

The Notion of Relevance in Cybersecurity: A Categorization of Security Tools and Deduction of Relevance Notions

Philipp Kuehn*

kuehn@peasec.tu-darmstadt.de
Science and Technology for Peace and
Security (PEASEC), Technical
University of Darmstadt
Darmstadt, Germany

Julian Bäumler

julian.baeumler@stud.tu-
darmstadt.de
Science and Technology for Peace and
Security (PEASEC), Technical
University of Darmstadt
Darmstadt, Germany

Marc-André Kaufhold

kaufhold@peasec.tu-darmstadt.de
Science and Technology for Peace and
Security (PEASEC), Technical
University of Darmstadt
Darmstadt, Germany

Marc Wendelborn

marc.wendelborn@stud.tu-
darmstadt.de
Science and Technology for Peace and
Security (PEASEC), Technical
University of Darmstadt
Darmstadt, Germany

Christian Reuter

reuter@peasec.tu-darmstadt.de
Science and Technology for Peace and
Security (PEASEC), Technical
University of Darmstadt
Darmstadt, Germany

ABSTRACT

Proper cybersecurity requires timely information to defend the IT infrastructure. In a dynamic field like cybersecurity, gathering up-to-date information is usually a manual, time-consuming, and exhaustive task. Automatic and usable approaches are supposed to be a solution to this problem, but for this, they require a notion of information relevance to distinguish relevant from irrelevant information. First, on the basis of a literature review, this paper proposes a novel cybersecurity tool categorization based on corresponding tool types with their respective definitions and core features. Second, it elaborates information used in each category and deduces notions of relevance. Third, it outlines how these findings informed the design of a security dashboard to guide computer emergency response team staff in identifying current threats in open source intelligence sources while mitigating information overload.

KEYWORDS

cybersecurity, relevance assessment, tool categorization, usability, human-computer interaction

1 INTRODUCTION

Staying up-to-date with cybersecurity (CySec) information is crucial to secure one's system. It is one of the tasks security operations centers (SOCs) and computer emergency response teams (CERTs) are doing on a daily basis. Newly found vulnerabilities, which

remain undetected in one's IT infrastructure, can be the source for tremendous damage [48]. While traditional security checks like penetration testing and system hardening are good investments to identify known vulnerabilities, approaches like cyber threat intelligence (CTI) are necessary complements for a proper cyber defense [42]. CTI facilitates information sharing on incoming threats in partner organizations and automates the processes of information gathering and correlation from open source intelligence (OSINT) sources. This requires reliable sources of information on cyber threats, which, however, are usually identified manually [24].

Automated tools for gathering CTI have already been developed, but they either use information sources known beforehand (e.g., security advisories or blogs) [42] or are limited to Twitter [62]. Although these sources might be relevant for the specific use-cases, other scenarios in cybersecurity (CySec) require different sources. A systematic identification of relevant information, in known and unknown sources, is not possible, because, to the best of our knowledge, there is currently no systematically identified notion of relevance. This leaves the domain of CTI, with a focus on OSINT sources, to experts who either use known information sources or have sufficient experience in the field to quantify the relevance of sources themselves.

Furthermore, in order to achieve usability and the successful adoption of tools, they often must be tailorable according to the individual or organizational needs, preferences, or definitions of actionable or relevant information [57, 58, 68, 76]. This paper aims to close these gaps by (i) providing an overview of different CySec categories and corresponding types of tools, (ii) deducing relevance notions for each tool category, and subsequently (iii) presenting their application for an automated and usable CySec dashboard.

Thus, we seek to answer (i) what different categories of CySec tools are used in research and practice (RQ1) and (ii) which security related relevance notions are found in these categories (RQ2).

*Corresponding author

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Veröffentlicht durch die Gesellschaft für Informatik e.V.

in K. Marky, U. Grünefeld & T. Kosch (Hrsg.):

Mensch und Computer 2022 – Workshopband, 04.-07. September 2022, Darmstadt

© 2022 Copyright held by the owner/author(s).

