They are often explicitly tailored to an organization and professionally executed [1]. Thus, they are tough to detect for traditional, automated security mechanisms like Security Information and Event Management (SIEM) systems or Intrusion Detection Systems.

These challenges make collaboration between people and machines essential for effective cybersecurity [25, p. 6]. Unknown, multi-layered attack patterns in complex infrastructures can only be unraveled by security experts with their domain knowledge. Their involvement is a key success factor for maintaining and improving an organization's security posture [24]. Nevertheless, automated analyses are crucial for processing the tremendous quantities of data. Thus, cybersecurity needs an integration of security experts into relevant processes by supporting collaborative analyses, including humans and machines. An established tool for realizing such collaborations and integrating human expertise into IT processes is Visual Analytics (VA) [34].

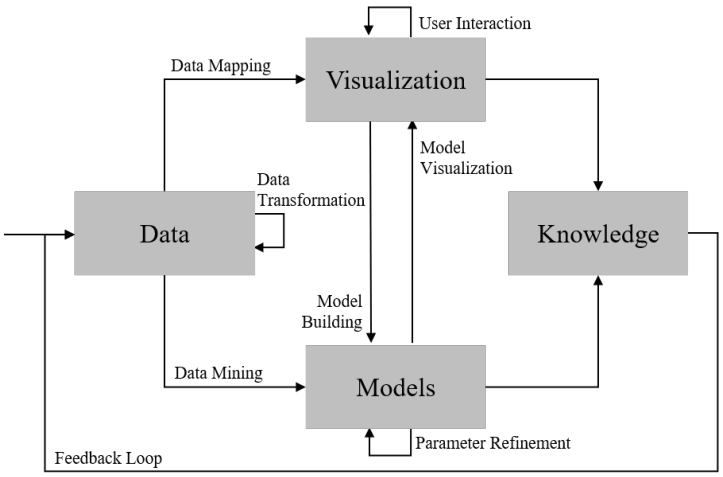


Figure 1: Visual Analytics Process based on Keim et al. [22].

VA combines humans and machines through interactive, visual representations of large amounts of data to enable effective and efficient analyses [36, p. 28]. The advantage of this combination is a delegation of tasks within a best-of-both-worlds approach [23, pp. 10-11]. On the one hand, IT is irreplaceable for automated data management, processing, and statistical analyses. Besides this fast processing and preparation of large amounts of data, IT can efficiently generate visual representations. On the other hand, humans can process visually presented information extremely fast and partly unconscious [39, p. 2]. However, a fundamental condition for this ability is domain knowledge. This prior knowledge allows domain experts to interpret visual representations of information and derive new knowledge [34]. It is exactly these capabilities that enable the combination of humans and machines with the help of visual representations. This idea of interactive and integrative visual analysis is specified by Keim et al. [22] in their VA process depicted in Figure 1. An essential component of this process is the feedback loop, which continuously feeds newly gained insights into new process instances [7, 15]. Due to the interactive nature of the visual representations, the VA process offers possibilities for users to directly

or indirectly influence all other process steps [10]. Thus, parameters of automatic analysis procedures can be adjusted, and changes in the visual representations can be made. With this interactivity, VA enables an exchange of knowledge between humans and machines and integrates human knowledge into automated analyses [9]. At the same time, this improves the integration of humans into the analysis processes [34, 38]. These core capabilities of VA can support modern cybersecurity operations to cope with imminent threats.

While the fields of cybersecurity and Visual Analytics have already brought forth considerable research, only a limited amount of work has been carried out in the area of leveraging Visual Analytics for cybersecurity [6, 14], which is referred to as Visual Security Analytics (VSA) throughout this dissertation [2]. Nevertheless, given the outlined challenges that security faces, incorporating visualization appears to be a logical and beneficial approach. Furthermore, existing work suffers from a shortcoming described as the dichotomy of VSA, which remains valid until today despite its early first mention [25, pp. 7-8]. The dichotomy criticizes the fact that many existing approaches are developed by security experts, who have little knowledge of VA or Information Visualization (IV) theory. The others are developed by visualization experts who have little knowledge about cybersecurity. This mismatch results in tools that lack one of the two aspects: either security domain knowledge or visual efficiency. However, for efficient and usable VSA approaches, knowledge from both worlds is required. This dissertation contributes to bridging the gap but cannot solve this complex problem entirely. The corresponding application-oriented research applies methodologies and approaches from the VA domain to address pressing issues in several cybersecurity domains. This highlights the advantages of leveraging VA for security and creates a transfer of knowledge from VA and adjacent fields into security.

The remainder of this work is structured as follows. Section 2 introduces the main research question and the related sub-ordinate questions, which are addressed throughout the dissertation project. Ensuring comprehensive and compelling research contributions, the dissertation rigidly follows the methodologies introduced in Section 3. The results of the research efforts are presented within Section 4. This section first gives an overview of the research papers and places them into the overall context of this dissertation. Afterward, each paper is summarized in a separate subsection while highlighting its main contributions to answering the research questions. Finally, this section introduces a series of complementary publications which are not directly part of this dissertation. The first chapter is concluded in Section 5 by summarizing the contributions made throughout this dissertation. Additionally, directions for future research are highlighted. In Part II, all original research papers contributing to this dissertation are imprinted.

2 Research Questions

VA provides methods and tools to improve the integration of expert domain knowledge into processes and mechanisms of cybersecurity. However, it is not a universal remedy for all problems. It instead needs to be applied with a clear objective in mind. From an abstract perspective, this objective is often to support activities dealing with large amounts of data. These activities benefit the most from the application of VA. In the context of information security, respective activities mainly occur within the central aspects of "Incident Detection" and "Incident Response" [31]. While the first aspect encapsulates tasks to identify threats, the latter comprises actions taken in response to an identified security incident. However, they both require monitoring and analysis of large amounts of data originating from a plethora of devices integrated into the complex architecture of an organization. Thus, they are a perfect match to leverage Visual Analytics to integrate humans and their domain knowledge into analytical activities. This hypothesis poses the main Research Question (RQ) of this dissertation:

Main RQ. *How can Visual Analytics be leveraged to integrate domain experts and their knowledge into cybersecurity activities?*

The main research question requires this dissertation to pursue a twofold objective. On the one hand, application-oriented research approaches show how domain experts and their knowledge can be integrated into key cybersecurity activities by leveraging VA approaches. This first objective shows the usefulness and applicability of VSA. On the other hand, an essential contribution of this work consists of a transfer of established methods and techniques from the research areas of IV and VA into the field of cybersecurity. This second objective helps to tackle the dichotomy of Visual Security Analytics.

Figure 2: Structural dependencies of Focus Areas within this dissertation.

The main RQ is broken down into four subordinate research questions to achieve the described objectives and provide a fruitful answer to the main RQ. First, it is necessary to establish profound definitions of relevant aspects and identify imminent challenges in this context. This provides solid **Foundations** (cf. Section 2.1) for the subsequent research efforts to build upon. Each of the following three subordinate research questions is devoted to leveraging VA for a specific cybersecurity activity. A first, highly relevant activity is the detection of threats by analyzing and correlating possible **Indicators** (cf. Section 2.2). If the analysis of these Indicators of Compromise (IoC) leads to the identification of actual **Incidents** (cf. Section 2.3), in-depth investigation of these incidents provides insights into exploited attack vectors, modus operandi, or perpetrators. This attribution can either be used as evidence in court or further enriched with contextual information and, if necessary, anonymized to be exchanged as **Intelligence** (cf. Section 2.4). Figure 2

visualizes the structure of these focal areas that are addressed within the dissertation. Within all these areas, VA can be leveraged to integrate the domain knowledge of security experts by providing concepts supporting experts in their tasks.

A broad bandwidth of security domains (e.g., Identity and Access Management, Cyber Threat Intelligence, or Digital Forensics) covers aspects of these focus areas. A transfer of methods, techniques, and solution approaches from VA research to cybersecurity can be realized by considering different security domains within the focus areas of this dissertation. Based on this contextual embedding, the following subsections provide further detail on the individual research areas and derive the respective subordinate research questions.

2.1 Focus Area 1: Foundations

VA can only be beneficial when the intended users and their tasks that need to be supported are known and well-defined. Thus, it is necessary to clearly understand the organizational context in which the relevant cybersecurity activities are embedded. This helps to understand the environment in which security experts with the required domain knowledge are working. Especially in larger companies, the respective activities are often carried out in a specialized organizational structure called Security Operations Center (SOC). This structure brings experts from various security domains together in a single organizational unit. Understanding the structure, tasks, and current challenges provides a foundation for well-informed support of SOC activities with VA. After identifying and defining the organizational environment, it is necessary to derive a high-level process connecting the core cybersecurity functions within this organizational environment. This process lays the basis for identifying contact points where VA can be leveraged to support security experts and integrate their knowledge into the respective process steps. As stated before, security experts' knowledge plays an integral role in cybersecurity, but knowledge itself is a highly overloaded term whose actual meaning is dependent on the respective context. Thus, a spectrum of different types of knowledge can be considered relevant for this dissertation. These have to be defined and formalized profoundly to allow a shared vocabulary across related research.

These aspects bundled within this Focus Area form the first of four subordinate research questions:

RQ 1. *Which organizational environment and processes are cybersecurity activities embedded into and what types of knowledge play a role in them?*

2.2 Focus Area 2: Indicators

After defining these foundations, the second focal area is devoted to the identification of possible threats and malicious behavior from various indicators of compromise. This is a first activity, which can benefit from an improved collaboration between humans and machines leveraged by VA. Modern automated analysis methods already contribute

greatly in revealing possible security threats within this activity. However, automated analyses are only of actual use to detect indicators of known attacks or point out suspicious behavior. Security experts and their domain knowledge are indispensable to interpret indicators and, if necessary, analyze the underlying raw data in-depth. However, they cannot manually deal with the tremendous amount of data to be monitored. Thus, combining the advantages of humans and machines through VA improves the effectiveness of separating actual threats from normal behavior within Indicators of Compromise.

Focusing on the analysis of indicators leads to the second subordinate research question encapsulated in this Focus Area 2:

RQ 2. *How can VA be leveraged to help security experts interpret indicators of compromise and identify imminent cybersecurity incidents?*

The analysis of indicators is an essential task within several security domains. Thus, to support the intended transfer of methods and concepts from VA to cybersecurity, this RQ should be discussed from the perspective of different domains.

2.3 Focus Area 3: Incidents

When malicious behavior has been detected, it is first and foremost necessary to contain the threat. In addition, however, it is crucial to analyze and understand the incident in greater detail. Thus, two vital objectives are pursued in this focal area: First, the collection of evidence to be used in court, and second, the preparation of threat intelligence containing indicators, possible threat actors, exploited attack vectors, and other in-depth information. One domain that deals with corresponding activities is Digital Forensics (DF). Of particular interest in this dissertation is the aspect of Live Digital Forensics (LDF), in which infected devices (those affected by a threat) are isolated and digitally examined in an active state. This procedure requires domain knowledge of forensic analysts. The corresponding methods are primarily manual, requiring forensic experts to use various tools. Thus, there are several aspects where VA can be leveraged to support them. Decisions on which tools to use must be made under time pressure in a dynamic environment. In addition to the available IoC that contributed to the detection of the incident, more extensive information is sometimes required for an LDF investigation. However, this information can often only be accessed using tools explicitly installed for this purpose, although installing tools endangers the integrity of relevant evidence on the device. Therefore, an approach is needed to visually present relevant data to experts to support their analysis decisions as LDF investigations remain mainly manual work. Additionally, this data needs to be extracted without significant interference from the device under investigation.

In this challenging environment, numerous aspects must be analyzed to understand the incident under investigation. The subordinate research question for this Focus Area can be defined as follows:

RQ 3. *Which VA approaches can appropriately support forensic experts in their in-depth analysis of cybersecurity incidents?*

2.4 Focus Area 4: Intelligence

The results of an in-depth incident analysis (e.g., via an LDF investigation) can prove valuable to an organization for improving existing security mechanisms. At the same time, these findings are relevant for exchanging information about the threat across organizational boundaries. This information is called threat intelligence or Cyber Threat Intelligence (CTI). Cooperative security approaches powered through sharing CTI have gained in popularity throughout the last years as it has become apparent that this is necessary to hold up against the modern attack landscape. The exchange of CTI has advanced significantly through several formats optimized primarily for machine readability and interoperability. While some elements of CTI can be generated automatically, this is restricted to mainly low-level intelligence. The domain knowledge of cybersecurity experts is still indispensable for enriching CTI with semantic details and contextual information. Applying VSA in this context allows combining the knowledge of machines (i.e., automatically generated CTI) with the domain knowledge of experts leading to more comprehensible descriptions of threats.

CTI can only be helpful for collaborative security approaches if a high quality of the exchanged information is ensured. However, the current lack of approaches to capture the quality of CTI is problematic and hinders its increasing adoption. Thus, it is necessary to assess relevant CTI quality dimensions. While some aspects of CTI quality can be assessed automatically, others require the domain knowledge of security experts. Thus, VA offers a way to enable the collaboration of humans and machines in order to measure the quality of CTI.

These two crucial aspects of Focus Area 4 are summarized in the last subordinate research question of this dissertation:

RQ 4. *How can security experts be enabled to enrich semi-structured CTI data and assess its quality?*

3 Research Methodology

The research for this dissertation was conducted at the Chair of Information Systems at the University of Regensburg, Germany. Accordingly, the research questions are answered within a framework of established methodologies from the respective research area, supplemented by methodologies from VA where possible and appropriate.

The following subsections give an insight into the applied methodologies. First, an overview of the research methodologies for information systems research is given. Then, the guidelines and the main research process, which shape this work methodologically, are presented. In a last step, the Nested Model for Visualization Design and Validation, an influential methodology from the field of VA applied in this dissertation, is introduced.

3.1 Information Systems Research Methodologies

The Chair of Information Systems is grounded in a research field summarized under the German term *Wirtschaftsinformatik*, which is often translated as "business informatics". With this literal translation, the original intention of this field as a link between business administration and computer science becomes clear [27]. Despite its development from comparable streams in these two research areas, *Wirtschaftsinformatik* has evolved into a distinct and equally accepted research discipline in the last decades. However, concerning the direction of the respective research, the Anglo-Saxon term *information systems research* is to be considered the equivalent [41]. Applied information systems research also reflects this as it is characterized by two central approaches: Behavioral Science Research and Design Science Research.

Behavioral Science Research has its origins in natural sciences, which is reflected in its basic research approach. It studies information systems to develop and verify (or falsify) hypotheses about the behavior of people, organizations, and technologies. Common to the approaches of Behavioral Science Research is the assumption that information systems should increase an organization's efficiency [17]. On the other hand, Design Science Research originates from engineering and describes a more problem-oriented paradigm. Strictly speaking, it is concerned with designing new information systems as research artifacts to extend human, organizational, and technical capabilities [17]. However, the fundamental difference between the two research approaches is ontological [18]. The behavioral approach studies preexisting phenomena, while the design approach creates new phenomena. Thus, these approaches are rather complementary than dichotomous.

3.2 Design Science Research Guidelines & Process

As the goal of this dissertation is to create new phenomena to address the imminent problems (cf. Section 1) and to answer the related research questions (cf. Section 2), it is strongly influenced by the central methodological foundations of Design Science Research. It is aligned with these guidelines to ensure well-conducted and valid Design Science Research as proposed by Hevner et al. [17]:

1. **Design as an artifact:** This guideline dictates any Design Science Research to produce a model, a method, or an instantiation as its purposeful artifact. Accordingly, this dissertation's contributions are designed to result in such artifacts. They can mostly be seen as models and instantiations, resulting from application-oriented research.

2. **Problem relevance:** Any research effort in Design Science Research should address important and unsolved business problems relevant to the constituent community. Thus, the relevance of the four Focus Areas in this dissertation is expounded through discourse with related research and practitioners.

3. **Design Evaluation:** The resulting artifact of Design Science Research must prove its utility, quality, and effectiveness through suitable evaluation methods. This

dissertation applies different evaluation methods to evaluate the resulting artifacts, ranging from questionnaires and expert interviews to exemplary use cases.

4. **Research Contributions:** Research activities must result in verifiable and valuable contributions in terms of design artifacts, foundations, or methodologies. As indicated by the Focus Areas, the research carried out within this dissertation results in valuable foundations for progressing the research domain and design artifacts for specific domain problems.

5. **Research Rigor:** The construction and evaluation of the designed artifact have to be carried out rigorously based on existing theoretical foundations and research methodologies. Any research carried out to provide answers to the research questions introduced in Section 2 has been fundamentally grounded on existing research and commonly accepted methodologies to design comprehensive artifacts.

6. **Design as a Search Process:** The design process of an artifact is generally a search process looking to discover the most effective or at least a sufficiently effective problem solution. As the design of a VA tool is a process with no objectively most effective solution, parts of this dissertation's contribution are guided by specific methodologies enabling a comprehensible search for a good design.

7. **Communication of Research:** Any research results of Design Science Research activities must be made available for and presented to both relevant technology-oriented and management-oriented audiences. Respectively, all contributions of this dissertation have been submitted to renowned and peer-reviewed academic journals and conferences, ensuring appropriate communication of the results.

These guidelines form the methodological foundation of this dissertation. The research process applied to address and answer the research questions introduced in Section 2 is aligned with these guidelines. It follows the six activities of the Design Science Research Methodology introduced by Peffers et al. [33]. The process starts with the identification of a problem and the description of its relevance. This motivation clarifies the need and the benefits of solving the problem. The goals of the intended solution are determined in the second activity of the process. This involves describing either how the solution will help improve the current state or to which extent the development of a new artifact is necessary as a solution. The third activity then deals with the design of the solution and the implementation of an artifact. The applicability of the solution to multiple instances of the defined problem is demonstrated in the fourth step. As the fifth activity of the process, the evaluation of the artifact takes place on this basis. The evaluation shows how well the artifact can contribute to solving the problem. The sixth and thus last activity of the process requires the communication of the research results. Here, the problem, its relevance, design decisions leading to the artifact, and its effectiveness in solving the problem are communicated and made available to the relevant audiences. In addition to sequential execution of the individual activities, the process allows several

iterations, e.g., to split the problem into individual sub-problems. Any research carried out in the course of this dissertation is oriented towards this research process.

3.3 Information Visualization Design

The problem-oriented approach for answering the raised research questions requires including another fundamental methodology for several research items. This specialized methodology provides an approach for the design and implementation of VA artifacts. Accordingly, it is used in this dissertation to extend the first three activities of the Design Science Research Methodology for the design of visualizations as research artifacts.

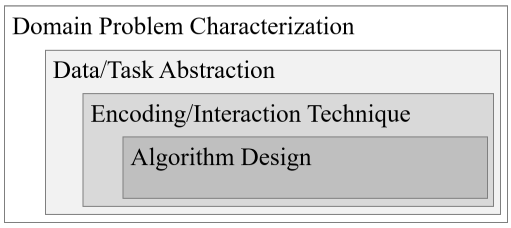


Figure 3: The layers of the NBGM based on Munzner [30] and Meyer et al. [29].

The Nested Model for Visualization Design and Validation [30] and its extension, the Nested Blocks and Guidelines Model (NBGM) [29], describe a systematic method for designing effective visual representations. This methodological approach enables transparent and comprehensible visual design decisions. Especially with regard to the design of visualizations, this is important since even methodically derived decisions remain subjective due to the plethora of different available encodings and interactions [26]. For this reason, the orientation towards a well-established design methodology is crucial for comprehensible design decisions. In addition, a vital objective of the NBGM is the collaboration of domain experts and visualization experts. Thus, it contributes to bridging the dichotomy of VSA [35, 37]. The central layers, i.e., activities, of the NBGM are shown in Figure 3 and are briefly summarized in the following paragraphs.

1. **Domain Problem Characterization:** This first activity of the model characterizes the specific domain problem which the visual design is intended to address. Therefore, the target group's situation and environment, the data available for solving the problem, and the tasks of the target group are taken into account. Since each domain considered in this phase (e.g., cybersecurity) uses its individual vocabulary, the characterization is framed in the vocabulary of the target group. This enhances the visualization experts' understanding of the domain experts' problems.

2. **Data/Task Abstraction:** The second step of the model abstracts the domain problem by translating it into a visualization-specific vocabulary. Therefore, this layer describes abstract visualization tasks (e.g., finding outliers or identifying temporal trends), which the artifact to be designed must support. Additionally, it defines the abstracted data types (e.g., items, attributes, grids) available to address the domain problem. On the one hand, this abstraction enables the selection of

suitable techniques in the next step. On the other hand, it allows the comparison of different designs across different domains.

3. **Encoding/Interaction Technique:** This step involves deciding on the visual representation and interaction techniques that are appropriate for the identified abstract tasks and data types. Encodings and interactions must be aligned and combine the first two layers of the model by instantiating the abstract visualization design for the domain problem.

4. **Algorithm Design:** Finally, the innermost layer of the NBGM is devoted to the particular implementation of the designed artifact with efficient and appropriate algorithms. In this dissertation, prototypes demonstrating the interactive, explorative possibilities of the visualization are developed within this step.

4 Results

In the course of this cumulative dissertation, each of the subordinate research questions (RQ1 – RQ4) is answered by two research papers. The research papers have been published in renowned journals or conferences to ensure adequate communication of the research results to a techno-scientific audience of researchers and practitioners. An exception is Paper P2, which was still under review at the time this dissertation was submitted. The addressed journals and conferences cover a broad spectrum of subject areas within information security and cybersecurity. All publications have been subject to a rigorous peer-review process by the respective venue prior to publication. By the time of writing this dissertation, the published research papers had been cited 56 times, highlighting the reference of the published work. The following subsections provide an overview of the papers and embed them into the context of the subordinate research questions.

4.1 Overview of Research Papers

Table 1 lists all papers published as part of this dissertation to answer the research questions RQ1 to RQ4. The table contains the full citations, each publication's status[1], the type of publication[2], and the number of citations. The information listed in the table represents the publications' status and number of citations as of December 14, 2021. As outlined in Section 2, the dissertation is structured following four primary focus areas. Figure 4 gives a visual overview of these areas with each of the eight publications assigned to them. The papers and more detailed information including the portion of each author's contribution can be found in Part II of this dissertation.

[1] *pub.* = published, *sub.* = submitted
[2] *C* = Conference, *J* = Academic Journal

Table 1: Overview of research papers.

No.	Full Reference	Status	Type	Cit.[3]
P1	VIELBERTH, M., BÖHM, F., FICHTINGER, I., AND PERNUL, G. Security Operations Center: A Systematic Study and Open Challenges. *IEEE Access 8* (2020), 227756–227779	pub.	J	8
P2	BÖHM, F., VIELBERTH, M., AND PERNUL, G. Formalizing and Integrating User Knowledge into Security Analytics. *Springer Nature Computer Science* (2021)	sub.	J	n/a
P3	PUCHTA, A., BÖHM, F., AND PERNUL, G. Contributing to Current Challenges in Identity and Access Management with Visual Analytics. In *Data and Applications Security and Privacy XXXIII*, vol. 11559 of *Lecture Notes in Computer Science*. Springer, Cham, 2019, pp. 221–239	pub.	C	3
P4	PUTZ, B., BÖHM, F., AND PERNUL, G. HyperSec: Visual Analytics for Blockchain Security Monitoring. In *ICT Systems Security and Privacy Protection*, vol. 625 of *IFIP Advances in Information and Communication Technology*. Springer, Cham, 2021, pp. 165–180	pub.	C	1
P5	BÖHM, F., ENGLBRECHT, L., AND PERNUL, G. Designing a Decision-Support Visualization for Live Digital Forensic Investigations. In *Data and Applications Security and Privacy XXXIV*, vol. 12122 of *Lecture Notes in Computer Science*. Springer, Cham, 2020, pp. 223–240	pub.	C	1
P6	BÖHM, F., ENGLBRECHT, L., FRIEDL, S., AND PERNUL, G. Visual Decision-Support for Live Digital Forensics. In *2021 IEEE Symposium on Visualization for Cybersecurity (New Orleans, 2021)*, pp. 58–97	pub.	C	0
P7	BÖHM, F., MENGES, F., AND PERNUL, G. Graph-based visual analytics for cyber threat intelligence. *Cybersecurity 1*,16 (2018), 1–19	pub.	J	24
P8	SCHLETTE, D., BÖHM, F., CASELLI, M., AND PERNUL, G. Measuring and visualizing cyber threat intelligence quality. *International Journal of Information Security 20*, 1 (2021), 21–38	pub.	J	19

P1 and P2 lay the necessary foundations for this dissertation by providing answers to RQ1. P1 provides a comprehensive definition of a SOC as an organizational structure that bundles many security functions and which is the working environment for security

[3]Citation count based on https://scholar.google.com/

experts. Addressing the lack of well-defined processes for SOCs, P2 introduces the Incident Detection Lifecycle (IDL). The IDL is an important process within a SOC covering activities like collecting raw security-relevant data, identifying indicators of compromise, analyzing detected security events, and utilizing acquired threat intelligence. For successfully implementing the IDL in a SOC, different types of knowledge need to be regarded, enabling effective and comprehensive security operations. P2 identifies and defines these notions of knowledge before illustrating their connection points into the IDL. The IDL provides a structure for core security activities, which need to be augmented with experts' domain knowledge. Thus, it serves as the structural foundation for the following focal areas.

In the context of Focus Area 2, research papers P3 and P4 investigate possible approaches for leveraging VA within the first phase of the IDL. More precisely, P3 and P4 propose the use of different VA approaches to detect imminent threats or malicious behavior as an answer to RQ2. P3 highlights how Visual Analytics can support domain experts recognize issues within an organization's Identity and Access Management (IAM). Furthermore, P4 focuses on the vital security domain of distributed ledger technologies. This research paper introduces a VA approach for monitoring a blockchain to identify anomalies and indicators of compromise.

The papers P5 and P6 are assigned to Focus Area 3. They deal with a subsequent activity of the Intrusion Detection Lifecycle after indications for an actual threat have been detected. Besides taking actions to contain it, the incident has to be analyzed in-depth to collect evidence, understand the exploited attack vector, and gather as much intelligence about the incident as possible. Often, DF experts carry out these activities. They have a wide variety of tools at their disposal to analyze infected devices or networks. While having this toolset, they need to carefully (and mostly manually) decide which tools to apply based on their expertise and the available indicators of compromise. P5 designs a visual decision-support system that helps DF experts make well-informed decisions which forensic tool to deploy and which indicators need their unrestricted attention. P6 improves this design and introduces a comprehensive prototype for the visual decision-support system. Thus, these two publications provide a suitable answer to RQ3.

Forensic analyses of incidents are the basis for the activities summarized within Focus Area 4. Intelligence gathered from these analyses is structured as CTI and can be exchanged between organizations in collaborative security efforts. The utilized data formats are optimized for machine readability which makes them cumbersome to deal with for humans. However, domain experts are required to enrich the CTI with semantic information and understand it as indicated by RQ4. P7 provides a Visual Analytics approach for making complex CTI accessible and editable for security experts. Additionally, when analyzing CTI, domain experts need to have insight into its quality. Thus, P8 introduces metrics to assess the quality of CTI and extends the VA approach from P7 to include relevant CTI quality aspects.

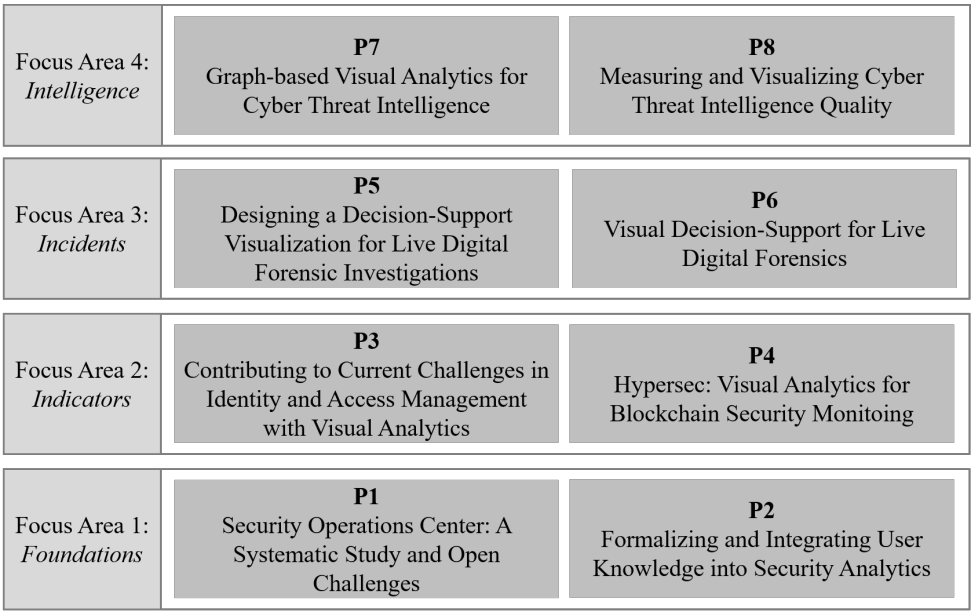


Figure 4: Overview of research papers and corresponding research areas.

4.2 Focus Area 1: Foundations

RQ1 is answered within Focus Area 1 by two research papers. As posed by RQ1, it is a prerequisite to any further work in this dissertation to define the organizational environment in which the considered security functions are embedded and in which the considered security experts are working. This includes different involved actors, processes, or technologies.

P1: Security Operations Center: A Systematic Study and Open Challenges

Especially in larger organizations, crucial security activities are often bundled within a SOC. The importance of SOCs for an organization's security posture has increased significantly over the last few years. SOCs are formed in an attempt to identify, detect, and mitigate major threats effectively. Despite this relevance, respective research lacks a precise definition of a SOC and its components, hindering the efficient deployment of SOC and their further innovative improvement. Thus, P1 addresses this issue by identifying and structuring the state-of-the-art of SOCs and compiling a collection of the most pressing issues they are currently facing. Thus, P1 covers a fundamental aspect of RQ1 by describing the organizational environment of security experts and respective core activities.

Existing literature about SOCs is fragmented and widespread. Academia agrees on some capabilities comprising a SOC but lacks a clear consensus. Furthermore, existing work focuses on single characteristics rather than the bigger picture. P1 summarizes the state-of-the-art for SOCs and identifies challenges to be solved in its context. At first, we carry out a comprehensive Structured Literature Review (SLR) to identify and synthesize relevant academic literature. In a second step, we derive challenges that need

to be solved to improve the effectiveness of SOCs from the findings of the SLR. The SLR to capture the state-of-the-art of SOCs follows the rigid methodology by Okoli et al. [32]. Throughout the SLR 158 relevant publications published between January 1, 2003, and June 31, 2020, are analyzed. The papers mainly deal with aspects from two categories: *General Aspects* encapsulating definitions, operating models, and architectures for SOCs, and *Building Blocks* describing the components that make up a SOC.

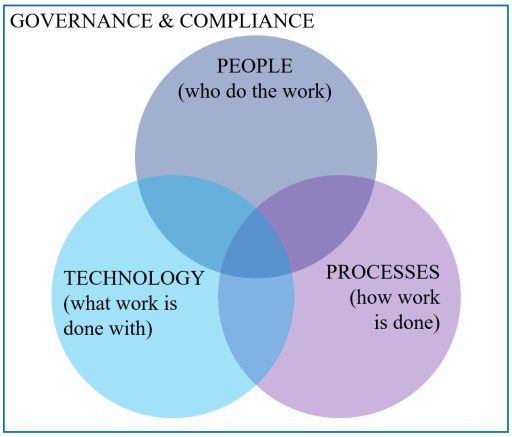


Figure 5: Structural building blocks of a SOC as published in P1.

Despite the lack of a generally accepted definition of a SOC, we were able to identify a common core among existing definitions. A SOC is seen as an organizational component of a company's security strategy. Therefore, it combines the processes, technologies, and people to manage and improve the security posture. In addition, the SOC provides governance & compliance as a framework in which the people of the SOC work and to which the processes and technologies are adapted (see Figure 5). Relevant literature classifies various architectures to implement a SOC based on three fundamental concepts: centralized, distributed, or fully decentralized SOCs. Additionally, a variety of operating models for how a SOC can be deployed are being discussed. However, the feasibility of a specific model depends on various factors, such as corporate strategy, available resources, security risks, and many more.

We summarize the building blocks of a SOC under the aspects of people, processes, technology, governance, and compliance. Security analysts working in a SOC are classified into three hierarchical tiers, depending on experience and skills. These are considered the technical roles within a SOC. However, a SOC requires additional management and supporting roles collaborating with the technical staff. All activities within a SOC are structured within several processes. At a high level of abstraction, these can be divided into four phases: *Preparation*, *Detection & Analysis*, *Containment, Eradication, & Recovery*, and *Post-Incident Activity*. A large body of existing work deals with technology for the SOC. Here, research has introduced a vast range of diverse applications and solutions. We structure these technologies according to the process phases mentioned above to present an overview of technology available for each process step. Regarding governance & compliance, different standards and guidelines applied in the context of a

SOC are identified. The analyzed literature provides a first concept for a comprehensive framework to determine the maturity of a SOC and metrics to monitor its effectiveness.

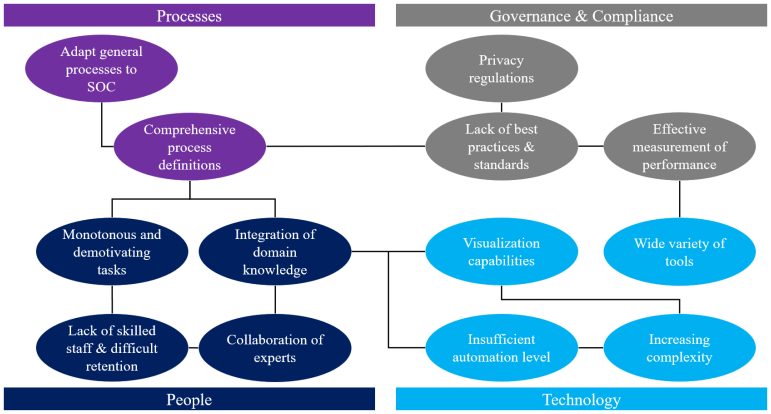


Figure 6: Overview over identified SOC challenges as published in P1.

Based on the state-of-the-art, we derive several challenges that need to be solved. Although thirteen different challenges are outlined in the paper (see Figure 6), only two challenges are mentioned here as they are fundamentally relevant in the context of this dissertation. A first challenge related to the people in the SOC is the integration of domain knowledge that needs improvement. To detect unknown threats, a SOC needs access to the knowledge of security experts and experts from other domains (e.g., engineers). The effectiveness of SOCs can only be improved by closely linking these people with the SOC's automated mechanisms. The second challenge is the lack of methodically developed visualization capabilities for the SOC. As the complexity and quantity of the monitored systems are constantly increasing, visualizations are becoming the primary working tool of the SOC staff. However, current visualization tools used in SOC are often designed without the necessary theoretical rigor.

This paper's comprehensive definition of a SOC and the corresponding building blocks lay a central foundation for the dissertation. Essential research questions such as which types of knowledge are important for security are derived from this paper. At the same time, the identified processes within a SOC provide the basic structure of the focus areas of the dissertation. Finally, the problems and research questions that the dissertation addresses can be derived directly from the identified challenges for SOCs.

Contribution of P1: P1's main contribution is the a comprehensive definition of state-of-the-art Security Operations Centers, including their main building blocks. A structured literature review allows to define the term "Security Operations Center" and derive people, processes, technologies, and governance as vital aspects for a holistic perspective. As a result of this definition, several major challenges that hinder further development and innovation for SOCs are derived.

P2: Formalizing and Integrating User Knowledge into Security Analytics

P1 identified the lack of human knowledge integration into SOC operations. However, there are numerous competing definitions of the term knowledge and even several sub-categories of knowledge discussed among scholars. Thus, a strict and formal definition of knowledge in the context of security is necessary. Although humans are often considered the weakest link, their expert knowledge is essential to identify unknown, targeted threats in modern security operations. Integrating users and their knowledge into security activities can only succeed based on a clear definition of knowledge as a common vocabulary. P2 paves the way towards an in-depth understanding of the different security-relevant knowledge types. Additionally, we introduce the Incident Detection Lifecycle, a process encapsulating core security activities, and we highlight the connecting points of different knowledge types with the IDL. Therefore, P2 provides a suitable answer to aspects of RQ1 not covered by P1.

In the first step towards an in-depth understanding of knowledge in the context of cybersecurity, we formally define different types of knowledge. An initial differentiation has to be made between explicit knowledge and implicit knowledge. We define explicit knowledge as knowledge that machines can read, process, and store. Analogously, implicit knowledge is held by humans and is very specific to the individual. In the context of this dissertation, three forms of explicit knowledge can be identified. First, machine learning models and the alike play an important role for anomaly-based detection mechanisms as machine-readable knowledge. Second, signatures and rules to detect indicators of compromise are highly relevant for signature-based detection mechanisms. CTI and other semantically rich information (e.g., forensic evidence) are the third form of explicit knowledge. Furthermore, we define three types of security-relevant, implicit knowledge: operational, situational, and domain knowledge. Operational knowledge in the context of cybersecurity describes an employee's ability to operate relevant security mechanisms. Situational knowledge is closely related to the concept of situational awareness, allowing employees to perceive unusual events or suspicious behavior. The last type of implicit knowledge is domain knowledge. Within P2, we differentiate between security domain knowledge and non-security domain knowledge. The first one directly encapsulates safety and security aspects. The latter, however, describes manufacturing or engineering knowledge and the like, which are not directly related to information security but highly relevant, e.g., in the context of cyber-physical attacks. Several knowledge conversion processes can be defined based on these formal definitions of security-relevant knowledge types. These processes describe how different forms of knowledge are translated and transformed into each other.

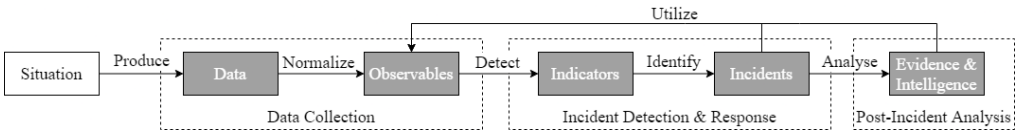


Figure 7: The Incident Detection Lifecycle as introduced in P2.

Subsequently, we introduce the iterative Incident Detection Lifecycle (see Figure 7), describing the logical interaction of core security activities. The IDL interprets any real-world situation as a producer of data. This data is normalized and, thus, available to be monitored in the form of observables. Within these observables, security analysts and mechanisms can detect indicators of malicious behavior or other threats. When several indicators are correlated, actual incidents and security events can be identified. In-depth analysis of incidents might reveal how they happened, which vulnerabilities were exploited, or even who is responsible for the attack. Besides actual evidence to be used in court, this information can be utilized as intelligence in further iterations of the IDL to improve and adjust existing security measures. The IDL is dependent on several different types of knowledge. To show which role knowledge conversion processes and knowledge types play, we extend the IDL by highlighting connection points of security mechanisms with different types of knowledge as indicated in Figure 8.

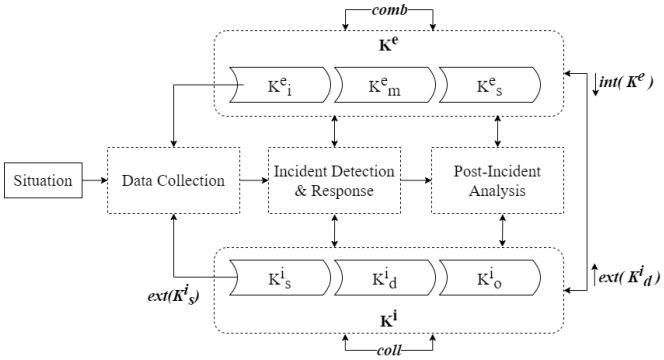


Figure 8: The IDL extended with relevant knowledge notions as introduced in P2.

Within the paper, we identify a severe mismatch in cybersecurity mechanisms: While security experts have comprehensive operational and security domain knowledge, engineers and others (i.e., security novices) hold a lot of non-security knowledge (which is crucial for comprehensive security operations), but they lack the operational knowledge to include this knowledge into security mechanisms. This issue prevents novices from contributing their non-security domain knowledge to a comprehensive Security Analytics (SA) approach. This problem can be tackled from two different angles. On the one hand, training of security novices could improve their operational knowledge and make them more capable of integrating their knowledge into existing security mechanisms. On the other hand, simplifying complex functionalities of the security tools could reduce the necessary operational knowledge to handle them. Our work introduces an early prototype following this second approach of simplifying the security mechanisms by abstracting complex core functionalities. The approach focuses on the use of a visual programming language to define rules that can be used in a signature-based detection mechanism, e.g., a SIEM system. The visual programming language provides a way for security novices to create signatures without having to know the complex syntax of the rules. The visual pattern builder is embedded in an easy-to-use web application that supports real-time monitoring of created rules and collaborative signature creation.

Contribution of P2: The main contribution of P2 is its formalization of security-relevant knowledge and the introduction of the Intrusion Detection Lifecycle. The lifecycle is augmented with the formal knowledge aspects to highlight the integration of explicit and tacit knowledge into core security activities. A research prototype provides a first approach to address crucial research knowledge gaps identified within this augmented version of the Intrusion Detection Lifecycle. This paper is a significantly extended version of A4 (see Section 4.6) that has been invited for submission based on positive reviewer feedback and the conference presentation.

4.3 Focus Area 2: Indicators

Within Focus Area 2, two research papers (P3 and P4) are addressing RQ2. These papers highlight how VA can be leveraged to support domain experts in detecting imminent threats. Regarding the IDL introduced in P2, the papers in Focus Area 2 show how VA can help domain experts analyze different indicators and detect imminent incidents or threats. The papers P3 and P4 consider two different security domains explicitly supporting the intended transfer of visualization knowledge into different information security domains.

P3: Contributing to Current Challenges in Identity and Access Management with Visual Analytics

P3 highlights pressing issues within the domain of IAM regarding the identification of threats or vulnerabilities and shows how visualizations can be used to support the detection of potential problems by engaging experts. Identity and Access Management is a crucial component of modern organizations enabling them to manage identities and grant access to resources. Besides traditional identities (i.e., human employees), more and more technical identities (i.e., sensors, machines, etc.) require access to company resources and, thus, need to be integrated into IAM systems. This leads to an explosion of heterogeneous identities that need to be managed. Therefore, numerous problems within IAM systems require an effective way to analyze identities and their entitlements. Providing a solution to those problems is only possible through including domain experts and their knowledge. P3 defines key challenges within IAM that can be solved by integrating experts' domain knowledge. We demonstrate how VA can be applied to integrate this knowledge in order to address current IAM challenges.

To identify the most pressing challenges, we review academic literature addressing specific challenges for IAM. Analyzing 19 relevant scientific papers enables us to derive an initial list of challenges. However, since IAM is strongly business-driven, this preliminary list has to be validated and, if necessary, updated to reflect the business perspective. We do this based on reports of IAM analysts and semi-structured interviews with IAM consultants and representatives of companies using IAM. Following this approach, we derive five critical challenges for IAM. These include, but are not limited to, the complex problem of identifying all relevant identities, the problematic management of heterogeneous identities, or the issue of maintaining and improving IAM data quality.

These problems also pose a challenge for security as many threats, especially those posed by insiders, would be identifiable within a well-managed IAM. However, the issues that we identified make it impossible to automatically analyze IAM data to detect threats. Mainly because the domain knowledge of IAM experts is necessary to address these challenges comprehensively. As VA has proven its potential in recent years in integrating domain experts into complex IT security tasks, it is also a promising approach in the context of the challenges for IAM. Thus, we design and implement a VA solution that can help to address some of the challenges. In close cooperation with IAM consultants and based on anonymized real-world data, we built a prototype as a proof-of-concept for the feasibility of VA to solve several of the identified IAM challenges.

The prototype's architecture processes information about identities within an organization from three different data sources: the central role-based IAM, additional applications that are not connected to the IAM, and a management system for technical identities not integrated into the IAM system. The different data models from these heterogeneous sources are merged into a single, high-dimensional table and annotated with additional information if necessary. The next step of data processing within the prototype is data transformation. In this process, the result of the previous data integration is restructured again to meet the requirements for visualizing the data. Additional fields are calculated, fields are split, and similar adjustments are made. In order to make the performance-intensive data mapping, in which the data is translated into the final structure to be visualized (e.g., circle segments with a specific radial angle and width), as efficient as possible, data filtering takes place beforehand. This interactive step reduces the high-dimensional data table to the columns that the users currently want to analyze.

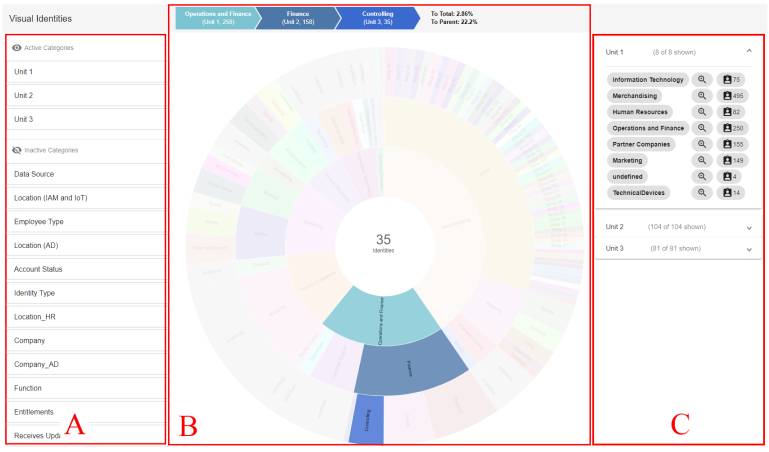


Figure 9: Sunburst diagram visualizing identity information as published in P3.

We designed the interactive visualization depicted in Figure 9 for making the complex, heterogeneous identity information accessible to users. The central component is a sunburst diagram (see Figure 9-B). It represents the hierarchical identity information using a series of concentric circles, where each circle corresponds to an attribute of the identities, and each circle segment corresponds to a respective attribute value. The interactive filter options of the prototype are highlighted in Figure 9-A. Here, experts

can drag and drop the boxes which correspond to the available attributes of the entities. The order of the boxes in this filtering mechanism defines the order in the sunburst diagram in that the top-most box, i.e., attribute, corresponds to the inner-most circle of the sunburst diagram. Figure 9-C provides further information as a dynamic list as well as the possibility to display all identities with a certain attribute value.

Several exemplary use cases, which represent relevant analysis steps to identify possible insider threats, demonstrate the applicability and usefulness of the prototype in the context of the IAM challenges that we identified earlier. For example, the visualization can be used to identify identities that are not yet managed in the central IAM. In addition, identities with an unusual number of entitlements could be identified in the underlying real-world dataset. Furthermore, the detection of quality problems, especially concerning attribute values, is possible with the help of the sunburst diagram.

Contribution of P3: In P3, a literature review and expert assessments point out five pressing challenges for IAM stemming from the ongoing explosion of identities. Several of these challenges require in-depth analysis of IAM data through domain experts. Subsequently, P3 introduces a proof-of-concept architecture including an interactive visual representation to highlight how IAM experts can be supported in analyzing heterogeneous identity information.

P4: HyperSec: Visual Analytics for Blockchain Security Monitoring

P3 deals with IAM, which has been of interest in information security research for more than three decades. P4 now focuses on Distributed Ledger Technology (DLT) as a domain that has only become a strong focus of research and practice in recent years. These systems' distributed and decentralized nature makes it challenging to identify attacks with automatic methods. Instead, ways must be created so that security experts monitoring a blockchain can easily identify anomalies present in the available indicators. One way to enable analysts to do so is the application of VA for the monitoring of security-relevant blockchain data. Thus, in P4, we design a VA tool that supports monitoring indicators of compromise related to a blockchain network.

Blockchain applications are increasingly being used, for instance, to improve transparency in complex business networks. At first glance, DLT improves the security of the underlying application by providing built-in availability, integrity, and non-repudiation assertions. However, like any software, it is subject to vulnerabilities and threats. The inherent complexity of blockchain technologies makes it particularly difficult to detect malicious activities. Each blockchain network consists of independent participants who have a limited view of the network. Human experts are indispensable for comprehensive incident detection mechanisms because their domain knowledge enables them to detect intricate attack patterns. Therefore, domain experts need to monitor heterogeneous indicators within a blockchain network. Visual Analytics solutions are very well suited to be applied in this context.

We follow the NBGM as a well-established design methodology to develop a com-

prehensible and reproducible design [29]. Aligned with this approach, we first define the addressed domain problem for a target-oriented use of visual representations. From a high-level perspective, we address the lack of a monitoring solution for blockchain security experts. To comprehensively describe the domain problem, an overarching process of blockchain security monitoring is described. This process is divided into five steps: *Identify*, *Protect*, *Detect*, *Respond*, and *Recover*. Especially the *Detect* phase, which encapsulates the detection of security breaches related to the monitored blockchain, currently lacks adequate visualization and analysis tools for domain experts. Subsequently, we set out to define and describe the intended users for the visualization design, their tasks within the monitoring process, and the available data to be monitored. The intended users for the visualization design are blockchain security experts working on the *Detect* phase to identify and analyze malicious incidents. The overarching task of these users is to analyze blockchain data to identify potential threats or incidents. We derive eight more specific tasks based on possible threat vectors for a blockchain network. Domain experts carry out these tasks (such as detecting vulnerable smart contracts or comparing transaction metrics) to detect security incidents. Finally, we elaborate on the available data sources for performing blockchain security monitoring. At first glance, only data concerning the blocks and the corresponding transactions are available. However, different aspects such as network activity and respective metrics can be derived based on this information. Additional in-depth data such as log files can be available if domain experts have access to specific network nodes.

In the next step, we derive a set of generally applicable design requirements based on the intended users, their tasks, and the available data. These requirements describe three fundamental views any Visual Analytics system supporting the security monitoring of blockchains needs to include. Primary, general security information (the blockchain configuration, smart contract information, etc.) should be accessible to users (R1). Furthermore, a view regarding the blockchain network, including peers, identities, and network connections, is necessary (R2). Finally, a transaction view gives experts insight into the blocks and transactions handled by the blockchain (R3). Wherever possible, all available details should be made interactively accessible to the experts (R4).

We develop our prototype named Hyperledger Security Explorer (HyperSec) based on these requirements enabling the monitoring of a Hyperledger Fabric network. HyperSec extends the open-source project Hyperledger Explorer to include additional data sources (such as in-depth metrics) and views relevant for security monitoring. Our prototype implements each requirement R1-R3 as a standalone view within the extended Hyperledger Explorer web interface. These views are fully interactive, allowing experts to explore the available data (cf. R4). For example, this includes the Transactions View shown in Figure 10 (cf. R3). This view consists of five interactively connected parts. Figure 10-A allows selecting an analysis period for which the corresponding data is loaded into the other views (Figure 10-B, 10-C, 10-D, and 10-E). The bar chart depicted in Figure 10-B shows the basic activity (transaction count) of the blockchain network for the selected period. The bar chart also allows for further, more fine-granular selection of

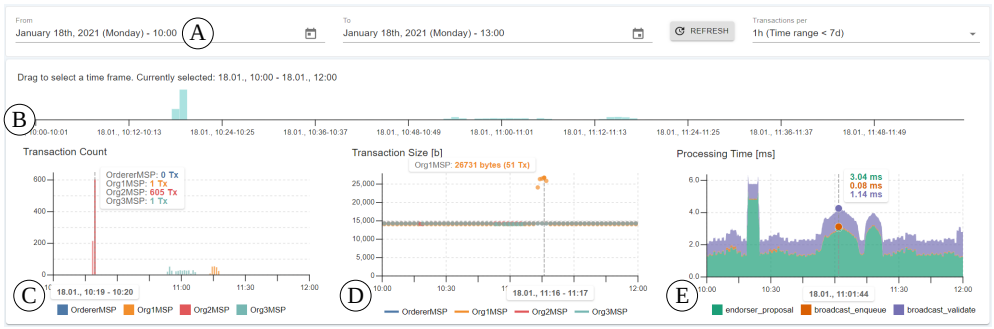


Figure 10: Transaction View of HyperSec as published in P4.

a focus period. The charts in Figure 10-C, 10-D, and 10-E represent specific metrics that are of high importance for security monitoring. The evaluation of the prototype as the result of this design study is done by simulating three attack scenarios on a Hyperledger Fabric test network. All three attacks are detectable using the HyperSec prototype so that countermeasures can be initiated quickly in an operational case.

Contribution of P4: P4 contributes a problem-oriented design and prototypical implementation of a VA security monitoring tool for Hyperledger Fabric. The underlying domain problem addressed by the proposed design is the lack of support for analyzing IoC within blockchain data. Thus, based on the involved users, their specific tasks, and available data elements, we derive design requirements for the VA tool design. Our prototype shows how these design requirements can be implemented in an open-source tool.

4.4 Focus Area 3: Incidents

After a threat has been detected, possibly with the support of the approaches introduced in Focus Area 2, the IDL intends two vital steps to happen. First, fast reactions need to be taken to contain the attack and recover affected systems. Subsequently, the incident has to be analyzed in detail to identify the exploited attack vector, the perpetrators' techniques, applied tools, and other relevant information. These activities (among others) are encapsulated under the term of Digital Forensics. The primary goal of forensic investigations is to obtain reliable evidence and reconstruct the crime scene. Therefore, the research carried out within Focus Area 3 aims to support forensic experts in their incident analysis as part of a Live Digital Forensics investigation. Thus, the papers summarized in this focal area provide an answer to RQ3.

P5: Designing a Decision-Support Visualization for Live Digital Forensic Investigations

In this context, Paper P5 introduces a design for a VA approach to help forensic experts make well-informed decisions about forensic tools to be applied. LDF techniques are part of security incident analyses to investigate and understand certain types of malware,

especially Fileless Malware (FM). FM exists exclusively in memory-based sections of a computer (such as the RAM). This makes it challenging to use traditional forensic techniques for reconstruction and evidence acquisition because they are mainly applied to switched-off systems. However, due to FM's characteristics, shutting down the compromised device can result in an extensive loss of evidence. LDF techniques allow to analyze running systems, identify artifacts, gather intelligence, and secure evidence. During an LDF investigation, forensic experts must make fast and well-informed decisions about the forensic tools to use. Poor or slow decisions can compromise or even destroy critical forensic artifacts. As LDF investigations are mainly carried out by experts and cannot be automated due to the domain knowledge involved, there is a strong need to support the decision-making process of domain experts. Therefore, we propose the application of VSA, helping to decide which tools to use or which indicators to investigate in more detail.

To design a feasible VA solution, we first need to have a clear understanding of the domain in whose context the tool will be used. The domain problem, which we aim to address, is that forensic investigators struggle in selecting the right LDF techniques, tools, and artifacts. However, before we can achieve an appropriate design, it is necessary to understand the forensic process in more detail. A simplistic process model for conducting LDF investigations serves as our starting point here. The key steps of this process are *Collection*, *Examination*, *Analysis*, and *Reporting*. During Collection, relevant data sources for incident information are identified; and the respective data is collected and pre-processed. In the next step, Examination, the data and its suitability for the intended analysis is evaluated. The actual analysis then takes place in the following step. Here, indicators and other data points are correlated to answer specific, relevant questions for the DF analysis. In the Reporting step, the analysis results are prepared and presented. All these process steps need forensic experts to perform decision-making tasks based on vast amounts of data. Thus, forensic experts can benefit significantly from visual decision support.

Therefore, we identify several specific tasks that forensic investigators must perform as part of the process during an LDF investigation and in which a VSA solution provides support. They can be summarized under the following key concepts: *Data Acquisition* (e.g., identification of suspicious devices and media), *Establish Intelligence* (e.g., identifying currently existing network connections), *Memory & Data Analysis* (e.g., detecting anomalies in persistent and volatile memory), and *Documentation* (e.g., documenting evidence processing). Compared to traditional DF investigations, only data that can be retrieved without interfering with the analyzed device is available during an LDF investigation. This is because any intervention in the running system could concurrently compromise evidence. For this reason, only a subset of traditional forensic artifacts is relevant for LDF: File accesses, network packets, process lists, event logs, and system statistics.

In order to make comprehensible decisions regarding specific visual encodings and interactions, we abstract the users' tasks and available data. Through abstraction, the

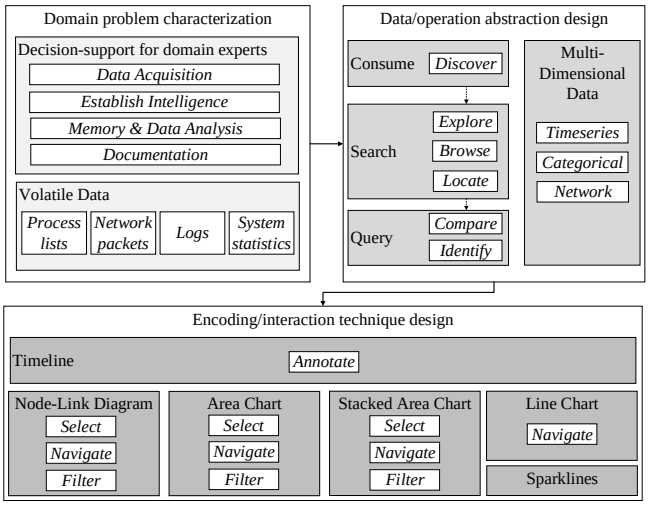


Figure 11: Design concept for the decision-support system as published in P5.

descriptions of tasks and data are translated from a specific forensic terminology to a more generic visualization terminology. This abstraction ensures comparability throughout different visualization designs across different domains. It also enables us to derive suitable visual encodings for the abstract visualization tasks encapsulated within the forensic tasks we aim to support. This methodical design approach is depicted in Figure 11 which illustrates the abstraction and decision process underlying our design. The only relevant data type considered for the abstract data is multi-dimensional data. The forensic experts' tasks can be abstracted into different visual tasks. The main task here is *Discover*, which confines the formulation and verification of hypotheses. This task is decomposed into further abstract sub-tasks in order to be able to use more target-oriented visualization techniques.

Figure 12: Design sketch for the visual decision-support system as published in P5.

The domain problem, the abstract tasks, and the data types shape the final design of the decision-support visualization. The design, shown in Figure 12, comprises five interactive views. The overarching Investigation Timeline at the top allows users to select a time period for analyzing the data. The other four views show only data from that selected period. The Network Activity view (top left of Figure 12) provides an overview

of the external communication of the monitored device. The Read/Write Entropy (top right of Figure 12) shows the entropy of both read and write operations on the investigated device's hard drives. In the lower-left side of Figure 12, system activity is visualized in terms of system log events. Finally, system performance (both for the overall system and for each process) is plotted at the bottom right of Figure 12. All views are designed to have a high degree of interactivity (e.g., filtering and zooming).

The resulting artifact of this paper is not a working prototype but the methodologically derived design for a VA approach to support decisions of forensic scientists during an LDF investigation. To show the applicability and usefulness of the visualization design, we demonstrate how indicators of a fileless malware (i.e., Poweliks) are identifiable in the design. Additionally, we point out which conclusions in terms of decision-support forensic experts can draw from the prototype. P5 gives a first indication towards an answer for RQ3 by highlighting a possible VA design to support forensic experts.

Contribution of P5: P5's main contribution is the design of a decision-support system for Live Digital Forensic investigation utilizing Visual Analytics. The proposed VA design enables forensic investigators to make fast and well-informed decisions on the tools to be deployed for the specific investigation. This paper was awarded the Best Student Paper Award at the Data and Application Security and Privacy XXXIV (DBSec 2020) conference.

P6: Visual Decision-Support for Live Digital Forensics

Paper P6 builds on the design from P5 and significantly improves it in several aspects. P6 refines the domain problem, derives generalized requirements, which provide a foundation to develop other decision-support tools for LDF investigations. Furthermore, in this paper, we improve the VA design and implement a working prototype to visualize incident information retrieved from a mobile phone. Thus, P6 can be interpreted as a possible answer to RQ3.

LDF is not only necessary to investigate FM that resides exclusively in memory-based sections of the infected devices but also for mission-critical systems that cannot be shut down. During respective LDF investigations, the targeted use of forensic tools is paramount to avoid compromising evidence. After indicators of an incident are identified, forensic investigators must quickly decide which forensic tools to use. This decision is mainly based on their experience. Therefore, domain experts must be supported to quickly understand where the incident manifests in the system.

In this work, we address a similar domain problem as defined in P5. However, the specification of this problem is carried out in more detail in the form of a requirement analysis. In order to define target-oriented requirements for an LDF decision-support VA system, it is necessary to define the data to be visualized. In the context of LDF, highly volatile data from a system are in focus. Moreover, it must be possible to access this data without installing any (or as little as possible) additional software to avoid evasive actions by the potential attacker. Therefore, we identify the following volatile

data points as relevant for LDF expert decision support: *System logs* can help identify unusual events and guide forensic analysis. *Process information* supports detection of irregularities in running applications. *In- and outgoing network traffic* is essential to monitor, as it can be used, for example, to detect communication with Command & Control-servers. *File-system activities* give a deep insight into what data is being used and modified. Additionally, we employ a unique mechanism to extract versions of modified files and persist them in a forensically sound manner.

The intended users for the designed VA approach are DF experts. DF investigations require considerable domain knowledge, only owned by these experts. Therefore, we assume they can decide which forensic tool to use if adequately supported. An additional, important characteristic of the target group is the time pressure under which decisions have to be made in an LDF investigation. This definition leads directly to the tasks that this target group should fulfill with the visualization. The overarching, general task of the users is to make well-informed and thoughtful decisions about the use of forensic methods, tools, and artifacts during an LDF investigation. This task is broken down into a series of fine-granular tasks along a process for forensic analysis. The detailed tasks show that forensic experts must regularly make decisions about additional acquisition of data and the use of specialized tools. This discussion serves as the basis to derive generalized requirements for LDF decision-support.

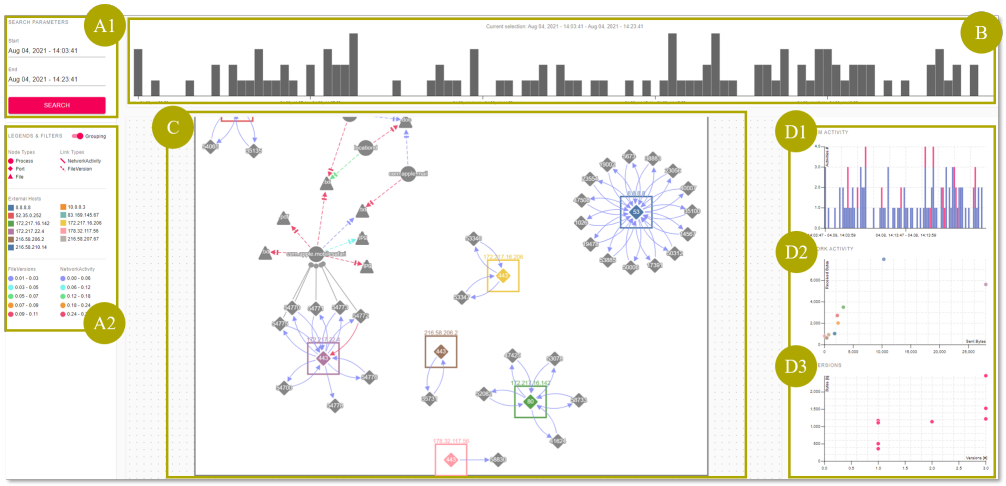


Figure 13: Prototype of the visual decision-support system as published in P6.

An exemplary implementation of these requirements in a prototype is also a goal of this paper. The enriched and implemented design of the VA prototype is shown in Figure 13. The frontend of this application consists of four interactive linked views. The time period that the expert intends to analyze is set in the search parameters (Figure 13-A1). Users can utilize the Overview (Figure 13-B) to get a first impression of the overall activity of the analyzed device within the time period. More detailed information about the device's activities and the entities involved (processes, ports, files, IP addresses, and their relations) can be analyzed using the Node-Link diagram (Figure 13-C) with the help of additional filters (Figure 13-A2). The details-on-demand views (Figure 13-D1, 13-D2,

and 13-D3) support the overview of activities with additional visualizations. Details about an entity replace these views once an appropriate selection is made in the node-link diagram. The prototype for implementing this design uses a client-server architecture. Raw data acquisition is performed with minimum interference with the analyzed device. This requires a trade-off between the level of detail of the data and installing additional software on the device. Our approach enables data acquisition of sufficiently detailed information from a running mobile device using only functionalities already available on that device. The automated data analysis of our prototype correlates the data from different sources to provide a meaningful and helpful visual representation of the data. The data filtering based on the selected time window takes place entirely on the server-side of the prototype to reduce the amount of data for the frontend as much as possible. A final mapping step translates the raw data into the data model finally displayed to the user.

The prototypical implementation of the design in this paper contributes to supporting the decision-making process of forensic scientists during an LDF investigation. The applicability and usefulness of our prototype are illustrated by describing the core functionalities in the context of an exemplary use case. The use case considers investigating an attack pattern based on the "Jeff Bezos Hack". Using the available reports and information about this attack, we show that forensic experts can use the prototype to identify relevant activities of the incident and apply the necessary tools for deeper forensic analysis.

Contribution of P6: In P6, a research prototype is developed to provide DF experts with decision-support during an LDF investigation. The design follows a problem-oriented approach with requirements derived from data, users, and tasks. The prototype extract indicators from a running mobile device and provides interactive visual representations of this data through several interlocked views. This helps forensic experts to identify targets for further forensic analyses when working to understand an incident.

4.5 Focus Area 4: Intelligence

After detection, containment, and in-depth analysis of an incident, the final phase of the IDL is concerned with activities to create, enrich and share intelligence about the incident and possible threats. This Cyber Threat Intelligence is a crucial enabler for any collaborative security activities. However, valuable and actionable CTI contains semantic information about complex aspects of a threat. This information is contributed mainly by domain experts. Thus, the last focal area of this dissertation encapsulates two papers discussing the integration of experts' domain knowledge in CTI-related activities and, thus, answers RQ4.

P7: Graph-based Visual Analytics for Cyber Threat Intelligence

Paper P7 directly addresses making complex CTI accessible for analysis and enrichment by domain experts using VA. The ever-increasing number of cyberattacks has led to different efforts to improve the efficiency of activities along the Incident Detection Lifecy-

cle. One of them is an intensification of cooperative cybersecurity efforts within the last step of the IDL. CTI is used in this context to describe and exchange important insights about attacks or threats. However, CTI formats are complex and optimized primarily for machine readability. These formats are not accessible to humans, which means that, although highly valuable, security experts can hardly contribute their knowledge about an incident to the exchanged CTI.

In P7, we present our research prototype Knowledge-assisted Visual Analytics for STIX (KAVAS), satisfying three crucial requirements for making CTI accessible for security analysts. At first, KAVAS needs to handle complex CTI and ensure its integrity across multiple processing steps. Second, the tool must enable a visual representation of CTI so that experts can understand the complex relationships described by the CTI. As a final requirement, our prototype should enable the enrichment of CTI with the knowledge of experts. Specifically, this means that the visualized CTI in the prototype must be efficiently and effectively editable. To highlight a possible solution fulfilling these requirements, the prototype works with the Structured Threat Information Expression Version 2 (STIX2), which is the de-facto standard for CTI exchange formats.

The first of the above requirements (i.e., dealing with complex CTI) is met by one of the two major building blocks of KAVAS. This building block, the CTI Vault, is a concept for persisting CTI described using STIX2 and preserving its integrity. Since STIX2 is conceptualized as a graph, which allows arbitrary relationships between the corresponding nodes, the CTI Vault is implemented using a graph database. It distinguishes between *inventory data* and *appended data*. The former represents the original intelligence captured during the analysis of an incident. This information is defined to be read-only and must not change during the use of KAVAS. However, if users change, add, or even delete information, these adjustments are captured by the concept of appended data. Thus, a new vertex containing the updated information is integrated into the graph and linked to the original element whenever an object changes. This concept ensures that the inventory data is preserved and that any changes create a traceable, integrity-assured history of all versions resulting from interaction with experts.

The second component of KAVAS is an interactive, visual display which makes the STIX2 format accessible and editable for domain experts. Thus, this visual display fulfills the remaining two requirements mentioned earlier. STIX2 is an expressive but text-heavy and semi-structured CTI format, where the description of a single incident can easily be several thousand lines long. This complicates any manual documentation and analysis process aiming to enrich the CTI with expert knowledge. An interactive visualization facilitates this process for users. As STIX2 is designed as a connected and directed graph, we also use a directed node-link diagram to represent the CTI. This visualization technique supports the understanding of the complex relationships described by the CTI. A user's analysis process starts with selecting a STIX2-based incident description from the CTI Vault. Within the visual representation, the domain objects used in STIX2 are displayed as vertices in the node-link diagram, and the relationships between domain objects are displayed as edges. If the object has been edited (i.e., appended data versions

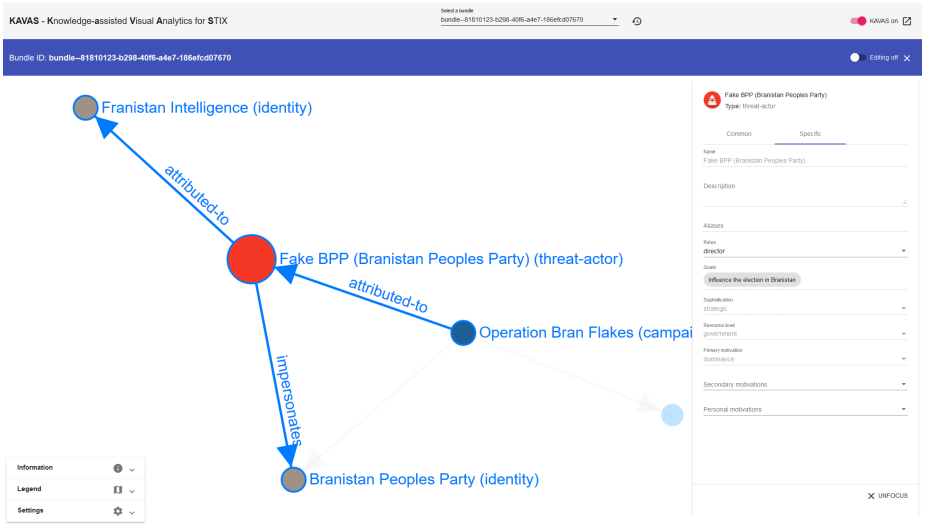


Figure 14: KAVAS's visual display with details of a CTI report as published in P7.

exist), only the most recent version (i.e., the version with the most recent timestamp) is displayed.

The node-link diagram of an incident description can be very complex due to many highly interconnected nodes. KAVAS's visual display is equipped with interactive functionalities to simplify understanding these complex incidents for users. On the one hand, there are several dynamic filter options, e.g., hiding specific node and edge types, to reduce the number of displayed objects. On the other hand, the graph's automatically generated force-directed layout can be interactively adapted to the analyst's needs by dragging individual nodes. Selecting a node or edge displays detailed information of the underlying STIX2 object. In order to enable the enrichment of an incident description in KAVAS and, thus, integrate expert knowledge into the existing CTI, experts can edit information in the Visual Analytics component and add new nodes and edges. Figure 14 shows the visual display of KAVAS, where a node of the analyzed incident description is selected to get detailed information.

We evaluate the prototype in two phases. First, an anonymous survey among CTI analysts validates the usability and fitness for use of KAVAS. Weaknesses, which were identified by this survey, are subsequently addressed. After that, we conduct semi-structured interviews with CTI experts. In these interviews, the experts work with KAVAS, and at the same time, their assessments of the different functionalities are evaluated. Both phases of the evaluation clearly demonstrated the usefulness of the prototype and the need for interactive visual representations of CTI.

Contribution of P7: P7's main contribution is a concept for interactive VA of Cyber Threat Intelligence. The presented approach persists CTI in a graph database with an integrity-preserving structure. The visual design facilitates integrating security experts' domain knowledge into structured threat intelligence. This is achieved with functionalities to explore the intelligence and add new knowledge allowing for more thorough incident documentations.

P8: Measuring and Visualizing Cyber Threat Intelligence Quality

Ensuring the quality of shared Cyber Threat Intelligence is central to successful cybersecurity cooperation. Comments from some interviewees during the evaluation interviews of P7 and existing related research indicate that inaccurate, incomplete, or outdated CTI is one of the biggest problems for enterprise collaboration. However, a foundation for ensuring quality is to create concepts to measure quality in the first place. In addition, the quality of CTI is an essential factor for security experts when prioritizing threat intelligence. At the same time, analysts' knowledge is a promising source for the usefulness of CTI. Paper P8 addresses these aspects by presenting an approach to measure relevant quality dimensions for a standardized CTI format and extending the prototype from P7 to visualize CTI quality. Additionally, the proposed solution allows experts to use interactive visualization to incorporate their judgment of quality aspects.

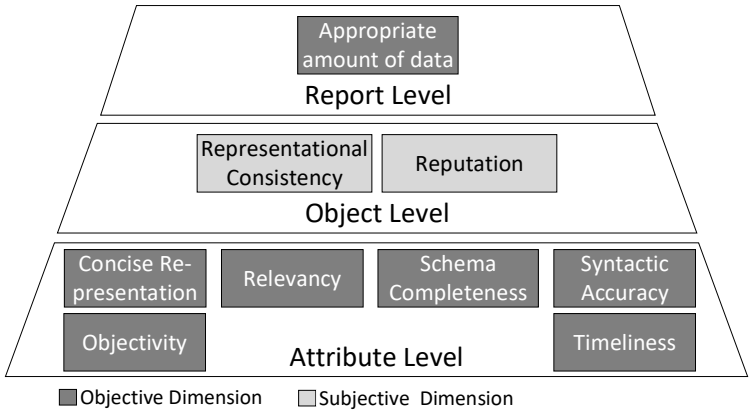


Figure 15: Hierarchical structure of CTI quality dimensions as published in P8.

Before we can assess the quality of CTI, relevant quality dimensions for CTI must be determined. To define the relevant quality dimensions, we rely on existing work that has already proposed initial collections of quality dimensions important for CTI. We complement these collections with additional dimensions that we identified from discussions with CTI researchers and practitioners. The dimensions considered in P8 are presented in Figure 15. Assessing these dimensions is only possible with a specific CTI data format in mind. Within this paper, we define respective quality metrics for the STIX2 format. Two reasons have led us towards this decision. On the one hand, STIX2 is a standardized and very well-established CTI format. On the other hand, focussing on STIX2 ensures compatibility of our work with P7 in order to incorporate the quality assessment into P7's prototype.

The quality dimensions identified relevant for the STIX2 format can be structured hierarchically as shown in Figure 15. Some dimensions can be quantified by evaluating unique attributes of STIX2 objects, while others can be determined by including additional attributes or considering the object as a whole. However, the dimension of "appropriate amount of data" can only be measured regarding an incident description in its entirety. Moreover, the hierarchical structuring allows to aggregate the different dimensions,

enabling a quick assessment of the overall quality of a CTI report. Furthermore, we distinguish between subjective and objective data quality dimensions. While explicit formulas for quantification can be determined for the objective dimensions, which are based exclusively on the underlying CTI incident description, the assessment of security experts is required to measure subjective quality dimensions. Our work defines a set of relevant quality dimensions and proposes metrics to assess them for a STIX2-based incident description. We apply mechanisms typically used within reputation systems for the subjective dimensions, which can only be quantified by incorporating experts' knowledge. This allows us to normalize the subjective assessments and translate them into quantifiable measurements. For aggregating the quality dimensions into a single value for a STIX2 report, we define a weighted average across all relevant dimensions and the three hierarchy levels.

Three steps are necessary to integrate these CTI quality aspects of P8 into the prototype from P7. The first step is to incorporate the quality indicators in the STIX2 format. In its original specification, STIX2 does not provide a way to describe its quality but allows the definition of custom objects. Thus, we define a custom object to persist the quality indicators for a native STIX2 object according to the specification in a STIX2 report. The next step is to extend the CTI Vault from P7 with a Quality Vault. This additional database persists the custom object for the quality indicators. Separating native STIX2 objects and the custom quality objects allows us to maintain the performance of the CTI Vault despite the rapidly increasing amount of data since the actual STIX2 report and the associated quality elements can be loaded independently from each other. In the last step of evolving the prototype from P7, we adapt the visual display to include the CTI's quality indicators. For instance, an extension of the detail view allows an insight into the quality indicators for a single STIX2 object. The experience and assessment of analysts on the subject quality dimensions can be directly and interactively incorporated into the system by ratings from one to five stars, which are known from reputation systems.

We evaluate the relevant data quality dimensions with the designed metrics and the extended KAVAS prototype through three interviews with CTI experts. In these interviews, the participants confirm the high relevance of the chosen dimensions and the high need for actual metrics to quantify these dimensions. Especially as the increasing importance of CTI makes it more evident that only high-quality CTI can be helpful.

Contribution of P8: The main contribution of P8 is a concept to measure CTI quality and include expert knowledge into this assessment. Our approach defines a relevant set of quality dimensions and configures respective metrics. Our approach is integrated into a visual representation of CTI, making the quality assessment transparent for CTI analysts and enabling the assessment of subjective quality dimensions based on analysts' input. Thus, we propose the first concept to measure CTI quality, including actual metrics allowing the assessment to be implemented in respective technologies.

4.6 Complementary publications

In addition to the eight publications that represent this dissertation's central contribution (P1 - P8), complementary research was carried out. Although the respective publications are not directly included in this dissertation, they have influenced or complemented the core contribution. For this reason, they are briefly summarized within this section. Table 2 provides an overview of the complementary publications. For each paper, the complete reference, the submission status[4] of the publication, the publication type[5], and the number of citations as of December 14, 2021 are shown.

Table 2: Overview of complementary research papers.

No.	Full Reference	Status	Type	Cit.[6]
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 *SICHERHEIT 2018* (Konstanz, 2018), Gesellschaft für Informatik e.V., pp. 257–260	pub.	C	11
A2	BÖHM, F., RAKOTONDRAVONY, N., PERNUL, G., AND REISER, H. Exploring the role of experts' knowledge in visualizations for cyber security. In *Posters - IEEE Symposium on Visualization for Cyber Security* (Berlin, 2018), IEEE	pub.	C	1
A3	BÖHM, F. Visual Security Analytics. *Encyclopedia of Cryptography, Security and Privacy* (2021), 1–3	pub.	E	0
A4	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	pub.	C	3
A5	BÖHM, F., DIETZ, M., PREINDL, T., AND PERNUL, G. Augmented Reality and the Digital Twin: State-of-the-Art and Perspectives for Cybersecurity. *Journal of Cybersecurity and Privacy 1*, 3 (2021), 519–538	pub.	J	0

Publication **A1** introduces a high-level architecture as a blueprint for modern SIEM systems. The work was presented as a short paper at the *SICHERHEIT 2018* conference's practitioners track [28]. Despite the paper's brevity, it served as a starting point for the considerations leading to innovative prototypical architectures like those presented in P2, P6, or P7.

Within Publication **A2**, tools presented in the research area of visualizations for

[4] *pub.* = published
[5] *C* = Conference, *J* = Academic Journal, *E* = Encyclopedia Entry
[6] Citation count based on https://scholar.google.com/

cybersecurity are analyzed with respect to their degree of integration of domain knowledge. This work was presented as a poster at the *IEEE Symposium on Visualization for Cyber Security* [4]. The recognized shortcomings of the analyzed research have been a strong motivation for the work presented in this dissertation. Furthermore, the lack of a common vocabulary regarding knowledge within the information security domain ultimately has led to the contributions made in P2 and A4.

In Publication **A3** the term of Visual Security Analytics is defined concisely. The article is part of the *Encyclopedia of Cryptography, Security and Privacy* [2]. A3 gives brief insights into the definition, background, theory, open problems, and future directions of VSA. This publication defines the understanding of VSA underlying this dissertation.

Publication **A4** is the initial version of P2. It was presented at the *7th International Conference on Information Systems Security and Privacy 2021* [5]. A4 focuses on an initial definition of knowledge types relevant for cybersecurity, introduces a model of knowledge-based Security Analytics and a prototype that allows the integration of domain knowledge of security novices. This work was invited to be submitted as an extended version based on positive reviewer feedback and the valuable contribution of the work.

Finally, Publication **A5** defines the state-of-the-art and derives possible use cases regarding the combined application of augmented reality and digital twins in the context of cybersecurity. The paper was published as an article as part of the *Journal of Cybersecurity and Privacy*'s special issue titled "Cyber Situational Awareness Techniques and Human Factors" [3]. With augmented reality advancing traditional two-dimensional visualizations, this publication explores possible application spaces of VSA within an ever-changing security domain. A5 can be seen as a glance into the future of VSA.

5 Conclusion and Future Work

Visual Analytics is a key enabler to improve existing security mechanisms by integrating domain experts into vital activities which cannot be fully automated. Only domain experts and their knowledge can reveal unknown attack patterns or advanced threats, making them invaluable for any security operation. Despite the importance of domain knowledge and the role of Visual Analytics to integrate this knowledge into automated mechanisms, leveraging Visual Analytics for cybersecurity is an area with a relatively limited research corpus. Thus, the purpose of this dissertation is to contribute novel approaches to the existing works by highlighting the advantages of VSA for improving the collaboration between security experts and automated security mechanisms as well as supporting experts in their manual tasks. A significant proportion of the existing work either lacks security domain knowledge or visual efficiency. Therefore, this dissertation supports resolving this dichotomy of Visual Security Analytics through a transfer of knowledge between previously distinct research communities.

In order to achieve these goals, this dissertation makes several essential contributions in four focal areas related to the field of Visual Security Analytics. In the first Focus Area, profound definitions of the SOC as security experts' organizational environment, as

well as relevant processes and related challenges, had to be delineated. In the SOC as an organizational unit, domain experts have to perform some of the most critical security activities. Understanding the work environment, underlying processes, and different roles of employees within a SOC is of central significance for the targeted use of VSA. Based on these findings, an iterative process for security analytics, the Incident Detection Lifecycle, was defined. This precisely describes the activities and tasks of the experts, which VSA can support. The use of VSA also enables the integration of expert knowledge into the respective phases of the process. In the dissertation, this knowledge was formally defined to establish a common vocabulary for the further course of this work and, thus, the respective research area.

In addition to these fundamental contributions, VSA was applied in major phases of the IDL and different security domains. The first relevant process step involves detecting and analyzing threat indicators and is considered in Focus Area 2. In this context, the dissertation highlights the feasibility of VSA for the security monitoring of distributed ledgers and the identification of potential security issues in Identity and Access Management. These contributions showcase possible applications of Visual Analytics to support security experts in the analysis of Indicators of Compromise.

When a concrete suspicion of a threat becomes apparent analyzing these indicators, it must be investigated in more detail. Therefore, the third focal area of the dissertation was dedicated to supporting domain experts in the detailed analysis of security incidents using VSA. The work focuses primarily on supporting the manual activities of experts during Live Digital Forensics investigations. In this context, the use of VSA facilitates decision support for the targeted and well-informed use of specialized forensic tools. Visualizations offer an advantage because they can simultaneously and interactively depict various aspects that serve as a basis for the experts' decisions. This enables them to analyze these indicators in an exploratory manner. The results of an incident analysis are documented as intelligence and used for cooperative security purposes.

In the fourth Focus Area of the thesis, possibilities were investigated on how VSA can be utilized to integrate experts' domain knowledge into Cyber Threat Intelligence. Since CTI is mapped in standardized formats optimized for machine readability, these formats are often not intuitively accessible to human experts. At the same time, however, it is precisely the domain knowledge of these experts that is of high importance for the informative value of CTI, so the dissertation proposed the use of visual security analytics to make CTI accessible and processable for experts.

In addition to these immediate findings, the work done in this dissertation has raised starting points for further developments and research. Three aspects, in particular, open up interesting and promising avenues for future contributions:

The first important issue emerged in connection with the foundations defined in the first Focus Area. Here it became clear that, in addition to integrating the domain knowledge of security experts, the integration of experts from other domains is essential for comprehensive security operations, especially in the light of current developments such as the increasing implementation of the Industry 4.0 paradigm. While security

experts have extensive knowledge of concrete cybersecurity aspects, such as the operation of firewalls or the analysis of network traffic, they lack the expertise to assess the security situation of cyber-physical systems. Engineers and similar professions must be called upon for this purpose. Only they can interpret whether, for example, a turbine used to generate electricity is behaving correctly, or it may have been the victim of a cyber-physical attack. However, these experts from other domains lack the necessary operational knowledge to handle and operate relevant security mechanisms. Therefore, they should be classified as security novices. However, visual approaches can help to facilitate the operation of security mechanisms and thus enable the integration of the non-security domain knowledge held by novices. The dissertation shows first possibilities in this direction, but there is a need for further research for innovative approaches and their comprehensive evaluation.

A second aspect that needs further research concerns the support of experts in Live Digital Forensics investigations. In this context, most of the work, and especially the decisions about the forensic tools to be used, are currently done manually by experts. A support option for experts that goes further than a visualization of relevant indicators proposed in this dissertation would be a more automated decision-support. For instance, the automatic generation of tool suggestions for experts would be conceivable. The basis for these tool suggestions (together with the associated uncertainty) could be visualized for the experts. This would support the experts' decision-making and, at the same time, keep the automation transparent. Visualization would also make it possible to individually parameterize the tool suggestions so that they are as tailored as possible. Based on the tool actually used by the experts, the automated suggestions could also be improved iteratively in the sense of a feedback loop (as envisaged in the VA process).

The third aspect for further research arises concerning the work with Cyber Threat Intelligence in this dissertation. While an initial approach to measuring the quality of CTI, along with ways to incorporate expert opinion, has been proposed in this thesis, there remains a great need for further research in the area of CTI data quality. The approach from this dissertation relates to a specific CTI exchange format. While some dimensions and corresponding metrics can be applied to other formats, further research is needed to create a generalizable and preferably universal framework for measuring and improving CTI quality. Only with this foundation, CTI can be a basis for effective cooperative cybersecurity in the long term.

To sum up, this dissertation shows, in addition to vital underlying aspects, how Visual Analytics can be leveraged for cybersecurity to involve experts and their knowledge in security analysis processes and provide support for their manual work. At the same time, the dissertation takes first steps towards bridging the dichotomy of VSA by applying and communicating VA methods for problem-solving in cybersecurity domains and thereby building a base for future research directions.

Part II

Research Papers

1 Security Operations Center: A Systematic Study and Open Challenges

Current status:	Accepted & Published
Journal:	IEEE Access, Volume 8, December 2020
CORE Ranking:	**B** (http://portal.core.edu.au/jnl-ranks/665/)
Date of acceptance:	December 13, 2020
Date of publication:	December 31, 2020
Full citation:	VIELBERTH, M., BÖHM, F., FICHTINGER, I., AND PERNUL, G. Security Operations Center: A Systematic Study and Open Challenges. *IEEE Access 8* (2020), 227756–227779
Authors' contributions:	Vielberth Manfred 35% Böhm Fabian 35% Fichtinger Ines 20% Pernul Günther 10%

Journal Description: IEEE Access® is a multidisciplinary, applications-oriented, all-electronic archival journal that continuously presents the results of original research or development across all of IEEE's fields of interest. Supported by article processing charges, its hallmarks are a rapid peer review and publication process with open access to all readers.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

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 VIELBERTH, FABIAN BÖHM, INES FICHTINGER, AND GÜNTHER PERNUL, (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 SOCs 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.

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

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 occurrs. However, there are a few research gaps and challenges associated with SOCs. 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 SOCs and researchers from further adding to the innovation of SOCs. 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

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	*Security Operation Center* **OR** *Security Operations Center* **OR** *Security Operations Centers* **OR** *Security Operation Centre* **OR** *Security Operations Centre* **OR** *Security Operations Centres*
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.

[8]For 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	∑
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.

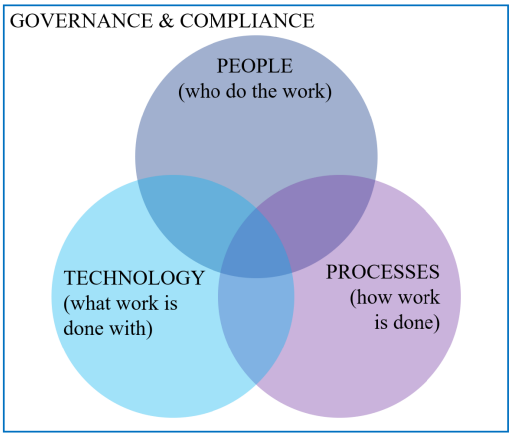


FIGURE 2. The People, processes and technology, governance & compliance (PPTGC) framework based on [70].

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 SOCBox [40]–[42]. These three architectures highlight the shift from centralized to distributed SOC setup over time. The original SOCBox 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.

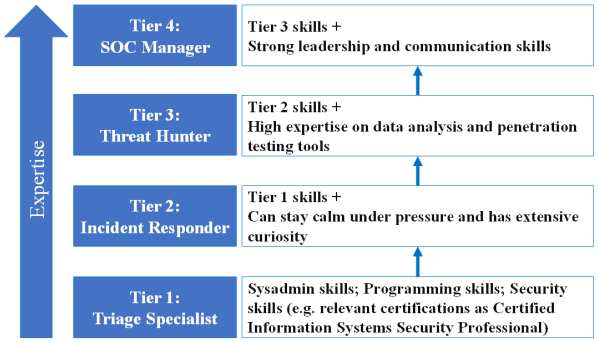


FIGURE 3. Necessary skills among SOC roles [54], [66], [75], [100], [101].

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.

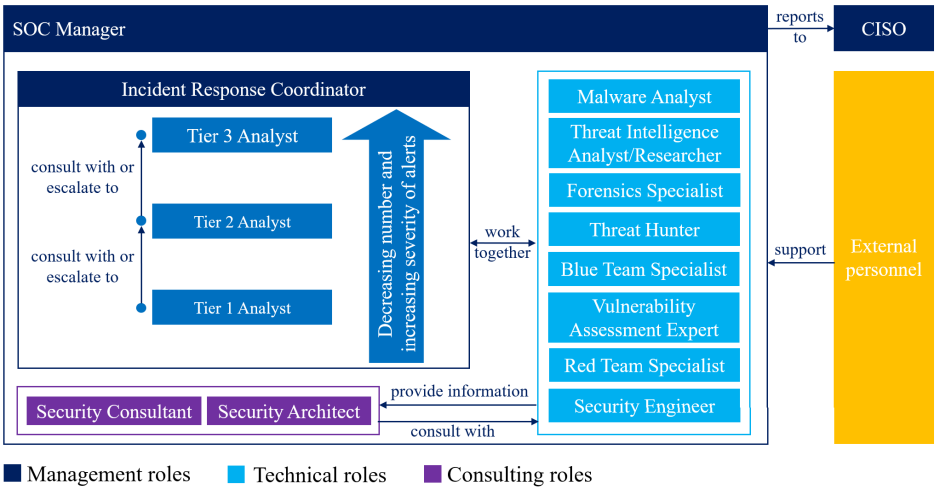


FIGURE 4. **Interaction of different roles within a SOC [79].**

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 SOCs 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 SOCs 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 SOCs. Hàmornik 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].

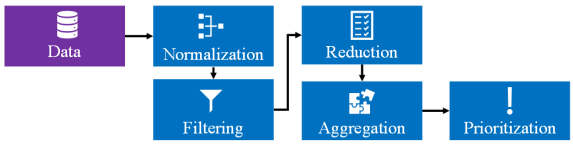


FIGURE 5. The data collection process.

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 unharmful 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].

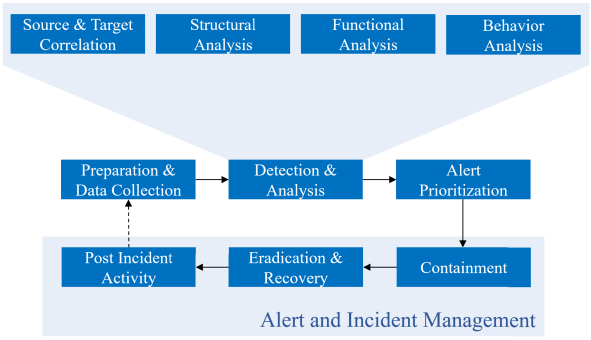


FIGURE 6. The SOC incident analysis, detection and management process.

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

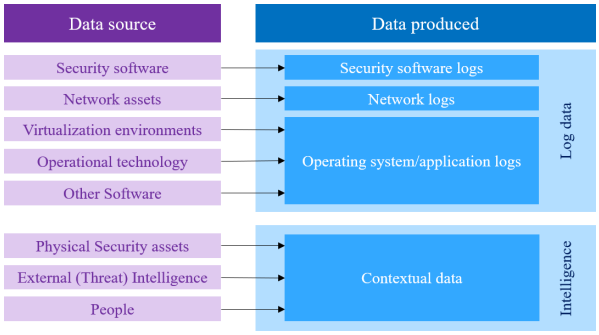


FIGURE 7. Data sources and the type of data they produce.

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	Specification	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]	✓			✓			✓	
∑	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

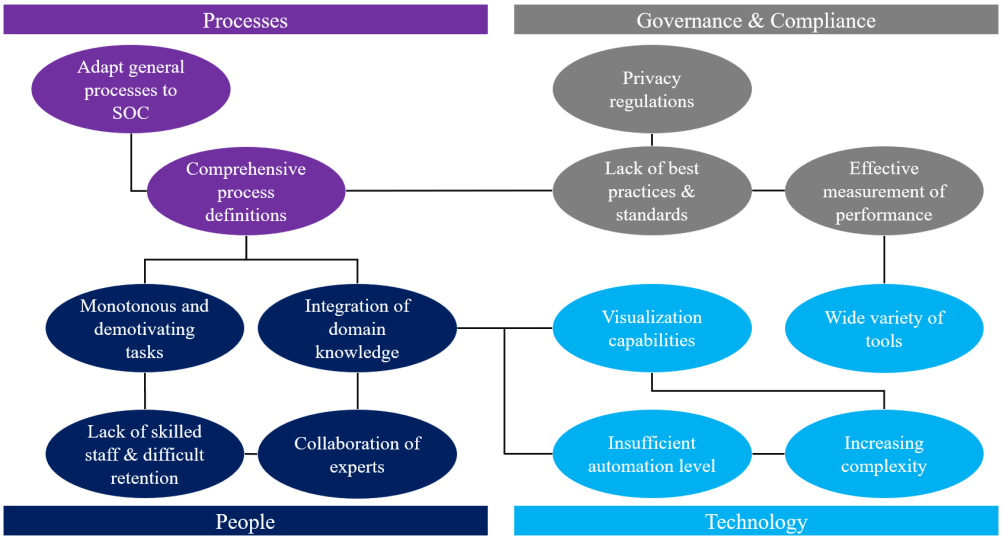


FIGURE 8. Challenges for SOC research.

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 SOCs' 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 analyses 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 SOCs. 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 volume 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, nos. 1–2, pp. 30–47, Mar. 2008.

[7] F. B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupé, 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. Kotenko, 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. Aijaz, 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. Chamiekara, 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 (ST-SIEM)," 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. Netw. Function Virtualization*, New York, NY, USA, 2019, pp. 7–12.

[58] D. Crooks and L. Valsan, "Wlcg security operations centre working group," *Proc. Sci.*, vol. 1, no. 1, pp. 1–25, 2017.

[59] D. Crooks, L. Vâlsan, K. Mohammad, S. McKee, P. Clark, A. Boutcher, A. Padée, M. Wójcik, H. Giemza, and B. Kreukniet, "Operational security, threat intelligence & distributed computing: The wlcg security operations center working group," *EPJ Web Conferences*, vol. 214, p. 15, May 2019.

[60] D. Crooks and L. Válsan, "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. Schrogl, Ed. Cham, Switzerland: Springer, 2020, pp. 921–939.

[97] C. Zhong, A. Alnusair, B. Sayger, 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. Sailhan, 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. Akhunzada, 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. Lakbabi, 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. Laih, "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. Savaä, 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 signature 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!" *Itnow*, 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. (AsiaJCIS)*, 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.
[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 remote 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.
[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'sa 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.

2 Formalizing and Integrating User Knowledge into Security Analytics

Current status:	Under Review
Journal:	SN Computer Science
CORE Ranking:	n/a
Date of submission:	September 30, 2021
Full citation:	BÖHM, F., VIELBERTH, M., AND PERNUL, G. Formalizing and Integrating User Knowledge into Security Analytics. *Springer Nature Computer Science* (2021)
Authors' contributions:	Böhm Fabian 45% Vielberth Manfred 45% Pernul Günther 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 Vielberth[1] and Günther Pernul[1]

[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

[1] Since 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.

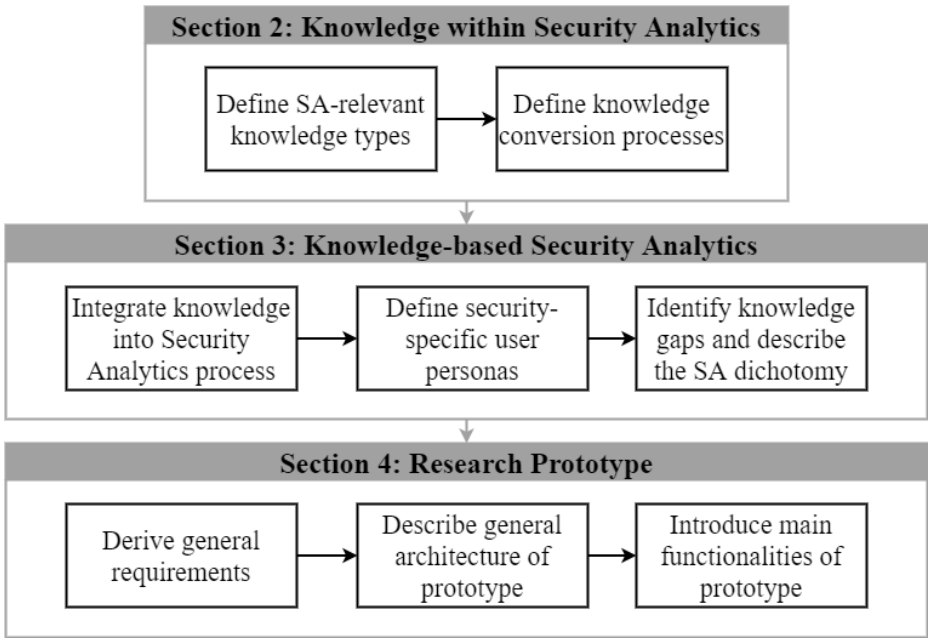


Fig. 1 Schematic of this work's main contribution.

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

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.

$$K^e = K^e_m \cup K^e_s \cup K^e_i \tag{1}$$

Models (K^e_m): 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^e_s): 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^e_s): 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

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.

$$K^i = K^i_d \cup K^i_s \cup K^i_o \tag{2}$$

Domain Knowledge (K^i_d): 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^i_d in Equation 3 in a more detailed way as a combination of two disjoint subdomains $K^i_{d(sec)}$ and $K^i_{d(nonSec)}$. $K^i_{d(sec)}$ comprises security-related domain knowledge, which is mainly part of the tacit knowledge of security experts. The components of $K^i_{d(sec)}$ 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 $K^i_{d(nonSec)}$. 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,

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

$$K^i_d = K^i_{d(sec)} \cup K^i_{d(nonSec)} \tag{3}$$

Situational Knowledge (K^i_s): 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^i_s 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^i_d about the SA of the enterprise is not required to make these inferences.

Operational Knowledge (K^i_o): 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^i_o 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^i_d and K^i_s 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^i_o represents the barrier to entry for this integration. Only with the necessary K^i_o can employees, for example, define an appropriate SIEM correlation rule based on their K^i_d.

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

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.

$$int : (K^e \overset{K_o^i}{\mapsto} K_{d(sec)}^i \cap K_s^i) \tag{4}$$

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].

$$ext : (K_d^i \cap K_s^i \overset{K_o^i}{\mapsto} K^e) \tag{5}$$

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 ($K_i^e \mapsto K_s^e$).

$$comb : K^e \mapsto K^e \tag{6}$$

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 ($K_d^i \mapsto K_s^i$).

$$coll : K^i \overset{K_o^i}{\mapsto} K^i \tag{7}$$

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

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.

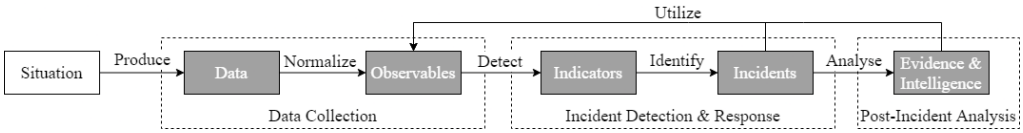


Fig. 2 The Incident Detection Lifecycle.

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^e_s) for signature-based detection, but also

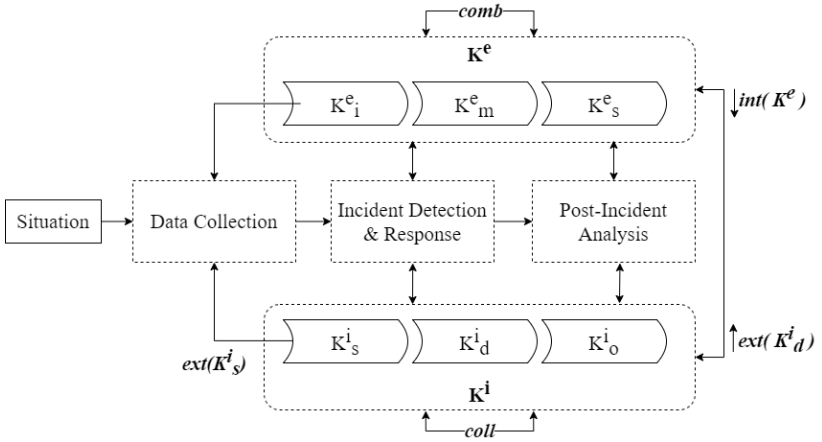


Fig. 3 The Incident Detection Lifecycle extended with Knowledge Types and Conversions.

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.

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: $ext(K^i_s)$ and $ext(K^i_d)$. They are each an instance of ext, but require a closer, contextualized look.

Externalization of situational knowledge K^i_s $(ext(K^i_s))$ 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^i_s into observables by $ext(K^i_s)$ 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 $ext(K^i_s)$, 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^i_d, but a more general knowledge that comes primarily from situational awareness K^i_s. 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 $ext(K^i_d)$. This activity basically comprises two processes: $ext(K^i_{d(sec)})$ and $ext(K^i_{d(nonSec)})$. 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^i_d 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 $ext(K^i_{d(nonSec)})$, 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 $ext(K^i_d)$ 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.

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

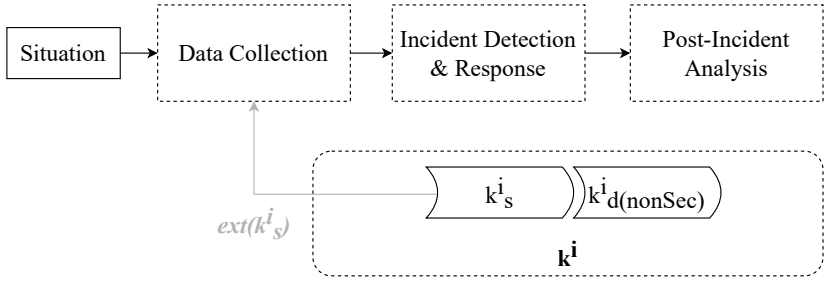
Based on the various security-related knowledge types, different groups of users can be distinguished. As shown in Equation 8 the knowledge of users can be seen in this context as different instances of K^i.

$$k^i_{d(nonSec)}, k^i_{d(sec)} \in K^i_d, k_s \in K^i_s, k^i_o \in K^i_o \tag{8}$$

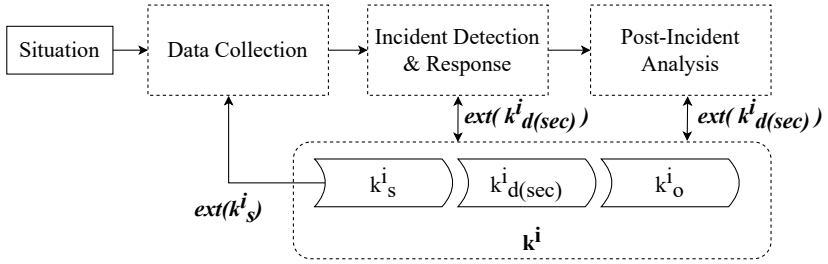
In an organizational context, employees can essentially be assigned to two roles from an SA perspective, which can be referred to as security personas: security novices S_n and security experts S_e.

- 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: $k^i_{d(nonSec)}$. 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^i_s of S_n is particularly relevant, as it enables them to judge a situation in combination with their unique $k^i_{d(nonSec)}$. Thus, they can contribute to the Incident Detection Lifeccycle 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^i_o. $ext(K^i_s)$ is possible if the respective system is sufficiently simple to use, thus, this process is present in the figure but grayed out.

$$S_n = \{k^i_{d(nonSec)}, k^i_s\} \tag{9}$$



(a) Perspective of S_n



(b) Perspective of S_e

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 $k^i_{d(sec)}$. 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 $k^i_{d(sec)}$ in combination with k^i_s 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^i_o 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 $ext(K^i_s)$ and $ext(K^i_d(sec))$ are possible, since the expert has the necessary k^i_o to comprehensively operate the systems involved.

$$S_e = \{k^i_{d(sec)}, k^i_s, k^i_o\} \tag{10}$$

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^i_d

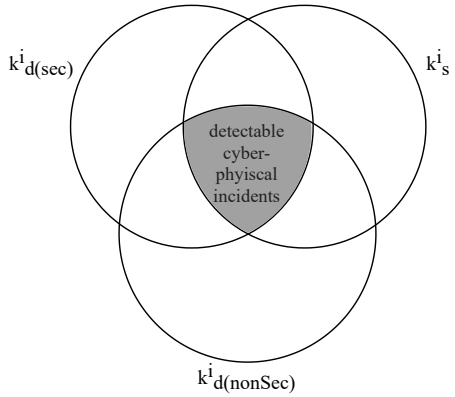


Fig. 5 Required knowledge types for incident detection.

differs ($k^i_{d(sec)}$ vs. $k^i_{d(nonSec)}$) and S_n has no or very little operational knowledge k^i_o.

As already defined in Equation 5 the externalization of implicit knowledge is the intersection of k^i_d and k^i_s. However, when considering cohesive incident detection in the era of cyber-physical attacks, the necessary domain knowledge k^i_d is distributed between S_n and S_e in the form of $k^i_{d(sec)}$ and $k^i_{d(nonSec)}$. Cyber-physical incidents are only detectable if knowledge about security incidents in general ($k^i_{d(sec)}$) and knowledge about the physical aspects in particular ($k^i_{d(nonSec)}$) are combined. In addition, situational knowledge k^i_s 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^i_o 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. $\boldsymbol{ext(K_S^i)}$: 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. $\boldsymbol{K^i_{d(nonSec)}}$: 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. ($ext(K_S^i)$) and 2. ($K^i_{d(nonSec)}$) 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

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* $\mathbf{ext}(\mathbf{K_d^i})$*:* 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:

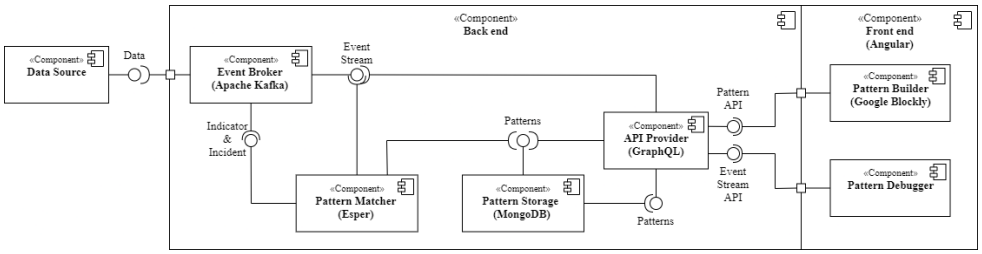


Fig. 6 Component diagram of the architecture for visual collaborative pattern definition [9].

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

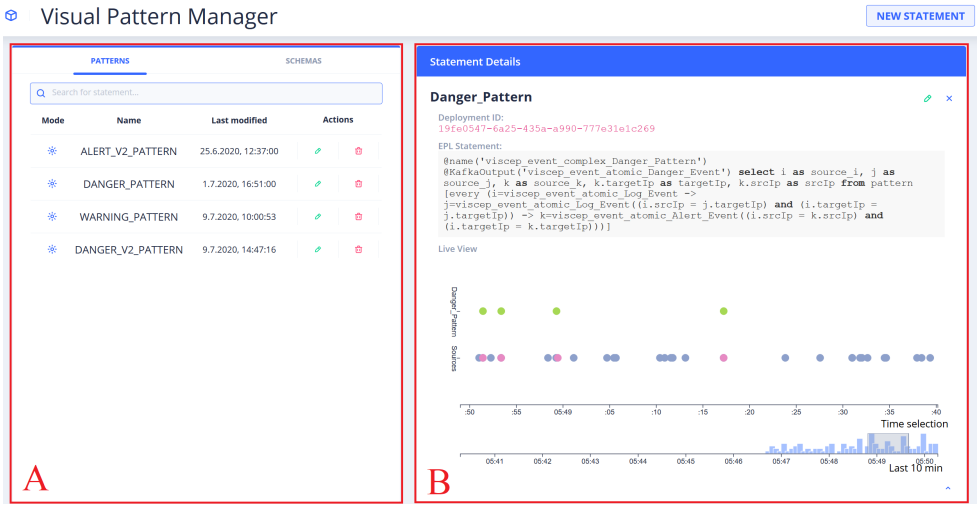


Fig. 7 Screenshot of the front end's landing page with a selected statement [9].

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 Esper[3]. The actual events are provided by various data sources that reflect the current situation of the Incident Detection Lifecycle. With the help of Apache Kafka[4], 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 (MongoDB[5]) 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 GraphQL[6] 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.espertech.com/esper/
[4] https://kafka.apache.org/
[5] https://www.mongodb.com
[6] https://graphql.org/
[7] https://angular.io/

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.

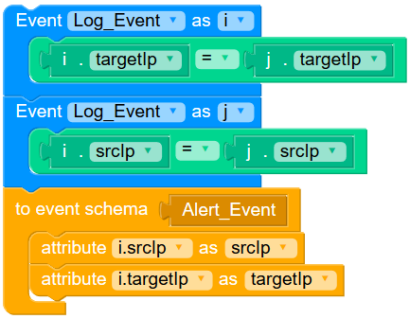


Fig. 8 Screenshot of EPL statement built with Blockly [9].

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

Blockly [8]. 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

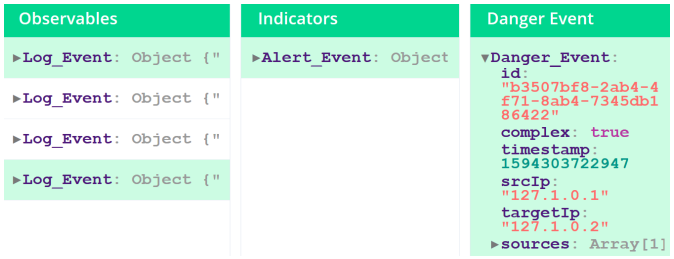


Fig. 9 Screenshot of the Pattern Debugger [9].

Using the arrow on the lower right side of component B, the *Patter 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

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

[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

(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., Grandclaudon, 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

[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 Contributing to Current Challenges in Identity and Access Management with Visual Analytics

Current status:	Accepted & Published
Conference:	33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy, July 15-19, 2019
CORE Ranking:	**B** (http://portal.core.edu.au/conf-ranks/2067/)
Date of acceptance:	April 14, 2019
Date of publication:	June 11, 2019
Full citation:	PUCHTA, A., BÖHM, F., AND PERNUL, G. Contributing to Current Challenges in Identity and Access Management with Visual Analytics. In *Data and Applications Security and Privacy XXXIII*, vol. 11559 of *Lecture Notes in Computer Science*. Springer, Cham, 2019, pp. 221–239
Authors' contributions:	Puchta Alexander 45% Böhm Fabian 45% Pernul Günther 10%

Conference Description: DBSec is an annual international conference covering research in data and applications security and privacy. The 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec 2019) was held in Charleston, SC, USA. 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.

Contributing to Current Challenges in Identity and Access Management with Visual Analytics

Alexander Puchta[1(✉)], Fabian Böhm[2(✉)], and Günther Pernul[2(✉)]

[1] Nexis GmbH, Franz-Mayer-Str. 1, 93053 Regensburg, Germany
alexander.puchta@nexis-secure.com
[2] University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany
fabian.boehm@ur.de, guenther.pernul@ur.de

Abstract. Enterprises have embraced identity and access management (IAM) systems as central point to manage digital identities and to grant or remove access to information. However, as IAM systems continue to grow, technical and organizational challenges arise. Domain experts have an incomparable amount of knowledge about an organization's specific settings and issues. Thus, especially for organizational IAM challenges to be solved, leveraging the knowledge of internal and external experts is a promising path. Applying Visual Analytics (VA) as an interactive tool set to utilize the expert knowledge can help to solve upcoming challenges. Within this work, the central IAM challenges with need for expert integration are identified by conducting a literature review of academic publications and analyzing the practitioners' point of view. Based on this, we propose an architecture for combining IAM and VA. A prototypical implementation of this architecture showcases the increased understanding and ways of solving the identified IAM challenges.