<https://doi.org/10.18420/muc2022-mci-ws01-220>

2 RELATED WORK

2.1 Relevant Information for Cyber Situational Awareness

In order to prevent and respond to attacks effectively, CySec staff needs to establish situational awareness and undertake informed decisions [61]. Endsley [16] coined a distinction of three phases to attain situational awareness: the perception of the elements, the comprehension of their meaning, and the projection of their status. Cyber Situational Awareness (CSA) refers to a state of knowledge of actors that enables them to perceive the relevant elements in the cyber environment within a certain volume of time and space, to comprehend their meaning, and project their status in the near future [25]. Although it was established as a subset of situational awareness, it cannot be considered in isolation because events in cyberspace usually impact the physical world, e.g., financially, socially, or politically. Its growing importance within public administration has been reflected in numerous national CySec strategies [18]. Especially professionals in formalized security organizations such as CERTs, whose work is particularly challenging due to the necessity of coordination and information exchange within and beyond the team, benefit from a continuous improvement of their situational awareness [20].

The term CSA is often related to knowledge about occurrences in one's own network, but CySec staff have to look way further [61]. This includes information about current threats events, new vulnerabilities, possible mitigations, and new technologies. However, due to the steadily increasing number of attacks and security breaches, the volume and variety of potentially relevant data and data sources is steadily increasing. Thus, obtaining and ensuring adequate CSA is increasingly dependent on the properties and capacities of the tools in use [15]. Husák et al. [25] identify both the joint aggregation, analysis, and visualization of data from as diverse sources as possible in real-time, as well as the support of operators in assessing the veracity and credibility of the provided information as novel challenges that have so far been addressed only very fragmentary in technology development. Even recently developed dashboards to support CSA focus on data from within the organization, such internal network security data and vulnerability management [45], while external data sources are rarely included and credibility assessment of social media data is not supported [5].

Though several notions of relevant information for CTI are discussed in literature [13, 42], there is a lack of a systematic investigation of these notions based on a comprehensive overview of tools in the field of CySec rather than on insights into individual aspects.

2.2 Data Sources & Tools for Cyber Threat Intelligence

At present, several data sources are available to obtain relevant information and enhance CSA. Skopik et al. [67] identified vulnerability databases, specialized search engines, and alerting systems as key data sources for cyber incident response. One of the most well-known vulnerability databases is the *National Vulnerability*

Database (NVD) [6], but other well-known alternatives are the commercial tool *Vulners*¹ or the open-source database *Open CTI*². In these databases, the Common Vulnerabilities and Exposures (CVE) have become established as the standard. All publicly disclosed security vulnerabilities in software or hardware components are cataloged there, supplemented by an assessment of their criticality in the form of the Common Vulnerability Scoring System (CVSS) score. This score takes into account various factors, e.g., the complexity for the attacker to exploit the vulnerability. However, vulnerability databases should not be used as the only source for attaining CSA, since they contain only publicly disclosed vulnerabilities [62], entries may be inconsistent [36], and vulnerabilities are not the only threats [42]. Furthermore, many manufacturers publish information about vulnerabilities and fixes on their own channels first [40] by means of security advisories.

Besides vulnerability databases and security advisories, CTI platforms have been established to allow both the collection and analysis of cyber threat data [46]. To support this task, several tools, like the Malpedia database [55], exist. Indicators of compromise (IoCs) are also often used to improve attack prevention, and tools such as MISP [73], ThreatFox [1], and Pulsedive [56] were designed to allow the collection and processing of IoCs. In recent years, social media has become another important data source for CTI. CySec experts can exchange threat information more quickly on those platforms than by using more formal channels [49]. A lot of social media intelligence (SOCMINT) tools have been developed to facilitate the collection and analysis of data from social media, also in other domains [29]. Regarding cyber incident response, Mittal et al. [49] created a Twitter-based warning framework for CySec incidents, and Rodriguez and Okamura [59] developed a CSA system aimed at retrieving and classifying security relevant information from the same platform. Still, the available tools are not designed to support CySec staff through the whole process of multi-platform data acquisition, data analysis, and cyber threat communication [5].