Keywords: Identity and access management · Identity management · Visual Analytics

1 Introduction

Identity and access management (IAM) has become a vital component of modern companies as it enables the management of identities and grants access to necessary resources. IAM also assures compliance with regulations like SOX [41] or Basel III [3]. To achieve this, IAM systems consist of manifold policies, processes and technical solutions [13]. The core of IAM are identities like employees and their access rights to resources maintained within the system. Besides human identities, new technologies like the Internet of Things (IoT) require the integration of technical identities (e.g. sensors and machines) into IAM [27]. Thus, the number of elements maintained in the system is constantly rising. This will ultimately lead to an identity explosion where a vast amount of heterogeneous

Published by Springer Nature Switzerland AG 2019
S. N. Foley (Ed.): DBSec 2019, LNCS 11559, pp. 221–239, 2019.
https://doi.org/10.1007/978-3-030-22479-0_12

identities has to be managed in a single system. This results in numerous problems to be addressed in the next years to ensure IAM systems remain an effective part of companies' IT landscapes.

A solution for those problems needs an effective way to manage and analyze the huge quantity of information with often thousands of identities and hundreds of thousands of entitlements. To decide whether information about an identity is wrong or redundant access rights are assigned to it, the knowledge of domain experts with experience and deep understanding of an enterprise's individual IAM landscape is needed. In this work we investigate how this domain knowledge can be integrated into an IAM landscape by leveraging Visual Analytics (VA) as VA is one of the central methods to include domain experts' knowledge and utilize their feedback [11]. In order to reach this goal, this work investigates three research questions:

- **RQ-1:** What are current and upcoming key challenges within IAM to be solved by integrating domain knowledge?
- **RQ-2:** How can VA be integrated into an existing IAM architecture and which steps are necessary?
- **RQ-3:** What could an exemplary VA solution for IAM look like and which challenges could be solved?

By answering these research questions our work focuses on two main contributions. We provide a list of challenges for current and future IAM. This list is an outcome of a structured analysis taking both academic and practice viewpoints into consideration. We also demonstrate how VA can be applied helping to integrate domain knowledge in tasks to identify IAM anomalies and possible erroneous configurations (e.g. over-authorization or wrong identity attributes). Therefore, we develop a prototypical visualization designed in cooperation with experienced IAM practitioners.

The remainder of this work is structured as follows. Section 2 introduces some background on IAM systems as well as related work regarding the integration of VA into IAM. Next, Sect. 3 follows a structured, two-fold approach to identify current challenges for IAM system as seen from academia and practice to answer *RQ-1*. An architecture to integrate VA into IAM (*RQ-2*) as well as a corresponding proof-of-concept visualization (*RQ-3*) are presented in Sect. 4. The benefits of this prototype regarding the identified challenges are highlighted with exemplary use cases in Sect. 5. Section 6 concludes our work and highlights possible future research directions as well as current limitations.

2 Background and Related Work

In this chapter we define key concepts of IAM and introduce related work regarding the integration of VA into IAM.

2.1 Background

IAM consists of two main fields which are managing identities and granting them access to resources. According to Pfitzmann and Hansen [33] an identity is a subset of attributes uniquely identifying a person. An identity is either real or exists as a digital identity like profiles in social media. Real and digital identities are often linked, and a real identity may own multiple digital personas. However, in the following we assume each entity to have exactly one digital identity as the scope of this work is limited to a single company's context. Currently, IAM regards employees, contractors or customers as identity because they all need to have access to certain resources [45]. In addition to humans having digital identities, technical equipment like machines or sensors are entities which need access to resources, too. Thus, these technical identities also are relevant for maintaining them within an IAM [12].

Digital identities in an IAM are managed from their creation to their deletion when not needed anymore. During this life cycle, access control is used to provide access to applications, data or other information [35]. Enterprises often employ role-based access control (RBAC) in order to grant access [37]. In RBAC, roles are utilized to bundle single access rights and consequently assigned to identities. On the contrary, attribute-based access control (ABAC) leverages identities' attributes and predefined access policies for dynamic access management [16].

To maintain landscapes with thousands of identities, enterprises employ IAM systems which are able to support the identity life cycle and provide identities with the correct entitlements. Besides that, modern IAM systems offer a variety of other functionalities (e.g. Single Sign-on) which are not detailed any further in this work.

2.2 Related Work

There are some existing publications applying visual representations for IAM problems. The earliest integration of VA to the best of our knowledge is the "role graph model" by Nyanchama and Osborn [32]. It is based on RBAC and is used to optimize existing roles for a company. In addition to that, several authors propose a matrix-based approach to visualize users and their entitlements [5,28]. Based on that, VA can be applied to identify suitable roles or outliers with extensive entitlement assignments. Recently, Morisset and Sanchez introduced a tool to visualize ABAC policies [30].

These approaches are focusing mostly on interactive visual techniques for Access Control. To the best of our knowledge there is no existing work taking Identity Management into consideration to build a more cohesive visual solution. Therefore, we try to fill this gap by identifying general IAM challenges where domain knowledge of experts is needed to solve them. After identifying those challenges, we build a prototypical visual approach to demonstrate how domain experts can be integrated.

224 A. Puchta et al.

3 IAM Challenges

This section defines current or future IAM challenges where domain knowledge of human experts may play a vital role. They can serve as a starting point to deduce requirements for any type of solution trying to tie experts and IAM systems closer together. In Sect. 4 we introduce a proof-of-concept visual solution to tackle some of the herein defined challenges.

For identifying the challenges, existing academic literature as well as practitioners experience within the field of IAM are taken into consideration. We are aware that there are far more challenges than the five proposed by us. However, based on the results of our structured analysis and the domain knowledge of practitioners, we chose the most relevant ones with respect to the necessity to integrate domain experts. An *IAM challenge* in the context of this work is a current or future problem with the need to be solved for IAM. Challenges already being tackled or focusing only on parts of an IAM system (e.g. access control) are not considered in this work. Neither do we consider problems where the inclusion of domain expert knowledge is not vital. To identify challenges, we follow a structured approach introduced in the Sect. 3.1.

3.1 Approach for Identifying Challenges

We derive current challenges following a structured approach depicted in Fig. 1. We ensure to include both the scientific and the practitioners' view as IAM is an active research field as well as a highly relevant topic in enterprises.

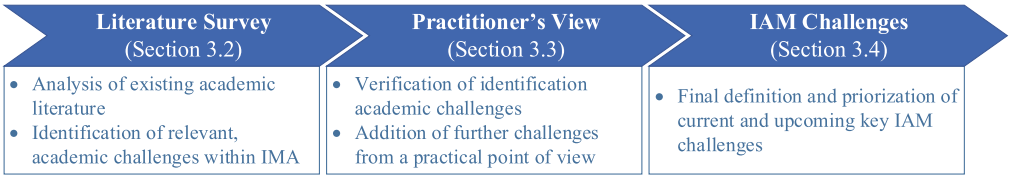


Fig. 1. Approach for defining the key IAM challenges.

During the literature survey we analyze existing academic literature published in the last ten years regarding IAM and respective challenges or problems. This scientific viewpoint allows us to derive a first set of IAM challenges. However, IAM is highly business-driven and there are numerous practical approaches outside the academic world. Thus, we also include the perspective of practitioners.

The goal of this second analysis is twofold. We verify the academic IAM challenges but also identify further challenges not yet considered by scientific literature. Three different sources of information are leveraged in order to minimize subjectivity of different business opinions:

1. Analyst reports and surveys from the IAM industry

2. Interviews with IAM consultants with 3 to 15 years of experience
3. Interviews with companies applying IAM solutions.

In a last step, we integrate all inputs from the analysis into five IAM challenges. The list of identified challenges is not exhaustive for the IAM field of research. Our work is focusing only on current challenges that can strongly benefit from integrating experts' knowledge.

3.2 Literature Survey

In order to identify relevant literature, we follow a structured approach by defining keywords to review relevant IAM literature. As we are defining challenges for the entire IAM system we only take resources into consideration which are dealing either with *"identity and access management"* or specific problems and challenges within *"identity management"* or *"access management"*. We transform these phrases into suitable search terms[1,2] and applied them to the dblp computer science bibliography[3]. Dblp is a service indexing relevant academic journals and proceedings of peer-reviewed conferences from computer science. Searching dblp results in a feasible number of results with a high suitability. Therefore, we can ensure to get only relevant academic publications. Dblp serves as a quality gate for our scientific analysis as it returns a manageable amount of entries compared to other engines like Google Scholar with nearly 10.000 results for the second search term. We manually filter the results based on title, abstract, and key sections to remove findings not mentioning any challenges or problems.

We apply a second, more unstructured search to identify additional relevant entries. In this step we include further academic databases (IEEE XPlore, ACM, Google Scholar) to find additional literature not listed within dblp. This results in a total of 19 academic publications mentioning or clearly defining relevant challenges for IAM. We group the identified problems and define the first four challenges (cf. *C1* to *C4* in Sect. 3.4).

3.3 Practitioner's View

We now conduct a business analysis to include the practitioners' point of view. Information received in this process step is often hard to generalize as it reflects subjective opinions. However, by including various sources of information we try to overcome this deficit. In the business analysis we look at reports from specialized IAM analysts namely KuppingerCole[4]. This company is focused on IAM and technologies around that sector and thus has accumulated valuable knowledge in this area [26,40,43]. Additional input is generated by Gartner [9], Forrester [7] and IDG Research Services [20].

[1] *identity—access management challenge/problem.*
[2] *identity-and-access-management.*
[3] https://dblp.uni-trier.de/.
[4] https://www.kuppingercole.com.

Furthermore, we conduct interviews with three different IAM consultants with several years of practical experience in the field of IAM projects. Besides that, four companies already applying IAM solutions are inquired regarding possible challenges. While the four previously defined challenges are verified throughout the interview, a fifth one (*C5*) arises as a current problem of IAM from a business viewpoint.

Table 1. Results of literature survey on ten years of academic work.

Source	Year	C1	C2	C3	C4	C5
Hovav and Berger [15]	2009	x	x			
Mahalle et al. [27]	2010	x	x	x		
Bandyopadhyay and Sen [2]	2011	x	x	x		
Jensen [22]	2012		x	x	x	
Kanuparthi et al. [23]	2013	x	x		x	
Fremantle et al. [12]	2014	x		x		
Xiong et al. [46]	2014		x			
Hummer et al. [18]	2015	x		x	x	
Kunz et al. [24]	2015				x	x
Hummer et al. [19]	2016	x		x	x	x
Moghaddam et al. [29]	2017		x			
Servos and Osborn [38]	2017		x			
Asghar et al. [1]	2018		x			
Damon et al. [8]	2018	x		x		
Hummer et al. [17]	2018	x		x	x	
Indu et al. [21]	2018	x	x	x		
Nuss et al. [31]	2018	x		x		
Povilionis et al. [34]	2018		x		x	
Kunz et al. [25]	2019			x	x	x

3.4 IAM Challenges

Within this section the identified IAM challenges are described in detail. A mapping of all relevant academic publication to the challenges is provided in Table 1. Table 2 maps the results of our analysis with practitioners to the challenges.

Challenge 1 - Identification of All Relevant Identities (C1): For current and future IAM systems the identification of all relevant identities may sound like a simple task. However, especially in practical application it is not. One of the major reasons for this is the integration of various types of identities

Table 2. Analysis results from practitioners' view.

Source	Year	C1	C2	C3	C4	C5
IDG Research Services [20]	2017	x		x		
KuppingerCole and CXP Group [26]	2017	x				
Tolbert [43]	2017	x		x		
Diodati et al. [9]	2018	x		x		
Small [40]	2018	x		x	x	
Cser and Maxim [7]	2018	x	x	x		
Interviews (IAM consultants)	2019	x	x	x	x	x
Interviews (Companies applying IAM)	2019	x		x	x	x

into IAM. Currently, mainly employee and contractor identities are maintained in an IAM system. A recent trend, customer IAM or shortly CIAM, strives to add customer identities into these systems as well [7]. Additionally, the Internet of Things requires integrating even more identities, mostly technical ones [31]. Furthermore, numerous IT systems are not even connected to IAM. Nevertheless, such systems also contain various identities with the need to be identified for IAM in order to prevent identities not being centrally manageable. These trends hinder IAM to establish a central view of all relevant identities. However, this view is vital for any further analysis to be done within IAM (e.g. identification of unnecessary accounts or entitlements).

Challenge 2 - Privacy Within IAM (C2): As modern IAM systems offer a centralized view on nearly all employees, contractors and even costumers including their attributes the need for privacy arises. Especially business solution power users like IAM administrators can easily retrieve personal information from the identities. Based on our practical experience this could be a simple mail address but may also uncover more sensitive information like wage brackets or entitlement usage information. In order to protect this information in compliance with regulations, privacy mechanisms are needed to grant access to such information only when necessary and for authorized users. This challenge is mainly focused by scientific research and not by practitioners at the moment. However, as the European General Data Protection Regulation (GDPR) came into effect in 2018, it certainly will have an impact on the business sector of IAM. Please note that this challenge is limited to the application of privacy mechanisms on IAM systems and does not include the application of IAM systems for enhancing GDPR compliance within companies.

Challenge 3 - Heterogeneity of Various Identities (C3): As there are various identities within an IAM system, they are not identical. In fact, they differ quite a bit as identities consist of various attributes (e.g. first name, department). Considering *C1*, it gets clear that not all identities have the same kind

of attributes. Technical and human identities are likely to have a completely different set of attributes. For example, technical devices do not have a first name, but instead have an attribute indicating their software version. This, on the one hand, rises a technical challenge to integrate this variability of identities into one underlying data set for IAM. In addition, IAM mechanisms like provisioning of entitlements still need to be working for all of these identities. On the other hand, it also hinders the analytic part of IAM as domain experts need to browse through an enormously large, heterogeneous database. By applying VA, domain experts could be supported as various attributes can be displayed in a more accessible way than in currently deployed table-based reports.

Challenge 4 - Data Quality and Data Management (C4): When it comes to attributes and other data existing in IAM, data quality and the underlying data management in IAM system needs to be considered. Attributes are often manually entered by different people; thus, wrong or inconsistent values are very likely to occur. For example, the current business location of an employee may be added by HR employees. If the employee moves to another department of an enterprise, the location also needs to be changed. Manual processes for attribute modifications exacerbate data quality issues as one can forget to adjust the location attribute. Therefore, IAM mechanisms like provisioning of entitlements based on the attribute *location* might fail. Additionally, wrong attribute values limit the possibilities of IAM analytics. Although an approach to improve attribute quality management was lately introduced [25], algorithms can only detect anomalies but can neither confirm nor reject whether it is a real data error. To do so, domain experts are needed, and VA can be highly beneficial to support related decisions by integrating domain expert feedback.

Challenge 5 - Transformation from Role-Based IAM to Attribute-Based IAM (C5): Challenge 5 was identified during the interviews with IAM consultants as it is not explicitly defined as an upcoming challenge in academic literature. It comprises the enterprise IAM transformation from a role-based approach to an attribute-based one. As mentioned before, enterprises mainly depend on an RBAC approach. However, this can lead to an increasing number of existing roles and requires increasing effort regarding role management [10]. In order to overcome these limitations, ABAC can be applied [16]. However, as this is a fundamental change of approach for IAM companies have to consider various factors (e.g. processes, technologies and policies [13]). Changes needed for this transformation are therefore not limited to access control, but existing research is mainly focused on the transformation of the access control model [36,47]. To the best of our knowledge there is no overarching approach how an enterprise IAM can be transformed from a role-based approach to an attribute-based one.

Tables 1 and 2 compare the results and show that *C1*, *C3*, and *C4* are found in both worlds and can easily be identified as relevant IAM challenges. Privacy in IAM and therefore, *C2*, is mainly embraced by academic literature and not explicitly mentioned in the business sector. *C5* is not described explicitly in

academic literature but only mentioned very shortly by 3 articles. We identified this challenge by conducting interviews with IAM consultants and companies.

4 Applying Visual Analytics to IAM

Any of the previously identified challenges can benefit from including domain experts and their knowledge. VA has proven its capabilities to help integrate domain expert knowledge in complex and data-intensive cyber security tasks throughout the last years [6,44]. Additionally, decision makers can be supported with VA by making highly technical data sources more accessible. Therefore, we argue that leveraging concepts from VA to solve the identified challenges in IAM is a reasonable approach. As described in Sect. 2, there is some existing work that has shown the feasibility and utility of VA in the context of IAM. However, none of the challenges identified in Sect. 3 has been explicitly tackled with visual approaches yet. We try to fill this gap as we describe the architecture and design of our new visualization approach. The visualization design cannot support all the identified challenges as they are far too different in requirements. However, our approach shows how heterogeneous information about human and technical identities can be integrated into a single visual representation. The resulting view allows identifying existing identities (c.f. C1) and their structures (c.f. C3) as well as users can detect problems regarding data quality (c.f. C4).

The visual representation is designed and implemented in close cooperation with IAM practitioners which were also part of our interviews during the challenge identification. By including them in development, we ensure that the representation that is helpful for practical use. The participating experts are IAM consultants working for numerous clients and with years of experience in practical work with IAM projects. While the current visual tool is at a proof-of-concept stage, we are planning to continue our fruitful cooperation with these experts to develop a solution that can be used in their day-to-day work. Our cooperation also allowed for the development of the prototype based on the adaption of anonymized real-world identity data from a medium-sized company in the manufacturing sector with around 1.200 employees.

The underlying architecture for our prototypical application is depicted in Fig. 2 and its main components - *Data Sources*, *Data Preparation*, and *Data Visualization* - are described in more detail throughout the following sections. This architectural design is based on the Information Visualization Pipeline [4] which is a widely accepted structural design concept for any interactive visualization approach. The applied architecture shows how identity-related information can be collected and integrated from different sources and how the information needs to be prepared for VA concepts supporting domain experts. The identification of different data sources and their integration into a single, displayable data set are a starting point for any visual representation of identity data. Therefore, the main part of our architectural design, the *Data Preparation*, demonstrates how visual representations in general can be integrated into an existing IAM structure. The operations executed during the *Data Integration Engine* and the *Data*

Transformation step need only small adjustments for varying *Data Sources*. The last part of the architectural design, the *Data Visualization*, demonstrates how VA can contribute to the focused challenges by introducing an exemplary visualization of identity information. Interaction in this step is crucial as it ensures that experts can adjust the view for their personal needs and explore the data based on their own preferences to gain insight.

4.1 Data Sources

Our current proof-of-concept tool collects information about identities from three main data layers. Although the company representing the use case has a central IAM system with a role-based access control mechanism, not all information about the existing identities is fed into it. Only partial information from the *Application Layer* is integrated into IAM, while other applications, like the company's Active Directory (AD) to manage windows accounts, are not connected to it. Therefore, information from these systems needs to be collected separately. Technical identities representing IoT devices are currently not integrated into IAM but rather maintained separately (IoT-layer). The wide variety of different data sources storing information about the companies' identities and the missing integration of this information into IAM are the main reasons for the challenges we focus in this prototype. It becomes increasingly hard for any company to keep track of its identities when the information about them is so spread out. Additionally, the different information systems store the available data in different formats or data models. Furthermore, it is very important for any company to keep the quality of their identity information at a high level. Spread out data makes this very difficult, especially when data is maintained redundantly in different repositories.

Additional data sources can be plugged easily into our architecture via the *Data Integration Engine*. Our proof-of-concept system works with one source from each layer. This number of data sources is already enough to demonstrate

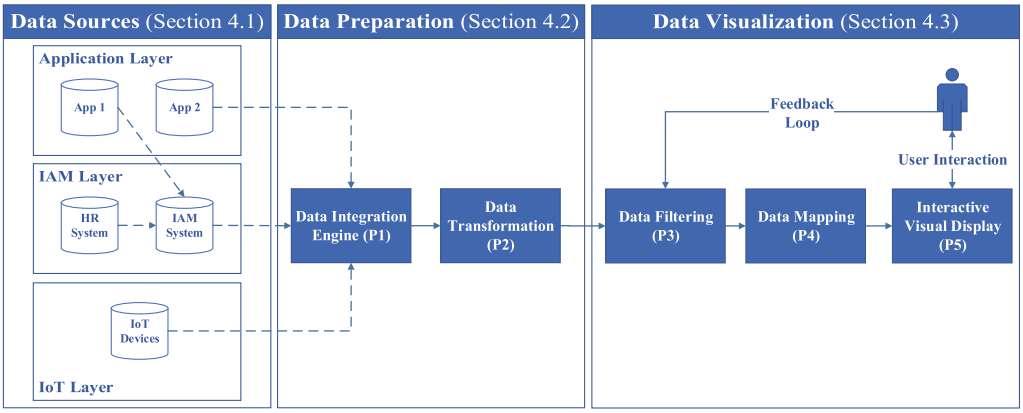


Fig. 2. Architecture for IAM Visual Analytics.

how VA helps to leverage experts' domain knowledge in the context of the aforementioned challenges as is demonstrated in Sect. 5.

4.2 Data Preparation

The main purpose of this part of the architectural design is to integrate and normalize the data from the sources into a single data model and format. Additional fields are added and calculated in this step. The resulting data is structured as a single table containing all relevant and necessary information about the identities. This step is essential for further visual display as it defines the level of detail available to the users. The operations applied to the data in this step are dynamic and can be changed whenever different information is of interest or new data sources are plugged into the architecture.

Data Integration Engine (P1): This part of the architecture is responsible for collecting data relevant from the data source and integrating them into a single cohesive data set. Our proof-of-concept work extracts CSV data from all data sources. However, each CSV export contains a different set of attributes. To preserve the information about the source of a data set, we annotate the data with a flag depicting the source. Additionally, we add a field describing whether the identity is a human or a technical identity. In our conceptual setting, this identity type is mainly dependent on the data source. For example, identities extracted from IAM are automatically annotated to be *"Human"* as only employees or costumers are integrated into IAM. In the same way, identities extracted from the IoT layer are annotated to be *"Technical"* identities. The cohesive data set is built as a union of the three data sets depicted in Fig. 2: $ApplicationLayer \cup IAMLayer \cup IoTLayer$.

Data Transformation (P2): After integrating all available data sources into a single, high-dimensional table, this data set is structured as needed for the visualization in this phase. This part of the architecture applies a variety of transformations. These include splitting a single field into multiple fields, replacing values in a specific field, calculating additional fields based on existing information. The result of this step is a cohesive data set containing all relevant and necessary information.

4.3 Data Visualization

This last part of the architecture is responsible for creating the interactive visual representation with the subset of the data selected by the domain expert. The interactions available to the users enable exploratory work with the visualization and the identification of inconsistencies, miss-configurations as well as structures and dependencies in the set of identities.

Data Filtering (P3): The interactive filtering assures the efficiency of the following steps and guarantees the expressiveness of the resulting view for the user. The subsequent *Data Mapping (P4)* can be CPU-intensive for very large data sets and, therefore, the input for this step needs to be as small as possible. It only contains the fields (columns) the user wants to see. The interactive selection of fields relevant for the user and the early integration of this interaction into the architectural design ensures that only relevant data is passed to the subsequent components. Up to this point the proposed architecture is generalized and can be applied to various visualization approaches within IAM. However, the *Data Mapping (P4)* and the *Interactive Visual Display (P5)* are highly dependent on the visualization technique selected for a specific VA solution. Therefore, the following considerations are specific to our exemplary solution.

Fig. 3. Screenshot of the prototype available under http://bit.ly/iam-vis. Please note that the current version of the tool is only working in Google's Chrome Browser.

Data Mapping (P4): This phase in the architecture maps the filtered identity information into a dynamic hierarchical data structure which is necessary for the prototype to visualize the data correctly. We will not elaborate this data structure any further as it is specific for the proof-of-concept visualization and is prone to change for different visualization types.

Interactive Visual Display (P5): Before we are able to build a visual representation for the data at hand, it is necessary to choose a suitable visualization technique. This technique needs to be capable of displaying the dimensions and structure of the underlying data properly. For our prototypical visualization the

technique must be able to represent multi-dimensional and hierarchical data. While there is a number of techniques (e.g. tree diagrams, circle packing, sunburst diagrams, or treemaps) which fulfill this requirement [39], each technique has its own advantages and disadvantages. It mainly comes down to the use case as well as the subjective preferences of the users which technique is most suitable. We used design sketches of the different suitable visualization techniques to interview IAM domain experts about their preferred visual representation. These interviews resulted in the *sunburst diagram* to be the most preferable technique to apply in the proof-of-concept application. However, any of the mentioned as well as a number of other techniques might be suitable, too.

The sunburst diagram displays a hierarchy using a series of concentric circles. Each ring itself corresponds to a level of the hierarchy. Therefore, underlying data structure is similar to a tree where the root node is depicted by the central ring and outermost circles represent the leaves of the tree-like structure. The sunburst's rings are sliced up and divided based on their hierarchical relationship to the parent slice. Therefore, the sunburst highlights structural, hierarchical relationships while being more scalable than other hierarchical visualization types. Figure 3 depicts the main view of our proof-of-concept application consisting of three main parts.

In (A) experts can drag-and-drop the boxes for the corresponding fields in the data set they want to be depicted in the sunburst diagram between two main lists. Each box represents an IAM employee attribute of the normalized data set. Exemplary fields which are contained in the proof-of-concept are the "Organisational Unit", "Data Source", or the number of entitlements of an identity. The upper list holds the currently active (i.e. displayed) fields and the lower one the inactive attributes. The first element in the list of active elements serves as the root element (innermost circle in the sunburst) when constructing the hierarchical data set. Accordingly, the second active attribute is displayed as the next outer circle. Logically, the last active element is included in the Sunburst diagram as the outermost circle.

The central part of the view is dedicated to the sunburst diagram (B). Each segment of the circles (attributes) depicts a single characteristic of an attribute in relative size to all existing identities. Hovering a segment brings up the number of identities depicted by this segment. The relative frequency with respect to the number of current root element (innermost circle) and the path to the hovered segment are displayed on top of the visual representation. Left-clicking a segment allows zooming in on this particular representation of an attribute. This improves the readability of specific hierarchy levels in very granular sunburst displays. When zoomed into a segment, clicking the white area in the middle of the sunburst diagram brings the zoom one hierarchy level upwards. Right-clicking a segment brings up a dialog. This dialog holds a table with identities in the rows and all attributes available for them in the columns. The identities displayed in the table are dependent on the clicked segment of the sunburst diagram as only the identities whose attributes fulfill the path to the segment are shown in the

table. The table allows filtering and sorting of the currently displayed identities. In the dialog identities can also be reported for further analysis if necessary.

(C) holds the description of the sunburst as a dynamic list containing all currently visible circles (attributes) and the respective visible ring segments (attributes values). Clicking the magnifier for a list element zooms in into the segment representing this element. A click on the counter badge brings up the table with the identities included in corresponding node in the hierarchy. Within this table identities can be marked for further analysis by using a *"Report"*-button for each identity in the details table-view (e.g. after identification of an anomaly), thus, providing the possibility for integration of domain expert feedback into other applications. However, further functionality beyond this notification is out of scope for this work and needs to be implemented in a following version of the prototype.

5 Exemplary Use Cases

The current prototypical implementation[5] of our visualization for IAM was developed in co-creation with experts as suggested by Staheli et al. [42]. We regularly conducted semi-structured interviews with the participating practitioners to ensure that the implementation fits their needs and requirements. This section explicitly highlights how the visual display can support domain experts. We therefore go through several problems and inconsistencies based on one use case and identified by IAM experts while exploring the data. These had not been noticed before applying the visualization.

As the different problems only become evident in the sunburst diagram with different actively visualized attributes, we added predefined scenarios of the sunburst to our publicly available version of the prototype. Using the drop-down menu in the top right corner, we provide a video showcasing each of the following subsections. We would recommend to look at the corresponding video for each subsection in order to grasp the connection between the IAM problem and the sunburst visualization for identification of the inconsistency.

The exemplary use cases are based on the data set from a manufacturing company with 1.200 employees mentioned in Sect. 4. The company recently introduced an IAM system and connected the HR system as well as some minor applications. However, the Active Directory (AD) is currently not under IAM control because of its complexity as it was one of the company's first IT system growing for two decades. Therefore, some employees are missing an AD account while some AD accounts from former contractors and employees are still active. These orphan accounts are not identified via the IAM system, but they are still active and can be used for malicious activities (cf. Sect. 5.1).

Furthermore, the company made some investments in automating specific process tasks. Thus, two assembly machines and some automated users were

[5] The prototype is available under http://bit.ly/iam-vis. Please note that the current version of the tool is only working on Google's Chrome Browser.

integrated within the AD and were provisioned by an AD administrator. However, there was no communication with the IAM department and, therefore, no access management or integration into the IAM system took place. This results in technical identities with excessive entitlements. As the company is not experienced with such technical identities, the risk for failures (e.g. deletion of data) resulting from misconfiguration is high (cf. Sect. 5.2).

During configuration and assignment of a location to the technical identities some flaws regarding the existing location attribute values were detected. As entitlements shall be assigned automatically within the new IAM system based on a policy, the identification and correction of these values is highly relevant. Otherwise, identities with an incorrect value for their location attribute are not assigned enough entitlements (cf. Sect. 5.3).

5.1 Identities Not Managed Within a Central IAM (C1, C4)

As stated before, some identities within the company are not integrated in the central IAM system. Identifying these is a hard task considering the spread-out information. Our approach integrates applications not connected to the IAM system. Taking a look at the *"Data Source"* attribute the sunburst shows in which layer the respective data originates. Identities in the *"IAM Layer"* segment of the diagram are collected directly from IAM. However, another 17 identities are not managed by the IAM system. Three of them are maintained in the *"AD Layer"* while 14 are gathered from the *"IoT Layer"*. Taking a look at the details view of those 14 identities brings out that they are technical devices. Adding another circle for the *"Identity Type"* to the Sunburst allows an analyst to see that none of the identities within the *"IAM Layer"* are technical devices. So obviously the company has not integrated its technical identities into the central IAM system.

5.2 Identities with an Unusual Number of Entitlements (C3)

Another use case needing the attention of domain experts are identities with anomalous high number of entitlements. The Sunburst Diagram facilitates the identification of relevant entities and a decision how to proceed. Displaying the *"Entitlements"*, *"Identity Type"*, and the *"Function"* a small set of identities becomes visible having more than 76 entitlements. This seems conspicuous as most of the entities in the company have 0 to 25 roles assigned to them. Zooming into the segment with 76 to 100 entitlements a technical device attributed with function *"Support"* becomes visible and an identity from the company's customer *"Brandmark"* has an anomalous number of entitlements. These findings do not indicate an error per se, but it might be necessary to carry out further analyses. By browsing through the Sunburst domain experts are enabled to find various of similar cases. Any identity which might be over-authorized has to be examined, if all the entitlements are still needed (e.g. via recertification). If not, this indicates a serious security breach as identities having excessive permissions are legitimately allowed to access classified resources.

5.3 Poor Data Quality in IAM Data (C4)

The Sunburst diagram allows for identifying data quality flaws via several means. A first possibility is to compare similar attribute fields which originate from different data sources. Exemplary for this used when comparing *"Location (AD)"* and *"Location (IAM and IoT)"*. Information about locations of identities is administered in both the application layer and the IAM layer. Usually the IAM layer which contains the HR should be the master system for attributes like the location. After analyzing the data, some quality issues included in this system become apparent. An example is visible by zooming to the value *"Berlin"* of the *"Location (AD)"* attribute. The IAM layer has in fact 3 different attributes for identities having this value in the AD system, namely the correct value *"Berlin"* but also *"BER"* and *"10249 Berlin"*. Presumably, this value was recorded manually by the HR employees resulting in inconsistent data.

6 Conclusion

The complexity of modern IAM systems is constantly rising (e.g. increasing number of identities, further IAM mechanisms). Therefore, new challenges emerge. Within this work we showed that VA can be integrated into IAM in order to solve some of them. To achieve this, we initially identified five central challenges through a review of academic literature and analysis of the experience of practitioners (*RQ-1*). Thereby, we discovered two challenges especially connected to the identification and management of identities. Furthermore, we expect more challenges within the topics *Privacy* and *Data Quality*. Besides that, there will be the future challenge to transform role-based IAM into an attribute-based architecture for enterprises. We do not claim that our list of IAM challenges is exhaustive. However, we focused especially on problems where the integration of domain expert knowledge is vital. We detected some additional challenges but excluded them as they are not in the scope of the paper (e.g. inclusion of trust management in IAM, identity as a service, compliance with regulations).

Based on these challenges we identified VA as a possible solution as it enables enterprises to integrate domain expert knowledge and utilize their feedback to solve upcoming IAM challenges. We proposed an architecture how IAM by leveraging concepts from VA in order to answer our previously defined *RQ-2*. Additionally, we implemented proof-of-concept visualization according to our architecture and based on real world data (*RQ-3*). By applying VA, we have shown that problems tightly connected to the defined IAM challenges can be identified. However, the implementation should be regarded as a first example how the architecture can be implemented and as proof that VA can support enterprises to solve central IAM challenges. Other visualization techniques might be applied to solve another subset of our identified challenges.

After proposing an architecture for integration of VA and a first proof-of-concept implementation we want to focus further on the process to choose a suitable visualization for the *Interactive Visual Display* component. Additionally, we want to introduce further VA implementations to solve the remaining

IAM challenges. Afterwards, we can orchestrate the single implementations to an overarching coordinated view [14].

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

References

1. Asghar, M., Backes, M., Simeonovski, M.: PRIMA: privacy-preserving identity and access management at internet-scale. In: Proceedings of the 2018 IEEE International Conference on Communications, pp. 1–6. IEEE Computer Society (2018)
2. Bandyopadhyay, D., Sen, J.: Internet of things: applications and challenges in technology and standardization. Wirel. Pers. Commun. **58**(1), 49–69 (2011)
3. Basel Committee on Banking Supervisions: Basel III: International Framework for Liquidity Risk Measurement, Standards and Monitoring (2010)
4. Card, S.K., Mackinlay, J.D., Shneiderman, B. (eds.): Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, Burlington (1999)
5. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.: Visual role mining: a picture is worth a thousand roles. IEEE Trans. Knowl. Data Eng. **24**(6), 1120–1133 (2012)
6. Crouser, R., Fukuday, E., Sridhar, S.: Retrospective on a decade of research in visualization for cybersecurity. In: Proceedings of the 2017 IEEE International Symposium on Technologies for Homeland Security, pp. 1–5. IEEE (2017)
7. Cser, A., Maxim, M.: Forrester - Top trends shaping IAM in 2018 (2018)
8. Damon, F., Coetzee, M.: The design of an identity and access management assurance dashboard model. In: Tjoa, A.M., Raffai, M., Doucek, P., Novak, N.M. (eds.) CONFENIS 2018. LNBIP, vol. 327, pp. 123–133. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99040-8_10
9. Diodati, M., Farahmand, H., Ruddy, M.: Gartner - 2019 planning guide for identity and access management (2018)
10. Elliott, A., Knight, S.: Role explosion: acknowledging the problem. In: Proceedings of the 8th International Conference on Software Engineering Research and Practice, pp. 349–355 (2010)
11. 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 2017 IEEE Conference on Visual Analytics Science and Technology (2017)
12. Fremantle, P., Aziz, B., Kopecký, J., Scott, P.: Federated identity and access management for the internet of things. In: Proceedings of the 2014 International Workshop on Secure Internet of Things, pp. 10–17. IEEE Computer Society (2014)
13. Fuchs, L., Pernul, G.: Supporting compliant and secure user handling - a structured approach for in-house identity management. In: The Second International Conference on Availability, Reliability and Security (ARES 2007), pp. 374–384. IEEE (2007)
14. Heer, J., Shneiderman, B.: Interactive dynamics for visual analysis. Queue **10**(2), 30 (2012)
15. Hovav, A., Berger, B.: Tutorial: identity management systems and secured access control. Commun. Assoc. Inf. Syst. **25**(1), 1–42 (2009)
16. Hu, V.C., et al.: Guide to attribute based access control (ABAC) definition and considerations. In: NIST Special Publication (2014)

17. Hummer, M., Groll, S., Kunz, M., Fuchs, L., Pernul, G.: Measuring identity and access management performance - an expert survey on possible performance indicators. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy, pp. 233–240 (2018)
18. Hummer, M., Kunz, M., Netter, M., Fuchs, L., Pernul, G.: Advanced identity and access policy management using contextual data. In: Proceedings of the IEEE International Conference on Availability, Reliability and Security, pp. 40–49. IEEE Computer Society (2015)
19. Hummer, M., Kunz, M., Netter, M., Fuchs, L., Pernul, G.: Adaptive identity and access management - contextual data based policies. EURASIP J. Inf. Secur. **2016**(1), 1–19 (2016)
20. IDG Research Services: Studies Identity- & Access-Management 2017 (2017)
21. Indu, I., Anand, P.M.R., Bhaskar, V.: Identity and access management in cloud environment: mechanisms and challenges. Eng. Sci. Technol. Int. J. **21**(4), 574–588 (2018)
22. Jensen, J.: Federated identity management challenges. In: Proceedings of the 2012 IEEE International Conference on Availability, Reliability and Security, pp. 230–235. IEEE Computer Society (2012)
23. Kanuparthi, A., Karri, R., Addepalli, S.: Hardware and embedded security in the context of internet of things. In: Proceedings of the 2013 ACM Workshop on Security, Privacy & Dependability for Cyber Vehicles, pp. 61–64. ACM (2013)
24. Kunz, M., Fuchs, L., Hummer, M., Pernul, G.: Introducing dynamic identity and access management in organizations. In: Jajodia, S., Mazumdar, C. (eds.) ICISS 2015. LNCS, vol. 9478, pp. 139–158. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26961-0_9
25. Kunz, M., Puchta, A., Groll, S., Fuchs, L., Pernul, G.: Attribute quality management for dynamic identity and access management. J. Inf. Secur. Appl. **44**, 64–79 (2019)
26. KuppingerCole, CXP Group: State of organizations - does their identity & access management meet their needs in the age of digital transformation? (2017)
27. Mahalle, P., Babar, S., Prasad, N.R., Prasad, R.: Identity management framework towards internet of things (IoT): roadmap and key challenges. In: Meghanathan, N., Boumerdassi, S., Chaki, N., Nagamalai, D. (eds.) CNSA 2010. CCIS, vol. 89, pp. 430–439. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14478-3_43
28. Meier, S., Fuchs, L., Pernul, G.: Managing the access grid - a process view to minimize insider misuse risks. In: Proceedings of the 11th International Conference on Wirtschaftsinformatik, pp. 1051–1065 (2013)
29. Moghaddam, F., Wieder, P., Yahyapour, R.: A policy-based identity management schema for managing accesses in clouds. In: Proceedings of the 8th International Conference on the Network of the Future, pp. 91–98. IEEE Computer Society (2017)
30. Morisset, C., Sanchez, D.: VisABAC: a tool for visualising ABAC policies. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy. Newcastle University (2018)
31. Nuss, M., Puchta, A., Kunz, M.: Towards blockchain-based identity and access management for internet of things in enterprises. In: Furnell, S., Mouratidis, H., Pernul, G. (eds.) TrustBus 2018. LNCS, vol. 11033, pp. 167–181. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98385-1_12
32. Nyanchama, M., Osborn, S.: The role graph model and conflict of interest. ACM Trans. Inf. Syst. Secur. (TISSEC) **2**(1), 3–33 (1999)

33. Pfitzmann, A., Köhntopp, M.: Anonymity, unobservability, and pseudonymity—a proposal for terminology. In: Federrath, H. (ed.) Designing Privacy Enhancing Technologies. LNCS, vol. 2009, pp. 1–9. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44702-4_1
34. Povilionis, A., et al.: Identity management, access control and privacy in integrated care platforms: the PICASO project. In: Proceedings of the 2018 International Carnahan Conference on Security Technology, pp. 1–5. IEEE Computer Society (2018)
35. Samarati, P., de Vimercati, S.C.: Access control: policies, models, and mechanisms. In: Focardi, R., Gorrieri, R. (eds.) FOSAD 2000. LNCS, vol. 2171, pp. 137–196. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45608-2_3
36. Sandhu, R.S.: The authorization leap from rights to attributes: Maturation or chaos? In: Proceedings of the 17th ACM Symposium on Access Control Models and Technologies, pp. 69–70. ACM (2012)
37. Sandhu, R.S., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-based access control models. Computer **29**(2), 38–47 (1996)
38. Servos, D., Osborn, S.L.: Current research and open problems in attribute-based access control. ACM Comput. Surv. **49**(4), 1–65 (2017)
39. Severino, R.: The data visualisation catalogue (2019). https://datavizcatalogue.com/index.html. Accessed 21 Feb 2019
40. Small, M.: Kuppingercole report - advisory note - big data security, governance, stewardship (2018)
41. SOX: Sarbanes-Oxley Act of 2002, pl 107–204, 116 stat 745 (2002)
42. Staheli, D., et al.: Visualization evaluation for cyber security. In: Proceedings of the 2014 IEEE Symposium on Visualization for Cyber Security, pp. 49–56. ACM (2014)
43. Tolbert, J.: Kuppingercole report - advisory note - identity in IoT (2017)
44. Wagner, M., Rind, A., Thür, N., Aigner, W.: A knowledge-assisted visual malware analysis system: design, validation, and reflection of kamas. Comput. Secur. **67**, 1–15 (2017)
45. Windley, P.J.: Digital Identity: Unmasking Identity Management Architecture (IMA). O'Reilly Media Inc, Newton (2005)
46. Xiong, J., Yao, Z., Ma, J., Liu, X., Li, Q., Ma, J.: PRIAM: privacy preserving identity and access management scheme in cloud. KSII Trans. Internet Inf. Syst. **8**(1), 282–304 (2014)
47. Xu, Z., Stoller, S.D.: Mining attribute-based access control policies from RBAC policies. In: Proceedings of the 10th International Conference and Expo on Emerging Technologies for a Smarter World. IEEE (2013)

4 HyperSec: Visual Analytics for Blockchain Security Monitoring

Current status:	Accepted & Published
Conference:	36th International Conference on ICT Systems Security and Privacy Protection – IFIP SEC 2021, 22–24 June, 2021
CORE Ranking:	**B** (http://portal.core.edu.au/conf-ranks/804/)
Date of acceptance:	March 22, 2021
Date of publication:	June 17, 2021
Full citation:	PUTZ, B., BÖHM, F., AND PERNUL, G. HyperSec: Visual Analytics for Blockchain Security Monitoring. In *ICT Systems Security and Privacy Protection*, vol. 625 of *IFIP Advances in Information and Communication Technology*. Springer, Cham, 2021, pp. 165–180
Authors' contributions:	Putz Benedikt 45% Böhm Fabian 45% Pernul Günther 10%

Conference Description: The IFIP SEC conferences aim to bring together primarily researchers, but also practitioners from academia, industry and governmental institutions to elaborate and discuss IT Security and Privacy Challenges that we are facing today and will be facing in the future. The conference seeks submissions from academia, industry, and government presenting novel research on all theoretical and practical aspects of security and privacy protection in ICT Systems.

HyperSec: Visual Analytics for Blockchain Security Monitoring

Benedikt Putz( ), Fabian Böhm, and Günther Pernul

University of Regensburg, Regensburg, Germany
{benedikt.putz,fabian.boehm,guenther.pernul}@ur.de

Abstract. Today, permissioned blockchains are being adopted by large organizations for business critical operations. Consequently, they are subject to attacks by malicious actors. Researchers have discovered and enumerated a number of attacks that could threaten availability, integrity and confidentiality of blockchain data. However, currently it remains difficult to detect these attacks. We argue that security experts need appropriate visualizations to assist them in detecting attacks on blockchain networks. To achieve this, we develop HyperSec, a visual analytics monitoring tool that provides relevant information at a glance to detect ongoing attacks on Hyperledger Fabric. For evaluation, we connect the HyperSec prototype to a Hyperledger Fabric test network. The results show that common attacks on Fabric can be detected by a security expert using HyperSec's visualizations.

Keywords: Distributed ledger · Permissioned blockchain · Information security · Visual analytics · Security monitoring

1 Introduction

New use cases of distributed ledger technology (DLT) are proposed on a daily basis by academia and practice, leading to an increasing number of projects and solutions. Beyond that, blockchain applications are increasingly being used in real large-scale supply chain environments, such as the TradeLens [10] and DLFreight [20] platforms. At first glance, DLT seems to increase an application's security or even solve existing applications' security issues. However, the task of securing the DLT itself is often neglected in practice due to its complexity and the number of serious challenges connected to it.

The complexity of blockchain technology makes it particularly challenging to identify malicious activities [4]. In any blockchain network, there are several independent peers operated by independent organizations, where each organization only has a limited view of the network. Each node also has various data sources from its components, making it difficult to obtain an overview of the

B. Putz and F. Böhm—Contributed equally to this manuscript.

Published by Springer Nature Switzerland AG 2021
A. Jøsang et al. (Eds.): SEC 2021, IFIP AICT 625, pp. 165–180, 2021.
https://doi.org/10.1007/978-3-030-78120-0_11

network's state [17]. Since blockchain is a networked database, it also requires monitoring both the host and the network, which results in a large volume and velocity of observable data.

Fully automated systems for live attack detection on blockchains do not yet exist. Even if respective technologies for blockchain security monitoring were available, human experts remain indispensable as their domain knowledge is crucial to identify and analyze intricate attack patterns [2]. Therefore, we need a way to make the heterogeneous data at hand available for domain experts. Visualizations offer a well-known path to achieve this goal. A visual representation can help a domain expert make sense of the displayed information and efficiently draw conclusions [12]. These observations lead to our work's research question:

RQ. *What are appropriate visualizations to assist security experts in detecting DLT threats?*

In this work, we make a two-fold contribution to this research question. We first characterize the domain problem: monitoring permissioned DLTs for attacks. This domain problem and derived general design requirements serve as the foundation for our visualization approach. The second part of our contribution is the task-centered design and prototypical implementation of *HyperSec*, a visual representation of security-relevant DLT information to support security experts' monitoring tasks.

The remainder of this work is structured as follows. Section 2 gives a brief overview of related academic work in the field of security visualizations in the blockchain domain. In Sect. 3, we flesh out the domain problem faced by security experts monitoring permissioned blockchain environments for immediate threats. Section 4 then introduces our visualization design and its prototypical implementation using open source technologies. Afterwards, we evaluate our visualization design by simulating attacks in Sect. 5. We discuss how an expert may proceed after an attack has been detected in Sect. 6. Finally, Sect. 7 concludes our work with a summary and possible future research directions.

2 Related Work

Recently, Tovanich et al. [23] carried out a systematic review to structure existing work on the visualization of blockchain data. Their research and previously conducted studies [19] identify several visualization approaches with a focus on criminal and malicious activity [8,13]. These surveys highlight that visualization tools for blockchains are on the rise. However, most of these existing visualization approaches for criminal or malicious activities in blockchains focus on historical analysis, i.e. detecting the events only after they have occured [23].

To effectively prevent attacks upfront, blockchain networks have to be actively monitored by blockchain security experts. Several studies discuss external and internal threats that could impair a blockchain network's functionality [9,18].

Zheng et al. propose a framework for monitoring the Ethereum blockchain's performance [24]. They introduce some respective metrics while using both node logs and Remote Procedure Calls (RPC) to gather data. Threat indicators to detect malicious activities in a blockchain network have recently been introduced by Putz et al. [17]. Based on this limited body of work from academia, blockchain metrics and threat indicators need to be made available to security experts for effective monitoring. Existing monitoring solutions like the dashboard by Bogner [3] focus only on transaction activity but do not consider other security-relevant data and metrics.

An approach pointing in this direction is the Hyperledger Explorer [21], the Hyperledger project's tool for monitoring Hyperledger blockchains. The Explorer connects to a local blockchain node and extracts data about blocks, transactions, peers, and more into a local PostgreSQL database. Additionally, a web application is available for inspecting blockchain data, including some basic visualizations of transaction data. However, these visualizations are not tailored to provide the necessary insights or indicators to detect threats. In addition, there is a Hyperledger Labs project integrating Fabric with ElasticSearch and Kibana, resulting in a Kibana dashboard able to display some transaction data [1]. Unfortunately, their visualizations are not very well suited to detecting blockchain threats in Hyperledger Fabric. In our experiments we found that the necessary integration and aggregation of additional data sources and custom visualizations are difficult to achieve in standard products like the Elastic stack.

Analyzing related work highlights an evident lack of dedicated and security-specific visualization approaches enabling security experts to monitor blockchain networks in real-time, while detecting common indicators of compromise or ongoing attacks on the network. Our work contributes a valuable solution approach to this issue.

3 Blockchain Security Monitoring

This Section addresses the first part of our contribution. Within our main contribution, we follow the user-centered and problem-driven Nested Blocks and Guidelines model (NBGM) for visualization designs [14,16]. This allows us to identify and address security experts' core problems and lay a foundation for a visualization design fitting their needs.

The first step of the NBGM is the definition of a domain problem. We characterize the problem at hand based on two primary sources of information. First, we consider reports from blockchain security professionals [11]. Second, we analyze literature on blockchain attacks to identify concerns for operators of a blockchain node [7,9,18]. We begin by outlining the overall blockchain security monitoring process in Sect. 3.1. The domain problem is then specified according to Miksch and Aigner's design triangle through more in-depth descriptions of specific users (Sect. 3.2), their tasks (Sect. 3.3), and data elements (Sect. 3.4) [15]. We address the second step of the NBGM (*Data/Operation Abstraction*) in

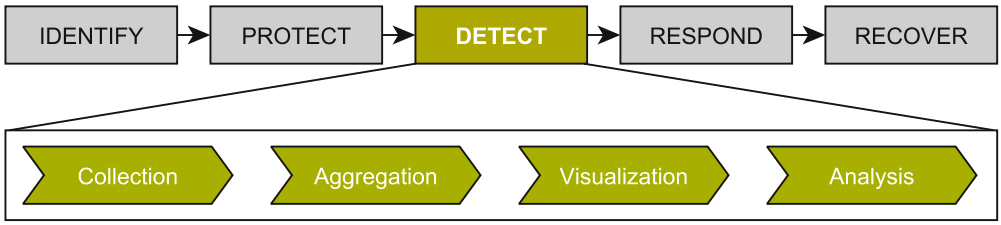


Fig. 1. Blockchain Security Monitoring process based on the NIST Cybersecurity Framework [5].

Sect. 3.5 by deriving general design requirements for a visualization approach to support blockchain security monitoring.

3.1 Blockchain Security Monitoring Process

Before we dive into users, tasks, and available data, we first need to understand the overall process underlying blockchain security monitoring. This subsection introduces our conceptual process based on the NIST Cybersecurity Framework for protecting critical infrastructures [5]. As shown in Fig. 1, the framework has five main functions: *Identify*, *Protect*, *Detect*, *Respond* and *Recover*. We apply these functions to a permissioned blockchain network. The *Identify* function serves to identify relevant assets and risks. This problem has been already addressed in prior work [17]. *Protect* involves a variety of protection measures applied to the system: identity management and access control, data security, secure configuration, and backups/log files, among others. These protection measures are usually part of the blockchain framework itself, with additional measures being applied at deployment time (such as secure configuration and appropriate backup procedures) [22]. The *Detect* function currently lacks appropriate visualization and analysis tools. It's the focus of this work and further developed in the following subsection. During the *Respond* phase, threats detected using our visualization approach are met with a response plan and appropriate mitigation actions. Finally, the *Recover* function provides appropriate tools to restore functionality after an attack has occurred. *Respond* and *Recover* are not specifically part of this work as attacks need to be identified before effective *Respond* and *Recover* can take place. Corresponding tools might be integrated into future work to permit swift threat response.

The *Detect* function can be subdivided into four smaller process steps. Relevant data needs to be collected (*Collection*) and aggregated to provide appropriate metrics if necessary (*Aggregation*). Data and metrics can then be visualized (*Visualization*) allowing domain experts to identify possible threats (*Analysis*). Please note that all steps beside *Analysis* can be performed automatically.

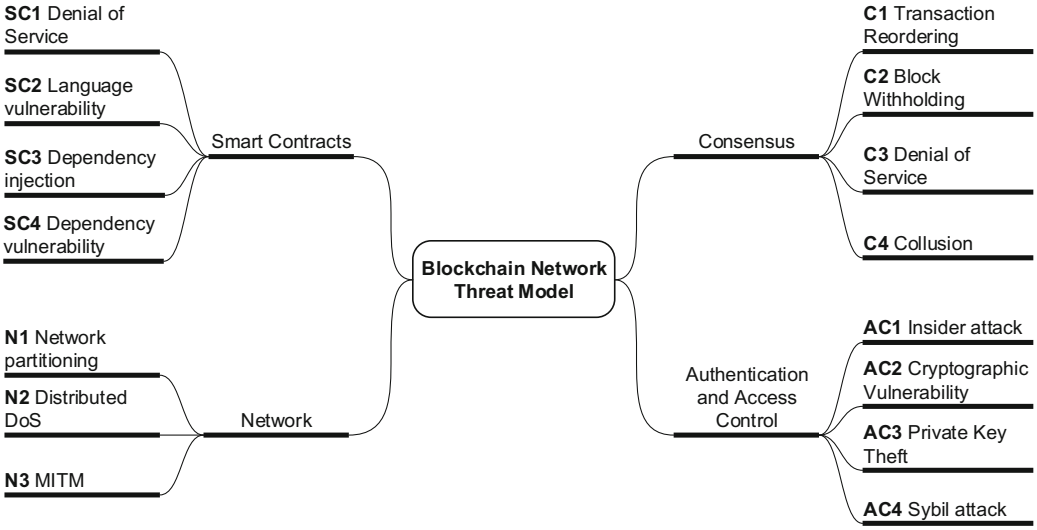


Fig. 2. Blockchain Networks Threat Model in attack tree notation.

3.2 Users

The intended users of visualization designs within the *Detect* function in the blockchain monitoring process are domain experts. These experts are responsible for analyzing blockchain data to identify malicious events within this function [11]. More specifically, we define the domain experts as security professionals knowledgeable in the cybersecurity domain. Therefore, we expect them to have the expertise to decide whether specific events or event series indicate an imminent threat to the blockchain. Within a permissioned blockchain, these security experts are responsible for monitoring the distributed network through the view of the local organization's blockchain node. Other organization's nodes could also be monitored, but data availability is likely limited due to access restrictions within the blockchain network.

3.3 Tasks

Visualizations or any other tool supporting the *Detect* function of the blockchain security monitoring process should be based on the tasks that the respective users need to carry out. Following the user characterization above, we derive the crucial tasks of the domain expert's work.

To illustrate the monitoring task's complexity, we show an overview of possible attacks in an attack tree notation in Fig. 2. The listed attacks are based on prior work [17,18] and related literature surveys [7,9]. For each leaf on the tree, there are various ways to successfully deploy the attack, which we did not include for conciseness. The attack tree focuses on the blockchain network and nodes. Therefore, it does not include application-specific attacks such as web

Table 1. Security expert tasks and related attacks (cf. Fig. 2).

Task	Description	Related attacks
T1	Identify vulnerable smart contracts	SC1, SC2, SC3
T2	Identify blockchain framework vulnerabilities	SC4, AC2
T3	Inspect log files of running services on demand	SC4, N1, N3, C3, C4
T4	Review networking activity	N1, N2, N3
T5	Compare transaction metrics over time	N2, C2
T6	Explore block and transaction history	SC1, SC2, C3, AC1
T7	Review configuration changes	C1, AC1
T8	Detect identity abuse	AC1, AC3, AC4

application vulnerabilities. Each of the shown attacks is indicated by different combinations of threat indicators [17]. Security experts need to identify threats based on these indicators as part of the *Analysis* process step. Visualizing the indicators provides the necessary overview to identify vulnerable components for in-depth analysis. Therefore, domain experts' overarching task is the *analysis of blockchain data to identify possible threats*, which is to be supported by visualizations. To allow domain experts to execute this work adequately, we have identified more specific tasks based on the attacks and corresponding threat indicators from prior work [17]. These tasks are shown in Table 1.

Each task comprises several sub-tasks that help accomplish the main task. To identify vulnerable smart contracts (*T1*), the expert may manually inspect smart contract code or scan smart contracts for vulnerabilities and inspect scan results. Identifying framework vulnerabilities (*T2*) can be accomplished by reading release notes for the framework and its dependencies. Since many anomalies can have multiple causes (i.e., low transaction throughput), log file inspection (*T3*) helps to identify the root cause of anomalies. To review networking activity (*T4*), the main indicators are the count of active connections to other peers, the activity level of those connections, and last seen times of offline peers. Transaction metrics (*T5*) include throughput, latency and unprocessed transactions. Block and transaction history monitoring (*T6*) implies watching the chain of blocks for inconsistencies such as changed blocks or missing transactions. Reviewing configuration changes (*T7*) includes both active and proposed changes to be able to intervene in case of manipulation attempts. Identity abuse (*T8*) is possible during all phases of an identity's lifecycle, so an expert must monitor issuance, usage in transactions, and revocation.

3.4 Data Elements

Blockchain Frameworks such as Ethereum and Hyperledger Fabric offer a number of data sources for monitoring. The most obvious data sources are blocks and associated transaction data [23]. These can be used to derive active users,

smart contracts, and the general level of activity on the network (i.e., transaction throughput). Numerical data on network activity is also provided through metrics, which can be used to raise alerts for anomalous behavior. On a more technical level, each component of the blockchain node also provides log files. These files give detailed information about smart contract execution, consensus protocol violations, and other node internals. They can be helpful to determine the root cause of an anomaly.

3.5 Design Requirements

To wrap up this first part of our contribution, we derive the following general requirements for visualizations aiming to support the *Detect* function of the blockchain security monitoring process. The requirements are based on the above user, task, and data characterizations. Although we follow these requirements in the remainder of this work to design our prototype, they can serve as a general collection for respective visualization designs. We summarize the requirements under several main views that a Visual Analytics system supporting the domain experts' tasks should comprise:

R1 - General Security Information: A view should allow users to overview a series of general, security-relevant information from the monitored blockchain. Attention should be drawn to any changes on the blockchain's overall configuration (*T7*). Whenever new smart contracts are deployed to the blockchain, they should be checked (automatically or manually) for vulnerabilities. The results of these checks need to be made available for the analysts (*T1*). Additionally, newly discovered vulnerabilities within the applied blockchain framework should be shown to users within this general view (*T2*).

R2 - Network View: Another view should provide access to any data and metrics related to the peers and their network activities. This includes displaying available information about the peers themselves and the respective identities that interact with the blockchain on behalf of the peers (*T8*). This view should also provide visual access to any network-related metrics that assess the overall network's health (*T4*).

R3 - Transaction View: Domain experts need to access a view displaying information about the blocks and transactions being handled by the blockchain. This includes detailed information on the blocks and transactions themselves (*T6*) as well as a time-based view on transaction-related metrics allowing to identify any changes in typical structure and processing of transactions (*T5*).

R4 - Interactivity and Details: Any of the previously mentioned views (R1–R3) needs to be fully interactive to provide the best possible support for domain experts' tasks and enable exploratory analysis. Whenever suspicious actions or threat indicators are identified, experts also need access to further details and underlying log files (*T3*).

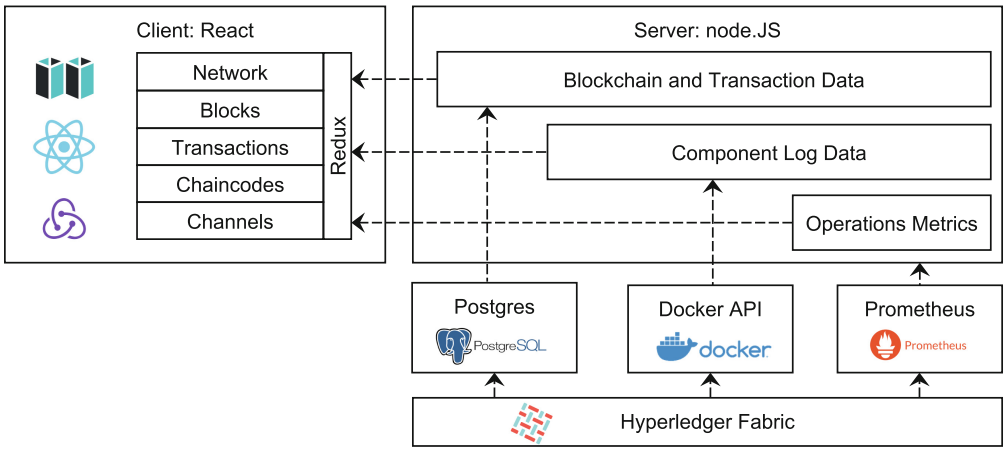


Fig. 3. Prototype architecture and data flows.

4 HyperSec: Hyperledger Security Monitoring Using Visual Analytics

We now introduce our prototype **HyperSec** (**Hyper**ledger **Sec**urity Explorer), a modified version of the open-source project Hyperledger Explorer based on the design requirements introduced in Sect. 3.5. The prototype is open-source and available online, along with a demo deployment[1]. Our modifications address the two remaining layers of the NBGM by designing our solution based on the domain problem and implementing it within a prototype.

4.1 Architecture and Technology

We choose Hyperledger Explorer as a starting point since it already provides a working synchronization architecture based on Hyperledger Fabric's block event subscription. We extend the existing architecture to allow for more comprehensive accessibility of relevant data and effective security monitoring. This results in the architecture displayed in Fig. 3. We keep the basic structure (data sources, server, and client) of the original architecture for interoperability and transparency reasons. However, in our previous study [17] we found that security-relevant information for Hyperledger Fabric must be retrieved from several data sources: the Hyperledger Fabric SDK, operations metrics, and the application logs available via Docker. Block data is already stored in Hyperledger Explorer's PostgreSQL database. We integrate additional metrics and log sources through server-side proxies to the respective Prometheus and Docker APIs. The React client accesses these through the API exposed by the Hyperledger Explorer server.

[1] https://github.com/sigma67/hypersec.

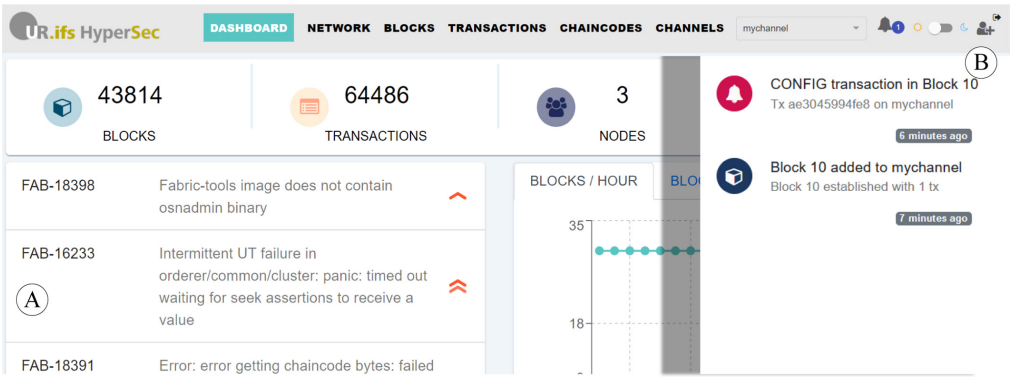


Fig. 4. *Dashboard* view: Security issues, alerts and general overview.

We implement the views defined in Sect. 3.5 by adapting existing views from the Hyperledger Explorer project. This allows us to retain the frontend structure while introducing new monitoring capabilities. Therefore, domain experts do not need to work with a completely new interface but rather get additional relevant information on the respective views. The updated views host a series of interactive visualizations based on the *visx*[2] visualization primitives for React. They all follow a similar structure: relevant data is retrieved from the client's *Redux* state handling, transformed for use in the visual display, mapped into visual primitives, and finally rendered [6].

4.2 Visual Representations and Interactions

We now go into more detail on our HyperSec prototype's visual representations addressing the requirements *R1–R3* and their interactivity (*R4*). As mentioned before, we integrate the visualizations into existing Hyperledger Explorer views to retain the familiar structure for domain experts. This Section is structured accordingly to the naming of the original Hyperledger Explorer views.

Views *Dashboard* and *Chaincodes*: To fulfill the Design Requirement *R1*, we adjust two views of the Hyperledger Explorer. First off, directly on the Explorer's landing page, called "Dashboard", we show a list of known Hyperledger Fabric issues of High/Highest importance from the Hyperledger JIRA[3] ordered by last updated (Fig. 4**A**). Any list item can be expanded to reveal additional information about the issue. Although there is no issue category directly reflecting security issues, this information is highly relevant for *T2 – Identify blockchain framework vulnerabilities*. Additionally, there is no other source for the respective information. In the side menu (Fig. 4**B**), an alert appears whenever the configuration of the monitored Hyperledger Fabric blockchain is changed (*T7*).

To allow domain experts to detect vulnerable chaincodes, we include available security scans in the respective "Chaincodes" view. Whenever a smart contract

[2] https://airbnb.io/visx/.
[3] https://jira.hyperledger.org.

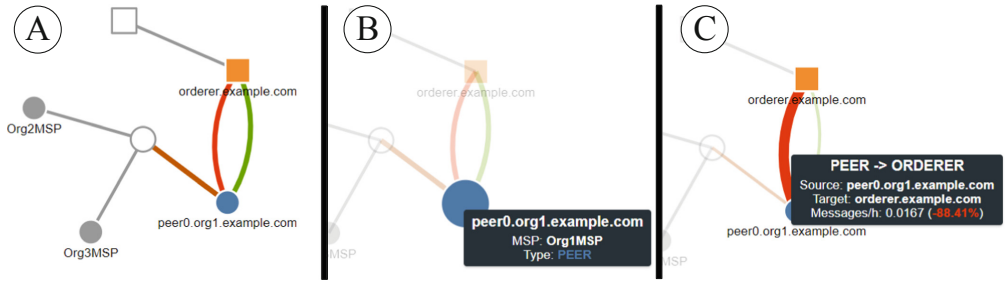


Fig. 5. *Network* view: Interactive visualization of network traffic between peers and orderers.

went through a security scan, analysts can directly check this scan's results in the HyperSec prototype (*T1*). We use the open-source static analysis tool revive-cc[4] to detect security vulnerabilities and store the scan result in the Hyperledger Explorer PostgreSQL database. To ensure the scans are up to date, we set up automated jobs to regularly generate security reports of deployed chaincodes.

View *Network*: The *Network* view targets design requirement *R2* intended for tasks *T4* and *T8*. The original Hyperledger Explorer shows a tabular list with basic information about the peers connected to the monitored Hyperledger Fabric network. In our HyperSec prototype, we extend this table with a force-directed node-link diagram to effectively visualize networking activity (Fig. 5**A**). The nodes' different shapes indicate different peer types within the network: Circles are used to display peers while rectangles represent orderers. Links between the glyphs are used to display known networking activities.

However, the unavailability of core information restricts this view's expressivity. While rich information about the peers can be easily retrieved from the Hyperledger Explorer, no data about the peers' network connections is provided. Therefore, HyperSec retrieves networking information directly from Hyperledger Fabric through the Prometheus API. By doing so, experts get at least some information about the peers' networking activity within the own Membership Service Provider (MSP). However, as the Hyperledger Fabric network is decentralized, it is not possible to get any information about other MSPs' networking activities. Because of this restriction, we introduce two empty nodes in the node-link diagram (uncolored nodes in Fig. 5**A**), which mark the border of the monitoring visibility regarding networking activities. Nodes within the owned MSP are colored; those within other MSPs are greyed out.

Links connecting the nodes in the graph represent known network connections. Again, outside the own MSP's borders, experts do not get much information. Therefore, we connect any foreign peer and orderer to the respective artificial node. The coloring of the links follows a continuous scale from −1 to 1. This scale measures the current deviation of the link's message traffic from the average, by comparing traffic in the last hour with traffic in the previous seven

[4] https://github.com/sivachokkapu/revive-cc.

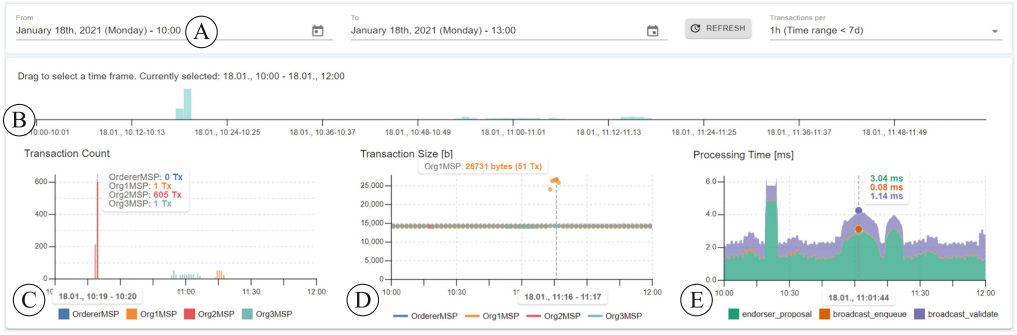


Fig. 6. *Transactions* view: Interactive visualizations for transaction count, size and processing time.

days. If this deviation is low, the link is colored in a green tone. A red link, in contrast, marks a high variation of message numbers.

The node-link diagram is fully interactive. Nodes are draggable to ensure that security analysts can adjust the layout to their own needs if necessary. Hovering over nodes (Fig. 5**B**) or links (Fig. 5**C**) highlights the hovered object and shows additional status information about it.

View *Transactions*: This view (*R3*) satisfies tasks *T5* and *T6*. Some modifications to the original simple table view ensure that the transactor identity and transaction size are visible. The primary adjustment we made to this view is introducing four visualizations (Fig. 6). To ensure a high performance of the visualizations even when dealing with several thousands of transactions, we implement an efficient data bucketing algorithm which allows easy and fast look up of relevant transaction data (see Algorithm 1).

We make small adjustments to the original timeframe selection (Fig. 6**A**). The selection defines the time range for which information about transactions should be displayed. On the right side of the timeframe selection, we added a dropdown menu to select the aggregation granularity (1 min, 1 h, 12 h, 24 h) for the visualizations. This helps security experts if they need to compare and contextualize available information.

The wide bar chart (Fig. 6**B**) always displays the entire selected date range. Each bar represents the number of transactions within a specific range of time specified through the aggregation granularity. This bar chart supports analysts in navigating the selected time range. A brushing interaction (horizontal dragging on the chart) selects an even smaller time range for detailed analysis. On interaction, the other visualizations (Fig. 6**C**, **D**, and **E**) and the transactions table are dynamically updated with data from this narrowed time range.

A stacked bar chart (Fig. 6**C**) visualizes the number of transactions per aggregation window. However, it does this only for the transactions selected through the brushing interaction on the visualization Fig. 6**B**. It shows the transaction count's composition based on which MSP contributed how many transactions. The scatterplot Fig. 6**D** shows the transaction size in bytes throughout the time

Algorithm 1: Transaction data bucketing

Input: Time Window from timestamp t_s to t_e with $t_s, t_e \in T$, $t_s < t_e$, and $T \in \mathbb{R}$. Aggregation granularity $s_b \in \mathbb{R}$

Output: Map M^{tx} with transactions sorted into the respective time-based bucket

function `generateTxBuckets`(t_s, t_e, s_b)**:**
  $L^{tx} \leftarrow getTransactionListForTimeWindow(t_s, t_e)$;
  $M^{tx} \leftarrow$ new $Map()$;
  $t_b \leftarrow t_s$;
  while $t_b < t_e$ **do**
    $i_b \leftarrow \lfloor t_b / s_b \rfloor$;
    $b \leftarrow$ new $Bucket()$; `// Object for transactions and meta-data.`
    $M^{tx}_{i_b} \leftarrow b$;
    $t_b \leftarrow t_b + s_b$;
  end while
  foreach $tx \in L^{tx}$; `// Find correct` $Bucket$ `and update with` tx`.`
  do
    $i_{tx} \leftarrow \lfloor (t_{tx} - t_s) / s_b \rfloor$;
    $M^{tx}_{i_{tx}} \leftarrow M^{tx}_{i_{tx}} \cup parse(tx)$
  end foreach
  return M^{tx};
end

range for each MSP. Each circle on the scatterplot represents the average size of transactions submitted by a specific MSP. This aggregation is performed to scale the chart for large numbers of transactions. During attacks, thousands of transactions can be submitted within just minutes, thus freezing the chart if each transaction were drawn individually. The stacked area chart Fig. 6**E** finally shows the development of three different metrics, which we identify as helpful to get an idea for the processing time in seconds that a transaction needs from proposal to validation. As information for processing times are not available distinctively per transaction but continuously per time unit, we choose to display this metrics with a continuous visualization technique.

The visualizations Fig. 6**C**, **D**, and **E** are again fully interactive. Hovering individual bars or hovering along the continuous sizes and times displays additional information as tooltips. Different metrics can also be toggled using the legend icons below the visual representations.

5 Evaluation

For our evaluation, we focus on three common attacks that cover the majority of the tasks outlined in Table 1: **SC2**, **N2**, and **AC1**. We simulate these attacks a Hyperledger Fabric test network, which the HyperSec prototype is connected to.

SC2 refers to a language vulnerability, i.e. a software bug that exposes chaincode to malicious exploits. A security expert may become aware of such an exploit by identifying vulnerable smart contracts (*T1*) and by inspecting transaction history (*T6*). For example, consider a read-after-write vulnerability detected by the chaincode scanner *revive-cc*. The security expert can inspect an automatically generated chaincode scan in the *Chaincodes* view. Intuitively, the experts check for past exploitations using the *Transactions* view. Thereto, the transactions table can be filtered using the chaincode name and applicable time frame. The filtered transactions can be inspected individually to find unusual read/write sets.

N2 refers to a distributed denial of service attack. If a peer or orderer is targeted by a traffic-based denial of service attack, its connection to other peers will be impaired as well. The *Network* view (Fig. 5**C**) shows high deviation in gossip communication traffic to the targeted peer during such an attack (*T4*). If the local peer is targeted, the metrics in the *Transaction* view (Fig. 6**E**) show increased transaction processing latency due to high peer load (*T5*). For attackers that can send transaction to the network, transaction-based DoS is more effective. Figure 6**C** and 6**D** show two such attempts using high transaction volume (**C**) and large transaction size (**D**). Figure 6**E** also shows spikes in processing latency during the time of attack (spikes 1 and 3 in that chart).

To investigate the source of the anomaly, experts can check the peer logs, which are available in the Network view (*T3*). They cross-reference any error messages with open issues in the Hyperledger JIRA, which are available in the Dashboard view (*T2*).

AC1 refers to an attack where an insider abuses valid credentials for malicious purposes. Consider an insider attempting to corrupt the blockchain network's configuration using a configuration transaction. Security experts are immediately notified about the configuration change in the notification sidebar (*T7*, see Fig. 4). Details of the attempted configuration change are available in the transaction history table (*T6*), where the full read-write set of the transaction is available by selecting the respective transaction.

6 Discussion

The evaluation has shown that the visualizations can assist a security expert in detecting ongoing attacks. If an attack is detected, the next steps in the Cybersecurity Framework (see Fig. 1) are *analysis*, *respond* and *recover* activities, which are discussed hereafter.

Analysis. Based on the present threat indicators the expert then proceeds with *analysis* of the root cause. The logs shown in HyperSec can be a starting point, but may only show symptomatic errors. In-depth analysis of application and network logs on the systems running blockchain components can yield further information. The expert must determine if it is a crash fault or a byzantine fault. At the same time, a communication channel should be available with other

organizations of the consortium to determine if it is a more widespread problem. Guidelines and checklists can help structure this process.

Respond. Once the cause is identified, the expert contacts operations teams to request mitigation actions. Local or network configuration changes can mitigate crash faults and network/consensus threats (see Fig. 2). Compromised smart contracts may require an upgrade, or even a ledger rollback if the consequences were severe. Hyperledger Fabric supports ledger snapshots for this purpose [22]).

Recover. After mitigation of an attack, evidence collection is another subject of interest. System and Docker logs are the primary source of evidence, complemented by ledger transaction data stored in HyperSec's PostgreSQL database. However, the forensic analysis of attacks on Hyperledger Fabric is a topic in need of further research.

7 Conclusion

This work introduced the task-oriented design and prototypical implementation of HyperSec, a visual analytics security monitoring tool tailored for Hyperledger Fabric. Throughout the design of HyperSec, we followed the NBGM design methodology. The domain problem describes the activities of the blockchain security monitoring process to be supported by visualizations. Subsequently, we identified the involved users, their specific tasks, and the available data elements. These considerations culminated in design requirements that apply to any visualization system aiming to support blockchain security analysts. Our prototype HyperSec picks up on these design requirements. It extends the open-source architecture of Hyperledger Explorer with additional security-relevant data sources. The data is aggregated, processed and displayed in appropriate visualizations supporting blockchain security analysts to detect potential attacks.

Our prototype might not cover every possible subtask of the defined tasks of blockchain security analysts. This is in part due to limited availability of data provided by Hyperledger Fabric itself. We plan to update our prototype as additional data sources become available in the future, and are open to contributions from the community.

The security of the monitoring tool itself is also important, as it should not contribute additional attack vectors by leaking blockchain data. During our implementation we found some bugs and vulnerabilities within Hyperledger Explorer, which we subsequently fixed and contributed to the upstream project.

References

1. Baset, S., Prehoda, B.: Hyperledger Labs Blockchain Analyzer, March 2021. https://github.com/hyperledger-labs-archives/blockchain-analyzer. 30 May 2019
2. Ben-Asher, N., Gonzalez, C.: Effects of cyber security knowledge on attack detection. Comput. Hum. Behav. **48**, 51–61 (2015). https://doi.org/10.1016/j.chb.2015.01.039

3. Bogner, A.: Seeing is understanding: anomaly detection in blockchains with visualized features. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, NY, USA, pp. 5–8. ACM (2017). https://doi.org/10.1145/3123024.3123157
4. Boshmaf, Y., Al Jawaheri, H., Al Sabah, M.: BlockTag: design and applications of a tagging system for blockchain analysis. In: Dhillon, G., Karlsson, F., Hedström, K., Zúquete, A. (eds.) SEC 2019. IAICT, vol. 562, pp. 299–313. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22312-0_21
5. Calder, A.: NIST Cybersecurity Framework (2018). https://doi.org/10.2307/j.ctv4cbhfx
6. Chi, E.: A taxonomy of visualization techniques using the data state reference model. In: Proceedings of the IEEE Symposium on Information Visualization 2000, pp. 69–75. IEEE Computer Society (2000). https://doi.org/10.1109/INFVIS.2000.885092
7. Dabholkar, A., Saraswat, V.: Ripping the fabric: attacks and mitigations on hyperledger fabric. In: Shankar Sriram, V.S., Subramaniyaswamy, V., Sasikaladevi, N., Zhang, L., Batten, L., Li, G. (eds.) ATIS 2019. CCIS, vol. 1116, pp. 300–311. Springer, Singapore (2019). https://doi.org/10.1007/978-981-15-0871-4_24
8. Di Battista, G., Di Donato, V., Patrignani, M., Pizzonia, M., Roselli, V., Tamassia, R.: Bitconeview: visualization of flows in the bitcoin transaction graph. In: 2015 IEEE Symposium on Visualization for Cyber Security (VizSec), pp. 1–8. IEEE (2015). https://doi.org/10.1109/VIZSEC.2015.7312773
9. Homoliak, I., Venugopalan, S., Reijsbergen, D., Hum, Q., Schumi, R., Szalachowski, P.: The security reference architecture for blockchains: towards a standardized model for studying vulnerabilities, threats, and defenses. IEEE Commun. Surv. Tutor. (2020). https://doi.org/10.1109/COMST.2020.3033665
10. Jensen, T., Hedman, J., Henningsson, S.: How TradeLens delivers business value with blockchain technology. MIS Quart. Execut. (2019). https://doi.org/10.17705/2msqe.00018
11. Kacherginsky, P.: Attacking and Defending Blockchain Nodes. In: DEFCON 2020, p. 54 (2020)
12. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 154–175. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70956-5_7. iSSN: 03029743
13. McGinn, D., Birch, D., Akroyd, D., Molina-Solana, M., Guo, Y., Knottenbelt, W.J.: Visualizing dynamic bitcoin transaction patterns. Big Data **4**(2), 109–119 (2016). https://doi.org/10.1089/big.2015.0056
14. Meyer, M., Sedlmair, M., Quinan, P.S., Munzner, T.: The nested blocks and guidelines model. Inf. Vis. **14**(3), 234–249 (2015). https://doi.org/10.1177/1473871613510429
15. Miksch, S., Aigner, W.: A matter of time: applying a data-users-tasks design triangle to visual analytics of time-oriented data. Comput. Graph. **38**, 286–290 (2014). https://doi.org/10.1016/j.cag.2013.11.002
16. Munzner, T.: A nested model for visualization design and validation. IEEE Trans. Visual Comput. Graphics **15**(6), 921–928 (2009). https://doi.org/10.1109/TVCG.2009.111
17. Putz, B., Pernul, G.: Detecting blockchain security threats. In: 2020 IEEE International Conference on Blockchain (Blockchain), pp. 313–320. IEEE (2020). https://doi.org/10.1109/Blockchain50366.2020.00046

18. Putz, B., Pernul, G.: Trust factors and insider threats in permissioned distributed ledgers. In: Hameurlain, A., Wagner, R. (eds.) Transactions on Large-Scale Data- and Knowledge-Centered Systems XLII. LNCS, vol. 11860, pp. 25–50. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-662-60531-8_2
19. Sundara, T., Gaputra, I., Aulia, S.: Study on blockchain visualization. Int. J. Inform. Visual. **1**(3), 76–82 (2017). https://doi.org/10.30630/joiv.1.3.23
20. The Linux Foundation: DLTLabs Case Study - Hyperledger (2020). https://www.hyperledger.org/learn/publications/dltlabs-case-study
21. The Linux Foundation: Hyperledger Explorer (2020). https://www.hyperledger.org/use/explorer
22. The Linux Foundation: Hyperledger Fabric 2.3 Documentation (2020). https://hyperledger-fabric.readthedocs.io/en/release-2.3
23. Tovanich, N., Heulot, N., Fekete, J., Isenberg, P.: Visualization of blockchain data: a systematic review. IEEE Trans. Visual. Compute. Graphics 1 (2019). https://doi.org/10.1109/TVCG.2019.2963018
24. Zheng, P., Zheng, Z., Luo, X., Chen, X., Liu, X.: A detailed and real-time performance monitoring framework for blockchain systems. In: Proceedings - International Conference on Software Engineering (2018). https://doi.org/10.1145/3183519.3183546, iSSN: 02705257

5 Designing a Decision-Support Visualization for Live Digital Forensic Investigations

Current status:	Accepted & Published
Conference:	34th Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy, June 25-26, 2020
CORE Ranking:	**B** (http://portal.core.edu.au/conf-ranks/2067/)
Date of acceptance:	April 14, 2020
Date of publication:	June 18, 2020
Full citation:	BÖHM, F., ENGLBRECHT, L., AND PERNUL, G. Designing a Decision-Support Visualization for Live Digital Forensic Investigations. In *Data and Applications Security and Privacy XXXIV*, vol. 12122 of *Lecture Notes in Computer Science*. Springer, Cham, 2020, pp. 223–240
Authors' contributions:	Böhm Fabian 45% Englbrecht Ludwig 45% Pernul Günther 10%

Conference Description: DBSec is an annual international conference covering research in data and applications security and privacy. The 34th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy (DBSec 2020) was held in Regensburg, Germany. 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.

Designing a Decision-Support Visualization for Live Digital Forensic Investigations

Fabian Böhm( ), Ludwig Englbrecht, and Günther Pernul

Universität Regensburg, 93053 Regensburg, Germany
{fabian.boehm,ludwig.englbrecht,guenther.pernul}@ur.de

Abstract. Fileless Malware poses challenges for forensic analysts since the infected system often can't be shut down for a forensic analysis. Turning off the device would destroy forensic artifacts or evidence of the fileless malware. Therefore, a technique called Live Digital Forensics is applied to perform investigations on a running system. During these investigations, domain experts need to carefully decide what tools they want to deploy for their forensic analysis. In this paper we propose a visualization designed to support forensic experts in this decision-making process. Therefore, we follow a design methodology from the visualization domain to come up with a comprehensible design. Following this methodology, we start with identifying and defining the domain problem which the visualization should help to solve. We then translate this domain problem into an abstract description of the available data and user's tasks for the visualization. Finally, we transform these specifications into a visualization design for a Live Digital Forensics decision-support. A use case illustrates the benefits of the proposed method.

Keywords: Digital Forensics · Visual Analytics · Live forensics · Visualization design

1 Introduction

Malware has been around since the early days of computers. While traditional malware relies on malicious executable files, there is one particularly evil type of malware: Fileless Malware (FM). This type is hard to detect as it hides itself in locations that are difficult to analyze [31]. It exists exclusively in memory-based areas like the RAM instead of being written directly on the target's hard drive. This complicates forensic investigations of FM as most traditional Digital Forensic analysis techniques are designed to work on computers after they got turned off [16]. However, as FM solely exists in memory, turning off the target would lead to significant loss of evidence. Although some evidence of FM can be acquired through traditional DF analysis techniques, keeping the potentially infected system running allows the investigator to gather additional evidence

Published by Springer Nature Switzerland AG 2020
A. Singhal and J. Vaidya (Eds.): DBSec 2020, LNCS 12122, pp. 223–240, 2020.
https://doi.org/10.1007/978-3-030-49669-2_13

occurring during an incident. Moreover, there are mission-critical systems that simply cannot be shut down in order to not disrupt business operations. Therefore, Live Digital Forensics (LDF) is necessary.

LDF allows domain experts to investigate a running system, identify artifacts and collect evidence. This helps to understand FM-based attacks but at the same time requires fast and careful decisions about the LDF tools used to carry out the analysis. A poor choice of the analysis tool could destroy or compromise important artifacts.

In order to support forensic analysts to make faster and better decisions upon which tools should be used during an LDF investigation or upon which indicators might need additional attention, we propose to apply Visual Security Analytics (VSA). VSA allows domain experts to interactively explore the data of the system under investigation. It supports the decision-making process by allowing the forensic investigators to assess the current situation with a tailor-made visualization approach for a specific situation [29]. Therefore, they can lead the attention towards possible indicators for FM and deploy the respective LDF analysis tools like *volatility*[1] or *SysAnalyzer*[2].

This paper shows our process of developing a visual representation aimed to help Digital Forensic experts with directing their attention throughout their analyses. We follow a methodological design approach to bridge the gap between domain (digital forensic) and visualization experts [17,30]. Our main contribution is the methodological design of a visual decision-support system aiding forensic experts to direct their further investigations during a live forensic analysis. We introduce the methodology, derive a design from the requirements and problems within the LDF domain, and evaluate our design by showcasing the identification of a fileless malware's artifacts within a live forensic analysis.

The remainder of this work is structured as follows: Sect. 2 identifies and summarizes related work within the digital forensic analysis domain and existing visualization approaches. We describe the applied methodology to design the visualization in Sect. 3. The first step of our methodology is a characterization of the domain problem in Sect. 4. Section 5 follows the remaining steps of the methodology to design a comprehensible visualization for the characterized domain problem. This design is afterwards evaluated in Sect. 6 by showcasing how artifacts of the fileless malware *Poweliks* can be identified and how this helps to guide further investigations. We conclude our work and point to further possible research in Sect. 7.

2 Related Work

A Live Digital Forensic analysis is performed on a running system during an ongoing incident. The data is collected and analyzed simultaneously. The focus is on the preservation and processing of semi-persistent or volatile traces. This could be the content of the RAM, active network connections or running processes and programs [1]. Since these traces are no longer available after a system

[1] https://www.volatilityfoundation.org/.
[2] http://sandsprite.com/iDef/SysAnalyzer/.

restart, they cannot be extracted from a disk image by post-mortem analysis [13]. Live analysis is therefore useful if volatile data is essential for reconstructing an incident. This is the case if the system cannot be shut down for reasons of availability or dependency, or if encrypted data systems can no longer be accessed after a restart, for example when analyzing a fileless malware [11].

A disadvantage of live analysis is that the process can often not be repeated after leaving the location of the seizure [11]. In addition, the analysis takes place in a potentially compromised environment, so that relevant traces can be hidden, for example by using rootkits [1]. Furthermore, in the context of live analysis, a modification of the system by the investigation activities is almost unavoidable [1]. These modifications should be as limited as possible and all activities in the system must be precisely documented [13].

It is challenging to prove in court that the data integrity of the digital evidence has been preserved throughout the entire digital investigation. This may lead to a reduction in the admissibility of the evidence or even to a prohibition of its use. However, there are methods for comprehensible documentation and differentiation between the actions of an Incident Response team and the activities of an active attacker [8]. Providing a profound and tamper-proof documentation of analysis steps reduces the possible impact on the admissibility of volatile evidence and/or its modification. Nevertheless, these methods usually have to be implemented in advance as Digital Forensics Readiness measures.

Additionally, post-mortem and live analysis are not competing approaches, but rather complement each other. Live analysis enables the extraction and processing of additional traces, which can considerably support post-mortem analysis and the reconstruction of the course of events [1].

We identify several related visualization approaches originating from both the Visual Analytics (VA) and the Digital Forensics research domains. Within the VA domain, the designs are often based on user-centered approaches to provide a solution for a specific, relevant task of forensic experts. These visualizations feature a broad variety of use-cases ranging from the forensic investigation of shadow volumes and directories [14,15] to live monitoring of network traffic [3,4]. Tools like EventPad [6] allow the interactive and explorative analysis of large, dynamic data sets to identify malware and its behavior. The KAMAS solution is a tool providing not only innovative automated malware analysis features but also the functionality for malware analysis experts to exchange domain knowledge with the automated analysis methods [28]. Although a variety of related visualization designs exists in the VA domain, none of these visual representations is specifically designed to support the decisions forensic investigators need to make during an ongoing live forensic investigation. The same applies to the VA approaches introduced in the DF research domain. Tools like Timelab [23], LogAnalysis [7], MalViz [22], Vera [26], or Devise [27] allow a visual representation of different types of data for static forensic investigations but are by no means capable to support fast, dynamic decisions for live forensics.

None of the above-described visualization approaches pays special attention to the decision-support required throughout an LDF investigation. Additionally,

to the best of our knowledge there is no existing work on bridging the gap between the domains of Live Digital Forensics and Visual Analytics by applying methodologies to develop comprehensive and reproducible visualization designs. Therefore, the knowledge from the Visual Analytics domain is beneficial as it pays attention to design aspects that are being neglected up to now in the LDF domain. We aim to contribute to a transfer of knowledge from the VSA towards the LDF domain in this research as it has been done in other security-related domains within the last years [25].

3 Methodology

This section summarizes the methodology which we follow throughout this work. A methodological approach allows our design decisions to be reproducible and comprehensible. Especially in visualization design this is of utmost importance because even methodologically based decisions remain subjective [18]. Therefore, we follow the Nested Blocks and Guidelines Model (NBGM) which is a well-established methodology for designing visualizations [19]. Another important aspect of the NBGM is that it is aimed to support the collaboration between domain and visualization experts and, therefore helps to close the aforementioned gap in LDF visualization designs [30]. The high-level layers of the NBGM are depicted in Fig. 1 and described in the subsequent sections [19,21].

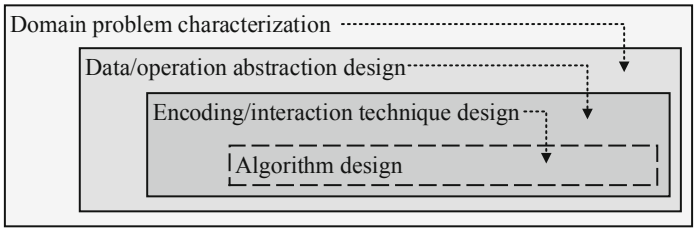


Fig. 1. Nested layers of the NBGM [19,21].

Domain Problem Characterization: The main task in this first layer is the identification of the specific situation and problem for which the visualization should be designed. The tasks and data of the target group are identified including their workflows and processes. Each target domain has its own descriptive vocabulary and it is important within this phase to work with the target users using their familiar vocabulary. This layer of the nested model bridges the gap between visualization experts and domain experts as it allows designers to understand the world the domain experts work in and which problems they face [32].

Data/Operation Abstraction Design: The second level abstracts the domain problem characterization using the vocabulary of the visualization

domain. Therefore, it describes visualization tasks and the data relevant to the design. Domain-specific tasks and data descriptions are translated into a visualization-specific vocabulary. This way, visualization designers identify what tasks (e.g. finding outliers, identifying trends) the domain experts have to solve from a visualization point of view. The tasks of the users in visualizations can be derived from a variety of existing task taxonomies [5,28]. Additionally, the data abstraction allows designers to describe data transformations of available data identified within the domain problem characterization into a different format if necessary, for subsequent encoding technique decisions.

Encoding/Interaction Technique Design: This layer describes the visualization (encoding) techniques and the necessary interactions for users. Both, encoding and interactions must be aligned together and are derived from the visualization tasks in combination with the data at hand from the data / operation abstraction design-layer. Encoding and interaction techniques combine the first two nested layers with a design that instantiates the abstract visualization for the domain problem.

Algorithm Design: The innermost layer of the NBGM requires to create appropriate algorithms carrying out the beforehand designed encodings and interactions. We do not consider this step in our current work and focus on the first three layers of the model. The final implementation of the design is part of our further research.

4 Decision-Support for Live Forensics

In the case of an LDF investigation, decisions can be directly linked to the risks involved. Therefore, it is important to make well-considered decisions when choosing the right techniques, tools, and artifacts. In this section, we characterize the domain problem to enable a suitable visualization design helping domain experts facing this domain problem. We emphasize the supporting effect of the visualization for a digital forensic examiner. In particular, the tasks during a live digital forensic investigation are discussed. The goal is to apply the design methods of visualization experts to support better decision-making for a domain expert.

4.1 Live Digital Forensics Process

The collection and analysis of digital evidence should be based on a defined comprehensive process model. A common description of a forensic investigation process is represented by the model of Kent et al. [13]. The investigation process is divided into four phases as depicted in Fig. 2. We have extended the original approach to include an overarching decision-support by an interactive visualization at every stage. The following paragraphs describe the different original stages, which need a decision-support:

Collection: Data related to the criminal activity are identified, labeled, recorded, and secured from all potential sources of relevant data [13]. Possibly relevant additional data sources might be identified, and respective data needs to be collected during an LDF analysis.

Examination: The data collected in the previous phase is evaluated. The aim is to identify and extract relevant data [13]. Since our approach is applied to a live investigation, a visual analysis of the data allows the decision to include additional data sources for the analysis. Consequently, the visual decision-support creates a return to a previous phase.

Analysis: The results of the previous phases are analyzed in depth and interpreted to establish connections between persons, places, objects, and events and to obtain useful information regarding specific questions [13]. Findings from the visual analysis are directly incorporated into the analysis process. Malicious activities can be better understood through a visual representation of the data.

Reporting: The results of the analysis are prepared and presented, including important information. The format and content of the report depend on the type of recipient [13]. Especially, visual representations can contribute considerably to the understanding of the incident. Particularly, if the attack is complex, spikes in network traffic or system performance can provide a good insight on the activities during the incident.

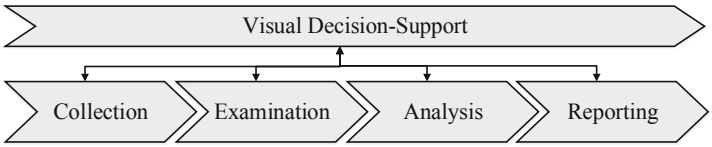


Fig. 2. A high-level process for Live Digital Forensics and Visual Decision-Support.

4.2 Tasks of Domain Experts in Live Digital Forensics

Mistry and Dahiya [20] discuss the volatile memory forensics approach in detail. Using live forensics, real-time data is analyzed and stored based on the system activities. The analysis of the memory (RAM) is very important while considering live computer forensics. The approach of live forensics plays an important role in identifying Indicators of Compromise (IoCs) and recording volatile data, which would be lost after shutting down the system. The authors use *memory forensics* to run through various challenging scenarios and prove their approach based on previously extracted and identified data in real-time. Since their approach provides a good description of the domain experts' workflow and it has been used by the authors in several scenarios, this is further considered. We abstract and extend the original approach as a baseline (see Table 1) to identify the main tasks during an LDF analysis:

Table 1. Summarized expert tasks in LDF.

Task	Details
Data Acquisition	– Identify suspected devices and media – Dump RAM, cache, and network traffic – Acquire an image of system (if possible)
Establish Intelligence	– Parse memory structure – Identify relevant memory segments – Identify loaded modules – Identify running processes and file accesses – Identify established network connections
Memory & Data Analysis	– Search outliers and irrelevant information – Extract additional relevant data – Verify findings for further decisions – Decide the next analysis steps
Documentation	– Document interesting findings – Document artifacts and evidence

– **Data Acquisition:** Within this task, investigators need to decide which data they export from the device under investigation. During an LDF analysis, only a limited amount of data can be extracted. An additional limitation for this task is often, that data only can be extracted with a-priori implemented functionalities.
– **Establish Intelligence:** This step is very much based on the present situation and requires that the investigator has a good sense of the specific case. Usually this is due to the prior knowledge of the investigator. It is important that in this step no analysis in the actual sense is carried out, but rather the region for possible purposeful evidence is identified. A graphical processing by means of VSA can contribute significantly to this. Especially decisions about the inclusion of further areas are very time-critical and a visual representation can contribute to a fast identification.
– **Memory & Data Analysis:** In this step, the previously identified data is examined for suspicious features. In addition, the findings are put into context to reconstruct the course of events. By a supporting effect of VSA, outliers and correlations can be better found.
– **Documentation:** The aforementioned tasks are documented during the whole digital forensic investigation to be used in the final report. This is an essential component to make the investigation comprehensible. During an analysis using VSA, findings based on a graphical preparation can be documented in the figures using markers (e.g. at peak values).

4.3 Available Data in Live Digital Forensics

Harichandran et al. [10] formed the term *curated forensic artifact (CuFA)* to specify the scope of forensic artifacts and their supervised attributes. The Artifact Genome Project (AGP), based on CuFA's principles, was launched in 2014 and has received 1099 forensic artifacts within the last few years [9]. It reached an acceptable level of maturity, as registered participants can contribute to this project by uploading artifacts along 19 categories.

Crimes are committed in several ways, and the expedient evidence is accumulated by different forensic artifacts. Depending on the peculiarity of a case, digital evidence either adds more value to an investigation or is completely inappropriate. The ontology of crimes by Kahvedzic et al. [12] provides a specification of past criminal cases and offers the possibility to specify almost every cyber case. We summarize the sub crime cases and focus only on cyber-crime cases.

The violation of the quality of forensic artifacts influences their admissibility at courts. Because of the fast-moving nature of digital evidence, we adopt the legal requirements by Antwi-Boasiako et al. [2] due to their overall completeness and applicability. This framework is appropriate for forensic investigations and reduces the overall scope of common data quality dimensions. These legal requirements cannot be circumvented, as admissibility in court is indispensable. AGP represents an open-source platform based on the CuFA principles. The following forensic artifacts categories have been extracted: Windows registry, memory, file, network packet, process, email message, address, code, disc partition, account, network socket, disk, user account, X509 certificate, user session, windows event log, volume, and Linux packages. These categories are further reduced since our concept focuses on LDF and not all categories are available in this type of investigation. To illustrate the possibilities of our approach, we will focus on the following categories of data sources that can be accessed during a live forensic analysis (without major interference due to the installation and execution of additional applications): *file access, network packets, process-lists, event logs (including PowerShell) and system statistics.*

VSA can support understanding and interpreting the context data in combination with stored data. Therefore, VSA allows experts to make better, context-based decisions for further investigations.

5 A Design for Visual Decision-Support in Live Forensics

Based on the domain problem which arises for forensic experts during a live forensic investigation (see Sect. 4) we derive appropriate visualization tasks, visual encoding and necessary interaction functionalities for a visual decision-support system within this section. Figure 3 depicts an overview over the respective, fully defined NBGM model for this problem domain.

The central contribution of this part of our work is the innovative application of different encoding techniques combined as an interactive, coordinated view where interactions in one view influence the representation in others. This allows a lateral, visual movement in the data enabling forensic analysts to browse

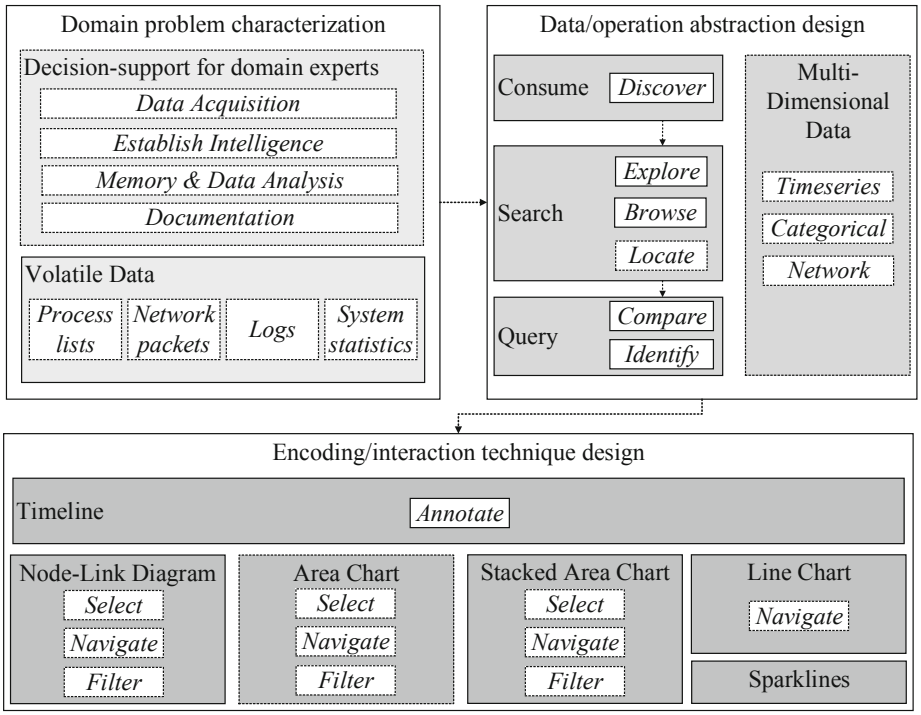


Fig. 3. NBGM applied for LDF covering the results Sects. 4 and 5.

through the data and identify artifacts that need further investigation with specific forensic analysis methods. However, without the visual decision-support they might not even have spotted the artifact. Therefore, we strengthen the necessity of a visual design as proposed by us in the following sections.

5.1 Data/Operation Abstraction Design

This section covers the abstraction of the domain problem characterization above. This step in the NBGM is carried out via designing and identifying data blocks and task blocks that describe the needs, requirements, and problems of the forensic experts. Defining these blocks, subsequently allows a comprehensible decision for specific visual encoding and interaction techniques.

Task Blocks. In Sect. 4.2 we identify several tasks that are important for forensic experts when they are performing an LDF investigation. A visual design for decision-support during these investigations needs to support the experts in these tasks. To be able to transfer the domain problems into a suitable visual design, we identify abstract visualization tasks from the *Why?* part of Brehmer and Munzner's task typology [5]. This typology describes why a specific task is performed in terms of which goal a user is pursuing.

The abstract, process-like overall task of forensic experts is to get an idea or indication where to further direct the ongoing LDF investigation and therefore,

which tools they should deploy (see Sect. 4.1). We summarize this high-level task as "Decision-support for domain experts" in Fig. 3. However, the process-based task of decision-support can be split into several abstract visualization tasks. The main goal for forensic experts from a visualization point of view is to *Discover* which is a task formed around the generation and verification of hypotheses. An exemplary hypothesis in this context might describe which malware is acting on a device or which forensic tools need to be deployed to continue the investigation.

The *Discover* task is further specified depending on whether the analyst has a hypothesis in mind when using the visualization or not. When the LDF investigation is carried out to find evidence for a specific malware with known indicators acting on the device, this corresponds with the *Locate* task. *Browsing* outlines actions to search through suspicious indications within the visualization to find out which investigative tool might help to continue the analysis. The remaining task at this level, *Explore*, represents an analyst exploring the displayed data to identify possible suspicious patterns in the data and therefore, to make decisions on possible malware types on the device or at least additional LDF tools to deploy on the device. Once a hypothesis about malware or LDF tools is made, the forensic expert continues to *Identify* additional characteristics of the malware or tries to further strengthen the need for the LDF tool. This is possible by using the visual representation of our design.

Data Blocks. The available volatile data described in Sect. 4.3 is mostly defined as multi-dimensional data from a visualization point of view. However, there are some relevant subcategories of data types and formats which significantly influence the decision for corresponding visual encoding.

All available data that needs to be visualized within the decision-support for LDF is time-based data as it relates to some characteristics or actions of the system at a specific point on time, e.g. the network connections at a specific time or the respective CPU and RAM workload. Additional data categories can be categorical data like the different severity types of collected logs and network data.

5.2 Encoding/Interaction Technique Design

This section connects the different aspects between the problems of the forensic experts and the abstract blocks derived in Sect. 5.1 by introducing the visual encoding and corresponding interaction techniques that are necessary to enable a visual decision-support for LDF investigations. The visual encodings described in the following section are mainly well-known and established visualization techniques. We decided to only use these to ensure easy and fast perception of the design for forensic investigators. The encoding techniques are derived from the available data blocks while the interactions are necessary for the forensic experts to follow their tasks using the visual encoding of the data. The design sketch for the decision-support visualization is shown in Fig. 4. It comprises five interactive main components that are further detailed within this section. In terms of more

abstract visualization tasks, the design allows *Navigation*, *Selection*, *Filtering*, and *Annotation* [5].

Fig. 4. Resulting design sketch of a decision-support visualization for Live Digital Forensics investigations.

Investigation Timeline. The first component is the overarching *Investigation Timeline*. This component is necessary since most of the data represented within the design is time series data (see Sect. 5.1). Therefore, a timeline allows navigating through different points in time of the collected data and analysts can select a specific time window for their analysis. The selected time window is indicated by the small white box-shaped overlay on the timeline and the time-range on the right side. In the design sketch of Fig. 4 a window of two minutes between 12:00:00 and 12:02:00 is selected. However, the box overlay can be moved across the timeline and can also be resized to allow the selection of different time ranges. The other four components of the visualization design display only data from the selected time window. An additional functionality of the *Investigation Timeline* is the event annotation. Forensic analysts can mark and label specific points in time when they identified possible evidence or interesting artifacts. This allows to come back to these events in a later investigation or even the collaboration of multiple analysts where one can pick up the investigation on a mark added to the timeline by another analyst.

Network Activity. The next component displayed in the upper left corner of Fig. 4 is the *Network Activity*. This view aims to give an overview over the

device's external activities regarding the endpoints and IP addresses it communicated with. In the center of this view the device under investigation and process IDs (PIDs) is shown. To keep this representation clearly laid out only PIDs with active connections during the selected time frame are displayed. The connection partners are illustrated with ellipses labeled by IP addresses. The connection targets are clustered by IP address range allowing to distinguish different networks. In the exemplary design sketch, for example, the local network of the device is clustered on the left of the *Network Activity* clearly separated from external connection targets on the right.

We include the connections between a process and its communication partner by adding directed links for both incoming and outgoing communication. The color coding of the links is dependent on the cumulative number of bytes sent through the connection with a scale from blue (i.e. few bytes or "cold connection") to red (i.e. many bytes or "hot connection"). The distinguishable and color-coded links for up-link and down-link connections allow to quickly detect large data flows and to identify the process responsible for this data flow. Examples for possible artifacts needing additional analysis with sophisticated LDF tools are the imaginary process with the PIDs 2345 and 3456. The first one is sending a lot of data to an external IP address while the other one is downloading numerous bytes from another address.

Clicking on a PID highlights the connections of the selected process in this view but also in the *System Performance* and the *Read/Write Entropy* views where the activities of the process are highlighted respectively. This allows a quick indication about a process's overall statistics including its network activity as well as its CPU and RAM activity. Hovering a connection opens a thumbnail with additional information on this specific network communication between a process and an external IP address. This additional information contains the exact number of bytes sent over the connection as well as the port and protocol used to open it. A similar hovering interaction is also provided for the nodes depicting the communication partners of the device under investigation. The thumbnail for these nodes contains the total amount of bytes sent from and to this node as well as ports that were used to connect to. Hovering a node simultaneously also highlights the processes that established a connection with the corresponding IP address.

Read/Write Entropy. In the upper right corner the decision-support visualization design features a *Read/Write Entropy* display. The area charts of this view show the entropy of both read and write operations on mounted drives of the investigated device. The x-axis of the charts encodes the selected time frame from the *Investigation Timeline* while the positive y-axis displays the entropy of the data read from the specific drive at a point in time on a range from 0 to 1 (0% to 100%). Analogously, the negative y-axis represents the same indicators but for data written onto the drive. Therefore, the entropy for write operations is indicated as a negative value in our design. This only serves to clearly distinguish positive and negative y-axis values in the area charts. In addition to

indicating the difference of read and write operations by indicating them with different vertical directions, they also are encoded with different colors. In the top right corner of this view, the drives for which the entropy values should be displayed can be selected via check-boxes.

This view allows a zooming interaction, preferably by mouse-wheel, where zooming in narrows down the displayed time window and zooming out analogously widens the time span. If experts zoom into this view, the time window is adjusted respectively for all other views. Possible artifacts that catching an expert's eye in this view are unusually high entropy scores for read or write operations. As an example, serves the increasing entropy scores in the design sketch towards the end of the selected time frame for both drives. This might indicate the writing of a lot of encrypted data on the two drives.

System Activity. The left view in the bottom row of our design is a visual representation is also an area chart, but a stacked version. It provides a visual encoding of *System Activity* by displaying a count of system events (e.g. Windows Event Logs, Powershell Events, Syslogs). The events are colored depending on their type or severity allows experts to detect a rising number of errors or similar indications of artifacts. The check-boxes on the top right of the view allow enabling different event types to be displayed. The x-axes of the area chart are like a timeline while the y-axes indicate the cumulative count of currently displayed event types at a specific point in time. The stacked area chart allows identifying trends and changes in the logged activities of the system.

This chart is also allowing a zoom interaction like the previous *Read/Write Entropy* views. Within this view, an unusually high number of error logs in the Windows Event Log that is constantly appearing throughout the whole two minutes currently under investigation could be an artifact for further analysis.

System Performance. The last view that is part of our visualization design located on the bottom right of Fig. 4. It is split into two smaller views which in combination give an indication of the *System Performance*. The upper part of this view is occupied by a line chart with two different lines. The blue line depicts CPU performance while the second, orange line indicates memory or RAM activity. Both lines are on a relative scale, meaning that the y-axis ranges from 0 to 1. Both lines on the chart are again displayed for the selected window of time. The line chart allows a zoom interaction similar to the interaction described within the *Read/Write Entropy* and the *System Activity* views.

The lower part of the display contains a table with active processes during the time which is currently defined for analysis. The table has four columns for the process name, its PID, and a spark line visualization allowing a fast perception of this process's CPU and RAM activities. A spark line is a special, word-sized type of line chart. They are not displayed with any axes and serve a single purpose: to give an indication about the trend of a single indicator. The table might be longer than the five exemplary rows from our design sketch and therefore, needs to be scrollable. Rows can also be selected leading to a highlighting of the

corresponding process in the table and in the *Network Activity* view. Selecting a process also changes the *Read/Write Entropy* view by now only showing the entropy of the read or write operations performed on behalf of this specific process. This enables forensic analysts to conclude on the influence of a process on the systems performance and possible correlations with network activities.

6 Use Case

To show how our visualization design can support forensic experts in their LDF investigations, we go through a short use case featuring a well-known and documented fileless malware attack. We describe how indicators of this malware become apparent within our visual decision-support and how we support the tasks of forensic experts during an LDF investigation identified in Sect. 4.2.

The use case features the fileless-malware *Poweliks* which attacks Windows-based systems. This malware became known as a file-based piece of malicious code but in 2014 it moved to a file-less variant. After computers are infected they are part of a click-fraud botnet where bots request advertisement data from a central Command-and-Control (C&C) server, load the ads and click them to generate revenue [24]. As a side effect, *Poweliks* often acts a door-opener for other malware as it clicks up to 3000 ads per day on a single computer and does not care about whether the ads are malicious or not. Although this malware attracted attention back in 2014, its design is special in two aspects. *Poweliks* acts without leaving a file on the computer's file system. It stores all the data it needs in the registry and memory by injecting code into legitimate processes currently running. Therefore, it is hard to detect once it gained a foothold on the system. The second interesting aspect of *Poweliks* is, that, despite being a fileless malware, restarting the infected device does not remove it as it reboots itself from altered registry keys. This makes *Poweliks* a very special and dangerous type of fileless malware [33]. Because of those characteristics, we choose to describe how indicators for the *Poweliks* malware are visible within our visualization design.

Based on publicly available details and threat hunting details about the ad-fraud variant of *Poweliks*, we describe indicators that can be detected within our design, helping forensic analysts to make decisions where to guide their attention for further analyses. We structure the indicators and their identification according to the different views of our visualization design (see Sect. 5.2).

Network Activity: Regarding the network activity of an infected system, there are several indicators becoming apparent within a visual display. First, *Poweliks* is known to download the Powershell as well as the .NET framework from official Microsoft download pages if not available on the computer. The respective connections might appear in the view as connections of processes to official Microsoft IP addresses and a high payload on the down-link transfer, i.e. the link between the Microsoft IP and the process turns red.
Additionally, as the malware acts as a botnet, it regularly connects to its C&C server. These are only short connections with a very limited payload.

However, as they appear on a very regular basis, they can be identified as an indicator for further analysis why the system is connecting to the respective IPs.

Another suspicious activity to be spotted via the proposed design is the ad-clicking component of *Poweliks*. The behavior of requesting ad data from the C&C server, contacting a search page for the URL of the ad, and clicking the loaded advertisements becomes recognizable as the network activity would show many small-scale connections to a lot of different, external IP addresses. This is all more suspicious when the respective network connections are originating from a single process.