Although we identified several data sources and tools for gathering CTI, we were unable to find a comprehensive categorization of CySec tools, which will be subject of the next section before a deduction of relevance notions follows.

3 CATEGORIZATION OF CYBERSECURITY TOOLS

In this section, we aim to give an overview over the different CySec tools used in research and practice, their respective use-cases, and their superordinate categories.

3.1 Method

First, in order to design a categorization of CySec tools for professionals, web searches were performed to get an overview of available categories, tools, and already conducted tool comparisons. A particular emphasis, in addition to established taxonomies, was put on systematic surveys and market studies of tools of one type in scientific publications, on websites for CySec practitioners, and on websites of CySec software vendors. Based on this, a list of

¹<https://vulners.com/>

²<https://www.opencti.io/en/>

relevant tool types and exemplary systems was derived and discussed among the authors. Our team of authors consisted of two researchers in human-computer interaction and CySec, respectively, one researcher and graduate student in CySec, and one graduate student in political science, all of whom were involved in the interdisciplinary research project CYWARN [32]. We excluded tools for non-professional end-users (e.g., internet security suites) and grouped tools with very similar purposes under a uniform designation. It should be emphasized that the delineated tool types represent ideal types.

Second, after mutual agreement on the scope of considered tools, another search was conducted. Google Scholar was used to collect scientific literature on the characteristics, core features, and definitions of the previously identified tools, resulting in 16 relevant references for analysis. We selected Google Scholar as search engine, since it indexes publications from numerous computer science relevant databases, e.g., IEEE-Xplore, ACM Digital Library, and USENIX. While no systematic literature review was conducted, the premise of the search and selection was to identify scientific and, if possible, peer-reviewed fundamental literature on tool types, which contains a concise tool definition. If multiple publications for one tool type were found, it was checked whether the definitions referred to different core features, and if this is not the case, the most cited publication was selected. The main rationale driving our high-level categorization was to assign various tool types to one of six key operational scenarios in CySec practice. Tool types that could not be appropriately allocated to one of the six thematic categories were grouped into the seventh category, *others*. Besides the descriptive (cf. §3.2) and visual presentation (cf. Fig. 1) of our categorization, we provide an overview of the tool definitions and core features found in scientific literature.

3.2 Cybersecurity Tool Hierarchy

Our categorization of CySec tools comprises the seven categories of (a) *management*, (b) *vulnerability*, (c) *incident*, (d) *treatment*, (e) *intelligence*, (f) *risk*, and (g) *others*.

The category of *management* comprises tools for security event management and security information management, which are often combined as security information and event management. A security event management tool is used for monitoring, processing, and correlation of log and event data to enable incident response [51]. Thus, it primarily assists analysts during the detection and real-time analysis of security-relevant information. A security information management tool, on the other hand, uses log data from local systems and applications to primarily support compliance reporting, internal threat management, and monitoring [51]. Often it allows for automatic reconciliation of log data with pre-set organizational standards.

The second category of tools is centered around hardware and software *vulnerabilities*, including vulnerability databases, vulnerability management tools, and vulnerability scanners. Vulnerability databases are platforms for the collection, dissemination, and maintenance of information about known security weaknesses in both computer system hardware and software [52]. Besides the vulnerabilities themselves, assessments of their severity and in-depth