Read/Write Entropy: Overall, activities on the file system is less apparent as *Poweliks* is a file-less malware. However, as the malware can request up to 3000 ads per day on a single computer it is very likely that the malware also "clicks" other malicious ads. The entropy of read and write operations on different drives of a computer shall light on possible ransomware being active due to *Poweliks*' activities. Increasing entropy values in this view indicate the transfer of encrypted data. This highlights the necessity for domain experts to further investigate this malware since it could have features of a ransomware.

System Activity: *Poweliks*' special fileless persistence method uses a watchdog and PowerShell scripts when it is establishing its foothold. It also modifies many key registry entries trying to lower or disable browser security settings to be able to perform the ad-clicking behavior. Both of these actions produce log events (Windows Event Logs, PowerShell Events, etc.) with different severity. However, as the performed behavior is rather uncommon, the *System Activity* view shows several warnings and errors to the domain experts. Therefore, they might for example decide to analyze the changes made to the key registry in depth.

System Performance: Also, the system performance is not too bad during the execution of the *Poweliks* malware. This is because the malware does not want to significantly affect the performance of the infected computer. However, with our concept it can be seen that the CPU and RAM are used when a web page is accessed in the background for a few moments. This is due to the fact that the browser has to interpret and render the website. If the observed computer is not running other programs, this is also a possible indication of the malware. These findings by using a visual display during a live investigation help the forensic examiner to better assess the current situation and to make a well-considered decision for the use of certain tools. Also, the display of the processes and their RAM as well as CPU indicates possible further investigation needs. In the case of *Poweliks* which is hiding in different common processes (e.g. cmmon32.exe, dllhost.exe, logagent.exe), unusual activity of those processes indicates further investigation potential. Especially, when these processes are involved in anomalous network activity as well.

However, the malware *Poweliks* will be detected by current virus scanners but a coming back by a modification of the malicious code or behavior is very likely since file-less techniques evolved in the last few years. Nonetheless, they are relevant artifacts helping forensic experts to guide their further analyses.

7 Conclusion and Future Work

Within this work we made a contribution utilizing Visual Security Analytics as a decision-support approach for Live Digital Forensic investigations. We describe and abstract the problem of forensic investigators which have a wide variety of tools at hand for their analyses but need to decide quickly which of them need to be deployed in the current situation. To support them in this decision-making process we applied a methodology derived from the visualization research domain. Contributing to this domain problem with a tailor-made visualization approach enables forensic investigators to make faster and well informed decisions. We described the proposed visualization design and evaluated the visual representation with a simple use case. Summarizing, we showcased how Visual Security Analytics could help to solve an existing problem on the domain of LDF.

For future work we mainly see two different directions to follow. First of all, we want to apply our visual security analytic approach to a more sophisticated malware using *Process Doppelgänging*[3] where current anti-virus software and forensic tools reach their limits. Process Doppelgänging refers to a file-less code injection that uses a Windows native function and an undocumented implementation of the Windows Process Loader. This technique leaves no traces and is very difficult to detect. Our approach can highlight malicious activities and assist the digital forensics examiner during a live forensics investigation. Furthermore, another path to pursue in future work is the generalization of our approach. This requires to identify inherent characteristics of FMs and their classification based on a subset of those characteristics. A more holistic and modular version of our design approach would allow to have a specific encoding for each malware characteristic. This would support the work of forensic investigators even further as they can define individual dashboards as subsets of the available designs fitting their need to identify known and unknown FM.

References

1. Adelstein, F.: Live forensics: diagnosing your system without killing it first. Commun. ACM **49**(2), 63–66 (2006)
2. Antwi-Boasiako, A., Venter, H.: Implementing the harmonized model for digital evidence admissibility assessment. DigitalForensics 2019. IAICT, vol. 569, pp. 19–36. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28752-8_2

[3] https://www.blackhat.com/docs/eu-17/materials/eu-17-Liberman-Lost-In-Transaction-Process-Doppelganging.pdf.

3. Arendt, D., Best, D., Burtner, R., Lyn Paul, C.: Cyberpetri at CDX 2016: real-time network situation awareness. In: 2016 IEEE Symposium on Visualization for Cyber Security (VizSec), pp. 1–4. IEEE (2016)
4. Boschetti, A., Salgarelli, L., Muelder, C., Ma, K.L.: Tvi: a visual querying system for network monitoring and anomaly detection. In: Proceedings of the 8th International Symposium on Visualization for Cyber Security - VizSec 2011, pp. 1–10. ACM Press, New York (2011)
5. Brehmer, M., Munzner, T.: A multi-level typology of abstract visualization tasks. IEEE Trans. Vis. Comput. Graph. **19**(12), 2376–2385 (2013)
6. Cappers, B.C., Meessen, P.N., Etalle, S., van Wijk, J.J.: Eventpad: rapid malware analysis and reverse engineering using visual analytics. In: 2018 IEEE Symposium on Visualization for Cyber Security (VizSec), pp. 1–8. IEEE (2018)
7. Catanese, S.A., Fiumara, G.: A visual tool for forensic analysis of mobile phone traffic. In: Proceedings of the 2nd ACM Workshop on Multimedia in Forensics, Security and Intelligence - MiFor 2010, p. 71. ACM Press, New York (2010)
8. Englbrecht, L., Langner, G., Pernul, G., Quirchmayr, G.: Enhancing credibility of digital evidence through provenance-based incident response handling. In: Proceedings of the 14th International Conference on Availability, Reliability and Security, ARES 2019, pp. 26:1–26:6. ACM (2019)
9. Grajeda, C., Sanchez, L., Baggili, I., Clark, D., Breitinger, F.: Experience constructing the artifact genome project (AGP): managing the domain's knowledge one artifact at a time. Digit. Invest. **26**, S47–S58 (2018)
10. Harichandran, V.S., Walnycky, D., Baggili, I., Breitinger, F.: Cufa: a more formal definition for digital forensic artifacts. Digit. Invest. **18**, S125–S137 (2016)
11. Hoelz, B., Ralha, C., Mesquita, F.: Case-based reasoning in live forensics. In: Peterson, G., Shenoi, S. (eds.) DigitalForensics 2011. IAICT, vol. 361, pp. 77–88. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24212-0_6
12. Kahvedzic, D., Kechadi, M.T.: Dialog: a framework for modeling, analysis and reuse of digital forensic knowledge. Digit. Invest. **6**, 23–33 (2009)
13. Kent, K., Chevalier, S., Grance, T., Dang, H.: Guide to integrating forensic techniques into incident response. NIST Spec. Publ. **10**(14), 800–886 (2006)
14. Leschke, T.R., Nicholas, C.: Change-link 2.0: a digital forensic tool for visualizing changes to shadow volume data. In: Proceedings of the Tenth Workshop on Visualization for Cyber Security - VizSec 2013, pp. 17–24. ACM Press, New York (2013)
15. Leschke, T.R., Sherman, A.T.: Change-link: a digital forensic tool for visualizing changes to directory trees. In: Proceedings of the Ninth International Symposium on Visualization for Cyber Security - VizSec 2012, pp. 48–55. ACM Press, New York (2012)
16. Mansfield-Devine, S.: Fileless attacks: compromising targets without malware. Netw. Secur. **2017**(4), 7–11 (2017)
17. Marty, R.: Applied Security Visualization. Safari Tech Books Online. Addison-Wesley, Boston (2009)
18. McCurdy, N., Dykes, J., Meyer, M.: Action design research and visualization design. In: Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV 2016, pp. 10–18. ACM Press, New York (2016)
19. Meyer, M., Sedlmair, M., Quinan, P.S., Munzner, T.: The nested blocks and guidelines model. Inf. Vis. **14**(3), 234–249 (2015)
20. Mistry, N.R., Dahiya, M.S.: Signature based volatile memory forensics: a detection based approach for analyzing sophisticated cyber attacks. Int. J. Inf. Technol. **11**(3), 583–589 (2018). https://doi.org/10.1007/s41870-018-0263-4

21. Munzner, T.: A nested model for visualization design and validation. IEEE Trans. Vis. Comput. Graph. **15**(6), 921–928 (2009)
22. Nguyen, V.T., Namin, A.S., Dang, T.: Malviz: an interactive visualization tool for tracing malware. In: Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis - ISSTA 2018, pp. 376–379. ACM Press, New York (2018)
23. Olsson, J., Boldt, M.: Computer forensic timeline visualization tool. Digit. Invest. **6**, 78–87 (2009)
24. O'Murchu, L., Gutierrez, F.P.: The evolution of the fileless click-fraud malware poweliks (2015). https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/evolution-of-fileless-click-fraud-15-en.pdf. Accessed 24 Feb 2020
25. Puchta, A., Böhm, F., Pernul, G.: Contributing to current challenges in identity and access management with visual analytics. In: Foley, S.N. (ed.) DBSec 2019. LNCS, vol. 11559, pp. 221–239. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22479-0_12
26. Quist, D.A., Liebrock, L.M.: Visualizing compiled executables for malware analysis. In: 2009 6th International Workshop on Visualization for Cyber Security, pp. 27–32. IEEE (2009)
27. Read, H., Xynos, K., Blyth, A.: Presenting devise: data exchange for visualizing security events. IEEE Comput. Graph. Appl. **29**(3), 6–11 (2009)
28. Rind, A., Aigner, W., Wagner, M., Miksch, S., Lammarsch, T.: Task cube: a three-dimensional conceptual space of user tasks in visualization design and evaluation. Inf. Vis. **15**(4), 288–300 (2016)
29. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G., Keim, D.A.: Knowledge generation model for visual analytics. IEEE Trans. Vis. Comput. Graph. **20**(12), 1604–1613 (2014)
30. Simon, S., Mittelstädt, S., Keim, D.A., Sedlmair, M.: Bridging the gap of domain and visualization experts with a liaison. In: Eurographics Conference on Visualization (EuroVis) - Short Papers. The Eurographics Association (2015)
31. Sudhakar, Kumar, S.: An emerging threat fileless malware: a survey and research challenges. Cybersecurity, **3**(1), 1–12 (2020)
32. van Wijk, J.J.: Bridging the gaps. IEEE Comput. Graph. Appl. **26**(6), 6–9 (2006)
33. Wueest, C., Anand, H.: Internet security threat report: living off the land and fileless attack techniques (2017). https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/istr-living-off-the-land-and-fileless-attack-techniques-en.pdf. Accessed 24 Feb 2020

6 Visual Decision-Support for Live Digital Forensics

Current status:	Accepted & Published
Conference:	18th IEEE Symposium on Visualization for Cyber Security, October 27, 2021
CORE Ranking:	**C** (http://portal.core.edu.au/conf-ranks/1540/)
Date of acceptance:	August 12, 2021
Date of publication:	December 14, 2021
Full citation:	BÖHM, F., ENGLBRECHT, L., FRIEDL, S., AND PERNUL, G. Visual Decision-Support for Live Digital Forensics. In *2021 IEEE Symposium on Visualization for Cybersecurity (New Orleans, 2021)*, pp. 58–97
Authors' contributions:	Böhm Fabian 30% Englbrecht Ludwig 30% Friedl Sabrina 30% Pernul Günther 10%

Conference Description: The 18th IEEE Symposium on Visualization for Cyber Security (VizSec) is a forum that brings together researchers and practitioners from academia, government, and industry to address the needs of the cyber security community through new and insightful visualization and analysis techniques. VizSec provides an excellent venue for fostering greater exchange and new collaborations on a broad range of security- and privacy-related topics.

2021 IEEE Symposium on Visualization for Cyber Security (VizSec)

Visual Decision-Support for Live Digital Forensics

Fabian Böhm* Ludwig Englbrecht† Sabrina Friedl‡ Günther Pernul§

Chair of Information Systems
University of Regensburg
Germany

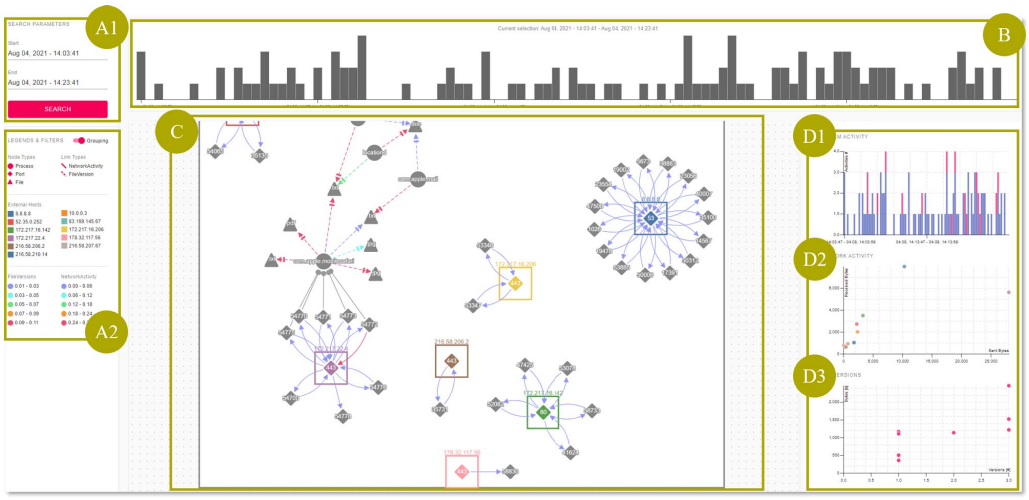


Figure 1: We introduce a visual analytics tool that supports cyber forensic experts to decide which parts of a system need further investigation. The tool provides four interactive views: *Search Parameters & Filters* (A1, A2), brushable *Overview* bar chart (B), interactive *Node-Link diagram* (C) to perceive the system's activities, and a supporting *Details-on-Demand* view (D1, D2, and D3).

ABSTRACT

Performing a live digital forensics investigation on a running system is challenging due to the time pressure under which decisions have to be made. Newly proliferating and frequently applied types of malware (e.g., fileless malware) increase the need to conduct digital forensic investigations in real-time. In the course of these investigations, forensic experts are confronted with a wide range of different forensic tools. The decision, which of those are suitable for the current situation, is often based on the cyber forensics experts' experience. Currently, there is no reliable automated solution to support this decision-making. Therefore, we derive requirements for visually supporting the decision-making process for live forensic investigations and introduce a research prototype that provides visual guidance for cyber forensic experts during a live digital forensics investigation. Our prototype collects relevant core information for live digital forensics and provides visual representations for connections between occurring events, developments over time, and detailed information on specific events. To show the applicability of our approach, we analyze an exemplary use case using the prototype and demonstrate the support through our approach.

*e-mail: fabian.boehm@ur.de
†e-mail: ludwig.englbrecht@ur.de
‡e-mail: sabrina.friedl@ur.de
§e-mail: guenther.pernul@ur.de

Index Terms: Applied computing—Computer forensics—; Human-centered computing—Visualization systems and tools—; Human-centered computing—Visual analytics—; Security and privacy—Intrusion detection systems—;

1 INTRODUCTION

Conducting live digital forensics (LDF) investigations is becoming increasingly important in the context of cyber forensics. While traditional forensic analysis methods are primarily applied to switched off devices (static analysis), LDF focuses its investigation on running devices [19]. The need for this alternative approach originates from a variety of recent developments. One of them is that live (i.e., real-time) forensics is imperative to analyze some specific malware types like fileless malware. This type of malware exists primarily in memory [24]. As soon as the infected device is turned off to perform a static forensic investigation, evidence of essential importance will be lost. Other examples where LDF is necessary are mission-critical or very large systems that cannot be shut down due to their importance to business operations. Thus, these reasons force forensic and security experts to act directly by collecting evidence while an attack is happening [15].

Experts for Digital Forensics (DF) conducting the live investigations have a wide range of forensic tools at their disposal, which they mainly apply according to their own experience. The tools are highly specified to collect and analyze a specific type of evidence and, therefore, differ widely [26]. During an LDF investigation, the targeted use of tools is essential to minimize the possibility of corrupting or even destroying evidence. For this reason, any use of a forensic tool on the device under investigation must be well-

2639-4332/21/$31.00 ©2021 IEEE
DOI 10.1109/VizSec53666.2021.00012

considered [8]. However, since the target device continues to be operational during the investigation, it is at the same time necessary to decide about the tools to be used timely. If the decision is made too slowly, this poses the risk that evidence will be corrupted or that the malware continues to spread. The live environment also leads to a high volume and velocity of data on which to base a decision for specific forensic tools.

As mentioned before, after having identified indicators for an attack or threat, forensic analysts need to decide which specialized forensic tools they should apply to work through the attack and collect evidence. This decision is based mainly on the experts' experience without any reliable automated solutions to support this decision-making. Therefore, in the context of a live forensic investigation, it is necessary to support domain experts in quickly getting an idea where on the device the attack manifests. From these locations, evidence can then be collected and analyzed. In cyber forensics, much of the respective evidence is related to the file system and changes in the file system. In a static forensic investigation, it is only possible to analyze user data on the persistent data storage, assuming that it has not yet been overwritten by further operation. However, in an LDF investigation, specific changes to persistent files, including the resulting file versions, can be incorporated into the analysis. Additionally, file system activities are often related to network communication (command & control server, leakage of information, or lateral movement of malware). It is, therefore, necessary for experts to be able to correlate the changes in the file system and the network activities of the device under investigation to identify indicators and decide on specialized forensic tools to apply. These observations lead to the main research question of our work:

RQ 1 *How can core information needed for Live Forensic investigations be acquired and systematically visualized for DF experts in order to support their decision-making?*

Our contribution to this research question is two-fold. In a first step, we derive general design requirements for an LDF investigation decision-support tool. Afterward, we present a prototypical visual decision-support tool that enables DF experts to analyze relevant volatile data from a compromised device. This approach differs from traditional DF as we provide a visualization pipeline that collects, pre-processes, and visualizes data that would no longer be available for post-mortem analysis. We collect the relevant data from a live and operative device, with physical access to the device being the only requirement. This information enables experts to make informed and justifiable decisions regarding the forensic tools to be applied.

2 Related Work

Several visual analytics (VA) tools exist that can be used for forensic purposes, such as malware or volume analysis. We show only a short selection of recent approaches that are related to our work.

The visual representations of the forensic analysis tool *Change-Link 2.0* allow experts to comprehend changes over time within shadow volume data [14]. Therefore, the authors propose metaphors to visualize directory structures, directories, file content, and directory metadata. While *Change-Link 2.0* provides insight into temporal changes of a file system, it does not show processes responsible for these changes and their external communication.

FIMETIS is a tool allowing to interactively explore file system snapshots [1]. The tool provides a security analyst with simple and straightforward analysis views for file system records, the temporal sequence of events, and data clusters. Although detailed insights into the file system are provided, the tool can't be applied for LDF as, in this context, file system snapshots are usually not available.

The research prototype *MalViz* is a tool for analyzing the behavioral patterns of malware [21]. Its intended use is to investigate the relationships and dependencies of a system's processes with an active malware. Using *MalViz*, DF experts can identify unusual patterns within the processes' activity more efficiently. *MalViz* can be applied for live forensics. However, users only gain an understanding of process activities without much-needed context about other relevant activities.

Another approach for forensic analyses of malware was introduced with *Eventpad* [4]. Eventpad's advantage is its capability to significantly reduce the complexity within network traffic samples to quickly understand the networking behavior of malware samples. Eventpad has proven to be highly effective for network data, but it does not help domain experts to identify internal indicators.

Furthermore, innovative approaches for the analysis of network packet captures (PCAPs) have been proposed [25]. They feature a web-based visualization design meant to support malware analysts and administrators in their tasks that frequently involve PCAP analyses. Although this tool is not explicitly designed with forensic analyses in mind, network traffic analysis is vital for live forensics but needs to be combined with information about the internal communications of a device.

Most of the existing visual designs in the context of forensic analysis focus on either static analysis, on a single data source, or at least on a specific type of data (e.g., network traffic or volume data). While this is undoubtedly helpful for the forensic analyst, especially concerning further investigation, there is no way to get a quick, initial overview of a system's activities during an LDF investigation. Thus, there is no support for domain experts in live forensics to help them make decisions on the forensic tool to apply.

The need for a separate visual tool is even more stressed through our design proposal within the cybersecurity field [2]. We partially build on this design idea as a reference point [20]. However, the existing design falls significantly short in several regards. First of all, it is a purely theoretical design. Thus, there is no proof that a respective tool would be feasible, especially concerning data availability. Second, our previous work is narrowed down to the proposed design and, therefore, not generalizable. It is possible and necessary to derive comprehensible requirements leading the design and development of respective visual tools.

3 Design Methodology

In this work, we design and implement a visual tool to support DF experts in deciding on the analysis tools to collect and analyze evidence. We follow the design methodology proposed by Meyer et al. [17]. Their method is strongly problem-oriented. Thus, it enables the derivation of appropriate design proposals. Additionally, an important reason for applying this methodology is to bridge the gap between domain and visualization experts [3, 23], and, to some extent, help to resolve the dichotomy of security visualization in the area of cyber forensics [16].

We identify the domain problem based on existing academic results, our own experience, and discussion with DF experts. All of these highlighted the need of experts for support in the tool selection during an LDF investigation. Again and again, the experts are confronted with not knowing which specialized forensic tool (e.g., volatility [1], autopsy [2], Cellebrite UFED [3]) they should use. They need to balance between acting as quickly as possible and at the same time not damaging or even destroying evidence by using the wrong tool. Although the DF experts are not directly involved in the design process, their precious input during informal discussions helped us determine their needs regarding data sources, possibly helpful visualization designs, and the tasks they need to fulfill and, thus, derive general requirements.

In Section 4, we firstly determine the data to be presented. In addition, in this step, we determine the concrete target group of the design proposal and define the essential tasks of this target group in

[1] https://www.volatilityfoundation.org/
[2] https://www.sleuthkit.org/autopsy/
[3] https://www.cellebrite.com/en/ufed/

the context of the domain problem. Finally, and building on these findings, general requirements for visual decision-support tools for LDF are derived.

Section 5 introduces the visual encodings we identify suitable to support the previously defined users' tasks. Our proposed design consists of multiple interactively interlocked visual representations allowing DF experts to comprise an investigated device's internal and external activities. Therefore, the prototype allows exploratory analysis leading to well-considered decisions regarding the forensic tools to apply in further and more detailed analyses.

Finally, Section 6 gives an insight into the technical implementation of the design proposal in the context of the research prototype created in this project. An agile development process is applied, in which the design is repeatedly discussed with the involved DF experts and, if necessary, adapted. An exemplary use case of the intended use of the prototype during an LDF investigation is highlighted in Section 7.

4 Requirements Analysis

In this section, we define the available data, the intended users, and their tasks as a basis for deriving requirements for actual visualization designs [18].

4.1 Data: System activities, File Versions, and Network activity

During an LDF investigation, a comprehensive picture of the current situation needs to be obtained. For this purpose, it is important to collect information about running applications, write and read operations on volatile as well as persistent storage media, and all outgoing and incoming signals (e.g., network traffic). This data must be retrieved in its raw form so that integrity is not compromised during the investigation process. In a subsequent step, this information must be correlated in a meaningful way so that the analyst can understand the relations.

We define the data at hand, which is relevant to provide a visual design supporting users in the problem domain. Analyzing a running and probably infected system is challenging. Available data during an LDF is limited to data acquired with little interaction and without installing additional software. Otherwise, the attacker could recognize that the system is being analyzed and initiate anti-forensic measures. Additionally, too much interference with the operating device might also cause evidence to be corrupted. Intending to support the decision-making process during a live investigation and to remain as undetected as possible, we identify the following data sources as relevant:

- **S1: System Logs.** Data about the activities of a running system can be used to support the decision-making process on the appropriate forensic tool. In this context, the analysis and evaluation of system logs play an important role. The system logs can be used to identify common and unusual events and to target the forensic analysis.
- **S2: Process information.** In addition to the system logs, continuous monitoring of the running processes is essential. This allows detecting irregularities in the running applications, such as a high CPU utilization or a high memory requirement. These are common indicators for an infection of the system.
- **S3: In- and outgoing network traffic.** Furthermore, information about the network activities of the device is important to detect communication with a Command & Control server or information leakage. This shows which process has sent data over which port to which destination address at what time.
- **S4: File-system activities.** Looking at the activities of the file-system provides deep insights into what files are used and modified during the execution of an application. The gathering of the file-system events does not include the actual content of the consumed or modified files but provides meta-information about the activities.
- **S5: Copy of modified files.** The acquired evidence is at the heart of a digital forensic investigation. The previously mentioned artifacts relate to volatile evidence. Those traces are not stored permanently on the disk and need to be captured and stored in a forensically sound way. We apply an additional tool for capturing the actual file content of modified files during the LDF investigation. This differs from the recording of metadata of the file-system by providing a snapshot of every file version. Although this form of backup can have an immense storage overhead, we use this data for the LDF. This makes it possible to examine the actual file contents for unusual changes in the content. Also, with this approach, a large base of usable digital evidence can be created, which can be used to prepare the DF report and consequently in court.

In order to identify suspicious activities on the system and the processes of an active attack, it is vital to understand the context of the various data sources. Only then is it possible for DF experts to apply the appropriate DF analysis tool in a targeted manner. The provided list of data sources highlights the most suitable ones but depending on the actual situation, it might be feasible to include additional data sources.

4.2 Users: Digital Forensics experts

After we are clear about the underlying data, defining the intended user group for the visual design is necessary. Subsequently, the target group has been delineated, defining the tasks they need to perform with the visual representation in the next step.

Since visual analytics is inherently user-driven, meeting the needs of users is critical. A thorough understanding of the users' needs, tasks, and work environment are required to achieve this. However, this is not easy to accomplish because problems in this area are ambiguous and, therefore users can be described in many ways.

We identify DF domain experts as the intended target group. There is a need for supporting their decision-making process, which forensic methods and tools apply during LDF. Forensic investigations are not to be carried out by security novices, as these analyses require considerable domain knowledge, which is only available among DF experts [3]. Thus, we expect the users to have the expertise to decide which forensic tool best fits the current activities on the device when supported adequately.

These experts are rarely exposed to enormous time pressure in the classic forensic investigation because they do not operate on a live system. Instead, they can perform the necessary static analyses based on memory dumps and similar static artifacts. With these copies, it is also possible to prevent evidence from being corrupted due to incorrect tool decisions by making a forensically sound copy of the hard drive apriori. However, this is not possible in the context of a live forensic investigation, and once corrupted, evidence is unrecoverable. In addition, any activity of the still-operational device can potentially corrupt evidence. This characterizes the target group of the intended visual decision-support tool as forensic domain experts who have to make a quick decision under time pressure in an ever-changing and possibly unfamiliar environment. In summary, the target group has solid domain knowledge but only limited operational knowledge regarding complex visualizations [6].

4.3 Tasks: Support forensic tool selection

The overarching task we want to support visually is to make a well-informed and thoughtful decision about the use of forensic methods, tools, and artifacts in the course of an LDF investigation. The goal is to accelerate and improve decision-making for domain experts. The

experts need to effectively identify potential indicators of malware or unusual activity to guide their further analysis process. The Live Forensic approach plays a vital role and recording volatile data or evidence that would be lost if the device were turned off. The analysis of volatile memory and the device's network activity is especially important in this as described in Section 4.1. For these tasks, experts have a wide variety of specialized forensic tools in their repertoire. In the real-time environment of an LDF investigation, a quick overview of the situation on the device plays a significant role. It is important to note that both from a technical perspective and increasingly from a legal perspective, continuous collection of the necessary data to support live forensics is not possible. Instead, the experts need a way to activate real-time data collection in case of suspicion and collect data from that moment on [16].

Figure 2: Forensic analysis process according to [13].

The related tasks of domain experts can be derived from the results of existing work [2, 19]. In this paper, we use a process model for LDF (see Fig. 2) to group the relevant tasks by commonly established process phases of a DF investigation. In the following, we describe the tasks in the context of the respective process phase:

1. **Data Acquisition:** DF experts need to *decide what data from the suspicious device or other systems needs to be acquired* for the investigation. This is not a simple task, especially in a live investigation since the decision needs to be made fast, and the situation can change quickly. Furthermore, the expert needs to *decide if additional software needs to be installed* to acquire appropriate data for the DF investigation. Any system interaction can be detected by an active attacker and lead to possible concealment actions. Usually, the amount of data that can be acquired without previously installed DF software is limited to the tools provided by the operating system. This leads to DF experts starting with limited data and successively acquiring additional information by selecting a suitable tool.

2. **Establish Intelligence:** Although DF investigations have a similar overall process, each attack confronts experts with different situations. During this process step, the skill is very much dependent on the expert's prior knowledge to initiate the correct next analysis step. No evidence is analyzed in-depth, but *it is decided which areas are helpful for the investigation*. There can also be a return to the previous task in which further sources are included to evaluate targeted areas. A visual representation of the current situation on the device under investigation supports analysts either *decide which areas to focus on for further analysis* or *which additional data sources need to be acquired.* A well-considered and well-informed decision at this stage is an essential success factor in resolving an attack.

3. **Memory & Data Analysis:** The previously *found and classified data is analyzed in detail* in this step. The data and information are put into context to *create a comprehensive picture of the current situation*. In particular, *unusual processes and correlations can be detected*. At this stage of a (live) forensic investigation, mainly specialized tools to are applied. Nevertheless, the domain experts still profit from a high-level overview of the device's live status to guide their analysis in the right direction.

4. **Documentation:** The *appropriate and accurate documentation of the analysis*, including every step of the expert, is a crucial aspect of the credibility of the acquired evidence. A comprehensive report supports the presentation of the facts in a court of law and the admissibility of specific evidence obtained after interacting with the compromised system.

In summary, these tasks clearly show that forensic scientists must make decisions regarding more advanced data acquisition or specialized tools at several points in the forensic analysis process. These decisions must be made repeatedly, especially in the second (*Establish Intelligence*) and third (*Memory & Data Analysis*) step of the process in an LDF investigation. It requires a clear overview of the data that can be collected without deep intervention on the investigated device. A visual representation of the data can support this overview and the associated decision-making process.

4.4 Requirements

Concluding the first part of our contribution, we derive a list of general requirements for decision-support tools within live forensic investigations. These requirements need to be fulfilled to support the users' tasks described in Section 4.3. They not only serve as a basis for a sound visualization design in the further course of this work but can be used for LDF decision-support applications in general:

- **R1: Retrieve and visualize only relevant, accessible data.** Each analysis has different prerequisites and, therefore, also different data is available in the *Data Acquisition* phase. Consequently, to keep the interference with the running system as low as possible, it is vital to retrieve only the data relevant and accessible within the DF analysis. This could be the usage of network activities, file-system activities, and information about running processes.

- **R2: Provide visual representations of the data in a timely manner.** As little time as possible should elapse between the retrieval of the data, its preparation, and the presentation. Thereby, a workflow of a real-time or (near-) real-time processing needs to be achieved to support the *Establish Intelligence* and *Memory & Data Analysis* phase.

- **R3: Document origin of and changes to the raw data.** The procedure for retrieving the files from the system must be explained in a comprehensible manner. If the analysis process results in modifications to the system and thus to the retrieved data, these must be meticulously documented. This strengthens the credibility of the data and allows the artifacts to be admitted as evidence. Additionally, it is a good approach to acquire and save the original data in its raw format in the *Data Acquisition* phase. All steps and findings must be recorded within the *Documentation* phase.

- **R4: Show significant behavioral changes over time.** All retrieved information has to be shown during the *Establish Intelligence* phase in a way that allows DF experts to detect significant changes (in- and decrease) in the device's behavior over time.

- **R5: Display available information in its situational context.** Information obtained from different data sources must be correlated with each other. The correlation can take place via timestamps and further attributes (e.g., process IDs). The resulting correlations should be visible during the *Establish Intelligence* step to the experts in the visual representation.

- **R6: Highlight conspicuous anomalies.** If, for example, processes or other entities of the monitored device behave conspicuously, this behavior should be visually identifiable. Conspicuous behavior can be when a process sends large amounts of data to a remote recipient via a wide variety of ports or

performs unusual processing of files. These findings move the focus from the *Establish Intelligence* phase to a deep *Memory & Data Analysis*.

- **R7: Allow exploratory analysis of the data.** Information is supposed to be visually abstracted for an analysis process so that an overview of a specific situation can be obtained within the *Establish Intelligence* phase. However, users should be able to explore the data and receive additional details when necessary through a *Memory & Data Analysis*.

5 Visual Design

This section describes the visualization techniques we employ for our proposed solution. The designs aim to make the data described in Section 4.1 accessible and understandable for DF experts based on the requirements from Section 4.4. With the help of the interactive design, DF experts can quickly grasp the essential indicators during an LDF investigation and decide on which other tools they need to employ for the analysis. Although data is captured and pre-processed in a real-time manner within our processing pipeline (see Sec. 6), the different views of our prototype are not updated automatically when new data from the monitored device becomes available. Instead, analysts must manually change the selected time window within A.1 to get the latest available data. In this context, the role of a DF expert differs from that of a security analyst who would have to react immediately. Therefore, we made this decision to ensure that the DF experts can focus on the analysis of the selected time range without being interrupted by live updates.

Thus, the design of the visualization techniques and the interactions follows the central guideline of the Information Seeking Mantra [22]. The overall design is depicted in Fig. 1. The period of interest is first determined by the search parameters (Fig. 1.A1). Users can then use the Overview diagram (Fig. 1.B) to get a first impression of the overall activity within the time period. More detailed information about the activities as well as the entities involved (processes, ports, files, etc.) can then be analyzed in more detail in the node-link diagram (Fig. 1.C) using the appropriate filters (Fig. 1.A2). The Details-on-Demand view (Fig. 1.D1 - D3) supports the overview as long as no entity is selected and otherwise displays detailed information about the selected object. The following sections are structured according to Fig. 1 and go into more detail on the designs regarding these visual representations.

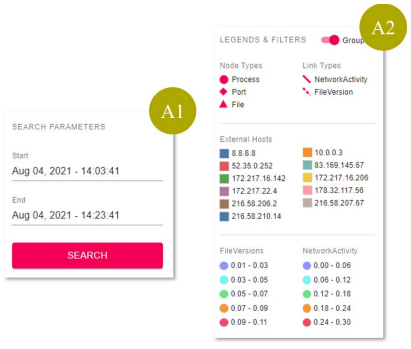


Figure 3: Search parameters (A.1) and available filters (A.2).

5.1 View A: "Search Parameters & Filters"

Fig. 3 shows search and filtering options available within our prototype. The *Search Parameters* (A.1) is the actual starting point of the design. Users select the start and end times for their analysis. Afterward, only the data that lies within this time window will be considered in the other views (**R2**, **R7**).

Fig. 3.A2 shows the *Legends & Filters* for the node-link diagram described in Section 5.3 including comprehensive filtering options (**R7**). As used in our prototype, a major disadvantage of node-link diagrams is that they tend to display only a "hairball" when the number of nodes is high or when the nodes are highly interconnected (high number of edges). Such networks can no longer be perceived by users and are thus almost useless. This shortcoming affects network diagrams regardless of the layout algorithm used. To address this drawback in some way, we implement a series of interactive filters that can be used to hide or show individual nodes, edges, or even entire classes of nodes or edges. This allows users to influence the graph's layout themselves and reduce the number of elements displayed if the automatic layout no longer produces satisfactory results. In addition to the possibility of explicitly filtering types of nodes and edges in the graph, we also offer users other interactive filters, which we describe in Section 5.3.

Figure 4: Bar chart for an overview of all activities with a specific time range brushed.

5.2 View B: "Overview"

The *Overview* chart presents the overall activity of the device in terms of file versions (**S4**) and network traffic (**S3**) for the time period selected in the *Search Parameters*. For this purpose, a simple bar chart is used in our prototype. Each bar represents the number of activities within a certain period of time. The size of this period is dynamically determined depending on the size of the total time window selected in the *Search Parameters*. The bar chart allows users to quickly perceive when the device has conspicuously high or very low activity (**R4**).

The bar chart for an overview of the device's activities is interactively interlocked with all overviews. A Brush interaction allows analysts to select a time window from the overall graph window. Fig. 4 respectively shows the *Overview* with an active Brush selection. The node-link diagram (see Section 5.3) and the detailed views (see Section 5.4) display only the activities within this brushed time frame (**R7**).

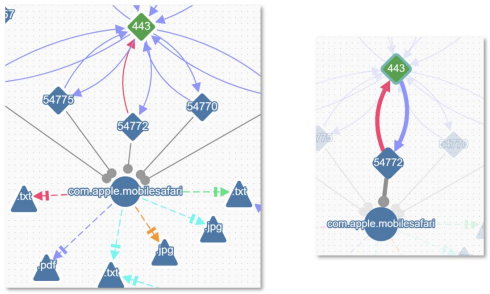


Figure 5: Detailed views of the node-link diagram. One showing the ungrouped display (left) and the other showing a highlighted node with its neighbors (right). Ports are displayed as rectangles, files as triangles, and processes as circles.

5.3 View C: "Node-link diagram"

The node-link diagram gives insight into the activities of the device during the time window selected in the *Overview*. If no brush selection is made, it displays the entire time range from the *Search Parameters*. The network-based encoding allows DF experts to perceive very active nodes (i.e., nodes with a high number of links) or suspiciously inactive nodes vice versa. Thus, users can identify possible indicators for further and more detailed analyses (**R5**, **R6**).

We apply a directed, compound node-link diagram that uses the fCoSe-layout to arrange nodes in the available display space. fCoSe is an improved version of the original CoSe layout algorithm [7]. While maintaining good results even for relatively large datasets through a force-directed layout scheme, fCoSe is more efficient and works well for directed graphs. To ensure a smooth user experience and enable exploration of parts of the graph for DF experts, the network diagram supports zooming and panning (**R7**).

Different data elements intended to be displayed are processes (**S2**), files (**S4**), and ports including the host the ports belong to (**S3**). These data elements are represented as vertices, and different glyphs indicate their type. They are colored based on the host to which they can be allocated. Most of the ports, all files, and all processes belong naturally to the device that is being monitored. However, other relevant hosts become apparent through the network connections as the device sends or receives network packages to or from their ports. When the *Grouping*-option (see Fig. 3.A2) is activated, the hosts are displayed explicitly as parent nodes for ports, processes, and files. When this option is switched off, the hosts are not displayed and indirectly influence the display through the categorical coloring of the vertices. Nodes are draggable to ensure that users can adjust the layout according to their needs or if the layout algorithm does not produce an apt result. To further give the possibility to reduce the number of nodes to be displayed, filter options are available to hide specific node types ("Node Types" filter in Fig. 3.A2) or to hide nodes of a specific host. Selecting a node displays more details (see Sec. 5.4) and highlights incoming and outgoing links from this node as well as its 1st-degree neighbors (**R5**, **R7**).

Three different types of links connecting the nodes are present as can be seen in Fig. 5 (**R5**). Each link type represents a specific activity on the investigated device:

1. "FileVersion"-links from a Process-node to a File-node indicate that the respective process edited at least one new version of this file (**S2**, **S4**).

2. "NetworkActivity"-links between two Port-Nodes represent network communication from one node to the other (**S3**).

3. "PortActivity"-links from a Process-node to a Port-node mark the attempt of a process to open a network connection via the respective port (**S1**, **S2**, **S3**).

Each type of link is displayed as a different line. Not every activity is drawn as a particular link but unique links where the respective activities are accessible through the *Details-on-Demand* window after selecting the link (**R7**). "PortActivities" are not colored, as no additional information is available. We only know whether a process used a specific port or not. However, "FileVersions" and "NetworkActivities" are colored according to their weight concerning all other displayed links of the same type (**R6**). The weight for "FileVersions" is calculated as the number of bytes edited by the process (source of the link) in the file (target of the link) divided by all file edits made within the selected time range. The weight for "NetworkActivities" is determined similarly by the sum of the package size of all packages that went through the respective link divided by the number of bytes sent between ports overall. We use a quantile color scale with only five different colors for each link type. Although this reduces the level of detail, it makes anomalous links stand out much more clearly. Types of links can, analogously to nodes, be hidden through a filter ("Link Types" filter in Fig. 3.A2), but also specific quantiles of link weights can be hidden to only view very high or very low weighted links ("Network Activity" and "File Version" filter in Fig. 3.A2).

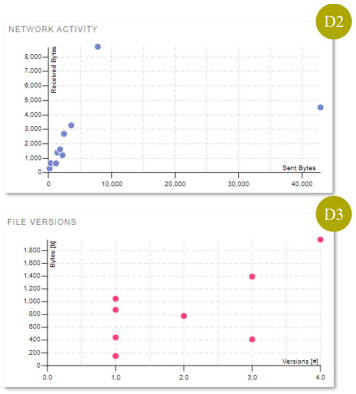


Figure 6: Scatterplot to support users in identifying abnormal network and file system activities.

5.4 View D: "Details-on-Demand"

This part of the prototype's interface serves a two-fold objective. First, it supports experts in gaining an overview of the ongoing activities as long as no specific element in the node-link diagram is selected. Second, when experts click a node or link, it displays further details about this element.

When supporting the overview task, the *Details-on-Demand* view displays three additional charts:

1. System Activity (Fig. 1.D1): This stacked bar chart gives a close-up view of the overall system activity within the brushed time period. File version activities and network activities are distinguishable through distinct colors in this chart (**R4**).

2. Network Activity (Fig. 6.D2): This scatter plot renders one dot for each host the investigated system has communicated with. This dot indicates a sent-receive-ratio in terms of bytes sent to or received from the respective host. The x-axis of this plot maps the total amount of bytes sent to a host, while the y-axis marks the sum of bytes received from it. This helps users to very efficiently spot conspicuous connections either which differ from benign behavior (**S3**, **R6**). One example is the dot on the far right of Fig. 6.D2 indicating that the device sent a lot more data to this host than to others.

3. File Versions (Fig. 6.D3): This scatter plot displays additional information about the file versions. Each dot depicts a specific file. The x-axis shows the number of times this file was edited, while the y-axis shows the sum of overall added or deleted bytes. This is calculated simply by summing up the byte differences from one version of a file to the next one. This again helps DF experts to explore the data and spot anomalous activities (**S4**, **R6**). A high number of files touched one time might indicating an ongoing data leakage action.

For details on a selected element, the charts D1, D2, and D3 are replaced with the view shown in Figure 7. This view shows all available attributes of the selected element. When the analyst clicks a link, all underlying events are accessible through this view (**R7**).

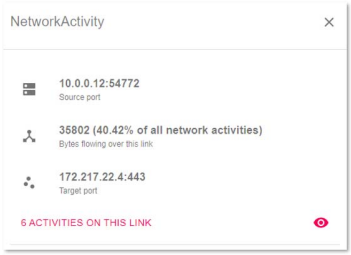


Figure 7: Displayed details after a link is selected. Clicking the "eye" symbol brings up a modal with further details on all relevant activities.

File versions can also be downloaded for further investigation with specific forensic tools (**S5**).

6 Research Prototype

We finally take a closer look at the implementation of our prototype. This includes a significant amount of work to acquire the relevant data with as little interference as possible to fulfill both **R1** and **R2**. This section describes our data acquisition approach, how the information is processed and finally sheds light on the actual implementation of the above-described visual representations. The structure of this section is based on the visualization pipeline [5]. The visualization design introduced in Section 5 represents the technical outcome of this pipeline rendered on the client-side of our prototype.

Fig. 8 shows the basic architecture of our prototype, including the applied technologies. Before we explicitly discuss the individual steps of the pipeline within our prototype, we briefly describe its overall structure. It is designed as a client[4]-server[5] application with an underlying document-oriented database[6] for persisting the raw data acquired from the *Data Collection* and the pre-processed data. The pre-processed data is the output of the *Data Analysis* step, in which a Python script continuously prepares newly available raw data. Further data analysis is performed by the specialized application SauvegardeEx, and the results are available to our prototype through an interface. A Node.js application represents the central component of the server for *Filtering* and partial *Mapping* (i.e., correlation) of the pre-processed data. The client's interface to the server is implemented through a GraphQL API. The client itself is a React-based application. Two different frameworks are used for the client-side *rendering* of the data in the form of interactive visualizations. On the one hand, the framework VisX[7] is used to display the bar chart, scatterplot, and alike. On the other hand, due to the potential complexity of the network graphs to be displayed, we resort to the specialized graph analysis framework cytoscape.js [12].

6.1 Raw data collection

A general requirement in the context of live forensics is that the system under investigation should be affected as little as possible by the investigation (**R1**). This poses a significant challenge, especially for the collection of data. For our prototype, we focus on appropriate data collection for mobile devices. In discussions with the experts involved, it became clear that these devices, in particular, are currently causing difficulties for DF experts, as there are few established procedures for collecting data for live forensics from active mobile devices. In most cases, the appropriate software must be integrated into the device before the investigation. With our approach, we trade-off between a highly detailed data collection and the need to install additional functionality or application on the device. Our prototype's complete data collection process can be performed using already available functionalities on a mobile device. The only requirement for this is physical access to the device to apply various forensic hard- and software.

Our prototype acquires data from an Apple iPhone operating on iOS 14.5. It allows extraction of information about active processes, file editing activities, and information about the ports used by the processes for communicating over the network from this smartphone. The respective information is collected using different appliances:

syslog: First of all, we acquire all system-logs by using *libimobiledevice* [8] and save them in JSON format (**S1**).

ps: Second, meta-information (like the underlying services' names, CPU usage, among others) of processes are collected using the Linux-tool *ps* which we use to extract a list of active processes every 5 seconds. *ps* is a lightweight tool and does not affect the device's performance significantly. It fulfills the data-source (**S2**).

PCAP: Besides this detailed information about the processes' activities (i.e., internal activities), we connect the iPhone to a network access point collecting network activity information (i.e., PCAP files). Acquiring and incorporating the in- and outgoing network traffic refers to data (**S3**).

syslog: Since the *netstat* service to receive information about the network communication of individual ports is not available on iOS 14.5, we include the *syslog*-appliance in our data collection process. *Syslog* can be used to collect Syslog information from the system. Besides much other information, they also contain information about processes opening and closing a network connection over a specific port. Although we do not get any information about the destination of the network connection, we at least can collect some rudimentary data about the processes network communications (**S3**).

fsmon: Highly volatile information about the file editing activities of processes and the resulting file versions is extracted using the *fsmon*-appliance (**S4**) and forwarded to *SauvegardeEx* which is a specialized application to process *fsmon* information. This mechanism creates a copy of every modified file on the system (**S5**). We will describe its functionalities within Section 6.2.

The above-described approach for data collection targets the use of existing data sources (**R1**) in a direct and timely processing of the data (**R2**) in an indirect manner. To fulfill the requirement (**R3**) of recording the origin and all changes of the data for proper traceability and documented we store the data in the following manner: All the collection processes can be triggered via a command-line instruction on the device and simply connecting the iPhone to the PCAP-collection access point. Besides the information extracted with *fsmon*, all raw data from this process step is forwarded in JSON format into a central MongoDB database.

6.2 Data Analysis

The *Data Analysis* step of our prototype mainly comprises a correlation of the previously collected raw data. This is necessary because the raw data itself does not allow visualization of connections between processes, file edits, and network communication. Therefore, two components in our architecture perform the data analysis: the *Python pre-processing* script and the *SauvegardeEx* application.

Python pre-processing : We apply an iterative, continuous preprocessing, which includes several steps for every raw data source. We will only describe exemplary ones briefly in this work. Any new document (i.e., new raw data element) that is pushed to the database by the command line tools applied in the *Data Collection*-step is further processed to ensure the data can be adequately visualized. As *ps* does only allow to extract its data encoded, we need to decode this data and reduce the amount of respective information by only persisting the relevant information for the processes (i.e., process

[4] https://github.com/bof64665/LDF_ReactFrontend
[5] https://github.com/bof64665/LDF_GraphQLServer
[6] https://www.mongodb.com/
[7] https://airbnb.io/visx
[8] https://libimobiledevice.org/

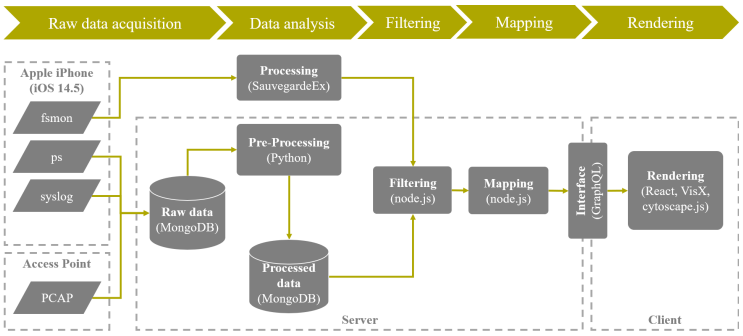


Figure 8: Architecture of the prototype.

id, name, CPU, and memory usage at a given point in time). Similarly, we parse and filter the *syslog* messages as only the messages containing information about processes opening or closing network connections via a specific port. A data reduction similar to the one applied to *ps*-data is also performed on the PCAP information. Although PCAPs contain a highly detailed description of each package leaving or reaching the iPhone, we are only interested in a comparatively small part of this description. We thus pre-process every PCAP transferred to our raw data storage and remove any currently unnecessary fields. Note that raw data is not overwritten in our process, as it might become relevant for further analysis. Thus, all pre-processed data is persisted separately from the raw data (**R3**).

SauvegardeEx: To gain useful insights into the generation and alteration of relevant files during the operation of a live system, we incorporate the tool *SauvegardeEx* [9]. *SauvegardeEx* has been used and extended in our previous works [9, 10] to support a digital forensic investigation. The client version of *SauvegardeEx* and sends every single file alteration with the actual file content to the server. Thereby, a specific file can be restored at any captured point of time, and the challenge of overwriting a freed storage area (due to deletion or updating a file on the file system) is addressed. This information can be retrieved from the *SauvegardeEx* web server through a well-defined API. Having this mechanism in place during the LDF investigation, content-specific information of every file-version can be obtained and used for the visualization.

6.3 Filtering

Filtering in the prototype is done entirely on the server-side. Based on the user's input at the start and end time of the time window to be analyzed, the relevant information is retrieved from the database and the SauvegardeEx API in this step.

6.4 Mapping

Due to the potentially tremendous amount of data that can still be relevant after the filtering, we also perform the *Mapping* almost entirely in the backend. This keeps the computational load for the frontend as low as possible and ensures that the interactive visualizations can be operated smoothly. The first step in the mapping process is thus the correlation of the data sets. Here, we correlate all data sources and obtain a data set with information about the file versions, the processes responsible for them, the ports used by the processes, and the network activities originating from these ports. The correlation is done based on the timestamps and various attributes, which allow a clear assignment of the individual events. The second sub-step is an aggregation of the correlated information into containers. Currently, we use a fixed number of 100 containers, whose size is dynamically determined depending on the selected time window.

The actual mapping and thus the last step maps the now available filtered and aggregated information to nodes and edges and all other geometric forms needed for the visualizations. However, each element additionally retains its original data to ensure access to details on-demand. This geometric information is then made available to the client via the GraphQL API.

7 Use Case

The prototypical implementation of the visual designs (see Section 6) contributes to the decision-making process according to an optimal first-hand tool by displaying the available information in its temporal and situational context (cf., R4 and R5 in Section 4.4). By applying our tool during an LDF investigation, preserving important evidence can also be supported by storing volatile data in a dedicated database. To illustrate the applicability of the prototype, we highlight its key features through an exemplary use case.

The attack pattern in this use case is based on the "Jeff Bezos Hack", where attackers applied various fraudulent techniques to obtain data from the personal iPhone X of the Amazon founder and CEO Jeff Bezos. After a meeting between Bezos and the crown prince of Saudi Arabia, Mohammad bin Salman in 2017, they exchanged phone numbers and wrote ordinary messages via WhatsApp. Shortly after Bezos received a video sent by bin Salman in 2018, his smartphone started sending large amounts of data via the Safari Mobile browser and the Apple mail program. The subsequent investigation and forensic analysis of the smartphone [11] brought to light that the compromised video (probably) contained malicious code. To this date, this was a zero-day vulnerability. The Pegasus and Galileo spyware were the most likely tools used in this attack.

We use this incident as a potential scenario for our prototype and describe the procedure during a forensic investigation from an expert's point of view with the help of the decision-support tool. We rely on the publicly available report from FTI Consulting [11] published in 2019 and extend it with additional details concerning activities of the file-system. Please note that there is no official data available from the "Jeff Bezos Hack" that would allow the comprehensive reproduction of the incident. We instead derive a set of artificial data which is available within the code repositories of our prototype. Thus, the following sections do not describe an in-depth evaluation of our prototype but rather a possible scenario where the prototype could have been applied. In the following subsections, we describe relevant steps that must be performed during an LDF analysis of similar cases. Where appropriate, relevant indicators (and their recognition) are presented and linked to the visual representations of our prototype.

[9] github.com/LudwigEnglbrecht/sauvegardeEX

7.1 Initialization

First, the device's user might have noticed suspicious events about half a year after receiving the WhatsApp message with a video attached. This manifested itself in strangely ambiguous WhatsApp messages (GIFs, pictures, videos) from Mohammad bin Salman, which reflected current situations from Jeff Bezos' private life that were not known to the public at that time (e.g., divorce from his wife). Ultimately, such messages prompted the victim to initiate a forensic analysis of his device. This initial suspicion is used as a starting point in the exemplary application of our prototype. After the initialization of the investigation, the core process steps (see Fig. 2) are explained and related to our prototype.

7.2 Data acquisition

In a first analysis step, the smartphone is acquired and physically available for an investigation. An initial analysis of the suspected video attachment in WhatsApp does not highlight any indications of active malware. However, there still is a possibility that Advanced Persistent Threats (APT) are used. Therefore, DF experts decide to perform a further forensic investigation. Based on these findings, a live forensic investigation is applied to extract both a decrypted and encrypted forensic image of the iPhone X. Since the forensically-sound copy of the iPhone X requires higher rights (root), the software *Cellebrite UFED 4PC* is used to gain root access without the necessity to reboot the device. After a forensically-sound copy of the current state has been stored, an in-depth live analysis of the running (and eventually compromised) system can be conducted. At this point, the raw data collection of our prototype is initiated.

7.3 Establish Intelligence

Narrow the analysis time-frame. In our prototype, in the *Search Parameters* (see Fig. 3.A1) the last 20 minutes are selected. With no additional filters applied, the views display all respective data. Meanwhile, data will be pushed continuously into the database and is available for further visual analysis. In this step, the system is intentionally left running without performing any actions. This enables to capture as much background activity (including unusual, suspicious activity) as possible. Since the attacker can be active during this time, valuable indicators and traces can be obtained. At this stage, the expert can spot and select a time-frame with a high amount of system activities in the *Overview* (Fig. 4.B).

Identify conspicuous network activity. In this step, the network scatter plot (Fig. 6.D2) reveals that the device is sending an unusually large amount of data to a specific IP address. This is illustrated by the far right dot on the referred figure. A further analysis of the traffic using *Wireshark*[10] provides more details about the amount of the data that is transmitted. Consequently, an expert can choose in this situation *Wireshark* as the following suitable, DF tool to apply.

Correlate network activity with processes and file system activity. The high-level insight by using the scatter plot (Fig. 6.D2) brings up the need for a deeper analysis of the actual data content of the connections as it seems that the device is sending a lot of data to one specific host. In the node-link diagram (Fig. 5), an expert can see that a process establishes connections to an endpoint using port 443 and sends a considerable amount of data to this site. The related application at the iPhone is the web browser *Safari*. Since the device is not used at that time of investigation to browse websites that correspond to the displayed connection, it can be concluded that this is an unusual occurrence. To investigate this, an expert could decide to use, for example, the *Telerik Fiddler*[11] tool for a detailed, in-depth inspection of the traffic. By doing this, it is possible to confirm the previously gained knowledge with *Wireshark* and to increase the level of certainty of the evidence.

[10] `https://www.wireshark.org/`

[11] `https://www.telerik.com/fiddler`

Determine files for in-depth analysis. By observing and confirming large output data on specific connections that emerged from the network analysis in the steps before, the expert now goes further by taking a close look at the file system and created file versions. In Fig. 6.D3 the scatter plot shows that more files have been created but only few files have been modified during the period of investigation. This procedure makes it possible to see whether data has been copied to a temporary directory, are compressed (zip file) or split up, to be exported via exfiltration vectors (e-mail client, safari mobile). Such file operations and the copies of all created file versions with their content can further be investigated using the tool *The Sleuth Kit* which is the expert's decision for the next tool.

7.4 Memory & Data Analysis

Based on the selected forensic tools, the actual in-depth forensic analysis takes place here. In the considered use case, the memory and the extracted data of the iPhone X are examined with the tools. This enables the expert to find out more precisely what happened (e.g., fileless malware sent via WhatsApp). The interactions with the DF tools are not in the scope of our prototype.

7.5 Documentation

Our prototype brings together relevant data for a specific point in time to help experts deciding what DF tool is the most suitable in this specific situation. In our exemplary application of the prototype to the use case, only a small selection of tool decisions was presented. In the documentation phase, the data and the investigation proceedings are to be summarized in a report and conclude the analysis. This task is not directly addressed within our approach, but the supporting aspect of our prototype was illustrated.

8 Conclusion

The proposed and implemented research prototype provides cyber forensic experts with decision-support during an LDF investigation. Thus, the developed tool provides a solution that supports the initial selection of more specific forensic tools.

The design of the tool followed a problem-oriented approach. Further, the requirements of the application are sharpened through a requirements analysis which is divided into the areas (1) Data, (2) Users, and (3) Tasks. This methodical approach allows us to derive general design requirements applying to visual decision-support systems for live forensics. We implement the requirements in a prototype showcasing how they can support forensic analysts in investigating a mobile device. We extract relevant data from an iPhone within the prototype, pre-process this data, and display it in an interactive web application. The user interface features different possibilities for the experts to explore the data and identify targets for further, in-detail forensic analysis. Several bar charts and scatter plots are arranged and interactively interlocked with a node-link diagram to ensure this support of users' tasks. An exemplary use case underlines that the research prototype fulfills the requirements.

Only little empiric evidence of the design's effectiveness is currently available. It is yet to be evaluated in a real-world setting. To do this in an appropriate way, we are integrating our prototype into an existing professional learning, hands-on workshop for DF experts. The existing workshop will be extended to include our decision-support tool. This requires a considerable amount of additional work. However, in this way, workshop participants come into contact with possible visualizations supporting their work. Thus, empirical data can be collected about the prototype. Based on this data, further development of the prototype can then be carried out.

Acknowledgments

This work is partly performed under the BMBF DEVISE project which is supported under contract by the German Federal Ministry of Education and Research (`https://devise.ur.de/`).

REFERENCES

[1] M. Beran, F. Hrdina, D. Kouril, R. Oslejsek, and K. Zakopcanova. Exploratory Analysis of File System Metadata for Rapid Investigation of Security Incidents. In *2020 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pp. 11–20. IEEE, Salt Lake City, UT, USA, 2020. doi: 10.1109/VizSec51108.2020.00008

[2] F. Böhm, L. Englbrecht, and G. Pernul. Designing a Decision-Support Visualization for Live Digital Forensic Investigations. In A. Singhal and J. Vaidya, eds., *Data and Applications Security and Privacy XXXIV*, vol. 12122, pp. 223–240. Springer International Publishing, Cham, 2020. Series Title: Lecture Notes in Computer Science. doi: 10.1007/978-3-030-49669-2_13

[3] F. Böhm, M. Vielberth, and G. Pernul. 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. doi: 10.5220/0010225400980108

[4] B. C. Cappers, P. N. Meessen, S. Etalle, and J. J. van Wijk. Eventpad: Rapid Malware Analysis and Reverse Engineering using Visual Analytics. In *2018 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pp. 1–8. IEEE, Berlin, Germany, 2018. doi: 10.1109/VIZSEC.2018.8709230

[5] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. The Morgan Kaufmann series in interactive technologies. Morgan Kaufmann Publishers, San Francisco, Calif, 1999.

[6] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. van Liere, K. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. Data, information, and knowledge in visualization. *IEEE Computer Graphics And Applications*, 29(1):12–19, 2009.

[7] U. Dogrusoz, E. Giral, A. Cetintas, A. Civril, and E. Demir. A layout algorithm for undirected compound graphs. *Information Sciences*, 179(7):980–994, 2009. doi: 10.1016/j.ins.2008.11.017

[8] W. G. Eckert. *Introduction to forensic sciences*. CRC Press, Boca Raton, Fla., 1997. OCLC: 824195425.

[9] L. Englbrecht and G. Pernul. A Combined Approach for a Privacy-Aware Digital Forensic Investigation in Enterprises. *Journal of Cyber Security and Mobility*, 2021. doi: 10.13052/jcsm2245-1439.1012

[10] L. Englbrecht, S. Schönig, and G. Pernul. Supporting Process Mining with Recovered Residual Data. In *The Practice of Enterprise Modeling*, vol. 400, pp. 389–404. Springer International Publishing, Cham, 2020. doi: 10.1007/978-3-030-63479-7_27

[11] A. J. Ferrante. Project CATO. Technical report, 2019. https://assets.documentcloud.org/documents/6668313/FTI-Report-into-Jeff-Bezos-Phone-Hack.pdf.

[12] M. Franz, C. T. Lopes, G. Huck, Y. Dong, O. Sumer, and G. D. Bader. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, p. btv557, Sept. 2015. doi: 10.1093/bioinformatics/btv557

[13] K. Kent, S. Chevalier, T. Grance, and H. Dang. Guide to Integrating Forensic Techniques into Incident Response, 2006.

[14] T. R. Leschke and C. Nicholas. Change-link 2.0: a digital forensic tool for visualizing changes to shadow volume data. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security - VizSec '13*, pp. 17–24. ACM Press, Atlanta, Georgia, 2013. doi: 10.1145/2517957.2517960

[15] S. Mansfield-Devine. Fileless attacks: compromising targets without malware. *Network Security*, 2017(4):7–11, 2017. doi: 10.1016/S1353-4858(17)30037-5

[16] R. Marty. *Applied security visualization*. Addison-Wesley, Upper Saddle River, NJ, 2009. OCLC: ocn227921903.

[17] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model. *Information Visualization*, 14(3):234–249, 2015. doi: 10.1177/1473871613510429

[18] S. Miksch and W. Aigner. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics*, 38:286–290, 2014. doi: 10.1016/j.cag.2013.11.002

[19] N. R. Mistry and M. S. Dahiya. Signature based volatile memory forensics: a detection based approach for analyzing sophisticated cyber attacks. *International Journal of Information Technology*, 11(3):583–589, 2019. doi: 10.1007/s41870-018-0263-4

[20] T. Munzner. A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009. doi: 10.1109/TVCG.2009.111

[21] V. T. Nguyen, A. S. Namin, and T. Dang. MalViz: an interactive visualization tool for tracing malware. In *Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis*, pp. 376–379. ACM, Amsterdam Netherlands, 2018. doi: 10.1145/3213846.3229501

[22] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*, pp. 364–371. Elsevier, 2003. doi: 10.1016/B978-155860915-0/50046-9

[23] S. Simon, S. Mittelstädt, D. A. Keim, and M. Sedlmair. Bridging the Gap of Domain and Visualization Experts with a Liaison. In *Eurographics Conference on Visualization (EuroVis) - Short Papers*, pp. 1 –5, 2015. Artwork Size: 5 pages ISBN: 9783038680291 Publisher: The Eurographics Association. doi: 10.2312/EUROVISSHORT.20151137

[24] Sudhakar and S. Kumar. An emerging threat Fileless malware: a survey and research challenges. *Cybersecurity*, 3(1):1, 2020. doi: 10.1186/s42400-019-0043-x

[25] A. Ulmer, D. Sessler, and J. Kohlhammer. NetCapVis: Web-based Progressive Visual Analytics for Network Packet Captures. In *2019 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pp. 1–10. IEEE, Vancouver, BC, Canada, 2019. doi: 10.1109/VizSec48167.2019.9161633

[26] T. Wu, F. Breitinger, and S. O'Shaughnessy. Digital forensic tools: Recent advances and enhancing the status quo. *Forensic Science International: Digital Investigation*, 34:300999, 2020. doi: 10.1016/j.fsidi.2020.300999

7 Graph-based Visual Analytics for Cyber Threat Intelligence

Current status:	Accepted & Published	
Journal:	Cybersecurity, Volume 1, Number 16, December 2018	
CORE Ranking:	n/a	
Date of acceptance:	December 5, 2018	
Date of publication:	December 28, 2018	
Full citation:	BÖHM, F., MENGES, F., AND PERNUL, G. Graph-based visual analytics for cyber threat intelligence. *Cybersecurity 1*,16 (2018), 1–19	
Authors' contributions:	Böhm Fabian	45%
	Menges Florian	45%
	Pernul Günther	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.

Böhm *et al. Cybersecurity* (2018) 1:16
https://doi.org/10.1186/s42400-018-0017-4

Cybersecurity

RESEARCH Open Access

Graph-based visual analytics for cyber threat intelligence

Fabian Böhm*, Florian Menges and Günther Pernul

Abstract

The ever-increasing amount of major security incidents has led to an emerging interest in cooperative approaches to encounter cyber threats. To enable cooperation in detecting and preventing attacks it is an inevitable necessity to have structured and standardized formats to describe an incident. Corresponding formats are complex and of an extensive nature as they are often designed for automated processing and exchange. These characteristics hamper the readability and, therefore, prevent humans from understanding the documented incident. This is a major problem since the success and effectiveness of any security measure rely heavily on the contribution of security experts.

To meet these shortcomings we propose a visual analytics concept enabling security experts to analyze and enrich semi-structured cyber threat intelligence information. Our approach combines an innovative way of persisting this data with an interactive visualization component to analyze and edit the threat information. We demonstrate the feasibility of our concept using the Structured Threat Information eXpression, the state-of-the-art format for reporting cyber security issues.

Keywords: Cyber threat intelligence, Visual analytics, Usable cybersecurity, STIX

Introduction

Over the last years the number of IT security incidents has been constantly increasing among companies. In order to keep pace with this development, there is a necessity for ever-improving protective measures. As single entities are no longer able to handle the vast amount of possible attack scenarios acting collaboratively against such attacks is an emerging trend. It is widely believed that cooperative approaches, in particular those based on the exchange of threat intelligence information, can contribute significantly to improve defensive capabilities (Shackleford 2015). A key factor for realizing cooperative approaches are the underlying threat intelligence data formats. They offer a semi-structured representation of identified threats and ensure a common understanding of security-related observations. As they document incidents using general mark-up languages, a common characteristic of these formats is a good machine-readability.

However, text-intensive and semi-structured data is of very little use for security experts due to its extent and lack of human-readability. This is a major problem when taking the role of security experts in today's companies into consideration. As the success and effectiveness of incident prevention, detection, and reaction rely heavily on the knowledge of security experts (Shackleford 2016; Luttgens et al. 2014), they need to understand what happened, how to react appropriately, and how to prevent new outbreaks of cyberattacks.

Structured threat intelligence is of great value for experts as it enables them to understand threats and attacks. However, this is only possible when experts are able to read and analyze this information. It is further crucial for experts to easily edit it in order to

* Correspondence: fabian.boehm@ur.de
Department of Information Systems, University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany

include any additional or missing information. The interaction requires an integrity-proof approach to persist original data in order to ensure the availability of untampered evidence for possible subsequent court cases.

We propose KAVAS, a knowledge-assisted visual analytics concept for the Structured Threat Information eXpression (STIX). KAVAS enables security experts to analyze and enrich cyber threat intelligence (CTI) data. We combine a novel way of persisting this semistructured data in a graph-based database with an interactive visualization. To demonstrate the feasibility of KAVAS we utilize the state-of-the-art format for structuring CTI information, STIX 2. Our work aims to improve the accessibility of cyber threat intelligence for security experts and to include them in the process of creating a comprehensive documentation for security incidents.

The remainder of this paper is structured as follows. Section 2 introduces the background of our work with regard to related research fields. In Section 3 we analyze related work and reach out for introducing the addressed research gap. This chapter is followed by the description of applied concepts and design decisions we made for KAVAS in Section 4. After introducing the main concepts of KAVAS we proceed to showcase how our approach works in Section 5. Section 6 qualitatively evaluates the applied approach to make threat intelligence accessible to security analysts. We conclude in Section 7 by discussing our concept and identifying future work.

Background

This section provides an overview of the Structured Threat Information eXpression format STIX, which is the state-of-the-art project for semi-structured representation of cyber threat intelligence information. Furthermore, a general view on knowledge and its role in the field of visual analytics is given.

Structured threat information eXpression (STIX)

As argued above, structured formats are a key element within the threat intelligence exchange process because they pre-define which information can be shared. Additionally, these formats define requirements for the information density of the data to be shared. Depending on the specific use-case and the required contentual extent, the literature provides several formats that support structuring threat intelligence information. Examples for such formats are IODEF,[1] VERIS,[2] and STIX.[3] The primary focus of IODEF is the exchange of incident information between Computer Emergency Response Teams (CERTs), whereas VERIS focuses the measurement and management of risks involved in incidents. STIX 2, in contrast, is not bound to a specific use case and provides a comprehensive tool set for the representation of various information about incidents. As it is the format with the broadest possibilities in application (Menges and Pernul 2018), we focus our work on STIX 2 as the most recent version of STIX. This choice is further substantiated by STIX being the de-facto standard format for the exchange of threat intelligence information at present, which can also be anticipated for its successor STIX 2 in the near future (Shackleford 2015; Sauerwein et al. 2017). It provides the most extensive data structures among the available formats as shown by Asgarli et al. (Asgarli and Burger 2016) as well as by Menges and Pernul (Menges and Pernul 2018). This allows a wide ranging integration of expert knowledge into the analysis process. STIX 2 also provides highly flexible data structures allowing interactions of domain experts with very few limitations.

Regarding the content, STIX 2 provides a holistic representation for incident information, which is structured using the lightweight JavaScript Object Notation (JSON) file format. The data format provides two core component types: A STIX Domain Object (SDO) describing the characteristics of an incident and a STIX Relationship Object (SRO) describing relationships between those characteristics.

In its current version, STIX 2 specifies SDO elements for the representation of the attacking entity, event data describing the occurred incident as well as countermeasures initiated by the victim entity. The representation of the attacking entity includes information about the threat actor, the objectives, tools and attack patterns used within an attack. It also supports the description of entire attack campaigns and the attribution of attackers to such campaigns. The lateral movement of an incident can be represented using information such as exploited vulnerabilities, detected malware or digital identities involved in the incident. Actions taken to prevent an attack as well as responses to an attack can also be represented and associated to corresponding incidents afterwards.

Furthermore, STIX 2 specifies SRO elements to dynamically connect SDO elements. These connections can be realized using Relationship and Sighting Objects. Relationship objects indicate dependencies between SDOs, whereas Sighting objects refer to observed occurrences of SDOs. This allows building highly flexible representations for incidents only limited by the SDO definitions that are available within the data model (Piazza et al. 2017a; Piazza et al. 2017b). To encapsulate fully captured incidents, STIX 2 specifies an additional bundle element encapsulating all SDO and SRO elements captured in the course of an incident. Listing 1 gives a short example of a STIX 2 bundle.

```
{
  "type": "bundle",
        "id": "bundle--44af6c39-c09b-49c5-9de2
      -394224b04982",
  "spec_version": "2.0",
  "objects": [
    {
      "type": "threat-actor",
      "id": "threat-actor--9a8a0d25
          -7636-429b",
      "created": "2015-05-07T14
          :22:14.760144Z",
      "name": "Adversary Bravo",
      "description": "Is known to use phishing attacks",
      "labels": [
        "spy", "criminal"
      ]
    },
    {
      "type": "malware",
      "id": "malware--d1c612bc-146f-4b65 ",
      "created": "2015-04-23T11
          :12:34.760122Z",
      "name": "Poison Ivy Variant d1c6",
    },
    {
      "type": "relationship",
      "id": "relationship--ad4bccee-1ed3
          -44f5-9a56",
      "created": "2015-05-07T14
          :22:14.760144Z",
      "source_ref": "threat-actor--9 a8a0d25-7636-429b",
      "target_ref": "malware--d1c612bc
          -146f-4b65"
    }
  ]
}
```

Listing 1 Exemplary STIX 2 bundle

This listing shows the two SDO elements *threat-actor* and *malware* as well as the SRO element *relationship*, which connects the SDO elements using its properties *source ref* and *target _ref*. This example intends to illustrate the notation for objects and dependencies within the format as well as to give an impression of the possible complexity considering more extensive STIX 2 files.

Whenever the term "STIX" is used in the following sections, we actually refer to STIX 2.

Knowledge-assisted visual analytics

Visual Analytics (VA) is a combination of two important analytic reasoning processes: interactive visualization and automated analysis both striving to gain new insights (Keim et al. 2010). Keim et al. (Keim et al. 2008) define the creation of insight or knowledge as the final step in their widely accepted process for VA. This definition and other VA processes describe knowledge as a solely human artifact. However, not only humans own knowledge but a specific type of knowledge also exists for any automated analysis method included in VA (Fayyad et al. 2002; Sacha et al. 2014).

Therefore, knowledge-assisted visual analytics distinguishes the terms explicit and tacit knowledge (Nonaka and Takeuchi 1995; Polanyi 1983). Explicit knowledge can be defined as machine knowledge which can be read, processed, and stored by machines. Tacit knowledge is very specific to the individual and specialized as only humans are able to extract this knowledge type. In the context of knowledgeassisted visual analytics, tacit knowledge can be subdivided into smaller notions: 1) operational knowledge and 2) domain knowledge (Chen 2005). By having the appropriate operational knowledge a user knows how to interact with a visual analytics system. Domain or context knowledge is the ability of a user to interpret the visual representation regarding a specific context. Only a combination of these two types of knowledge enables users to understand the message told by a visual analytics system and thus to derive new knowledge (Chen 2005). Knowledge-assisted visual analytics aims to support the exchange of all these different knowledge types.

These exchanges can be formally described using knowledge conversion processes (Nonaka and Takeuchi 1995). Chen et al. (Chen et al. 2009) adapt these processes for information visualization. Wang et al. (Wang et al. 2009) as well as Federico et al. (Federico et al. 2017) further substantiate the concept of knowledge conversion to visual analytics with a special focus on explicit knowledge. The four conversion processes are namely: Internalization, Externalization, Combination, and Collaboration.

Internalization in knowledge-assisted visualization encompasses the transformation of explicit knowledge to tacit knowledge through visual interfaces. It supports humans in order to understand and transform explicit knowledge into domain knowledge (Wang et al. 2009). From a visualization perspective, this process is similar to the concepts of sensemaking (Pirolli and Card 2005) and insight or knowledge generation (Sacha et al. 2014; Chang et al. 2009). Internalization in terms of visualization can be described as follows: explicit knowledge is visually represented and through interactive exploration users gain tacit knowledge. Internalization is a high-level description of the generation of insight which is the primary goal and process of any visualization (Chen et al. 2009; Chang et al. 2009).

Externalization describes the transfer of knowledge along the opposite direction in contrast to internalization. It is a process where tacit knowledge is translated to explicit knowledge based on the insight of a user. There are existing prototypes in the visualization community showing that visualization tools taking externalization into consideration is suitable and effective for persisting and making use of experts' domain knowledge (Federico et al. 2017). Externalization can be applied using two main approaches. First, the more frequently applied approach is enabling users to directly transfer their knowledge. There exists a range of possibilities for implementing direct externalization. Examples are adjusting machine learning algorithms' parameters (Theron et al. 2017), adding patterns and rules to a knowledge database (Wagner et al. 2017) or changing an ontology used by automated analysis methods (Wang et al. 2009). Second, the other way to implement externalization is an implicit one by inferring explicit knowledge based on interactions of users with the visualization (Endert et al. 2012; Zhong et al. 2018). For example, dragging a node to a different location could be used to update and adjust the model of a clustering algorithm to fit the new position of the node.

Collaboration characterizes the exchange of tacit knowledge between humans (Wang et al. 2009). This process does not explicitly rely on computers and visualization as the most common form of sharing tacit knowledge is direct communication. However, collaboration can be supported through visual interfaces and the possibilities to externalize tacit knowledge and therefore, making it accessible for others at any time, supporting them to improve their own knowledge (Coleman et al. 1996).

Combination is a process where explicit knowledge from different sources is incorporated into an existing explicit knowledge system. It helps to improve available knowledge and to combine different bodies of explicit knowledge. This process is mostly independent from any visual representation of the explicit knowledge (Wang et al. 2009). However, users are integrated into this process by supporting the combination, identifying relations and finding inconsistencies or redundancies.

The development of knowledge-based interfaces and the representation of knowledge generated throughout the entire analytical process has been declared a key challenge for visual analytics research (Thomas and Cook 2005; Pike et al. 2009). However, in the domain of cyber security this is still underdeveloped.

Related work

Only few scientific publications tackle the problem of making threat intelligence information understandable for security experts by using visual interfaces. Even less work is available in the area of visual analytics systems specifically designed to display STIX.

Leichtnam et al. (Leichtnam et al. 2017) introduce a visualization approach for heterogeneous data sources. To transform the diverse data into a normalized model they derive a proprietary data model inspired by STIX. They build a visualization for their proprietary format. However, a visual representation for complex threat intelligence information documented with STIX itself is not provided.

A visualization displaying STIX in its full comprehensiveness is built by the STIX community itself.[4] This visualization builds a visual representation of a STIX bundle but lacks clear and structured design principles. Especially the functionality for security experts to convert their domain knowledge into machinereadable threat intelligence knowledge is missing.

While there is ongoing research in the area of structured formats for cyber threat intelligence (e.g. STIX) (Sauerwein et al. 2017) as well as knowledge-assisted visual analytics (Federico et al. 2017), there are, to the best of our knowledge, no efforts towards combining these two concepts in order to make threat intelligence information accessible for security experts.

In order to address this research gap, we define the following three requirements for our solution:

- **R1 - Handling complex threat intelligence data:** Enable integrity preserving storage and management of STIX as a notion of explicit knowledge in an appropriate database system rather than processing JSON files.
- **R2 - Visual representation of STIX:** Create an interactive visualization for STIX-based CTI information allowing security experts to derive knowledge and gain insights from an incident documentation.
- **R3 - Conversion of experts' knowledge:** Allow the exchange of explicit knowledge and security experts' tacit knowledge. Domain knowledge can be made available in the semi-structured STIX description of an incident by externalization. Therefore, the incident can be described more comprehensively and experts can benefit from each other's knowledge.

Our concept can be interpreted as a knowledge view in the information visualization framework introduced by Shrinivasan and van Wijk (Shrinivasan and van Wijk 2008) in 2008 to support analytical reasoning.

Concept and design

This section introduces the concept and design decisions made for the two main components of KAVAS: its persistence layer called Cyber Threat Intelligence Vault

(CTI Vault) to store and manage STIX as well as the corresponding visual analytics component to enable users to understand and interact with complex threat intelligence information. These concepts are aligned to the previously defined key requirements for KAVAS.

CTI vault

Hereinafter, we propose a concept for the persistence and handling of STIX cyber threat intelligence information.

R1 - handling complex threat intelligence data

STIX is designed as a graph-based model, which defines its domain objects as graph nodes and their relationships as edges. Therefore, we have chosen a graph database, as underlying technology in order to persist intelligence data appropriately.

> The CTI Vault serves as an extensible knowledge base, providing access for domain experts to the threat intelligence information, which can be seen as a notion of explicit knowledge. It represents a structured data storage for gathering captured incident data, which originate from individual files in JSON format. It serves as a technical foundation for storing incident information and additional domain expert knowledge, such as perceived similarities, differences and relationships between the different incidents.

Due to the dynamic data structures of STIX, the storage needs to provide capabilities for persisting data in a way that allows the integration of arbitrary relationships between the stored entities. Another essential requirement for the data storage is to assure integrity for the captured incident information. This is of special importance as the threat intelligence information could serve as piece of evidence in possible subsequent court cases. Therefore, it has to be ensured that interactions with domain experts will not distort any of the captured data, while preserving capabilities for enriching the captured data with additional information simultaneously.

To achieve these requirements, a differentiation between *inventory data* and *appended data* has to be made within the data storage. The inventory data, which represents the data foundation for incident information, describes all data that has been captured within an incident. The threat information contained in the stored entities as well as their relationships may not be changed after their initial storage and can consequently be considered constant. Therefore, this data has to be read-only. However, this is different for the use of appended data. These entities may be inserted, altered and deleted at any time and are intended to be connected with inventory data. Whenever information is edited, it has to be ensured that none of the operations performed on appended entities will influence the integrity of the inventory data.

The proposed concept is influenced both by the defined data structures within the STIX specification and the requirements for an interaction of domain experts with these data structures. However, the base requirement for the concept is the alignment to the STIX specification, to ensure the compatibility with the STIX data structures. This preserves the ability to exchange threat information with any endpoint compatible to STIX. Considering the requirements defined above, we firstly introduce an approach for persisting inventory data. This will be achieved by mapping the data available in the STIX data format, into a database representation.

The concept is subsequently extended by an approach for enriching the inventory data with appended data allowing the association of threat information to domain expert knowledge. Summarizing, the concept for handling complex threat intelligence data is based on the following two requirements, which will be specified in more detail afterwards.

- **R1.1 - Structured storage for threat intelligence data:** The collected data is stored in a structured way within a graph database as inventory data. The data storage has to be aligned to the STIX specification, allowing arbitrary relationships between the stored entities.
- **R1.2 - Integrity-proof storage and enrichment of persisted data:** A further requirement for the storage of threat intelligence data is to guarantee data integrity from insert operations onward. Moreover, subsequent update operations of the inventory data must not endanger its integrity. Therefore, it is mandatory to introduce a provenance process for every performed enrichment.

R1.1 - structured storage for threat intelligence data

To realize a concept of storing inventory data into the database, it is necessary to take a closer look at the STIX specification as well as to consider possibilities for the representation within a graph database.

The specification of STIX defines SDOs for the representation of threat intelligence information on the one hand and SROs defining relations between domain objects on the other hand. Both SDO and SRO are specified as stand-alone objects in STIX that allow to store multiple properties. According to the specification, SRO objects represent the relationships within the model by holding additional properties pointing to a source and target reference, each of which has to be a SDO. The combination of SDOs and SROs builds a directed graph,

in which the first ones represent graph node objects and the latter ones represent edges connecting these nodes.

In contrast to this type of representation, graph databases allow the use of object types for creating nodes, whereas edges cannot be represented using object types. This leads to the necessity of adjusting the type of representation within the database in order to properly translate STIX into the database representation. Our approach for adjusting these interrelations between the STIX objects is visualized in Fig. 1(a) and described in more detail afterwards.

Our concept defines the representation of both SDO and SRO as physical nodes within the graph database. While SDOs act as self-sufficient nodes, SROs represent the relationships between SDOs and, therefore, act additionally as logical STIX edges. Finally, information about the source and target attributes of the relationship is transformed into physical edges realizing these relationships within the database. This leads to a representation that fully maintains the structural integrity of the STIX data model on the one hand and allows to map relationship properties into logical edges on the other hand. Conclusively, this results in a logical representation for the directed-graph structure of STIX, which is stored using a physical non-directed graph structure within the database.

In addition to this, the STIX specification defines detected incident information to be pooled in relation to a root *bundle* element. Since the physical graph is non-directed, the bundle element can be connected to every SDO contained within an incident to achieve the pooling. This element can serve as an entry point for the traversal of incident information at the same time.

R1.2 - integrity-proof storage and enrichment of persisted data

Within the process of storing data into the CTI Vault, the integrity of captured data is essential to preserve its evidential significance for any subsequent forensic analyses or even for court cases. The proposed concept provides two different mechanisms to guarantee the integrity for stored incident information. On the one hand, the integrity of incident information has to be ensured when it enters the system for the first time, on the other hand, changes on persisted information have to be conducted in an integrity preserving manner. The integrity of inserted information is preserved using controlled redundancies. Inserted information will intentionally not be checked for redundancies to prevent any possible distortion of this data. The insertion of redundant data is possible, since the graph database assigns an internal unique identifier for every element inserted. This, in turn, prevents objects with the same content from producing collisions. However, delimitation for redundant

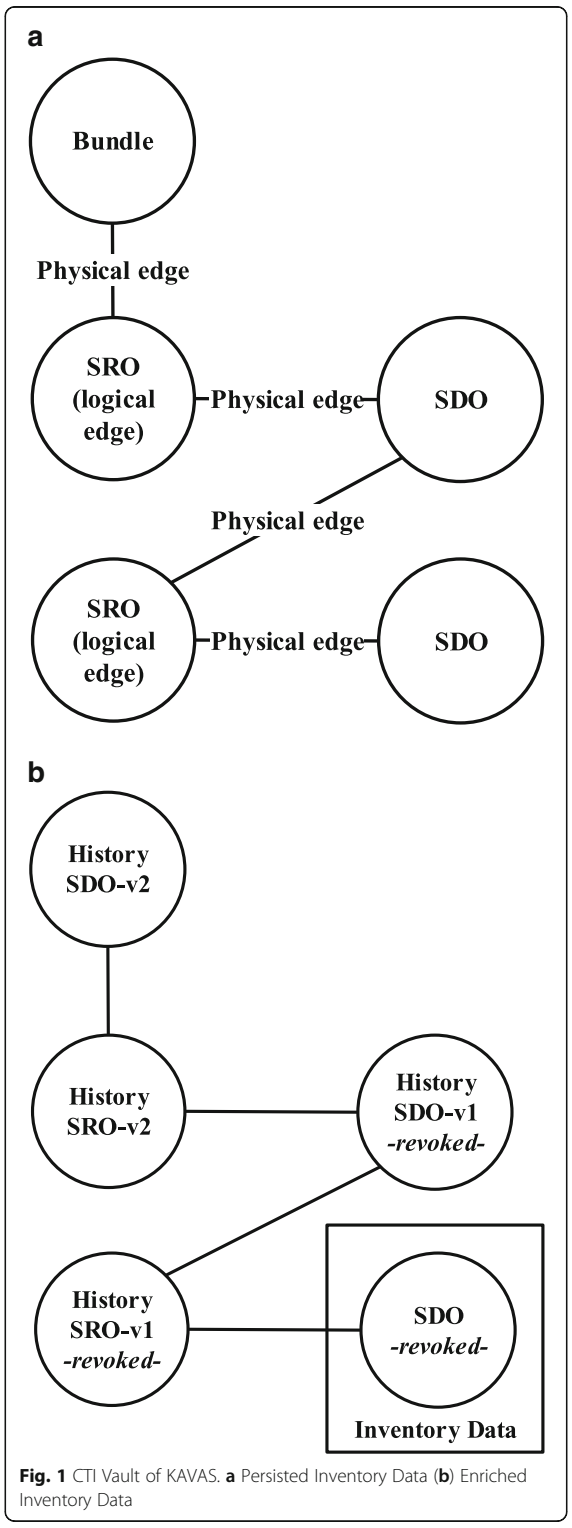


Fig. 1 CTI Vault of KAVAS. **a** Persisted Inventory Data (**b**) Enriched Inventory Data

objects remains still possible due to the pooling of elements and their affiliation to their root element, namely their bundle. The only exception for this are insertions of redundant elements within one bundle. However, this would only be the case if the elements contain identical STIX unique identifiers, which makes them both syntactically and semantically identical and consequently leads to a unification of these elements.

In addition to the concept of integrity-proof persisting for inventory data, the CTI Vault is designed to provide capabilities to store additional data that enriches the available information with domain knowledge of experts. Therefore, it needs to enable the extension of existing objects and relationships of inventory data. Since the enrichment of data with domain expert knowledge is not necessarily a singular event, the database also needs to provide capabilities for historicization of all performed changes.

As stated above, the concept of enriching inventory data is based on two main requirements. It has to be ensured that the inventory data will not be altered at any time and that the enriched data is still fully compatible to the STIX 2 specification. Consequently, the concept for enriching inventory data is also based on the STIX data structures.

According to this, only valid SDO or SRO elements that meet the STIX specification may be appended to the inventory data. Similar to the persistence of inventory data, appended data is also structured based on SDO nodes that are connected using logical and physical edges respectively. This results in a consistent database structure.

Figure 1(b) shows an exemplary SDO element within the inventory data extended by two subsequent changes, which are realized using a versioning structure within the database. In this process, supplementary nodes are added for each change. To indicate that nodes have been overwritten, the CTI Vault flags the respective former versions as “revoked” according to the STIX specification (Piazza et al. 2017b).

The first change is realized by creating a version SDO-v1 that extends the information within the original SDO, which is part of the inventory data. SDO-v1 in turn is connected to its base entity using a newly created relationship object SRO-v1. The second change is realized by creating a further version SDO-v2 and a corresponding relationship SRO-v2. It is important to maintain the order of succession for all changes performed. As a result, this concept enables every node within the inventory data to carry its own chain of edited data.

The presented concept for persisting cyber threat intelligence information in the STIX format fulfills therefore our requirement **R1**. This concept is the basis to support the *Combination* process as we interpret the STIX information stored in the CTI Vault to be explicit knowledge (Chen et al. 2009; Ackoff 1989).

Visualization design

The visual analytics component enables security experts to analyze, understand, and edit threat intelligence information. As described in Section 2.1, STIX is a powerful but text-intensive and semi-structured threat intelligence format. A single bundle can easily reach thousands of lines for complex incidents. This makes the documentation very hard to analyze and understand for security experts. This gets even worse when an expert appends information to the STIX file. In order to externalize domain knowledge, the complex structure of the format including all possible objects, relationships, their attributes, and allowed values for the attributes has to be known. To support the tasks of analyzing and enriching threat intelligence documented in STIX, we developed a visual analytics component on top of the previously introduced CTI vault.

Figure 2 shows the visualization component in the overall context of the system and defines the relations between KAVAS and security experts: the visualization uses the explicit knowledge stored in the CTI vault and maps this knowledge into an interactive view using the specification. The security experts can perceive the displayed knowledge to gain insight and situational awareness (Yen et al. 2014). At the same time they can use their operational knowledge to interact with the visualization in order to adjust the view specification or to enrich the information stored in the CTI vault.

R2 - visual representation of STIX

As STIX is designed to be a connected and directed graph of nodes and edges we are using a directed node-link diagram to represent knowledge persisted in the CTI Vault (Piazza et al. 2017a). This visualization technique is well suited for understanding threat intelligence as it reveals interconnections using nodes and edges (Severino 2018; Heer et al. 2010). Revealing the relationship between specific nodes (e.g. threat actors, used attack patterns and the targeted entities) is a crucial task of experts analyzing STIX. This makes the node-link diagram appropriate for the data structure at hand. However, Marty (Marty 2009) as well as Card et al. (Card et al. 1999) identify two main challenges when using node-link diagrams. To address those and to ensure the design of a suitable visual representation of STIX, we need to fulfill the following more specific requirements:

- **R2.1 - Render complex threat intelligence:** The cyber threat intelligence persisted in the IoC Vault

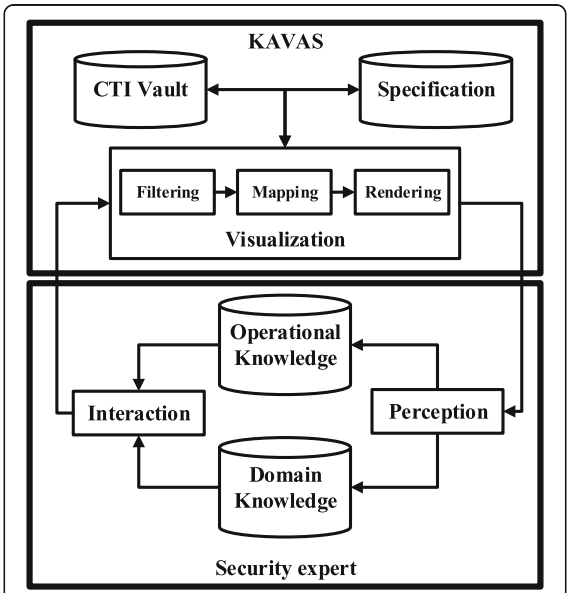


Fig. 2 High-level structure of KAVAS

is displayed in a suitable visual representation. The visual representation is fully capable to parse, map, and render all information provided in STIX bundles according to the STIX specification.

- **R2.2 - Scalable visual display:** As STIX bundles can contain hundreds of objects and even more links between these objects, the visual display has to be scalable. This can be assured by an appropriate layout algorithm and interactions for the users to adjust the layout.
- **R2.3 - Exploratory analysis:** To allow that users can deduce tacit knowledge from the displayed explicit knowledge, the visual representation must provide interactions supporting the analytical process of users.

R2.1 - render complex threat intelligence

The first challenge is to identify an appropriate way for positioning the nodes and links in the visualization space. SDO and SRO are abstract data constructs and do not have any natural position like on a geographical map. The InfoVis pipeline introduced by Card et al. (Card et al. 1999) is a process for creating views based on abstract data. By applying this process to SDOs and SROs, we are able to generate a visual representation of STIX. The following paragraphs describe our adaption of the InfoVis pipeline.

Originally, the pipeline starts with a data analysis responsible for data cleansing or interpolating missing values. We omit this step in our visual component as the CTI Vault is designed to persist only semantically and syntactically correct STIX bundles. Therefore, our view generation process starts with filtering the data to be visualized, as shown in the *Visualization*-box in Fig. 2. *Filtering* is realized by receiving a single userselected STIX bundle from the vault. This ensures that the analyst only sees information related to the bundle of interest. According to the InfoVis pipeline this single STIX bundle and the corresponding objects are referred to as focus data (Card et al. 1999).

The STIX objects in the focus data do not have any available positioning in the visualization space yet. Therefore, we need to transform the STIX-specific data structure into displayable nodes and edges in a mapping-step. As the STIX format defines SDOs to be nodes and SROs to be links in its graph-based structure, we adopt this definition. However, we had to make adjustments to improve the comprehensibility of a visually represented STIX bundle. We are displaying not only SROs as links in the node-link diagram but also important relationships embedded into SDOs referencing other objects. These embedded relationships are important to understand underlying connections in the threat intelligence information. For example, when an incident report is documented with STIX, embedded relationships of the report highlight which objects the report refers to. This and similar information can be important to an expert when analyzing an incident. To allow a fast perception of embedded links, we decided to include embedded relationships of SDOs as specially denoted edges into the diagram.

Additionally, we had to adjust the way STIX Sighting objects are represented in our visualization to retain a visually understandable way of representing STIX. These objects denote the insight that an attack, threat actor, campaign or other domain object was seen (Piazza et al. 2017b). They are used whenever an already documented attack is identified at another entity as well. Therefore, they are applied to track who was targeted as well as which attacks were performed. A Sighting object is specified to be a relationship. This means it would appear as a link in the visual representation although a Sighting is only connected to other SDOs via embedded relationships. We decided to include Sightings as nodes which are connected to SDOs via their different embedded relationships in the visual STIX representation to improve the perception of Sightings. These design decisions enable rendering all STIX objects as nodes and links on the canvas.

R2.2 - scalable visual display

Another issue of node-link diagrams is their limited scalability in terms of large numbers of highly connected nodes. They tend to resemble hairballs which makes it hard for users to understand the displayed information.

STIX bundles with large numbers of SDOs and SROs hamper a fast visual perception of relationships between the objects. However, a well-chosen layout algorithm and interactive functionalities for experts to adjust the layout can reduce this problem (Marty 2009). These functionalities are of great importance to ensure that a user is able to customize the visual representation of the STIX bundle. To arrange the information appropriately on the visualization canvas we apply a force-directed graph layout (Kobourov 2010). This algorithm creates a node-link diagram driven by different forces (e.g. gravity of node clusters, strength of links), which avoids overlapping as far as possible. However, due to the possible size and complexity of highly-interconnected STIX incident representations, it is necessary to provide interactive functionalities for security experts to adjust the layout themselves. This is especially necessary, when the automated force-directed algorithm is not capable to render a feasible layout anymore. In KAVAS we implement interactions allowing users to drag and drop single nodes and pin them to the desired position. Additionally, users can browse into specific parts of the STIX bundle by zooming. If the amount of nodes is overstraining the user, filters can be applied to show and hide the different types of SDOs and SROs.

R2.3 - exploratory analysis

Our concept allows security experts to interactively explore visually represented incident documentation. This exploratory analysis follows the Information Seeking Mantra defined by Shneiderman: "Overview first, zoom and filter, details on demand" (Shneiderman 1996). The *Overview* is provided by the initially generated node-link diagram based on the STIX intelligence information. With common interaction patterns like Pan-and-Zoom, hovering actions, filtering and Drag and Drop, security experts can adjust the view (Heer and Shneiderman 2012). This fulfills the *Zoom and filter*requirement of Shneiderman's mantra. *Details on demand* are displayed when an element of the node-link diagram is selected. By analyzing the visual STIX representation users broaden both their operational knowledge and their domain knowledge (Chen et al. 2009).

By implementing **R2.1**, **R2.2**, and **R2.3** in our approach, we are able to provide an interactive visual representation of the explicit knowledge embedded in the threat intelligence.

R3 - conversion of experts' knowledge

KAVAS allows the enrichment and editing of cyber threat intelligence while preserving the integrity of the original information at the same time. The enrichment and editing is necessary to externalize any additional or missing information from the user's domain knowledge. Preserving the integrity throughout this editing action allows the intelligence to serve as piece of evidence. In our approach, security experts are able to externalize their domain knowledge either through changing the attributes of existing SDOs and SROs or through adding new nodes and links. This functionality covers the *Externalization* process as users are able to transfer their domain knowledge to the CTI Vault, where it is preserved as explicit knowledge.

Our concept supports the *Collaboration* of several security experts by transforming it to explicit knowledge. This explicit knowledge can then be displayed to other users, which could further support them in their analysis of the incident. Thus, experts editing existing intelligence implicitly make their domain knowledge accessible for other users.

Visualization architecture

We adopted the classical Model-View-Controller (MVC) design pattern for the visual analytics component (Krasner and Pope 2000). This divides the application into three main interconnected parts to separate the internal representation of information and business logic from the visual presentation to a user. Figure 3 shows a high-level view on the MVC structure of the KAVAS visualization component. The MVC-structure of KAVAS shown in the figure is also aligned with the different steps of the InfoVis pipeline described earlier.

The *Database Connector* is the interface towards the available web services of the CTI vault enabling the visualization to retrieve threat intelligence data. It also enables the visualization to send updates to the database in case a security expert edited the STIX documentation. The visualization exchanges STIX-based documentations in JSON format with the vault.

The *STIX Parser* receives the JSON file from the *Database Connector*. It is responsible for parsing the file into instances of the SRO and the SDO data models. Both these models inherit a number of common properties every STIX object must contain. The models are specified in accordance with the STIX 2 specification (Piazza et al. 2017b). In addition to the simple attribute values, our models define the data type of the property and a description for the properties. They also define whether a property is required. All this information is extracted from the STIX specification to be able to parse CTI information from the vault and to create valid STIX documentations based on changes made by security experts. The model instances are held by the parser in two different lists; one containing relationship objects and the other containing domain objects. Parsing JSON into object instances has two main advantages: easy mapping and

rendering of objects into a node-link diagram as well as assuring compliance of processed STIX objects with the specification.

As pointed out earlier the abstract STIX data has no position in the diagram yet. The *STIX Mapper* maps the parsed STIX objects onto the visualization canvas. It wraps every instance of the beforehand described STIX models with a *NodeType* or *LinkType*. These data models contain additional properties (e.g. position, movement speed, etc.) to enable the *NodeLink Controller* to render the *NodeLink View*, which displays the interactive visual STIX representation. The *View Specification* tells the *NodeLink Controller* important settings such as the current zoom factor, gravity, link length, node radius and others.

The details of any STIX object can now be shown by handing over the selected *NodeType* to the *ObjectDetails Controller*. This controller then queries the object lists of the *STIX Parser* to receive the corresponding STIX object instance. This instance is forwarded to the *ObjectDetails View* for displaying details-on-demand. When an expert edits the STIX description, the parser receives the changes from the controllers, changes the model if necessary and forwards the changes through to the *Database Connector* to the CTI vault.

Prototype

In the following paragraphs we explain applied technologies for implementing KAVAS and give some detail of its functionalities with a short and small-scaled working exemplary bundle. A prototype of KAVAS is available here: http://bit.ly/2v9mSna (Sauerwein et al. 2017). Please note that KAVAS is currently an academic prototype. The linked version serves as a proof of concept. We are aware of required improvements to allow the operative use of KAVAS. The most emergent improvements are scoped for further versions of KAVAS and are described at the end of this article.

Applied technologies

The KAVAS visual analytics component is exclusively based on open-source web technologies forming a client-server web application in combination with the CTI vault (see Fig. 4). The CTI vault serves as back-end, providing the underlying data storage as described in Section 4.1 in combination with an API that enables data access for the front-end application. The vault is realized using the Java-based graph database Neo4j (Asgarli and Burger 2016) as base technology. Consequently, we also chose Java as language for realizing the access to the database as well as the related business logic managing the access. This layer assures the compliance to the object constraints predetermined by STIX, such as the specified object definitions and relationships. This is necessary, since the graph database does not provide such capabilities. In order to provide web-based access to the storage application, the actual Java implementation is running on a JavaEE[5] based application server. This allows us to provide REST webservices that can be accessed from the front-end application. The main technologies on the front-end are Angular.io[6] and Angular Material[7] which are frameworks on top of

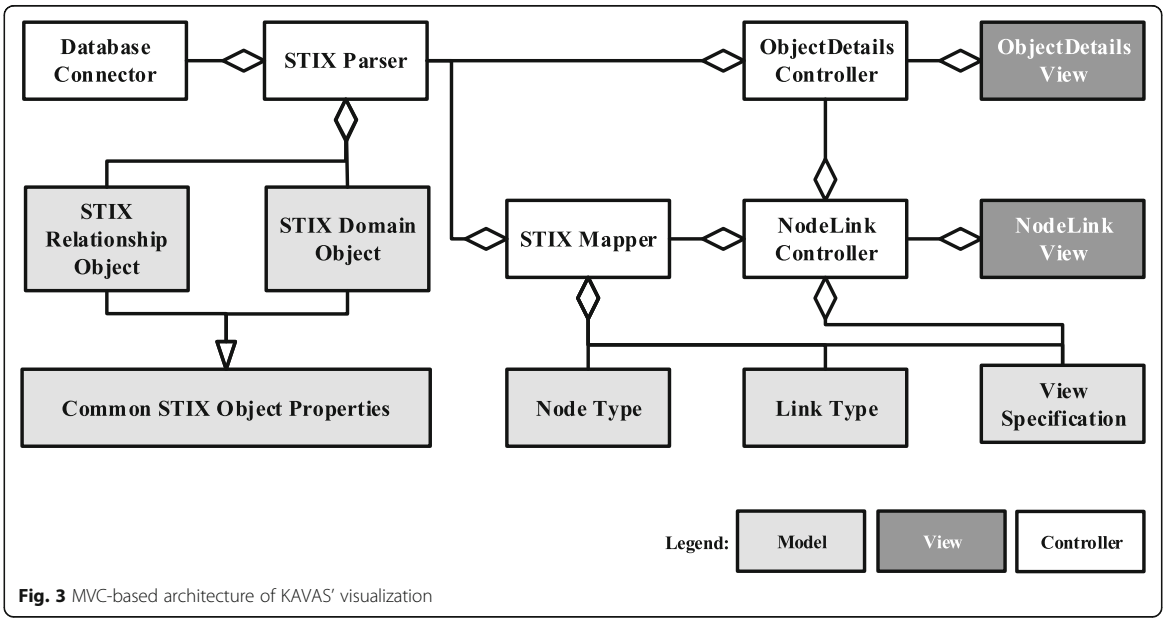


Fig. 3 MVC-based architecture of KAVAS' visualization

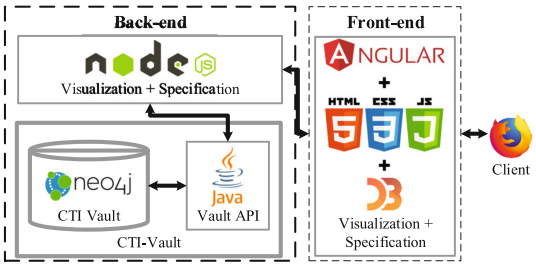


Fig. 4 Technologies used in KAVAS

HTML5, CSS3 and JavaScript. The interactive node-link diagram is implemented using the D3.js[8]-library.

Interactive exploration

Figure 5 displays different views of the visual interface of KAVAS. The bundle shown in the figure is part of the official example data sets for STIX 2.[9] Figure 5(a) shows the main view of the KAVAS visualization: an overview of a STIX bundle displayed as node-link diagram. The bundle itself documents an advanced persistent threat targeting the *Branistan Peoples Party (BPP)* which is one of the political parties of Branistan, a fictional country. The BPP's homepage is hit by an attack named *Operation Bran Flakes* where adversaries deploy *Content Spoofing* trying to insert false information into the BPP's web page. The campaign is rolled out by a *Fake BPP* which is most certainly sponsored by the *Franistan Intelligence* service, whereby Franistan is considered another fictional country. *The MITRE Corporation* detected and documented the attack.

An expert gets an overview (see Fig. 5(a)) of the STIX description in the node-link diagram after selecting the STIX bundle in the tool-bar's drop-down menu. The selected bundle is then received from the CTI vault, parsed and transformed for the visual display. To get a first glance of the documented incident, the expert can Pan-and-Zoom the diagram as well as drag and pin nodes to a fixed location on the canvas. Panning and zooming allows for interactive exploration. Dragging nodes across the canvas and pinning them to specific locations helps the analyst with adjusting the node-link diagram to be well arranged even for large numbers of nodes and edges. Whenever the mouse is moved over a node, KAVAS highlights the nearest neighbors of this node (see Fig. 5(b)). With enabling experts to select a node or link of the diagram and displaying the detailed properties of this STIX object (see Fig. 5(c)), KAVAS fully implements the Information Seeking Mantra for threat intelligence information.

Embedded relationships are not displayed as separated edges in Fig. 5(a). This is another functionality implemented in the visualization component. As described earlier, we map the embedded relationships of STIX objects as specially denoted edges. However, displaying all embedded relationships leads to incomprehensible diagrams very fast. Therefore, the embedded links as well as all other node or link types can be hidden or displayed interactively by the user.

Editing and enriching CTI information

Analysts can enable editing whenever they want to change or add any information to the bundle. When this mode is activated the view itself stays the same to keep the analyst in the existing visual metaphor. However, the interaction behavior is different. Clicking on the blank canvas now triggers the process for adding a node to the diagram. The first step in this process is selecting the STIX object type as it defines the properties of the SDO. KAVAS adds the node to the SDO list in the STIX parser and displays it on the canvas. Afterwards, the tool brings up the details-on-demand window and the user can now edit the information for the newly added object.

Instead of dragging a node as described earlier, clicking and moving the mouse with a node triggers the process of adding an SRO while in *Editing*-mode. If the mouse is released on a node, a new edge, with the starting node as source and the ending node as target is added to the canvas. From here on, the process for adding the SRO to the parser and the canvas is similar to adding a SDO. Finally, the user sees the newly created link highlighted and the editable details-ondemand window.

By clicking an existing node or link in editing mode the properties of this STIX object can be changed except for some properties, which by definition should not be element to any changes throughout the whole life-cycle of an object (e.g. its ID).

After the user clicks to save in the details-on-demand window, the input is checked for its conformity with the STIX specification. If the object is conform it is parsed into a compliant JSON. This happens regardless of whether a new object is added or an existing one is changed. Afterwards the JSON is sent to the CTI vault where the data is persisted.

When an expert starts editing a STIX bundle, this specific bundle is locked in the IoC Vault. Other users can still load the bundle from the vault to analyze the corresponding node-link diagram. However, they cannot switch to editing mode and they are notified that the bundle is currently edited by another user if they try to edit the bundle. When the editing user finishes the work on the bundle or closes the browser, the bundle is unlocked in the vault. This is possible as changes to the bundle are only possible on the level of SDOs and SROs which have to be saved separately after they were

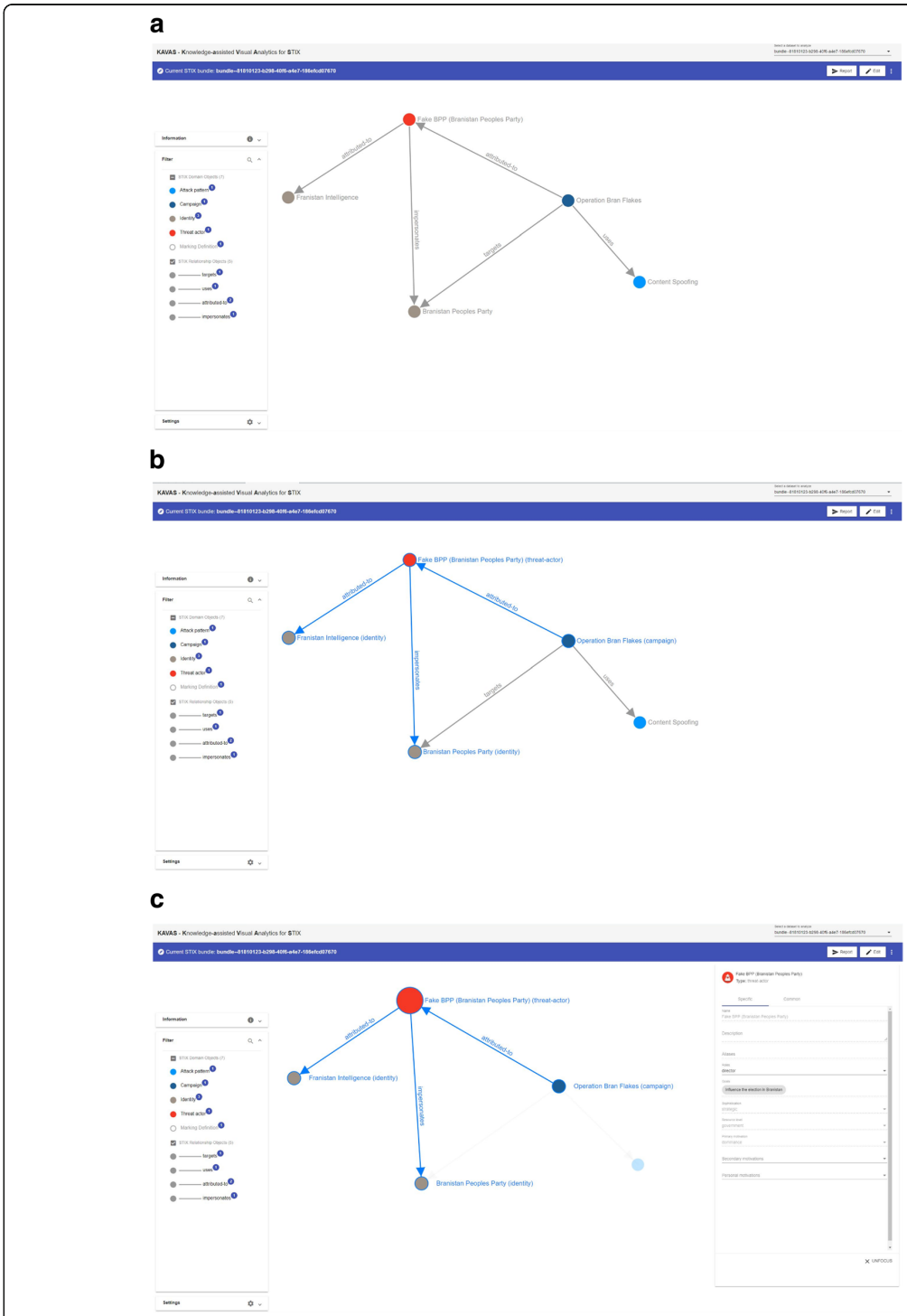


Fig. 5 Phases of Information Seeking Mantra in KAVAS A prototype of KAVAS is available at http://bit.ly/2v9mSna. The displayed bundle's id is *bundle–81,810,123-b298-40f6-a4e7-186efcd07670* and it can be accessed via the drop-down menu in the toolbar of KAVAS. **a** Overview (**b**) Zoom and Filter (**c**) Details on Demand

changed. Other users are now notified that the bundle is not locked anymore. When they activate the editing mode, the bundle is reloaded from the vault to ensure that they are working on the most recent version. They also can reload the bundle manually without switching the mode of action when they do not want to edit anything but still want to analyze the latest version of the STIX bundle.

Embedded knowledge processes

The KAVAS prototype is designed and implemented after a knowledge-assisted visualization approach. Therefore, the four knowledge conversion processes can be clearly identified within KAVAS' functionalities:

- *Internalization:* This knowledge conversion process describes the transfer of explicit knowledge into tacit knowledge through visual interfaces supporting humans to understand the explicit knowledge. KAVAS provides an interactive visual representation of explicit knowledge encompassed in the threat intelligence. In our system, internalization mainly happens through the interactive exploration of users. The node-link diagram and interaction functionalities aligned with the Information Seeking Mantra help users to inspect the knowledge and further support the discovery of unknown relationships and patterns which can become new domain knowledge.
- *Externalization:* Our concept allows tacit knowledge of domain experts to be externalized and persisted as explicit knowledge. Users can insert domain knowledge that does not yet exist in the threat intelligence information. Regardless of where the missing domain knowledge is originating, once acquired by the user, it can be directly inserted into the STIX bundle to augment threat intelligence. KAVAS allows this process through implementing means for users to directly edit the displayed STIX objects or add missing ones. Newly added information is persisted in the CTI Vault. After previously existing intelligence is changed, the original information is kept and linked to the updated version to ensure traceability of any changes to the STIX bundle.
- *Collaboration:* This process emerges when a user analyzes intelligence, which contains the externalized knowledge of other users. All available STIX information is persisted in the central CTI Vault and all intelligence displayed to the users is retrieved from this central intelligence storage. When one domain expert changes an incident description by editing existing intelligence or adding new pieces of information, this externalized knowledge is available for all other experts. Accordingly, having the CTI Vault as a centralized storage structure for all STIX intelligence and enabling users to externalize their domain knowledge, KAVAS supports the collaborative generation of tacit knowledge among its users.
- *Combination:* This process encompasses the insertion of new explicit knowledge into our existing knowledge base (CTI Vault), which is able to process any valid STIX bundle and to persist it. As a first step, it is highly important that the original bundle is stored regardless whether its information elements overlap with existing bundles. Hence, the bundle can be held in its original form and remains useful as possible evidence in court. After the initial storage of the original intelligence, further measures can be applied to detect and remove inconsistencies or redundancies. Currently, those measures are not yet part of the CTI Vault. However, the combination of existing explicit knowledge with new knowledge can be realized with our concept of the CTI Vault.

Evaluation

To validate our prototypical implementation of KAVAS and to provide first evidence of its usability and suitability to support knowledge conversion, we followed a two-fold research approach. An anonymous analyst survey validates the general suitability of the visualization approach for the addressed problem and eliminates usability issues of the interface. The survey is followed by expert interviews to confirm that KAVAS can facilitate knowledge conversions between domain experts and cyber threat intelligence.

Analyst survey

This survey intends to validate the relevance of the initial problem and the suitability of our design approach. Although, the survey cannot validate that the visualization facilitates all four knowledge conversion processes, it provides some hints whether the process of internalization is appropriately tackled.

Participants

The survey involved twelve security analysts from different academic institutions and companies such as internet service providers and security consultancies. The participants have a general understanding of threat intelligence. However, none of them is currently working with structured formats like STIX.

Design & Procedure

Staheli et al. (Staheli et al. 2014) propose a set of different aspects to evaluate visualizations for cyber security. Many of these aspects would need a more thorough user study. However, our survey is meant to give a first

indication on the suitability of KAVAS for making cyber threat intelligence accessible for human analysts. Based on the definitions proposed by Staheli et al. (Staheli et al. 2014) we assess the dimensions *User experience*, *Usability and Learnability*, *Insight generation*, and *Feature set utility*. The questionnaire encloses questions with informal character. Nevertheless, all questions are answered on an interval Likert scale ranging from 1 to 5 with the first and last numerical value being labeled with a textual description indicating the scale from *1: not at all* to *5: quite a lot*. The questionnaire includes the following five questions:

- **Q1:** Is the analysis and understanding of incidents relevant for your company/institution?
- **Q2:** Is the proposed visual tool effective for an investigation of threat intelligence information?
- **Q3:** Is the proposed visual tool clear and understandable?
- **Q4:** Is the proposed visual tool adequate to display and enrich the available incident information?
- **Q5:** Does the tool overall help to understand what happened during the described incident?

An additional open field allows participants to report any further comments or suggestions on the tool.

Before the beginning of the survey, the analysts are introduced to the tool, its features and our motivation to build it. Subsequently, a JSON representation of a synthetic incident as described in Section 5.2 is shown. By using the JSON representation we are able to highlight the main problem with STIX-based intelligence, which is the low readability and accessibility of the format. Afterwards, the participants explore the incident freely and are asked to fill out the questionnaire.

Results

Considering Fig. 6 and Table 1 we derive the fact that the addressed problem is relevant for the respective company or institution of the analysts. The high standard deviation leads to the conclusion that the need for sharing, exchanging, and analyzing threat intelligence is not prevalent throughout the participating organizations yet. The feedback on Q2 shows that a visual representation of threat intelligence is highly preferred over a text-based representation. From the answers to our third question about the usability of the proposed tool, we can conclude that the tool is indeed usable. However, we received some suggestions for improvement. Especially the analysts who answered Q3 with a score of 3 or lower, provided helpful feedback. For instance, one comment recommended that nodes should not bump back to their original position after dragging to adjust the layout of the node-link diagram permanently. This and further received feedback was implemented into the subsequent version of KAVAS after this survey and before the expert interviews. Feedback to the tool's suitability and adequacy with respect to editing threat intelligence information (Q4) is very positive, as well. Moreover, the feedback to Question Q5 shows that KAVAS improves the understanding of incidents within the target group.

Expert interviews

In order to get in-depth insight into the support of the knowledge-assisted concepts in KAVAS, we implemented the suggestions for improvement from the survey and used the revised prototype for interviews with security experts to conduct a more detailed evaluation. The main goal of these interviews is to validate that KAVAS helps security experts to understand threat intelligence and that existing information can be enriched with expert domain knowledge. By showing the fulfillment of our prototype in terms of these two requirements, we can confirm that KAVAS indeed facilitates the internalization and externalization knowledge conversion processes. The remaining knowledge conversion processes, combination and collaboration, both are implicitly implemented in KAVAS: Threat intelligence can be inserted into the CTI Vault at any time through an API (*Combination*). Additionally, experts can collaboratively gain knowledge through externalizing their knowledge and making it accessible for other users (*Collaboration*). Therefore, our interviews focus on the internalization and externalization knowledge conversion process.

Participants

The interviewees are represented by five security experts from different sectors. We conducted interviews with a Chief Information Security Officer and a security analyst of an international machine manufacturer, with a Chief Technology Officer of a SME operating in the area of secure cloud services, with a consultant from a security consultancy as well as with an academic researcher in the field of IT security. None of the experts participated in the previous survey. Each participant has a medium to high knowledge regarding threat intelligence, while three of them deal with threat intelligence and related structured formats like STIX on a daily basis. However, none of the interviewed experts obtains a visual representation to facilitate this work.

Design & Procedure

The interviews with the experts are designed to follow a semi-structured approach according to Lazar et al.

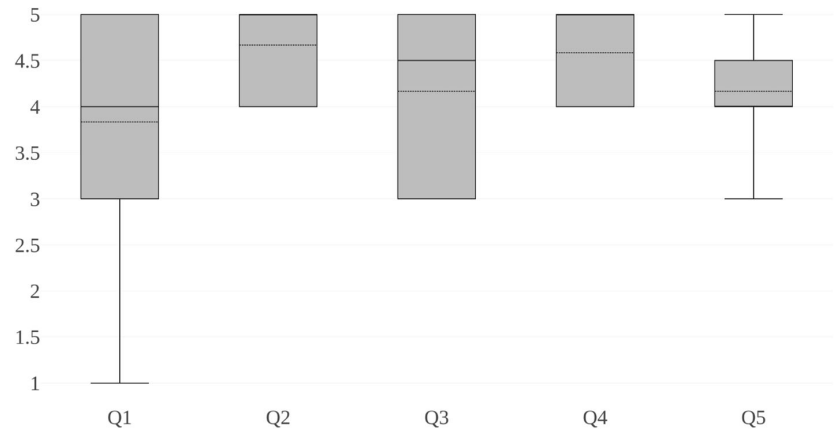


Fig. 6 Box plot reporting the KAVAS evaluation results. Solid lines mark the mean while dashed lines highlight the median

(Lazar et al. 2010). The interviews are separated into the following four phases:

- Phase 1) Introduction: At the beginning, every participant is questioned about their experience, such as their knowledge on CTI in general and on STIX. Afterwards, each expert receives a brief introduction into the STIX format and its problem of readability and accessibility. Thereby, the experts are asked to criticize any potential issues throughout the following interview phases. Next, each interviewee is guided to our prototypical web application. During the whole interview, the screen, of the participant using the tool, is shared with the interviewers.
- Phase 2) Internalization: To be able to test the intuitivity of the explorative analysis capabilities of KAVAS, the different interactive functionalities are not introduced in detail. The participants are asked to open a synthetic, previously designed STIX bundle (7 nodes, 8 links)[10] and to try to understand what happened in this bundle using the visual representation. In this phase, we pay special attention to the usage of interactions as well as to how the expert try to gain insight. After this first contact with KAVAS, the focus of the interview switches to a much more extensive bundle (65 nodes, 90 links).[11] With this bundle, we aim to discuss the scalability of the visual display in terms of the layout algorithm and the available interactions to adjust the layout. To conclude this phase of the interviews we ask for the experts' opinion on the tool so far and whether it supported them in understanding the threat intelligence information.
- Phase 3) Externalization: The focus of this phase is to test KAVAS' suitability to facilitate the externalization of domain knowledge, or more specifically, the insertion of new information and the modification of existing intelligence. To validate this with the interviewees, we provide a number of additional pieces of information and ask them to add this information to the previously explored smaller bundle. Again, we request them to give us feedback and criticize the tool whenever they have problems in understanding how it is working.
- Phase 4) Wrap-Up: The last phase of the interviews is dedicated to a summarizing discussion. Here, we discuss with the participant whether a more advanced version of KAVAS would be applicable to operative deployment and the conditions thereto. Finally, we collect a list of features and functionalities the interviewees find useful for improving the prototype.

Table 1 KAVAS survey results

	Q1	Q2	Q3	Q4	Q5
# Answers	12	12	12	12	12
Mean	3.83	4.67	4.17	4.58	4.17
Std dev	1.34	0.49	0.94	0.51	0.58
Min	1	4	3	4	3
Median	4	5	5	5	4
Max	5	5	5	5	5

Results

The interviews lasted between 45 to 70 min, which was mainly due to the summarizing discussion, where the experts brought up a lot of interesting points reaching from possible improvements of STIX itself to functionality features of KAVAS necessary for operative deployment in an organization. The results of the conducted

interviews are presented in the following, divided according to the four phases described before.

- Phase 1) Introduction: At the beginning of each interview the participants are asked general questions to obtain basic data about the interviewees. Therefore, they are asked about their company as well as their exact role within the company. Furthermore, they are asked about their knowledge of Cyber Threat Intelligence and the STIX format in particular to determine their level of expertise. This first phase showed, that even though interviewees are familiar with threat intelligence information in general, they are rather unfamiliar with the specifics of the STIX format in most cases. Table 2 gives an overview on these general information about the interviewees.
- Phase 2) Internalization: Within this phase, the interviewees are asked to take a look at a predefined STIX bundle and to understand the contents of the presented incident. The interviews showed that KAVAS supports users to quickly understand an incident without having any previous knowledge. Especially the included filter functions of KAVAS turned out to be particularly helpful in this context. The consistently positive feedback within this phase showed, that the chosen representation is both suitable for representing incident information and makes it easily available for the user.

  However, this phase also revealed some disadvantages and problems with the graph visualization in general and the realization in particular. While hassle-free usage was possible on large resolution displays, it turned out that problems arise when working on lower resolution displays, especially for handling larger datasets. The interviewees also missed some functionalities. For instance, they asked for advanced filter functions for different use-cases such as filtering the k-nearest neighbor nodes within specific tree sections. The interviews further revealed that existing filters and possible interactions with the user interface to re-structure the layout prove themselves as very useful features. It was also shown that the interface could be improved by implementing some additional features, such as on-demand windows displaying further information for objects with their associated relationships and an improved initial structuring of the presented graph representation. Altogether, the interviews show that KAVAS has a high utility for security specialists to convey and understand incident information. This manifested both in the assessment of the approach in general and the usability of the tool itself. However, it was also stated that a special training for employees might be necessary to cope with the complexity of STIX data. The interviewees also considered the tool to probably be helpful for practical usage. In this context they could for example think of a feed service to obtain incident information from a central authority, which could be used to understand attacks and prevent them from happening.

- Phase 3) Externalization: Within this phase, the interviewees are asked to use KAVAS to enrich the incident representation with additional, predefined knowledge made available by the interviewers. The process of editing information overall turned out to be mostly intuitive and easy to use for the experts.

  Adding and editing nodes was perceived as intuitive by all participants, whereas some participants argued that editing relationships was a bit counter-intuitive when working with the tool for the first time. The fact that KAVAS distinguishes between explore and edit mode was perceived differently by the participants. While some accentuated the benefits of this clear separation, others found it cumbersome. However, the tool could be helpful to collect and enrich forensic evidence in e.g. CERT or incident response teams reconstructing how an incident compromised an organization. In this context, it was envisioned that this tool could especially be helpful within team meetings to collaboratively collect and edit threat intelligence information. It was also accentuated that there is most likely a need for integrity-proof intelligence data in the foreseeable future. Altogether, the enrichment of intelligence data was overall easy to use for the participants and mostly intuitive. The interview reveals that editing intelligence information is equally important to analyzing it. Moreover, the interviewees highlighted that there is an actual need for this feature within companies.

- Phase 4) Wrap-Up: Within the last phase, possible scenarios and conditions for an operative deployment of KAVAS and possible improvements for the prototype were discussed.

Table 2 General information on the interview participant

	Position	Business Branch	Organization's size	CTI Knowledge	STIX Knowledge
#1	Security Researcher	Academia	ca. 5.000	high	medium
#2	Chief Information Security Officer	Manufacturing	ca. 15.000	high	high
#3	Security Analyst	Manufacturing	ca. 15.000	medium	low
#4	Chief Technology Officer	Secure Cloud Services	ca. 60	medium	medium
#5	Senior Consultant	Security Consultancy	ca. 20	low	low

One key problem revealed by the interviews is the question how threat intelligence data can be acquired. This concerns both the acquisition from external sources and the question how threat intelligence data can be produced within the company. In this context, it was also argued that there is a need for an automated generation of basic intelligence data that can be enriched by experts using tools like KAVAS afterwards. Integrating external intelligence feeds, cooperatively analyzing threat data as well as creating visual threat reports seems to be beneficial for companies. The interviewees also suggested several additional features to improve the user interface. These, for example, include improved highlighting for important and editable attributes or additional filter functions. Furthermore, the interviewees named some additional object properties that were necessary for practical usage, such as additional timestamps defining the point in time when the object was detected. These are not defined within the current STIX standard and consequently not available in KAVAS.

Discussion

The results of the conducted interviews show that KAVAS provides the ability for internalization and externalization of threat intelligence information. Given the fact, that it is still in the stage of a proof of concept prototype, the experts' feedback was already good. Furthermore, the experts provided several suggestions for future improvements of the tool.

The interviews also demonstrated that there is a strong interest for visualizing threat intelligence information among companies. The experts already have several use-cases for this kind of application in mind. However, the question of how to generate intelligence data in the first place remains.

Moreover, the interviews also showed that there are several weaknesses in the STIX standard, which became obvious while evaluating KAVAS. An example for this is the absence of a top-level element to represent and structure specific company assets such as IT systems affected by an incident.

Conclusion and future work

Conclusion

In this work we presented KAVAS, a concept for interactive visual analytics of threat intelligence information. Our approach persists information in a graph database to maintain an integrity-preserving data structure. This database is connected to a visual interface supporting security experts in understanding and analyzing incident descriptions. Additionally, the visual analytics component of KAVAS facilitates the process of including the knowledge of the security experts into CTI information. KAVAS achieves this with its functionalities to edit existing descriptions and adding new knowledge allowing for more thorough incident documentations.

While designing KAVAS, and especially its visual component, we aimed to follow the concept of knowledge-assisted visual analytics. More precisely we designed our concept to support the four main knowledge conversion processes which are essential to improve the collaboration of human and machines. *Internalization* is done in KAVAS by visually representing the incident documentations stored in the CTI vault. This way, the explicit knowledge in the CTI vault is accessible for security experts and they can gain knowledge using the visualization. KAVAS also supports *Externalization* as it allows for editing the STIX bundles. The tacit knowledge is externalized when the expert edits the threat intelligence information visually displayed in KAVAS. Being implemented as graph database the CTI vault has the essential functionalities to support the *Combination* knowledge conversion. This process is implemented in KAVAS as the CTI vault can be fed with new threat intelligence information and it includes this newly available knowledge into the existing knowledge base. A similar process in KAVAS supports the *Collaboration*. As externalization of an expert's tacit knowledge is possible, other security experts can profit from the externalized knowledge of each other providing an implicit form of collaboration.

The application KAVAS described throughout this work, clearly fulfills the three requirements we started with:

- **R1 - Handling of complex threat intelligence data:** The CTI Vault persists STIXbased threat intelligence information in a graph database. It additionally provides the possibilities to store

externalized user knowledge in its knowledge base, while the integrity of the original information is preserved and ensured. Moreover, any data stored in the vault is compliant with the STIX format at any point in time.
- **R2 - Visual representation of STIX:** KAVAS' visual component can display threat intelligence and enables security experts to interactively explore incidents and gain insight about what happened.
- **R3 - Conversion of experts' knowledge**: As described above, KAVAS provides functionalities for each of the four knowledge conversion processes.

Fulfilling all the stated requirements, KAVAS offers a flexible platform for sharing, analyzing, annotating and visualizing cyber threat intelligence information based on the STIX data format.

Future work

Although we met the previously defined requirements for KAVAS, some challenges remain, which have to be addressed in future work.

A key challenge for future work regarding the CTI Vault will be the analysis of STIX data to find interconnections and redundancies between different bundles, which currently are standalone object pools, not attached to each other. Enabling the interconnections between and the merging of bundles could contribute greatly to the usage of STIX features. Additionally, this would improve the quality of available threat intelligence information. Examples for this are the merging of different incidents into a whole campaign of attacks and the determination of correlations between observed events within different incidents. The process for merging bundles and finding redundancies has to be subject for further research as it is a challenging task to identify interconnections and quality problems across independent bundles.

Additionally, there are some potential improvements regarding the functionalities of the visual component. During the interviews, the participants highlighted the need for a number of different advanced filters as well as some other features, which would help them even more to work with complex threat intelligence. Furthermore, experts should be included into the process of merging and connecting bundles. KAVAS could also be extended to support more sophisticated collaboration features for security experts like annotating CTI information to exchange domain knowledge in a more direct manner.

Another important future challenge regarding our proposed visual analytics tool is a comprehensive user study to quantify its effects on the work of security experts. These effects need to be quantified. Also the tool's impact on the quality of threat intelligence documentation has to be measured as expert knowledge can be externalized with KAVAS. Currently, KAVAS is only validated in terms of being able to work with the very limited examples provided by the OASIS committee and by a qualitative evaluation to show its feasibility. The main reason for this small-scaled evaluation is the lack of available real-world threat intelligence data being documented with STIX 2 up to this point in time. Its predecessor STIX 1 is the industry-wide state-of-the-art for documenting this type of information and we presume that it is very likely for STIX 2 to achieve the same amount of acceptance in the near future. Since the specification of STIX 2 is still under development, it is not reasonable to evaluate the effectiveness and efficiency of KAVAS in a comprehensive and quantitative manner yet.

Another topic for future work has to be the analysis and assurance of data quality among STIX bundles. As STIX supports collaborative efforts to maximize the number of prevented cyberattacks, the data quality of the incident descriptions is crucial. This is becoming even more true when the information is analyzed and enriched by human operators. High quality information is essential to ensure trust. Therefore, existing data quality metrics have to be applied on STIX-based descriptions to assess the added value they provide. Moreover, visual metaphors for these metrics have to be added to the KAVAS visual representation helping analysts to assess the trustworthiness of the information.

Endnotes

[1] https://trac.tools.ietf.org/html/rfc7970
[2] http://veriscommunity.net
[3] https://stixproject.github.io
[4] https://github.com/oasis-open/cti-stix-visualization
[5] https://www.oracle.com/technetwork/java/javaee/overview
[6] https://angular.io
[7] https://material.angular.io
[8] https://d3js.org
[9] https://oasis-open.github.io/cti-documentation/example/defining-campaign-ta-is/
[10] http://bit.ly/2NLDn3W
[11] http://bit.ly/2xX74EO

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).

Funding
Not applicable.

Availability of data and materials
Source code - CTI Vault

- Project name: CTI Vault
- Project home page: http://bit.ly/2LKFcgT

- Archived version: 1.0-SNAPSHOT
- Operating system(s): Platform independent
- Programming language: JavaEE
- Other requirements: Glassfish Application Server 4.1.1, JavaEE 6 or higher
- License: GNU GPL v3

Source code - Visual analytics component

- Project name: Visual analytics component
- Project home page: http://bit.ly/2LVn6YM
- Archived version: 1.1.0
- Operating system(s): Platform independent
- Programming language: HTML, Typescript
- Other requirements: Apache Webserver or similar, NPM 6.2.0 or higher
- License: GNU GPL v3

Authors' contributions

FM carried out the design and implementation of the CTI Vault. FB carried out the design and implementation of the visual analytics component. FM and FB conducted the evaluation and drafted the manuscript to equal parts. GP participated in the design of the different components and the study. GP also helped to draft the manuscript revising it critically for important intellectual content. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 6 August 2018 Accepted: 6 December 2018
Published online: 28 December 2018

References

Ackoff RL (1989) From data to wisdom. Journal of applied systems analysis 16(1): 3–9
Asgarli E, Burger E (2016) Semantic ontologies for cyber threat sharing standards. In: IEEE Symposium on Technologies for Homeland Security (HST)
Card SK, Mackinlay JD, Shneiderman B (eds) (1999) Readings in information visualization: using vision to think. Morgan Kaufmann, Burlington
Chang R, Ziemkiewicz C, Green TM, Ribarsky W (2009) Defining insight for visual analytics. IEEE Comput Graph Appl 29(2):14–17
Chen C (2005) Top 10 unsolved information visualization problems. IEEE Comput Graph Appl 25(4):12–16
Chen M, Ebert D, Hagen H, Laramee RS, van Liere R, Ma K, Ribarsky W, Scheuermann G, Silver D (2009) Data, information, and knowledge in visualization. IEEE Comput Graph Appl 29(1):12–19
Coleman J, Goettsch A, Savchenko A, Kollmann H, Wang K, Klement E, Bono P (1996) Teleinvivo™: towards collaborative volume visualization environments. Computers & Graphics 20(6):801–811
Endert A, Fiaux P, North C (2012) Semantic interaction for visual text analytics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York
Fayyad U, Grinstein GG, Wierse A (2002) Information visualization in data mining and knowledge discovery. Morgan Kaufmann, Burlington
Federico P, Wagner M, Rind A, Amor-Amorós A, Miksch S, Aigner W (2017) The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In: Proceedings of IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE Computer Society Press, Los Alamitos
Heer J, Bostock M, Ogievetsky V (2010) A tour through the visualization zoo. Communications of the ACM 53(5):59–67
Heer J, Shneiderman B (2012) Interactive dynamics for visual analysis. Queue - Microprocessors 10(2):30
Keim D, Andrienko G, Fekete J-D, Görg C, Kohlhammer J, Melan.con G (2008) Visual analytics: definition, Process, and challenges. In: Information visualization. Lecture notes in computer science, vol 4950. Springer, Berlin, Heidelberg
Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. (eds.): Mastering the information age: solving problems with visual analytics, Goslar (2010)
Kobourov SG (2010) Force-directed drawing algorithms. In: Tamassia R (ed) Handbook of graph drawing and visualization. CRC Press, Boca Raton
Krasner GE, Pope ST (2000) A description of the model-view-controller user interface paradigm in the smalltalk-80 system. Journal of object oriented programming 1(3):26–49
Lazar J, Feng JH, Hochheiser H (2010) Research methods in human-computer interaction. Morgan Kaufmann, Burlington
Leichtnam L, Totel E, Prigent N, Mé L (2017) Starlord: Linked security data exploration in a 3d graph. In: IEEE Symposium on Visualization for Cyber Security (VizSec)
Luttgens JT, Pepe M, Mandia K (2014) Incident Response & Computer Forensics, 3rd edn. McGraw-Hill Education Group, Whitby
Marty R (2009) Applied security visualization. Addison-Wesley, Boston
Menges F, Pernul G (2018) A comparative analysis of incident reporting formats. Computers and Security 73:87–101
Nonaka I, Takeuchi H (1995) The knowledge-creating company: how Japanese companies create the Dynamcis of innovation. Oxford University Press, Oxford
Piazza R, Wunder J, Jordan B (2017a) STIX™ version 2.0. Part 1: STIX Core concepts. OASIS committee
Piazza R, Wunder J, Jordan B (2017b) STIX™ version 2.0. Part 2: STIX objects. OASIS committee
Pike WA, Stasko J, Chang R, O'Connell TA (2009) The science of interaction. Information Visualization 8(4):263–274
Pirolli P, Card S (2005) The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: Proceedings of International Conference on Intelligence Analysis McLean, VA, USA
Polanyi M (1983) The tacit dimension. University of Chicago Press, Chicago
Sacha D, Stoffel A, Stoffel F, Kwon BC, Ellis G, Keim D (2014) Knowledge generation model for visual analytics. IEEE Trans Vis Comput Graph 20(12): 1604–1613
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
Severino, R.: The data visualisation Catalogue (2018). https://datavizcatalogue.com/index.html. Accessed 2018-08-03
Shackleford D (2015) Who's using Cyberthreat intelligence and how? SANS institute, Swansea
Shackleford D (2016) SANS 2016 Security Analytics Survey. SANS Institute, Swansea
Shneiderman B (1996) The eyes have it: A task by data type taxonomy for information visualizations. In: Proceedings of the 1996 IEEE Symposium on Visual Languages. IEEE Computer Society Press, Los Alamitos
Shrinivasan YB, van Wijk JJ (2008) Supporting the analytical reasoning process in information visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York
Staheli D, Yu T, Crouser RJ, Damodaran S, Nam K, O'Gwynn D, McKenna S, Harrison L (2014) Visualization evaluation for cyber security. In: IEEE Symposium on Visualization for Cyber Security (VizSec). ACM, New York
Theron R, Magán-Carrión R, Camacho J, Fernandez GM (2017) Network-wide intrusion detection supported by multivariate analysis and interactive visualization. In: IEEE Symposium on Visualization for Cyber Security (VizSec). IEEE Computer Society Press, Los Alamitos
Thomas JJ, Cook KA (eds) (2005) Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE Computer Society Press, Los Alamitos
Wagner M, Rind A, Thür N, Aigner W (2017) A knowledge-assisted visual malware analysis system: design, validation, and reflection of Kamas. Computers &Security 67:1–15
Wang X, Jeong DH, Dou W, Lee S-W, Ribarsky W, Chang R (2009) Defining and applying knowledge conversion processes to a visual analytics system. Computers & Graphics 33(5):616–623
Yen J, Erbacher RF, Zhong C, Liu P (2014) In: Kott A, Wang C, Erbacher RF (eds) Cognitive Process. Springer, Cham
Zhong C, Yen J, Liu P, Erbacher RF (2018) Learning from experts' experience: toward automated cyber security data triage. IEEE Systems Journal:1–12

8 Measuring and Visualizing Cyber Threat Intelligence Quality

Current status:	Accepted & Published
Journal:	International Journal of Information Security, Volume 20, February 2021
CORE Ranking:	**C** (http://portal.core.edu.au/jnl-ranks/587/)
Date of acceptance:	February 8, 2020
Date of publication:	March 2, 2020
Full citation:	SCHLETTE, D., BÖHM, F., CASELLI, M., AND PERNUL, G. Measuring and visualizing cyber threat intelligence quality. *International Journal of Information Security 20*, 1 (2021), 21–38
Authors' contributions:	Schlette Daniel 35% Böhm Fabian 35% Caselli Marco 20% Pernul Günther 10%

Journal Description: The International Journal of Information Security is an English language periodical on research in information security which offers prompt publication of important technical work, whether theoretical, applicable, or related to implementation.

International Journal of Information Security (2021) 20:21–38
https://doi.org/10.1007/s10207-020-00490-y

REGULAR CONTRIBUTION

Measuring and visualizing cyber threat intelligence quality

Daniel Schlette[1] · Fabian Böhm[1] · Marco Caselli[2] · Günther Pernul[1]

Published online: 2 March 2020

Abstract
The very raison d'être of cyber threat intelligence (CTI) is to provide meaningful knowledge about cyber security threats. The exchange and collaborative generation of CTI by the means of sharing platforms has proven to be an important aspect of practical application. It is evident to infer that inaccurate, incomplete, or outdated threat intelligence is a major problem as only high-quality CTI can be helpful to detect and defend against cyber attacks. Additionally, while the amount of available CTI is increasing it is not warranted that quality remains unaffected. In conjunction with the increasing number of available CTI, it is thus in the best interest of every stakeholder to be aware of the quality of a CTI artifact. This allows for informed decisions and permits detailed analyses. Our work makes a twofold contribution to the challenge of assessing threat intelligence quality. We first propose a series of relevant quality dimensions and configure metrics to assess the respective dimensions in the context of CTI. In a second step, we showcase the extension of an existing CTI analysis tool to make the quality assessment transparent to security analysts. Furthermore, analysts' subjective perceptions are, where necessary, included in the quality assessment concept.

Keywords Cyber threat intelligence · Threat intelligence sharing · Data quality · Threat intelligence formats · Information security visualization

1 Introduction

The last years have seen the emergence of sharing information about threats, cyber attacks, and incidents by organizations. The urge to join forces in the fight against cyber criminals originates from an ever-increasing number of attacks and the related risks for organizations [1,2]. Not only the number but also the complexity of attacks has increased over the years resulting in successful intrusions with more severe forms of security breaches. For individual organizations, it is an almost impossible task to detect these complex and decentralized attacks on their own. Thus, organizations share their available information about incidents and attacks. This information is referred to as cyber threat intelligence (CTI).

However, investigations show that inaccurate, incomplete, or outdated threat intelligence is an important challenge for collaborating organizations [3,4]. More recently, empirical studies with domain experts emphasize that ensuring CTI quality throughout the collaboration process is crucial for its continuing success [5,6]. The exchange and utilization of meaningful threat intelligence depends on measuring and ensuring its quality. This necessity is strengthened as the quality of shared information is stated to have an impact on the required time to respond to an incident [7].

Additionally, it is important to inform stakeholders about the quality of individual CTI artifacts [5]. This can help analysts to narrow down available information to the intelligence actually requiring their attention. Therefore, analysts can come to better informed decisions how to react to incidents reported within the CTI. The other way around, the domain knowledge of security analysts is a very promising source for the "fitness for use" [8] of a CTI artifact. Including experts into the process of measuring quality of threat intelligence is a starting point to assess contextually depen-

Fabian Böhm
Fabian.Boehm@ur.de

Daniel Schlette
Daniel.Schlette@ur.de

Marco Caselli
marco.caselli@siemens.com

Günther Pernul
Guenther.Pernul@ur.de

[1] University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany

[2] Siemens AG, Otto-Hahn-Ring 6, 81739 Munich, Germany

dent data quality (DQ) dimensions. To leverage the domain knowledge of experts, it is necessary to make the data quality assessment transparent to them. In a further step, users should be allowed to contribute their own perception of threat intelligence quality which increases the trust into both platform and threat intelligence [9].

This work centers on two aspects making a contribution to measuring cyber threat intelligence quality. We present a first approach to assess relevant quality dimensions of a standardized CTI data format. For this purpose, we first derive relevant DQ dimensions for CTI and define metrics which allow to measure these dimensions. The metrics are then configured to the STIX format as they rely on its structure. We further differentiate metrics which can be calculated automatically and metrics where input of domain experts is needed. Thereupon, we extend our previously proposed open-source CTI analysis tool to convey CTI data quality to security analysts. The extension helps to provide an indication about the quality of the CTI artifact at hand. Our extension also demonstrates how security analysts can contribute to CTI quality assessment through an interactive visualization.

The remainder of this work is structured as follows: Sect. 2 gives an overview of related work in the field of cyber threat intelligence data quality. A brief introduction to the STIX 2 format can be found in Sect. 3. This section additionally provides an example to illustrate the format, the concept of CTI sharing, and related quality issues. In Sect. 4, we select and structure relevant DQ dimensions. Metrics for the assessment of these dimensions in the context of the specific format are configured in Sect. 5. In Sect. 6, we propose an extension of the STIX format for CTI quality and a possible approach to communicate this quality to users of a CTI analysis tool. This section also describes interviews we conducted with security experts to gain feedback on the proposed approach. Our article concludes in Sect. 7 with a short summary and possible future research directions.

2 Related work

Although CTI and especially quality of CTI are not yet extensively researched topics in the information security field, some related work has already been conducted. We give a short overview of this work hereinafter.

Dandurand and Serrano [10] are among the first to define requirements for a CTI sharing platform. The requirements for such a platform include some form of quality assurance and the provision of adjustable quality control processes. The authors, however, do not specify quality dimensions or metrics to assess the quality of the CTI in their proposed infrastructure.

In 2014, Serrano et al. [11] point out that there is missing support for quality control and management in existing CTI sharing platforms. The authors propose that organizations should install quality control processes to provide multiple measurable quality values. Although the need for quality assessment is discussed, it is not described how such an assessment could be implemented into a platform.

Sillaber et al. [5] perform a series of focus group interviews and discussions with threat intelligence experts. They derive a number of findings on how data quality dimensions influence threat intelligence. They do not identify fundamentally new data quality issues specific to the CTI area. However, the authors give several recommendations for future research and for possibly relevant data quality dimensions. This work does not propose an explicit approach to measure DQ in the CTI context but rather stays on a generic level.

In their survey investigating threat intelligence, Tounsi et al. [7] specifically call for methods to evaluate the quality of threat intelligence. This also applies to the wider organizational security operations center (SOC) context as low-quality CTI is identified to be a pivotal issue [12]. To the best of our knowledge, there is no respective academic work addressing these open issues. Furthermore, none of the currently available commercial threat intelligence sharing platforms is actively measuring CTI quality [7]. With this work, we aim to take a first step into this direction.

3 Structured threat information expression (STIX)

First, this section gives a brief overview of the STIX format. This is necessary as following sections rely on a fundamental understanding of format specifics. The second part introduces a motivational example which is intended to illustrate the STIX format and basic processes of a CTI sharing platform. This example highlights the importance of evaluating CTI quality in the context of a centralized sharing platform with multiple participants.

3.1 STIX format

We base our approach to assess CTI quality on the STIX 2 data format defined and maintained by the OASIS consortium.[1] According to recent analyses, STIX is the de facto standard used for CTI [13,14]. The successor of this format is called STIX 2. It is likely that STIX 2 will reach a similar popularity throughout the next years as it is the format with the most extensive application scenarios [14]. Therefore, our quality assessment is built upon this promising format. Whenever the term "STIX" is used in the remainder of this work, we actually refer to STIX 2.

[1] https://oasis-open.github.io/cti-documentation/.

STIX is a machine-readable, semi-structured format based on JavaScript Object Notation (JSON)[2] to structure and exchange cyber threat intelligence. The format provides two main object types:

1. STIX Domain Objects (SDOs) describing characteristics of an incident and
2. STIX Relationship Objects (SROs) describing the relationships between those characteristics.

SDOs and SROs contain a number of common attributes which are part of any STIX object and additional attributes specific to the respective object type. Common attributes are IDs or the type of the object, whereas exemplary-specific attributes are the motivation of an attacker or the version identifier of a tool.

The current specification of the format conveys twelve SDO types [15]. These allow to provide a holistic view of a cyber incident including both high-level attribution intelligence (e.g., the associated attack campaign or the threat actor) and low-level information (e.g., the data indicating the attack and exploited vulnerabilities).

There are two types of SROs. The first SRO type allows to connect any two SDOs with an explicit relationship highlighting e.g., the vulnerability exploited by a malware. Both can be modeled as SDOs, whereas the logical connection between them is expressed by an SRO. The second SRO type denotes that a specific SDO has been identified. It connects this SDO with an SRO describing the evidential data for this assumption.

SDOs and SROs relevant for a specific threat or incident can be encapsulated by a report. The SDO for this purpose is the *Report* object which references all, respectively, relevant SDOs and SROs.

3.2 Motivational example

In this section, we describe a fictional CTI sharing platform which is used by critical infrastructure providers (e.g., hospitals, energy operators, etc.) to exchange threat intelligence artifacts. Although the platform and the providers in our example are fictional, there is a number of real-world sharing platforms comparable to the described one. The specific characteristics and operation modes of the platform are not relevant to our example which is why we chose a fictional setting. The main goal of the following explanations is to describe the central idea and necessary processes of a CTI sharing platform.

Starting the example depicted in Fig. 1, we can think of a power plant operating a state-of-the-art security operations center (SOC). At some point in time, the alerting mechanisms

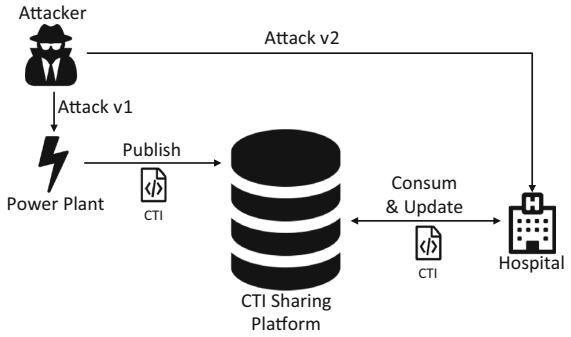


Fig. 1 Simplified CTI Exchange Platform structure

of the plant's intrusion detection systems (IDS) indicate an ongoing attack affecting various critical systems. Automated systems start the collection of related information through log file and network traffic analyses. Immediately, security experts start their analysis to protect the plant's cyber systems and to gain as much insight into the attack as possible.

The outcome of automated and manual analyses in the form of collected, attack-related data casts a light on what seems to be an unknown APT. Various machines of the power plant have been compromised and connected to several control units outside of the internal network. The related IP addresses as well as configuration files have been identified. Additionally, the attackers exploited known but unpatched vulnerabilities of a web server and a specific version of an operating system to spread their attack. This allowed them to conduct lateral movement in the organization's network without being noticed. To defend the network and remove the malware, security analysts applied appropriate countermeasures.

Part of the power plant's SOC is the active participation on a CTI sharing platform. On this platform, several operators of critical infrastructure collaborate to improve their cyber defense. Most of these collaborative efforts are based on exchanging intelligence about previously unknown threats or by sharing new insights about existing incidents. There are different roles of participants active on the platform: Publishers post CTI artifacts on the platform, whereas consumers process these artifacts. However, participants of a sharing platform usually hold both these roles simultaneously.

As the power plant's analysts did detect a new type of attack, they transform the gained insights into a STIX report which is published on the sharing platform. The CTI contains the identified threat actor, exploited vulnerabilities, and the deployed malware. Additionally, the analysts include indicators of compromise (file hashes, IP addresses, and the like) to help other participants to detect this attack. They also share the applied countermeasures.

[2] https://www.json.org/.

A simplified example of the STIX artifact shared by the power plant is shown in Listing 1. Please note that some aspects of the example are not fully aligned with the current STIX specification due to readability reasons.[3] However, the example allows to gain a better understanding of STIX. The shared CTI contains the identified *Threat Actor*, the deployed *Malware*, the exploited *Vulnerability*, and an *Indicator* referring to the respective malware file. Additionally, the *Relationships* between these entities are shown. For example, these relationships point out that the *Threat Actor* uses the *Malware* to target a *Vulnerability*.

Another user of the CTI sharing platform might be the operator of a hospital. The operator is leveraging the knowledge made available on the platform to improve the hospital's resilience to cyber attacks. Therefore, published indicators of attacks from the platform are automatically fed into the operator's intrusion detection systems. Additionally, security experts of the operator carry out manual analyses on the most relevant CTI artifacts to identify possible threats. The manual analysis of the artifacts is performed through a visual interface as the CTI format used by the platform is not easily readable for humans.

```
{
  ''type'': ''threat-actor'',
  ''id'': ''threat-actor--1'',
  ''created'': ''2019-04-07T14:22:14Z
     '',
  ''modified'': ''2019-04-07T14:22:14Z
     '',
  ''name'': ''Adversary Bravo'',
  ''description'': ''Is known to
     manipulate critical
     infrastructures, I suppose'',
  ''labels'': [ ''spy'', ''criminal'' ]
},{
  ''type'': ''malware'',
  ''id'': ''malware--1'',
  ''created'': ''2019-04-07T14:22:14z
     '',
  ''modified'': ''2019-04-07T14:22:14Z
     '',
  ''name'': ''Malware d1c6'',
},{
  ''type'': ''vulnerability'',
  ''id'': ''vulnerability--1'',
  ''created'': ''2019-04-07T14:22:14z
     '',
  ''modified'': ''2019-03-07T14:22:14z
     '',
  ''name'': ''A Webserver Vulnerability
     ''
},{
  ''type'': ''indicator'',
  ''id'': ''indicator--1''
  ''created'': ''2019-04-07T14:22:14Z
     '',
  ''modified'': ''2019-04-07T14:22:14Z
     '',
  ''labels'': [''malicious-activity''],
  ''pattern'': ''[ file:hashes.'SHA
     -256' =
     '4bac27393bdd9777ce02453256c5577c
       d02275510b2227f473d03f533924f877
       ']'',
  ''valid_from'': ''2019-04-07T14:22:14
     Z''
},{
  ''type'': ''relationship'',
  ''id'': ''relationship--1'',
  ''created'': ''2019-04-07T14:22:14Z
     '',
  ''modified'': ''2019-04-07T14:22:14Z
     '',
  ''source_ref'': ''threat-actor--1'',
  ''target_ref'': ''malware--1'',
  ''relationship_type'': ''uses''
},{
  ''type'': ''relationship'',
  ''id'': ''relationship--2'',
  ''created'': ''2019-04-07T14:22:14Z
     '',
  ''modified'': ''2019-04-07T14:22:14Z
     '',
  ''source_ref'': ''indicator--1'',
  ''target_ref'': ''malware--1'',
  ''relationship_type'': ''indicates''
},{
  ''type'': ''relationship'',
  ''id'': ''relationship--3'',
  ''created'': ''2019-04-07T14:22:14Z
     '',
  ''modified'': ''2019-04-07T14:22:14Z
     '',
  ''source_ref'': ''malware--1'',
  ''target_ref'': ''vulnerability--2'',
  ''relationship_type'': ''targets''
}
```

Listing 1 Exemplary STIX 2 artifact

The power plant's CTI artifact is analyzed by the hospital's security personnel only a few months after the respective incident. This is mainly because vast amounts of available CTI hinder the security experts to identify threat intelligence relevant for them. During the analysis of the artifact published by the power plant, the responsible security analyst of the hospital spots that the respective attack targets a software in use by the hospital as well. Subsequent network and endpoint analyses indicate that the hospital has been affected although the IDS seems to have not noticed the compromise as the binaries of the malware have changed in the meantime. In addition, although the same software is in use, the version number proclaimed to be exploited at the power plant seems to be invalid.

During the analysis of the incident at the hospital, analysts come across some changes and additional insights into the attack. Additionally, the proposed countermeasures are not sufficient to get rid of the attacker. Therefore, an updated ver-

[3] Object IDs are not in UUIDv4 format, and some mandatory schema structures are left out.

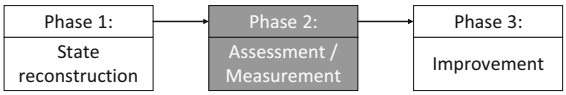


Fig. 2 Process steps of DQ methodologies [16]

sion of the CTI artifact is published to the platform to ensure each participant is informed about the advanced version of the cyber attack. However, during this process the information about the threat actor is unintentionally duplicated leading to redundant information.

The example above shows that the timely exchange of high-quality CTI is crucial for the effort of organizations to prevent cyber security breaches. However, there are numerous pitfalls regarding the quality of the shared threat intelligence. Examples from the above-described use case are: 1) inaccurate information caused by input errors made during the documentation of an attack (invalid version of exploited software), 2) outdated information caused by delays in CTI propagation (changed binaries of malware), or 3) duplicated information caused by collaboration (redundant description of threat actor). Even the overload of CTI available to human analysts and their incapability to determine the most relevant CTI can be seen as a data quality problem. Each of these examples stresses the urge to measure CTI quality and to visualize the results for human analysts.

4 Approach for CTI quality assessment

General DQ methodologies consist of three main process steps depicted in Fig. 2. Initially, the collection of necessary data is performed. Data sources and involved costs are fundamental building blocks for the following process steps. The second step includes the identification and measurement of relevant quality dimensions in the context where the methodology is applied. After quantifying data quality, the last process step strives to improve the quality following a fixed set of techniques and strategies. Although there is no cohesive methodology for information quality management of CTI yet, this work solely focuses on measuring DQ in the context of CTI as highlighted in Fig. 2. Up to now, existing work has mostly provided general advice for mainly the first and the last methodology step but has not described approaches to actually measure CTI quality [5,13,17]. We put explicit focus on the quality assessment. We thus assume the existence and availability of the necessary data for assessment.

Our work on selecting and structuring DQ dimensions relevant for CTI is the result of an iterative process in which we actively sought input and feedback from a number of CTI researchers and practitioners, e.g., domain experts from computer emergency response teams (CERTs). Throughout multiple evaluation iterations the relevant dimensions and their structure as described in the following two subsections were consequently adapted according to the input of the experts.

4.1 Selecting relevant DQ dimensions for CTI

Before introducing measurements for CTI quality, relevant DQ dimensions have to be selected. Extant work has already suggested a wide variety of different DQ dimensions referring to either the data values or the data schema [18]. The literature distinguishes three main approaches for proposing general and abstract quality dimensions: the theoretical approach [19], the empirical approach [20], and the intuitive approach [21].

Considering the different approaches and various DQ dimensions, it is not an easy task to select relevant and applicable dimensions for a problem at hand. Following the empirical approach by Wang and Strong [20], related research such as the work of Sillaber et al. [5], Sauerwein et al. [13], or Umbrich et al. [22] identify a first set of relevant dimensions which is refined throughout this work.

Our resulting set of dimensions is shown in Fig. 3. An interesting finding yielding from the discussion with the CTI experts is the high complexity of the *Appropriate amount of data* quality dimension. This dimension is meant to help experts to decide whether a CTI artifact by any chance could contain helpful information. In general, this decision can only be made by comparing the real-world artifact with its CTI description. However, this is rarely possible. Therefore, another approach is needed to give security analysts an indication for this dimension. Throughout our discussions, it turned out that experts are often basing their decision on the diversity of SDO types and their interconnection in a STIX report. Arguably, homogeneous SDO types and few relationships between them lead to experts' perception that the report does not describe the real-world incident properly. For the in-depth examination of the *Appropriate amount of data* quality dimension we refer to Sect. 5.3.

4.2 Structuring DQ dimensions for CTI

Our goal for DQ assessment in the context of CTI is to come up with measures to quantify the selected dimensions and aggregate them into a combined score for a STIX report. We therefore structure the dimensions in three different levels depending on the input data as shown in Fig. 3. The assessment of the dimensions on the "Attribute Level" operates on specific attributes of STIX objects, e.g., the dimension of *Timeliness* can be assessed using the *modified*- and *created*-attributes of STIX objects. The two dimensions located on the "Object Level" in Fig. 3 are not bound to predefined attributes

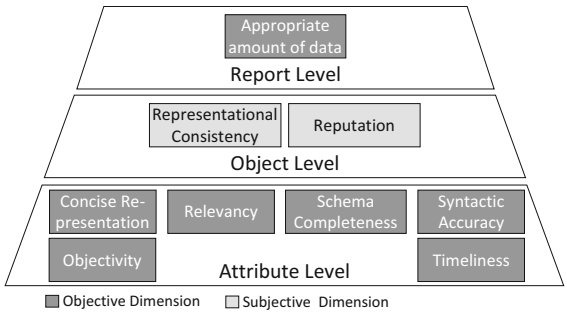


Fig. 3 Schematic of the structure of DQ dimensions

of the objects. In fact, they can be measured based on either varying attribute sets (*Representational Consistency*) or the object as a whole (*Reputation*). At the highest level ("Report Level"), we propose a final dimension to cope with experts' requirement to be informed about whether a report is likely to contain an *Appropriate amount of data* as described in the paragraph above.

Individual scores on both attribute and object level are then aggregated to a combined object quality indicator. This aggregation provides a quick and helpful insight for any user navigating through cyber threat information. Artifacts with a high-quality score are probably the ones to analyze first. Additionally, on a "Report Level" this aggregation allows to inform users about the average object quality in a given report. This is accompanied by an indication whether the report contains an appropriate amount of data. However, as DQ dimensions can be of varying importance for different users the aggregation has to be customizable [11]. Adjustable aggregation parameters enable CTI users to define the weight of each DQ dimensions in the procedure of calculating a quality indication for each STIX object. The corresponding metrics for aggregation are further outlined in Sect. 5.4.

Additionally, to these various levels for the DQ dimensions, we differentiate objective and subjective dimensions which are also indicated in Fig. 3. DQ has to be evaluated with objective measurements as well as from subjective perception [16,23,24]. Objective measurements rely on mathematical and quantitative measures to evaluate an artifact's quality. However, some dimensions of DQ are dependent on their contextual environment. It is thus necessary to incorporate the requirements, experience, and knowledge of domain experts. When it comes to the decision whether data is of high quality regarding a specific use case or context, objective DQ dimensions fail to provide reasonable quality scores [25]. At this point, it is necessary to incorporate subjective measures as a supporting concept. Here, the assessment of an artifact's quality is based on qualitative evaluation by data administrators and other experts.

In the context of a CTI sharing platform, the concept of subjective perception and domain knowledge to evaluate various DQ dimensions equally applies. While domain knowledge is a necessary input for subjective quality dimensions, it also supports assessment of objective DQ dimensions. The domain knowledge can be captured through a system similar to a reputation system where users provide their perception about the quality of an object or report [26]. The need for a reputation system to include subjective quality perceptions and to increase trust is also highlighted in empirical studies [5]. Subjective quality assessment in the CTI sharing context can originate from different stakeholders of a respective platform: On the one hand, consumers (security experts, analysts, etc.) contribute with their domain knowledge and their organization-specific background; on the other hand, a platform host can act as a trusted third party contributing to the quality assessment.

Overall, these three levels provide good and transparent indicators for the quality of a STIX-based cyber threat intelligence artifact. For indication of individual DQ dimensions, we adopt and extend existing naming conventions [20].

5 Measuring CTI quality

In this section, we elaborate on suitable DQ dimensions as the result of our studies. For each dimension, its applicability to the CTI context is described and respective metrics for assessment are configured. Those assessments are either of an objective or a subjective nature depending on whether they can be automated or need manual input. Subjective metrics are based on the perceptions and expressions of a CTI sharing platform's participants. Furthermore, there is a number of objective dimensions which benefit from additional manual input of domain experts. The ordering of the metrics follows the previously outlined structure of the dimensions in Sect. 4.2.

The proposed metrics in the following are again the result of an iterative process collaborating with CTI researchers and practitioners. Several metric configurations result from long discussions with domain experts where a lot of very valuable feedback was provided highlighting possible configurations to assess CTI quality.

Configurations for the metrics are based on the formal ground truth defined in Eqs. 1–5. We formally define two different attribute sets of STIX as A_r (i.e., Eq. 1) and A_o (i.e., Eq. 2). Required attributes a_r, for example, are unique IDs, names, labels, and types which are present in most STIX objects. As for optional attributes a_o, characteristics such as descriptions, versions, and external references are referred to Eq. 3 which defines any STIX Domain Object or STIX Relationship Object as a specific subset of both the available required and optional attributes. This subsequently allows us

to describe the objects O held by a CTI sharing platform as a set of objects where each object o is either a SDO or SRO (i.e., Eq. 4). STIX objects such as *Threat Actor*, *Malware*, or *Indicator* belong to the set of SDOs, while *Relationship* and *Sighting* objects are SROs. When an incident or an attack is reported to the platform, the resulting report r is defined by Eq. 5 to be a subset of all objects persisted in the platform.

$$A_r = \{a_r \mid a_r \text{ required in STIX 2}\} \tag{1}$$
$$A_o = \{a_o \mid a_o \text{ optional in STIX 2}\} \tag{2}$$
$$SDO,\ SRO \subseteq (A_r \cup A_o) \tag{3}$$
$$O = \{o \mid o \in (SDO \cup SRO)\} \tag{4}$$
$$R = \{r \subseteq O\} \tag{5}$$

5.1 Attribute level

This subsection defines the DQ dimensions we consider to be assessed at the attribute level, meaning that they rely on a subset of a STIX object's attributes.

Concise representation Concise representation addresses expressiveness of CTI and redundancies within the data [20]. Intensional and extensional are two distinct forms of conciseness. While the former is centered on the uniqueness of attributes and the schema level, the later emphasizes on unique objects. In the motivational example, the duplication of the information about the attacker links to the concise representation dimension as DQ is affected. It is worth noting that in the extant literature, concise representation sometimes only refers to compactly represented information [18]. In the context of STIX, the specification provides clear guidance how to implement a concise representation. It is explicitly stated that a unique identifier is assigned to each artifact. Additionally, each STIX object adheres to a specified JSON schema, and thus, optional and mandatory attributes are predefined. In general, the assumption holds that intensional conciseness is warranted through the schema definition. One exception in STIX is based on specifics of several STIX objects[4] as they contain lists referencing other objects. These lists are prone to redundant inputs, especially when defined manually.

With regard to extensional conciseness, the information within a CTI platform must be assessed for its respective quality. The main reason for this is that with a growing number of CTI producers, the probability of duplicated objects within the platform becomes likely. More precisely, there is a high chance that two or more objects on the platform are semantic duplicates. Even considering one single STIX report, semantically unique objects are not guaranteed as more than one person could work on the documentation of the incident and already existing information might be overlooked. Especially, when taking a look at the numerous free-text description fields defined in the current STIX specification, an indication whether these descriptions contain redundant information is important. However, comparing text for semantic redundancy is not an easy task. We encourage the application of methods for semantic similarity. The *Simhash* algorithm is one example proposed to approach this problem [27]. It allows for comparing two STIX objects regarding their uniqueness. An object o_1 is considered unique in a set of objects O if its $similarity$ to any other object $o_2 \in O$ is below a threshold t (see Eq. 6).

Objective metrics alone are not sufficient to assess concise representation in practical use. It is inevitable to include subjective perceptions through the utilization of domain knowledge. In this case, platform users conduct or support quality assessment and contribute by pointing out redundancies.

$$CR(o) = \begin{cases} 1 \text{ if } similarity(o_1, o_2) < t \\ 0 \text{ else} \end{cases} \tag{6}$$

Objectivity CTI is oftentimes created by multiple human actors during the analysis of an attack. These human CTI creators contribute not only objective threat information but might also introduce emotional or subjective perceptions. Most of the resulting descriptions are phrased in natural language. This is also the case in the motivational example in Sect. 3.2 and the threat actor description. There, the words "I suppose" indicate subjectivity and the context-depended observations of the security analyst. However, objectivity is a desirable characteristic of shared CTI artifacts as only objective information can be helpful for others. Natural language processing and sentiment analysis, therefore, can facilitate the assessment of unbiased and impartial CTI information as part of the objectivity DQ dimension.

Subjective descriptions of CTI information can be identified through the use of various subjectivity detection methods [28]. In the context of CTI and with regard to STIX, special focus is on attributes with free-text description fields in contrast to predefined enumerations and open vocabularies. This ultimately leads toward a sentence-level orientation for subjectivity detection as these fields contain only a limited number of words. Subjectivity detection methods in general can follow a syntactical approach or center on semantics. A thorough investigation into specifics of such methods must be considered during implementation to determine the best-fitting approach. Regardless of implementation, we classify relevant attribute values $v(a)$ of STIX objects into two distinct categories objective and subjective as shown in Eq. 7. Underlying this classification is the application of a suitable sentiment algorithm which yields a score for either objectivity or subjectivity. The results of the classification for chosen

[4] Examples are the *Report* object as well as the *Sighting* object.

attribute values are then aggregated to provide an objectivity metric for each object o based on Eq. 8.

$$OB(a) = \begin{cases} 1 \text{ if } v(a) \text{ classified as } \textit{objective} \\ 0 \text{ if } v(a) \text{ classified as } \textit{subjective} \end{cases} \tag{7}$$

$$OB(o) = \frac{\sum_{a \in o} OB(a)}{|o \cap (A_r \cup A_o)|} \tag{8}$$

Relevancy Relevancy forms a DQ dimension incorporating a user's perspective by comparing sets of property values to assess the usefulness of a CTI artifact for the consumer. This is an important aspect of CTI's fitness for use regarding an individual organization or analyst. For example, CTI describing an incident targeted at a specific industry sector is likely to be less relevant for other industry sectors. Also, security analysts might not be interested in threats targeting technologies not deployed in their organization. To illustrate this, the motivational example hints at the exploitation of vulnerabilities in software used at both the power plant and the hospital. Information about the relevance can be very helpful for analysts when prioritizing CTI artifacts to be analyzed.

Contextual information about the user can either be collected by the platform host or can be found in STIX objects describing the user. Specific characteristics (e.g., the industry sector) of a CTI publisher and those of a consumer are assessed for matches. In addition, attribute values for available STIX objects—for example, the *Vulnerability*—can be compared with the user's characteristics (e.g., the applied technologies), too. The coverage ratio expressed in Eq. 9 indicates relevance by taking the sets of all property values for consumer PV_c, publisher PV_p and relevant STIX objects PV_o into consideration. Congruent property values are set in relation to the total number of property values available for comparison.

The metric for the DQ dimension of relevancy could be further extended by inclusion of information contained in STIX *Sighting* objects. These objects incorporate a number describing how many times the referenced object has been identified. Therefore, this fosters the assessment of relevancy as frequently seen objects (e.g., an *Malware* object) might indicate a high relevance of these objects. This assumption can be expressed in a weighting factor added to the general metric and thus improve DQ assessment.

$$RE(o) = \frac{|PV_c \cap (PV_p \cup PV_o)|}{|PV_p \cup PV_o|} \tag{9}$$

Schema completeness The general completeness of data is confined to the assessment of schema completeness in the context of CTI. To distinguish this data quality dimension from syntactic accuracy, we focus on optional attributes and their values as the STIX JSON schemes already allow to assess the existence of required attributes. This aspect is covered by the DQ dimension of syntactic accuracy later on.

STIX-based threat intelligence can be assessed for schema completeness of individual optional attributes a_o. A missing optional attribute value $v(a_o)$ is identified and classified according to Eq. 10. A strict distinction between complete (i.e., with value) and incomplete (i.e., without value) attributes is enforced. Referring to the example in Sect. 3.2, the vulnerability could be described in more detail with an external reference to a specific Common Vulnerabilities and Exposures (CVE) entry. This optional information would help others to gain further information about the actual vulnerability, how it is exploited, and how it can be fixed. This would ultimately improve CTI quality significantly by making it easier for others to leverage the CTI. In a second step, schema completeness for an entire STIX object o builds upon the previously calculated completeness scores for included attributes. The ratio of filled optional attributes to the total number of optional attributes of an object represents the schema completeness metric as shown in Eq. 11.

$$SC(a_o) = \begin{cases} 1 \text{ if } v(a_o) \neq NULL \\ 0 \text{ else} \end{cases} \tag{10}$$

$$SC(o) = \frac{\sum_{a_o \in (o \cap A_o)} SC(a_o)}{|o \cap A_o|} \tag{11}$$

Syntactic accuracy The data quality dimension of accuracy contributes to the correctness of data. With focus on syntactic accuracy in the context of CTI, the data schema is of particular importance for quality assessment. Syntactic accuracy gives a first indication on the extend to which an object is aligned with its data format.

The OASIS consortium behind the STIX format provides a JSON schema for each object. This allows for an automated matching of objects against those schemes to assess syntactic accuracy. In general, this DQ dimension is measured based on the analysis of attribute values $v(a)$ with $a \in (A_r \cup A_o)$ being part of a domain D [16]. In application to STIX-based threat intelligence, we can use the existing JSON schemes and validate each attribute value against the schema definition. The domain D is derived from the JSON schema which provides data types and allowed values. The assessment for syntactic accuracy of each attribute value is expressed by Eq. 12. An overarching indicator for syntactic accuracy of an object o can, respectively, be calculated as shown in Eq. 13.

$$SA(a) = \begin{cases} 1 \text{ if } v(a) \in D \\ 0 \text{ else} \end{cases} \tag{12}$$

$$SA(o) = \frac{\sum_{a \in o} SA(a)}{|o \cap (A_r \cup A_o)|} \tag{13}$$

Timeliness In the context of CTI, time ascends to one of the crucial elements of CTI quality. As stated earlier, outdated intelligence is identified throughout the relevant literature as one of the core challenges [3,5,13]. It is quite evident that the most current and up-to-date CTI artifacts probably implicate the most value for any type of analysis.

Time-based information contained within CTI data builds the basis for the configuration of a timeliness metric applicable to the CTI context. In general, various metrics can be utilized to assess timeliness. Considering the STIX data format, a basic timeliness metric is described in Eq. 14. The two components of this metric—currency and volatility—are present in every STIX object or can be derived from inherent features of the CTI platform. Volatility in this setting is expressed by the number of modifications to the assessed STIX object. The number of modifications can be drawn if concepts like the historization from earlier work are implemented [29]. This concept allows to track changes and the number of changes applied to a STIX object. Currency is referring to the age of the information and thus the time since its last modification. However, this metric entails certain problems specifically with regard to interpretability as well as to other requirements [30].

Where statistical data about the decline of timeliness for specific CTI information does exist, the metric for timeliness must be adapted. Resulting values of a statistical timeliness metric shown in Eq. 15 can subsequently be interpreted as probability of up-to-date CTI information. Considering the example in Sect. 3.2, the decline for certain STIX objects is higher than for others. File hashes as in the *Indicator* of Listing 1 will likely have high decline values as, for example, malware binaries might undergo slight changes frequently leading to changed hash values. In contrast, information regarding the threat actor might not change in time, thus having no statistical decline at all.

In contrast to these metrics, specific assessment of STIX-based CTI for the DQ dimension of timeliness can also be based on characteristics of STIX objects. For example, *Sighting* objects can provide information about the time of occurrence of referenced STIX objects. It can be thus inferred that for the timeliness of referenced STIX objects, the concept of inheritance applies. STIX objects of type *Observed Data* can be assessed for timeliness following the same procedure. Our proposed metric described in Eq. 16 includes the current time, the time of last occurrence, and a predefined time-based threshold value to foster the applicability of timeliness to any given CTI use case. In general, we focus on objective metrics of timeliness. Subjective perceptions such as expert knowledge about threshold values assist the assessment and can be considered further during implementation. Referring back to the motivational example, the hospital's security analysts can define a threshold based on their experience that indicators are outdated after a specific amount of time.

$$TI_{Basic}(o) = \frac{1}{(Currency(o) \times Volatility(o)) + 1} \tag{14}$$

$$TI_{Statistical}(o) = \exp\left(-Decline(o) \times Currency(o)\right) \tag{15}$$

$$TI_{Assisted}(o) = \begin{cases} 1 \text{ if } t_{current} - t_{last} < threshold \\ 0 \text{ else} \end{cases} \tag{16}$$

5.2 Object level

On the object level, we consider two dimensions which rely on manual input and are therefore defined to be subjective dimensions. They center on object characteristics of a higher abstraction level and often follow a cross-object perspective.

Representational consistency In general, the assessment of representational consistency relies on a set of rules C and semantic conditions c_j contained therein for the underlying data [24]. This DQ dimension needs to be adjusted to the requirements of the individual context and the given use case. Analogous to schema completeness, representational consistency goes beyond aspects of syntactic accuracy. For the context of threat intelligence, representational consistency allows for the enforcement of additional formal requirements which are not addressed by the dimensions of syntactic accuracy or concise representation. These might originate from data format requirements or requirements imposed by a CTI sharing platform. In the following, we propose two exemplary conditions configured to the STIX data format. CTI platforms could define further conditions or adjust existing ones. This is part of an iterative approach to support an increasingly detailed assessment of representational consistency.

In the context of STIX-based threat intelligence, we suggest a first condition to represent the necessity of existence of referenced STIX objects. For all STIX objects, the following "inter-relation constraint" [16] applies: referenced objects of embedded relationships must exist. Moreover, considering individual STIX objects specific relationships must be verified. This applies for all SROs as they connect per definition two SDOs. A second exemplary condition takes time-based information and the chronological order of creation and modification of CTI into account. Hence, it must be verified on the "intra-relation constraint" level that the creation time of any object is prior or equal to the time of modification. Besides, SROs can connect two SDOs only after their creation. Creation time of the corresponding SDOs must be prior or equal to creation time of the SRO. Listing 1 reveals those two exemplary conditions for representational consistency, too. For the *Vulnerability*, modification time precedes creation time by a month. With regard to referenced objects, a *Relationship* (i.e.,

"relationship–3") points toward a nonexistent *Vulnerability* (i.e., "vulnerability–2").

The assessment of representational consistency on a condition basis is described in Eq. 17. A given STIX object is assessed for each defined condition $c_j \in C$ separately, and the results indicate if a condition is fulfilled. Representational consistency per object is aggregated over all defined conditions in the set of conditions C as seen in Eq. 18.

Please note that although the assessment of an object o regarding a condition c_j can be automated and therefore is objective, the definition of the respective conditions is fully in control of the responsible stakeholder. Thus, we interpret this dimension to rather be subjective than objective with respect to the definitions in Sect. 4.1.

$$c_j(o) = \begin{cases} 1 \text{ if } o \text{ fulfills condition } c_j \\ 0 \text{ else} \end{cases} \tag{17}$$

$$RC(o) = \prod_{j=1}^{|C|} c_j(o) \tag{18}$$

Reputation It is important to build trust in shared CTI environments. Trust and the assessment of trustworthiness can build upon the DQ dimensions of reputation, provenance, and believability. The introduction of two quality sub-dimensions for reputation—reputation of the publisher (i.e., provenance) and reputation of the data set (i.e., believability)—allows for a holistic coverage of the trustworthiness concept in the context of CTI exchange. Our proposed assessment is based on functionalities similar to reputation systems and external human input. Reputation scores for a given publisher p might adhere to a five-star rating system as shown in Eq. 19 as well as reputation scores of a STIX object o as shown in Eq. 20. Based on these reputation scores s contained in a set of scores S, an overall reputation $RS(x)$ for either publisher or STIX object is calculated according to a simple ratio function described in Eq. 21. Sample size $|S|$ supports data quality assessment further and constitutes a relevant additional data point. In the situational example, the hospital can articulate trust toward the power plant and its CTI by rating them accordingly.

While the above-mentioned configuration of reputation is purely subjective, possibilities exist to assist the quality assessment with objective metrics. For one, a list of trusted CTI publishers can be introduced as an indicator for the reputation of a publisher. An analogous indication for the reputation of an object is the number of access requests to a certain artifact set in relation to the number of CTI platform consumers having taken remediating steps upon the threat intelligence.

$$RS(p) = \{s \mid 1 \leq s \leq 5 \ \wedge \ s \in \mathbb{N}\} \tag{19}$$

$$RS(o) = \{s \mid 1 \leq s \leq 5 \ \wedge \ s \in \mathbb{N}\} \tag{20}$$

$$RS(x) = \frac{\sum_{s \in S} s_i}{|S|} \tag{21}$$

5.3 Report level

On the upmost level of Fig. 3, we place a single dimension which takes a complete STIX report including its contained SDOs and SROs into consideration.

Appropriate amount of data The requirement to include the appropriate amount of data quality dimension arose during our discussions with domain experts as described earlier. However, the application of a generic metric proves not feasible due to its semantic component in the form of needed data units. We therefore base our metric on the additional comments of security analysts. Homogeneous SDO types and very few relationships seemingly lead to the experts' perception that the report in general is not very helpful.

To distinguish between a report with homogeneous STIX objects and one with rather diverse objects is a matter of implementation and cannot easily be compressed into a metric. As described above, this is a rather complex task which needs further research efforts. As a first approach toward a feasible support of security experts, we propose a clear representation of occurrences of each STIX object in an artifact. This is achieved by simply counting the instances of the different SDO types within a report. Visualization can provide this relevant information at the report level and can aid DQ assessment at first glance.

Besides this, we take graph theory for the connectedness of the STIX report's SDOs into account. We argue that a metric based on the number of relationships can provide a basic indicator to assess this DQ dimension. In general, the metric for the DQ dimension of appropriate amount of data should yield a higher score for CTI which is densely connected. A given STIX report depicts a graph, and its contained SDOs represent vertices. SDOs are furthermore connected with each other through SROs which resemble edges from a graph perspective. The metric in Eq. 22 sets the number of existing SROs in a given STIX report in relation to the maximum possible number of SROs as defined by the number of SDOs for this report.

The metric for the appropriate amount of data is a challenge for future work. Our simplistic metric could be improved in different ways. A possible direction is a statistical comparison of all available reports. Calculating a report's score for the diversity of SDOs and the respective relationships as a comparison with a baseline diversity from other reports might be a feasible direction. However, the prerequisite to this approach is a sufficiently high number of reports included into the baseline.

$$AD(r) = \frac{|sro \in r|}{\frac{|sdo \in r|(|sdo \in r|-1)}{2}} \tag{22}$$

5.4 Aggregating quality indicators

The aggregation of DQ dimension scores for CTI has to be customizable as described earlier in Sect. 4.2. Adjustable aggregation parameters enable CTI consumers to define the weight of each of the DQ dimensions D in the procedure of calculating a quality indication for each STIX object. For this customizable aggregation, we propose a weighted average (see Eq. 23), where each dimensional score $d_i \in D$ is weighted with a parameter $w_i \in \mathbb{N}$. This parameter w_i can be adjusted by each platform consumer. If no custom value is provided for a dimension d_i, the default weight is $w_i = 1$.

To support consumers' decisions on which report available on a CTI sharing platform to analyze, we additionally propose a report quality indicator calculated following Eq. 24. This score contains the individual DQ object scores $DQ(o)$ and the additional report-level dimension of the appropriate amount of data $AD(r)$. Only the additional DQ dimension's score is weighted in this aggregation with $w \in \mathbb{N}$. The default value for w is again 1. Following this aggregation structure, the weight of each DQ dimension is adjustable by the platform users consuming the respective CTI to ensure that the quality scores represent their individual preference of the dimension's importance.

$$DQ(o) = \frac{\sum_{d \in D} d_i \cdot w_i}{\sum_{d \in D} w_i} \tag{23}$$

$$DQ(r) = \frac{(\sum_{o \in r} DQ(o)) + AD(r) \cdot w}{|r| + w} \tag{24}$$

6 Visualizing quality of CTI

Informing users of CTI about the quality of the intelligence at hand is of crucial importance. This is a vital task in the context of a sharing platform as it allows users to build trust toward the shared CTI. We argue that it is not enough to only inform users about the result of a CTI quality assessment. Instead, the assessment process itself must be transparent for security analysts. Thus, a visual interface should inform them "Why" a report has a specific quality score. As different aspects of CTI quality might also be of varying importance for users, the visual interface could also support parametrization of the quality aggregation as described earlier. Besides the need to inform users about the CTI quality and building trust, their subjective perception of a report's quality is highly relevant for the assessment process. Therefore, a solution is needed to allow them to share their opinion.

Providing a possible path to solve these requirements, we draw upon the idea to make complex threat intelligence exchange formats, like STIX, accessible for human experts through an interactive visual interface. The feasibility and applicability of this approach have been shown in earlier work [29]. In this work, we implemented and evaluated an open-source visual analytics prototype for STIX. We extend this proof of concept by including indicators about the CTI quality in the interface and by implementing functionalities for experts to share their subjective quality assessment where necessary. In the following sections, we briefly introduce the changes made to the original visual interface called Knowledge-Assisted Visual Analytics for STIX (KAVAS). Additionally, we extend the underlying database (CTI Vault) to integrate notions of threat intelligence quality. However, both the database and the visual representation are built to only handle data compliant to the STIX specification. Thus, before including CTI quality into the tool, a solution to represent CTI's quality in the STIX format is needed.

6.1 Integrating quality indicators into STIX

In its current specification, the STIX format has no object types or properties to model indications about the quality of CTI. However, the specification defines the format in a way which allows for the extension of the baseline specification [31]. This opens different possible ways to integrate CTI quality into this format. On the one hand, it is possible to define completely new types of STIX objects. On the other hand, additional properties could be added to the existing SDO and SRO types.

```
{
  ''type'':''x-quality-indicator'',
  ''id'':''x-quality-indicator--1'',
  ''created'':''2019-07-25T09:00:00Z'',
  ''modified'':''2019-07-25T09:00:00Z
     '',
  ''object_ref'':''threat-actor--1'',
  ''measures'': [
    {
      ''dimension'':''Syntactic
         Accuracy'',
      ''type'':''objective'',
      ''score'':0.8
    },
    ...
    {
      ''dimension'':''Reputation'',
      ''type'':''subjective'',
      ''score'':0.7,
      ''rating_count'':14}
    }
  ]
}
```

Listing 2 Exemplary *Quality Indicator* object

Table 1 Definition of the *Measure* custom data type for STIX 2

Property name	Type	Description
dimension (*required*)	String	The dimension for which the measurement is described
type (*required*)	String (*"subjective" or "objective"*)	Describes whether the dimension's score is based on a subjective or an objective metric
score (*required*)	Float	Double-precision number ranging from 0 to 1 describing the current result of the quality assessment for a quality dimension
rating_count (*optional*)	Integer	This property is only needed for "subjective" measures as it describes how many different ratings were given to produce the current score

Table 2 Definition of the *Quality Indicator* custom object for STIX 2

Common properties		
type, id, created_by_ref, created, modified, revoked, labels, external_references, object_marking_refs, granular_markings		
Quality indicator specific properties		
object_ref, measures		
Property Name	Type	Description
type (*required*)	string	The value of this property MUST be "x-quality-indicator"
object_ref (*required*)	identifier	Specifies the STIX Object that is referred to by this quality indicator
measures (*required*)	list of type measure	A list holding all measurements for the different quality dimensions available for the referred-to STIX Object

In any case for each STIX object, the calculated scores for the different quality dimensions need to be documented. To capture the necessary information in a STIX-conformant way, we therefore propose the custom data type *Measure* defined in Table 1. This data type consists of the name of a specific dimension and the object's respective score. It is worth noting that our proposal centers on float values. Nevertheless, scores on an ordinal scale are also possible. Respective conversions can be implemented by defining ranges of float values which refer to a specific ordinal scale (low, medium, and high). Additionally, the custom data type contains the type (subjective or objective) of the dimension. For subjective dimensions, the count of received ratings used to calculate the score can be stored.

We opt to attach a list of measures structured according to the proposed *Measure* data type to a new Custom STIX object. While it is also possible to include this list in any existing STIX object, our proposal aims to maintain a clear separation between actual threat information and the related quality information. Additionally, this proposal produces as less interference as possible with the existing data model. Neither the existing SDOs nor SROs need to be changed. In compliance with the specification, we follow the mechanisms and requirements given to introduce custom objects called *Quality Indicator*. Besides the mandatory *Common Properties*, a number of specific properties are established [31].

Table 2 defines the proposed STIX Custom Object. We include common properties of our *Quality Indicator* object which are mandatory for each SDO. These properties are followed by several specific properties defined for the object. The last part of Table 2 defines allowed data types and values for the specific *Quality Indicator* properties. The *type* attribute must not hold other values than "x-quality-indicator". The *Quality Indicator* object is not connected to any other objects with an explicit SRO but holds a property "object_ref" reflecting the ID of the SDO or SRO for which the object indicates the relevant quality measures. Finally, the object contains a list of "measures" which holds the scores for all the DQ dimensions. The list is formed of the custom *Measure* data type. An exemplary and simplified object is shown in Listing 2.

STIX is an actively maintained CTI standard. Recently, there have been developments that incorporate some aspects similar to our CTI quality concept within the newest STIX2.1 Committee Specification Draft.[5] Most notably, this draft includes an *Opinion* SDO to capture perceptions by CTI consumers about the correctness of a STIX object. The *Opinion* SDO aims to document the level of agreement with the referred-to STIX object(s) on a Likert-type scale ranging from strongly disagree to strongly agree. As can be seen by the purpose and the description of the *Opinion* SDO, this spe-

[5] https://docs.oasis-open.org/cti/stix/v2.1/stix-v2.1.html.

cific STIX object is another prospective option to implement elements of the *Reputation* data quality dimension. Nevertheless, in contrast to our proposed *Quality Indicator* SDO the draft and its *Opinion* SDO fall short to cover a larger CTI quality concept.

6.2 Persisting quality indicators in the CTI Vault

The original database for the CTI visualization is a graph-based approach based on Neo4J.[6] This is quite reasonable as STIX is based in graph-like structure itself. Additionally, the integrity-preserving storage concept proposed by Böhm et al. [29] is most efficiently implemented using this technology. We extend this approach by adding a new database to the architecture. This new database is solely supposed to persist the *Quality Indicator* objects introduced in Sect. 6.1. As described, these objects do not have any explicit connections to other STIX objects via SROs. Their integration would double the number of objects inside the existing database and would certainly affect the performance negatively. Therefore, we decided to avoid storing the quality object inside the existing vault.

Our newly added "Quality Vault" is a document-oriented database (MongoDB[7]) for performance reasons. This additional vault persists the JSON representations of the *Quality Indicator* objects which are directly related to a single SDO or SRO in the CTI Vault via the "object_ref" attribute.

6.3 Displaying quality indicators in KAVAS

Throughout this section, we describe the changes we made to the original visual interface to include visual indications about the quality of STIX artifacts. In Böhm et al. [29], the process of visually analyzing STIX-based CTI with KAVAS starts with a simple drop-down menu to select the report of interest. The drop-down menu contains only the name of the report given by its publisher. This does not disclose any additional information to the analyst whether the report might be of interest or not. We changed this initial view of the KAVAS interface to be more informative and also to give first insight into the quality of the report. The visual interface now contains an expandable list of all available reports from the CTI Vault. The expansion panel for each STIX report consists of three main sections depicted in Fig. 4 informing analysts on the contents and overall quality of a STIX artifact at first glance[8]:

1. At first, a description (if given by the report's producer) gives high-level information on what the report is about.
2. The second section shows which specific STIX objects, both SDOs and SROs, are contained in the report and how often they are present. Object types that are not present in the respective STIX artifact are grayed out. This view fulfills the requirement to provide a view on the homogeneity of a STIX artifact as described in Sect. 5.3 within the quality dimension of *Appropriate amount of data*.
3. The third section gives a very brief and high-level indication on the average quality of the STIX objects and their interconnectedness within the respective report using two gauge displays. This connectedness is represented by the score as described in Sect. 5.3.

After a STIX report is selected and its graph representation is loaded in the visual interface further changes become apparent, clicking a node or a link of the graph details its information in a details-on-demand card view. The original object card only contained a tab with the attribute values of the selected object and, for SDOs, a tab with its directly linked neighbors in the graph. We now add a quality badge in the header of the object card displaying the aggregated quality score of the dimensions from object and attribute level as described in Sect. 5.4. Furthermore, we add a new tab providing more detailed insight and transparency of the quality measuring. The new object quality tab on the details-on-demand view is shown in Fig. 5. Again, this component is divided into three sections:

1. A gauge visualization of the object's overall quality score aggregated from the scores at the attribute and object level.
2. A section with progress bars indicating the object's score for all described objective dimensions.
3. A third section that holds the indicators for an object's scores of subjective quality dimensions. For this part of the quality tab, we need to both inform the user about the current score and allow them to provide their own subjective quality measurement for the respective dimensions. To do so, we lend from reputation systems and display a rating bar ranging from one to five stars which is a well-known visual metaphor in reputation systems. These rating bars always show the current overall score for the quality dimension (blue stars) in relation to the possible highest rating while also allowing users to click each of the stars to provide their own rating. Numbers in parentheses besides the name of the DQ dimension indicate the count of ratings provided by other users (e.g., the number of subjective assessments on which the current score is based).

[6] https://neo4j.com/.

[7] https://www.mongodb.com/.

[8] Please note that the displayed information is computed based on a test data set which is different from the STIX example in Sect. 3.2.

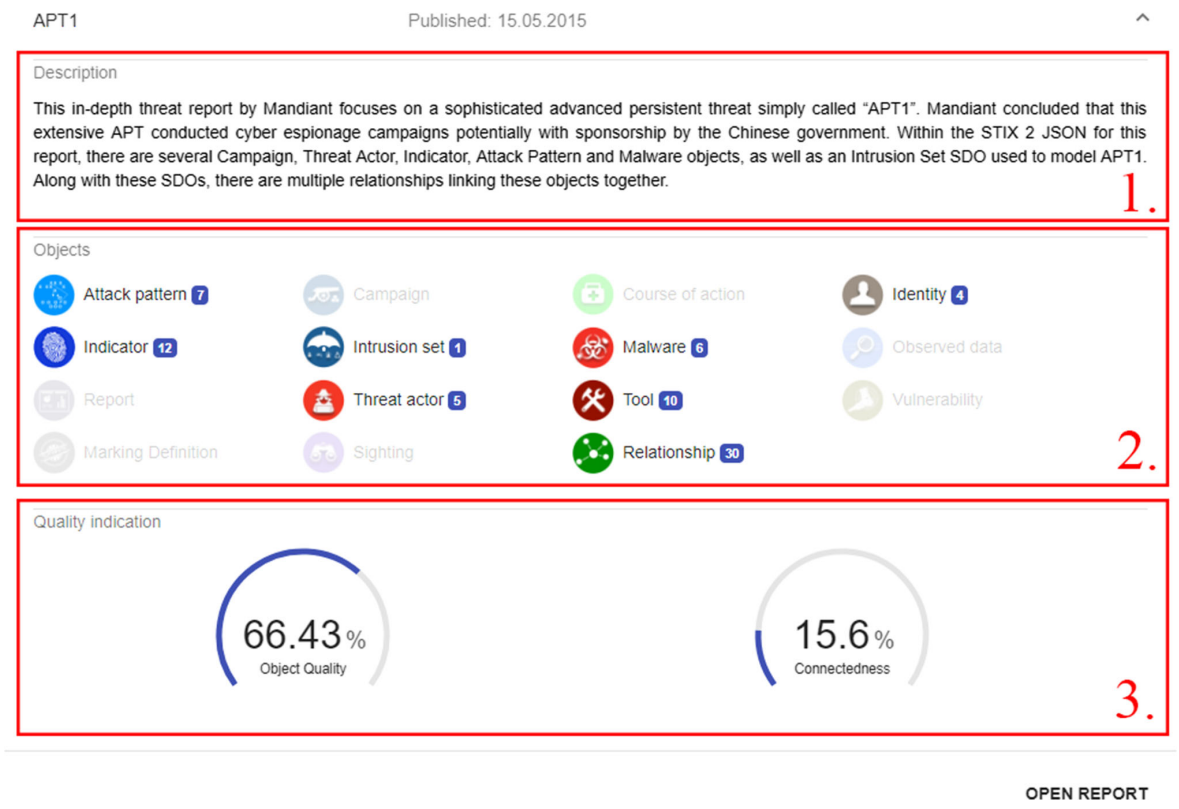


Fig. 4 View of report selection screen

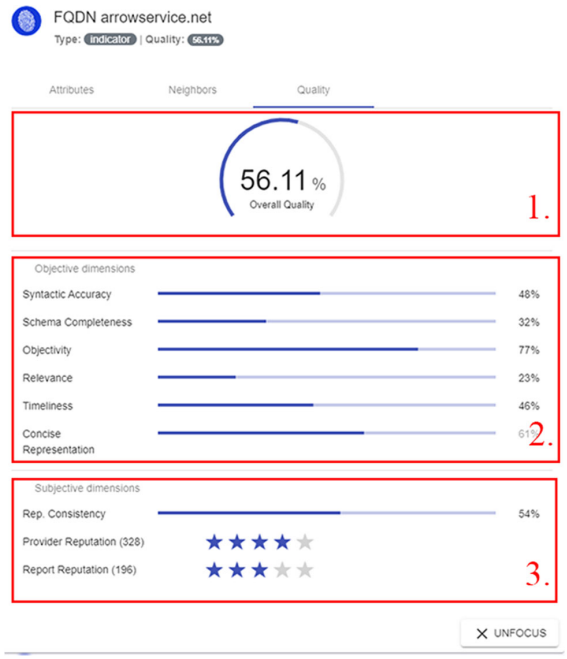


Fig. 5 Quality tab on object card (details-on-demand)

The quality tab fulfills a twofold goal: First, it makes the aggregation of the quality dimensions transparent, and second, it allows collecting user's input for subjective quality dimensions. We actively decided not to use any color-coding for the scores. Traditionally, respective scores are colored with red (low quality), orange (medium quality), and green (high quality). However, we only aim to inform CTI analysts about the quality scores and do not want to provide any kind of interpretation of low or high score for any quality dimensions. As described earlier, this is mainly because the quality dimensions might be of different interest for different consumers. Therefore, low scores for respective dimensions of an object do not automatically implicate that the object is irrelevant or of low overall quality for the consumer.

In order to allow users to customize the aggregation of quality dimension scores following our previously described bottom-up approach, analysts need a way to define the dimensions' weights. To provide this functionality, we extend the KAVAS settings dialog with a slider for each quality dimension as depicted in Fig. 6. The default configuration assumes that all dimensions are equally important (e.g., have a weight of 1). Analysts can use the sliders to customize the dimension aggregation according to their preference. If they do not want a specific dimension to have any influence in the aggre-

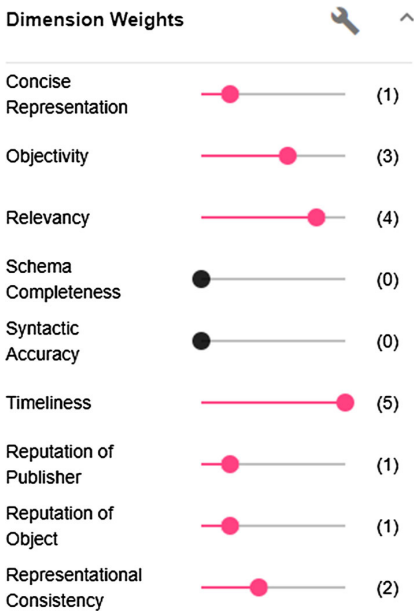


Fig. 6 Slider for dimension weights

gation, they can assign a weight of 0 and for a dimension with crucial importance, they can accordingly assign a weight of 5. Please note that the metric for dimension aggregation in Eq. 23 does not limit the range for the dimension's weight. However, we chose to limit them in the visual interface to a range from 0 to 5 for more practical feasibility.

6.4 Evaluating the visual display of CTI quality

To validate the visualization approach and to provide first evidence of its suitability, we conduct a number of expert interviews. The main goal of these interviews is to validate that the visual approach helps analysts to understand the DQ of the CTI artifact at hand.

Participants The interviewees are three security experts from different sectors and company sizes. We conduct interviews with two highly experienced security analysts from a big international conglomerate and a medium-sized manufacturing company. The third interviewee is a researcher focusing on CTI sharing formats. Each participant has a medium to high knowledge regarding threat intelligence as all of them deal with information security on a daily basis. None of the participants currently obtains a quality assessment on CTI.

Design and procedure The interviews with the experts are designed following a semi-structured approach and are splitted into the following four phases [32]:

1. *Introduction* Starting the interviews, each participant is questioned for some basic data, their experience, such as knowledge on CTI and DQ aspects. Afterward, each expert is introduced briefly to the STIX format (if necessary) and to the problem of measuring CTI quality. Thereby, the experts are actively asked to criticize any potential issues noticed throughout the following interview phases.
2. *Measuring CTI quality* In this phase, we aim to get additional feedback on the individual dimensions and the configured metrics for quality assessment of STIX artifacts (Sects. 4 and 5). Although the dimensions and the metrics are already the result of an iterative process where we collaborated with researchers and practitioners, an additional evaluation of these results is performed in this phase. The selected dimensions, their structure, and the configured metrics are discussed with the interviewees to identify whether they support the relevance of the proposed DQ measurement approach. We also ask the participants what aspects of the dimensions and metrics might need a more detailed explanation and whether they think that the metrics are comprehensible for security analysts without much prior knowledge in the DQ area.
3. *Visualizing CTI quality* The focus of this phase is to test the suitability of the proposed visualization approach. To enable the interviewees to work with the DQ visualization, we make use of sample STIX reports provided by the OASIS consortium. Prior to the interviews, these reports were manually fed into the existing KAVAS tool and enriched with the DQ measures. During the interviews, the participants can access the STIX reports through the extended KAVAS tool as described in Sect. 6. The main goal in this interview phase is to identify whether the proposed visualization elements to display the CTI quality are actually helpful for security analysts. We ask the interviewees whether the proposed DQ metrics are comprehensible with the chosen visualization elements and what further aspects they think would enhance the understanding of DQ assessment within CTI.
4. *Wrap-Up* The last phase of the interviews is dedicated to a summarizing discussion. Here, we discuss with the participants whether an implementation of the proposed metrics and the respective visualization approach would be applicable to operative deployment and the conditions thereto. Finally, we collect a list of ideas and features the interviewees find useful for improving our approach.

Results The interviews lasted between 45 to 75 minutes. The results of the conducted interviews are presented in the following, divided according to the four interview phases described before:

Table 3 General information on the interview participants

	Position	Business branch	Organization's size	CTI knowledge	DQ knowledge
#1	Senior security analyst	Manufacturing	ca. 400.000	High	Medium
#2	Head of security information management	Manufacturing	ca. 15.000	Low	Medium
#3	Security researcher	Academia	ca. 5.000	High	Medium

1. *Introduction* The results of the introduction phase are summarized in Table 3 giving an overview on general information about the interviewees.
2. *Measuring CTI quality* Above all, the interviewees unanimously stress the importance of metrics for quality within the field of CTI. Valuable and actionable CTI is stated to be highly dependent on quality and currently more often than not CTI is of low quality. A recurring theme mentioned in this phase by multiple interviewees is the interpretation of CTI quality. It is pointed out that the implementation of metrics for CTI quality by sharing platforms would benefit significantly from indication of low- and high- quality reference scores. Another identified theme is usability of DQ dimensions and metrics for CTI. Here, formally sound metrics, the chosen naming convention of DQ dimensions based on existing academic work and security analysts without DQ or mathematical background, stand opposite each other. Comprehensive explanations are seen as one approach to foster security analysts' understanding of the precise meaning of CTI quality dimensions and metrics.
3. *Visualizing CTI quality* All interviewees agree on the necessity to provide easy access to CTI quality through the use of visualization elements and validate our visualization approach. All interviewees agreed that the chosen visual representation allows for a quick recognition of CTI quality. They also uniformly considered the possibility to include subjective perceptions with means similar to reputation systems very helpful. Nevertheless, the interviewees name different extensions to the current visualization. For one, in-depth information about the DQ dimensions, the metrics, and possible interpretation is highlighted. Additionally, the showcased visualization includes percentages numbers and numeric weighting factors which could instead be visualized on a Likert-type scale. Another proposed extension targets the causal nature of low-quality scores. Visualization elements to detect improvements and eventually improve the CTI quality further are perceived as helpful. As one interviewee points out, user groups (e.g., system administrator or standard user) could be defined, given different permissions and thus see different visualizations.
4. *Wrap-up* In the final phase, the interviewees often come back to the timeliness dimension. The proposed metrics for this DQ dimension needed additional explanations with regard to STIX specifics (i.e., Sighting SDO). Ideas and features mentioned by the interviewees to extend our work cover guidance to improve CTI quality and quality filtering with visualization elements. For instance, visual recommendations to reach a higher CTI quality (with or without prior knowledge about quality details) might be added to the current reactive assessment.

Overall, the interviewees' feedback indicates the valuable contribution of measuring and visualizing CTI quality. In particular, the dual approach itself (measure and visualize) is assumed to reduce complexity, lower quality assessment barriers, and foster CTI utilization. With regard to the implementation within a CTI sharing platform, we draw the conclusions that 1) there needs to be discussion on usability and adequate naming of DQ dimensions, 2) reference values are crucial for CTI quality interpretation, and 3) visual elements and textual explanations must be combined to avoid ambiguity.

7 Conclusion and future work

This work shed light on the assessment of DQ dimensions in the context of CTI. Nonetheless, there are further areas where research needs to be intensified and extended to.

7.1 Conclusion

Recent developments in the cyber threat landscape urge organizations to join forces against the adversaries. Collaboration based on the exchange of available threat intelligence arises as one of their most effective weapons. CTI sharing leveraged by respective platforms helps to spread knowledge about current threats. However, respective formats are oftentimes complex and large leading to a lack of readability for domain experts. Therefore, it is a vital task to help experts understand the CTI, for example, by providing visual representations. CTI can only be effective when security experts are able to comprehend it quickly and efficiently. Another issue hindering the effectiveness of CTI is the missing quality control on sharing platforms. This lack of DQ management mostly

stems from missing proposals to measure CTI quality in the first place.

Our studies cumulated within this work constitute a necessary first step into this direction. This includes the two focal points of measurement and visualization of threat intelligence quality. Existing academic work proposed sets of possibly relevant quality dimensions as well as high-level requirements for CTI quality assessment. Although calling for an inclusion of quality assessment and assurance into the world of CTI sharing, up to now there are no proposals for actual quality metrics applicable to CTI. Therefore, proposing a relevant set of quality dimensions and configuring respective metrics for a specific CTI format is a necessary step toward actionable CTI quality assessment. The proposed dimensions and metrics can help to build a cohesive quality management methodology for CTI based on the STIX data format. Most of our findings regarding suitable as well as not applicable DQ dimensions or metrics can also be applied to other CTI formats. It is possible to think of additional, more specific dimensions which could be defined to assess quality of threat intelligence. However, in this work we define a base set of dimensions that originate from existing and widely agreed-upon DQ dimensions. This base set can easily be extended, and detailed metrics can complement our proposed ones if necessary.

Besides the definition of metrics to measure CTI's quality for relevant dimensions, we also showed how this quality assessment can be made transparent to users of a sharing platform. Transparency herein supports both building trust for the available information and making informed decisions about which CTI artifact is worth analyzing. This is important as current sharing platforms already hold an unmanageable amount of threat intelligence. Informing potential consumers of an artifact about its quality is a helpful decision support for the consumer. The visual display of an object's overall quality including the respective scores for individual quality dimensions helps consumers to understand how the DQ measurement result was reached. Additionally, it provides a way to collect important input from users for subjective quality dimensions. We therefore also show how human CTI analysts can be included into the quality assessment.

7.2 Future work

Our work can be seen as a first step into the direction of measuring CTI quality. However, we can identify several topics demanding additional research effort.

We are among the first to propose a cohesive set of applicable CTI quality dimensions. Therefore, these dimensions might be subjected to changes as more knowledge is gained about CTI sharing processes, platforms, and associated stakeholders. One dimension which needs further attention is the *Appropriate amount of data*. The proposed metric is a first approach toward a highly complex issue. It is difficult to define which amount of data—either data regarding STIX objects or the information described by these objects—is appropriate. Thus, we propose a simple metric to give domain experts an indication of the data contained in a STIX report. The DQ metric for the appropriate amount of data should be further detailed upon analysis and verification with CTI platform data. Furthermore, the metrics to evaluate quality should be reconfigured for other CTI formats and integrated into a cohesive data quality management methodology for CTI.

After formally configuring the metrics for the selected quality dimensions, those metrics should be implemented into an actual sharing platform. Up to now, we only tested them in a small scaled environment. A complete implementation will likely raise further issues about the selection of suitable algorithms and the control of user participation and intentions which go beyond the core DQ assessment and have not been addressed in this work. Warranted through an implementation, the extension of some proposed dimensions can become feasible as more information about the requirements will be available. Implementing and extending the dimensions and metrics are necessary steps to finally build a cohesive methodology for quality assessment of CTI including processes to assure and improve quality of artifacts on a sharing platform.

Acknowledgements Open Access funding provided by Projekt DEAL.

Funding Part of this research was supported by the Federal Ministry of Education and Research, Germany, as part of the BMBF DINGfest project (https://dingfest.ur.de). Part of this project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 830927.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Symantec Corporation.: Internet security threat report 2019 (2019). https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf
2. Riesco, R., Villagrá, V.A.: Leveraging cyber threat intelligence fora dynamic risk framework. Int. J. Inf. Secur. **18**, 715–739 (2019)
3. Ponemon Institute LLC.: Live threat intelligence impact report 2013 (2013). https://www.ponemon.org/blog/live-threat-intelligence-impact-report-2013-1
4. Ring, T.: Threat intelligence: Why people don't share. Comput. Fraud Secur. **2014**(3), 5 (2014)
5. Sillaber, C., Sauerwein, C., Mussmann, A., Breu, R.: Data quality challenges and future research directions in threat intelligence sharing practice. In: Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security - WISCS'16, pp. 65–70. ACM, New York (2016)
6. Sillaber, C., Sauerwein, C., Mussmann, A., Breu, R.: Towards a maturity model for inter-organizational cyber threat intelligence sharing: A case study of stakeholder's expectations and willingness to share. In: Proceedings of Multikonferenz Wirtschaftsinformatik (MKWI 2018), pp. 6–9. Springer, Heidelberg (2018)
7. Tounsi, W., Rais, H.: A survey on technical threat intelligence in the age of sophisticated cyber attacks. Comput. Secur. **72**, 212–233 (2018)
8. Juran, J.M., Gryna, F.M.: Juran's Quality Control Handbook, 4th edn. McGraw-Hill, New York (1988)
9. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decis. Support Syst. **43**(2), 618 (2007)
10. Dandurand, L., Serrano, O.S.: Towards improved cyber security information sharing. In: 2013 5th International Conference on Cyber Conflict (CYCON 2013). IEEE Computer Society Press, Los Alamitos (2013)
11. Serrano, O., Dandurand, L., Brown, S.: On the design of a cyber security data sharing system. In: Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security - WISCS '14, pp. 61–69. ACM, New York (2014)
12. Kokulu, F.B. 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 (Association for Computing Machinery, New York, NY, USA, 2019), CCS '19, pp. 1955–1970. https://doi.org/10.1145/3319535.3354239
13. Sauerwein, C., Sillaber, C., Mussmann, A., Breu, R.: Threat intelligence sharing platforms: an exploratory study of software vendors and research perspectives. In: Proceedings of the 13th International Conference on Wirtschaftsinformatik, pp. 837–851. Springer, Heidelberg (2017)
14. Menges, F., Pernul, G.: A comparative analysis of incident reporting formats. Comput. Secur. **73**, 87–101 (2018)
15. Piazza, R., Wunder, J., Jordan, B.: StixTM version 2.0. part 2: Stix objects (2017). https://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part2-stix-objects.html
16. Batini, C., Cappiello, C., Francalanci, C., Maurino, A.: Methodologies for data quality assessment and improvement. ACM Comput. Surv. **41**(3), 1 (2009)
17. 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 **60**, 154–176 (2016)
18. Batini, C., Scannapieco, M.: Data and Information Quality: Dimensions, Principles and Techniques. Springer, Cham (2016)
19. Wand, Y., Wang, R.Y.: Anchoring data quality dimensions in ontological foundations. Commun. ACM **39**(11), 86 (1996)
20. Wang, R.Y., Strong, D.M.: Beyond accuracy: What data quality means to data consumers. J. Manag. Inf. Syst. **12**(4), 5 (1996)
21. Redman, T.C.: Data Quality for the Information Age. Artech House Publishers, Norwood (1996)
22. Umbrich, J., Neumaier, S., Polleres, A.: Quality assessment and evolution of open data portals. In: 2015 3rd International Conference on Future Internet of Things and Cloud (FiCloud), pp. 404–411. IEEE Computer Society Press, Los Alamitos (2015)
23. Wang, R.Y., Storey, V.C., Firth, C.P.: A framework for analysis of data quality research. IEEE Trans. Knowl. Data Eng. **7**(4), 623 (1995)
24. Pipino, L.L., Lee, Y.W., Wang, R.Y.: Data quality assessment. Commun. ACM **45**(4), 211 (2002)
25. Batini, C., Palmonari, M., Viscusi, G.: The many faces of information and their impact on information quality. In: Proceedings of the 17th International Conference in Information Quality (ICIQ 2012), pp. 212–228. MIT, Cambridge (2012)
26. Sänger, J., Richthammer, C., Pernul, G.: Reusable components for online reputation systems. J. Trust Manag. **2**(5), 1 (2015)
27. Gascon, H., Grobauer, B., Schreck, T., Rist, L., Arp, D., Rieck, K.: Mining attributed graphs for threat intelligence. In: Proceedings of the 7th ACM on Conference on Data and Application Security and Privacy, pp. 15–22. ACM, New York (2017)
28. Chaturvedi, I., Cambria, E., Welsch, R.E., Herrera, F.: Distinguishing between facts and opinions for sentiment analysis: survey and challenges. Inf. Fus. **44**, 65 (2018)
29. Böhm, F., Menges, F., Pernul, G.: Graph-based visual analytics for cyber threat intelligence. Cybersecurity (Cybersecurity) **1**, 1 (2018)
30. Heinrich, B. Kaiser, M. Klier, M.: How to measure data quality? A metric-based approach. In: ICIS 2007 Proceedings pp. 108–122 (2007)
31. Piazza, R., Wunder, J., Jordan, B.: StixTM version 2.0. part 1: Stix core concepts (2017). https://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part1-stix-core.html
32. Lazar, J., Feng, J.H., Hochheiser, H.: Research Methods in Human–Computer Interaction. Morgan Kaufmann, Burlington (2010)

Bibliography

[1] BISSELL, K., LASALLE, R., AND DAL CIN, P. Innovate for Cyber Resilience. Accenture Third Annual State of Cyber Resilience Report, Accenture Security, 2020.

[2] BÖHM, F. Visual Security Analytics. *Encyclopedia of Cryptography, Security and Privacy* (2021), 1–3.

[3] BÖHM, F., DIETZ, M., PREINDL, T., AND PERNUL, G. Augmented Reality and the Digital Twin: State-of-the-Art and Perspectives for Cybersecurity. *Journal of Cybersecurity and Privacy 1*, 3 (2021), 519–538.

[4] BÖHM, F., RAKOTONDRAVONY, N., PERNUL, G., AND REISER, H. Exploring the role of experts' knowledge in visualizations for cyber security. In *Posters - IEEE Symposium on Visualization for Cyber Security* (Berlin, 2018), IEEE, pp. 1–2.

[5] 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.

[6] CROUSER, R. J., FUKUDA, E., AND SRIDHAR, S. Retrospective on a Decade of Research in Visualization for Cybersecurity. In *2017 IEEE International Symposium on Technologies for Homeland Security (HST)* (Waltham, MA, USA, 2017), IEEE, pp. 1–5.

[7] ENDERT, A., HOSSAIN, M. S., RAMAKRISHNAN, N., NORTH, C., FIAUX, P., AND ANDREWS, C. The human is the loop: new directions for visual analytics. *Journal of Intelligent Information Systems 43*, 3 (2014), 411–435.

[8] EUROPEAN COMMISSION. Commission to invest €14.7 billion from Horizon Europe for a healthier, greener and more digital Europe. Press Release, European Commission, 2021.

[9] FEDERICO, P., WAGNER, M., RIND, A., AMOR-AMOROS, A., MIKSCH, S., AND AIGNER, W. The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Phoenix, AZ, USA, 2017), IEEE, pp. 92–103.

[10] FEKETE, J.-D., VAN WIJK, J. J., STASKO, J. T., AND NORTH, C. The Value of Information Visualization. In *Information Visualization*. Springer, Berlin, 2008, pp. 1–18.

[11] FISCHER, ERIC. Cybersecurity Issues and Challenges: In Brief. CRS Report 7-5700, Congressional Research Service, 2016.

[12] GARTNER. Gartner Survey of Nearly 2,000 CIOs Reveals Top Performing Enterprises are Prioritizing Digital Innovation During the Pandemic, 2020.

[13] GARTNER. Gartner Forecasts Worldwide Security and Risk Management Spending to Exceed $150 Billion in 2021, 2021.

[14] GATES, C., AND ENGLE, S. Reflecting on Visualization for Cyber Security. In *2013 IEEE International Conference on Intelligence and Security Informatics* (Seattle, WA, USA, 2013), IEEE, pp. 275–277.

[15] HEER, J., AND SHNEIDERMAN, B. Interactive Dynamics for Visual Analysis: A taxonomy of tools that support the fluent and flexible use of visualizations. *Queue 10*, 2 (2012), 30–55.

[16] HERJAVEC GROUP. The 2020 Official Annual Cybercrime Report. Business report, Herjavec Group, Toronto, 2020.

[17] HEVNER, MARCH, PARK, AND RAM. Design Science in Information Systems Research. *MIS Quarterly 28*, 1 (2004), 75.

[18] HOLMSTRÖM, J., KETOKIVI, M., AND HAMERI, A.-P. Bridging Practice and Theory: A Design Science Approach. *Decision Sciences 40*, 1 (2009), 65–87.

[19] IBM SECURITY. Cost of a Data Breach Report 2020. Tech. rep., IBM Security, 2020.

[20] IBM SECURITY. Cost of a Data Breach Report 2021. Tech. rep., IBM Security, 2021.

[21] ISACA. State of Cybersecurity 2020 - Part 2: Threat Landscape and Security Practices. Tech. rep., Information Systems Audit and Control Association (ISACA), 2020.

[22] KEIM, D., ANDRIENKO, G., FEKETE, J.-D., GÖRG, C., KOHLHAMMER, J., AND MELANÇON, G. Visual Analytics: Definition, Process, and Challenges. In *Information Visualization*. Springer, Berlin, 2008, pp. 154–175.

[23] KEIM, D., KOHLHAMMER, J., ELLIS, G., AND MANSMANN, F., Eds. *Mastering the Information Age: Solving Problems with Visual Analytics*. Eurographics Association, Goslar, 2010.

[24] LAVIGNE, V., AND GOUIN, D. Visual Analytics for cyber security and intelligence. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology 11*, 2 (2014), 175–199.

[25] MARTY, R. *Applied Security Visualization*. Addison-Wesley, Upper Saddle River, NJ, USA, 2009.

[26] MCCURDY, N., DYKES, J., AND MEYER, M. Action design research and visualization design. In *Proceedings of the beyond time and errors on novel evaluation methods for visualization - BELIV '16* (New York, NY, USA, 2016), ACM Press, pp. 10–18.

[27] MENDLING, J., AND NEUMANN, G. *Wirtschaftsinformatik*. De Gruyter, 2019.

[28] 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 - Sicherheit, Schutz und Zuverlässigkeit* (Konstanz, 2018), Gesellschaft für Informatik e.V., pp. 257–260.

[29] MEYER, M., SEDLMAIR, M., QUINAN, P. S., AND MUNZNER, T. The nested blocks and guidelines model. *Information Visualization 14*, 3 (2015), 234–249.

[30] MUNZNER, T. A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 921–928.

[31] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1. NIST CSWP 04162018, National Institute of Standards and Technology, Gaithersburg, MD, 2018.

[32] OKOLI, C. A Guide to Conducting a Standalone Systematic Literature Review. *Communications of the Association for Information Systems 37* (2015).

[33] PEFFERS, K., TUUNANEN, T., ROTHENBERGER, M. A., AND CHATTERJEE, S. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems 24*, 3 (2007), 45–77.

[34] SACHA, D., STOFFEL, A., STOFFEL, F., KWON, B. C., ELLIS, G., AND KEIM, D. Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1604–1613.

[35] SIMON, S., MITTELSTÄDT, S., KEIM, D. A., AND SEDLMAIR, M. Bridging the gap of domain and visualization experts with a liaison. In *Eurographics conference on visualization (EuroVis) - short papers*. The Eurographics Association, 2015.

[36] THOMAS, J. J., AND COOK, K. A. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Department of Homeland Security, USA, 2005.

[37] VAN WIJK, J. J. Bridging the gaps. *IEEE Computer Graphics and Applications 26*, 6 (2006), 6–9.

[38] WANG, X., JEONG, D. H., DOU, W., LEE, S.-W., RIBARSKY, W., AND CHANG, R. Defining and applying knowledge conversion processes to a visual analytics system. *Computers & Graphics 33*, 5 (2009), 616–623.

[39] WARE, C. *Information Visualization: Perception for Design*, 4th ed. Elsevier, Philadelphia, PA, USA, 2020.

[40] WORLD ECONOMIC FORUM. The Global Risks Report 2020. Global Risks Report 15, World Economic Fourm, 2021.

[41] ÖSTERLE, H., BECKER, J., FRANK, U., HESS, T., KARAGIANNIS, D., KRCMAR, 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.

Appendix

Academic Curriculum Vitae

Fabian Konrad Böhm, M.Sc. With Honors

Chair of Information Systems (IFS)
Faculty of Business, Economics, Management Information Systems
University of Regensburg

Academic Experience

2017 - present	**RESEARCH ASSISTANT** *Chair of Information Systems, University of Regensburg*
2014 - 2017	**GRADUATE STUDENT RESEARCH ASSISTANT** *Chair of Information Systems, University of Regensburg*
2013 - 2014	**STUDENT RESEARCH ASSISTANT** *Chair of Information Systems, University of Regensburg*

Research Project Involvement

2021 - present	**INSIST (STMWI)** - Industrial IoT Security Operations Center *Bavarian Ministry of Economic Affairs, Regional Development, and Energy*
2021 - present	**DEVISE (BMBF)** - Data Quality Management for Improving Information Security *Federal Ministry of Education and Research*
2016 - 2020	**DINGFEST (BMBF)** - Detection, Visualization, and Forensic Analysis of Security Incidents *Federal Ministry of Education and Research*
2014 - 2017	**BILODEX** - Big Log Data Exploitation *Industry-sponsored*

Education

2017 - present	**PH.D. STUDENT** *University of Regensburg*
2014 - 2017	**MASTER OF SCIENCE WITH HONORS** *University of Regensburg & Universitat Politècnica de Catalunya*
2011 - 2014	**BACHELOR OF SCIENCE** *University of Regensburg*
2002 - 2011	**HIGH SCHOOL GRADUATION** *Burkhart-Gymnasium Mallersdorf-Pfaffenberg*

Industry Experience

2016	**INTERN (2 MONTHS)**, Management Consulting *Senacor Technologies AG, Munich*
2013	**WORKING STUDENT (6 MONTHS)**, Corporate Supply Chain *Infineon Technologies AG, Regensburg*
2013	**INTERN (2 MONTHS)**, Corporate Supply Chain *Infineon Technologies AG, Regensburg*

Teaching

2017 - 2021	**CO-LECTURER** - Security of data-intensive Applications *Graduate lecture at University of Regensburg*
2016 - 2021	**TUTOR** - IT-Security I *Undergraduate lecture at University of Regensburg*
2017 - 2021	**TUTOR AND MODULE RESPONSIBLE** - IT-Security *VAWi - Virtual Continuing Education in Information Systems*
2016 - 2017	**STUDENT TUTOR** - Information Systems *Graduate lecture at University of Regensburg*

Service to the Research Community

2022	**TECHNICAL PROGRAM COMMITTEE MEMBER** *2022 International Workshop on Graph-Based Network Security*
2021	**PROGRAM COMMITTEE MEMBER** *2021 IEEE Symposium on Visualization for Cybersecurity*
2021	**TECHNICAL PROGRAM COMMITTEE MEMBER** *2021 International Workshop on Graph-Based Network Security*
2020 - present	**JOURNAL REVIEWER** *ACM Digital Threats: Research and Practice*
2016 - present	**EXTERNAL REVIEWER** *ICISSP 2022, TrustBUS 2021, SECRYPT 2021, ARES 2021, ESORICS 2021, ISSA 2020, ESORICS 2020, ICCCN 2020, CPSS 20, SECRYPT 2020, CAiSE 2020, ISPEC 2019, ESORICS 2019, DEXA 2019, ARES 2019, IFIPTM 2019, CAiSE 2019, WISE 11, ISSA 2018, ESORICS 2018, IFIPTM 2018, CAiSE 2018, ICISSP 2018, SECRYPT 2017, ESORICS 2017, ARES 2017, IFIP SEC 2017, CAiSE 2017, ICISSP2017, ISSA 2016, ESORICS 2016*

Publications

[1] 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* (Konstanz, 2018), Gesellschaft für Informatik e.V., pp. 257–260

[2] BÖHM, F., RAKOTONDRAVONY, N., PERNUL, G., AND REISER, H. Exploring the role of experts' knowledge in visualizations for cyber security. In *Posters - IEEE Symposium on Visualization for Cyber Security* (Berlin, 2018), IEEE

[3] BÖHM, F., MENGES, F., AND PERNUL, G. Graph-based visual analytics for cyber threat intelligence. *Cybersecurity 1*,16 (2018), 1–19

[4] PUCHTA, A., BÖHM, F., AND PERNUL, G. Contributing to Current Challenges in Identity and Access Management with Visual Analytics. In *Data and Applications Security and Privacy XXXIII*, vol. 11559 of *Lecture Notes in Computer Science*. Springer, Cham, 2019, pp. 221–239

[5] BÖHM, F., ENGLBRECHT, L., AND PERNUL, G. Designing a Decision-Support Visualization for Live Digital Forensic Investigations. In *Data and Applications Security and Privacy XXXIV*, vol. 12122 of *Lecture Notes in Computer Science*. Springer, Cham, 2020, pp. 223–240

[6] VIELBERTH, M., BÖHM, F., FICHTINGER, I., AND PERNUL, G. Security Operations Center: A Systematic Study and Open Challenges. *IEEE Access 8* (2020), 227756–227779

[7] SCHLETTE, D., BÖHM, F., CASELLI, M., AND PERNUL, G. Measuring and visualizing cyber threat intelligence quality. *International Journal of Information Security 20*, 1 (2021), 21–38

[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] PUTZ, B., BÖHM, F., AND PERNUL, G. HyperSec: Visual Analytics for Blockchain Security Monitoring. In *ICT Systems Security and Privacy Protection*, vol. 625 of *IFIP Advances in Information and Communication Technology*. Springer, Cham, 2021, pp. 165–180

[10] BÖHM, F. Visual Security Analytics. *Encyclopedia of Cryptography, Security and Privacy* (2021), 1–3

[11] BÖHM, F., DIETZ, M., PREINDL, T., AND PERNUL, G. Augmented Reality and the Digital Twin: State-of-the-Art and Perspectives for Cybersecurity. *Journal of Cybersecurity and Privacy 1*, 3 (2021), 519–538

[12] BÖHM, F., ENGLBRECHT, L., FRIEDL, S., AND PERNUL, G. Visual Decision-Support for Live Digital Forensics. In *2021 IEEE Symposium on Visualization for Cybersecurity (New Orleans, 2021)*, pp. 58–97

[13] BÖHM, F., VIELBERTH, M., AND PERNUL, G. Formalizing and Integrating User Knowledge into Security Analytics. *Springer Nature Computer Science* (2021)