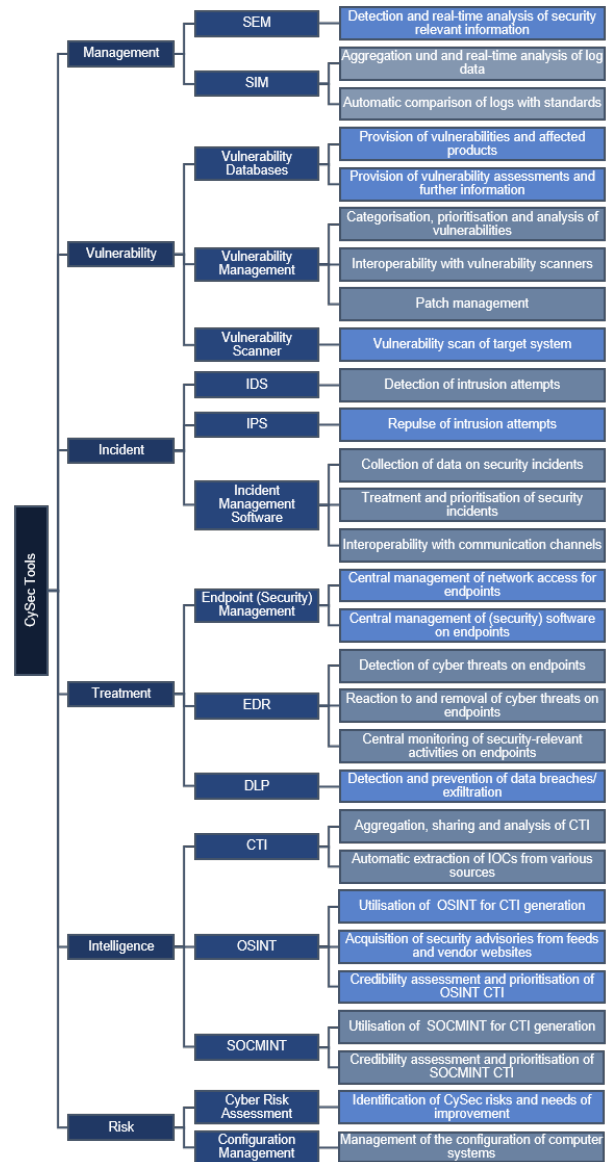


Figure 1: Categorization of cybersecurity tools. The category *others*, and respective tools, is omitted (Source: own research).

background information are often provided. Vulnerability management tools identify, assess, and fix vulnerabilities to minimize risks [53]. They typically incorporate the functionality to categorize, prioritize, and analyze vulnerabilities, are inter-operable with vulnerability scanners, and support during patch management. Finally, vulnerability scanners scan for security vulnerabilities in networks or host systems [7].

Another category of tools is designed around CySec *incidents*, including intrusion detection systems, intrusion prevention systems, as well as incident management software. intrusion detection systems are designed to support the detection and isolation of intrusions in computer systems [64]. intrusion prevention systems

feature an extended range of functions as they, *i.e.*, in addition to the features of an intrusion detection systems have capabilities to prevent intrusions [63]. Incident management tools, in contrast, primarily support the structuring and planning of workflows after the detection of security incidents, *e.g.*, through the collection of incident data and communication within the organization [34].

The fourth category is related to the *treatment* of cyber attacks, which comprises tools of endpoint (security) management, endpoint detection and response, as well as data loss/leakage prevention systems. Central endpoint (security) management tools enable inventory management, network access, and asset control of computer equipment of an organization as well as the monitoring and patching of installed software on endpoints, thus facilitating compliance with software license agreements, regulatory standards, and internal policies [74]. In contrast, endpoint detection and response systems monitor endpoints for suspicious behavior [23]. Beyond monitoring security-related activities, the tools also provide guidance for the reaction to and removal of cyber threats on endpoints [23]. Data loss/leakage prevention systems tools aim to detect and prevent potential data breaches in a timely manner by monitoring data while in use on endpoints, in motion as part of network traffic, or during storage [70].

The following category includes *intelligence* tools which are differentiated either as CTI, OSINT, or SOCMINT tools. CTI tools and providers provide low level IoCs or high level threat actor behavior information from different sources [37, 41]. One common feature is the extraction of IoCs from various data sources. In the context of CySec, the use of OSINT involves monitoring openly accessible websites, forums, and marketplaces in clear- and darkweb to collect CTI [71]. While OSINT can be extracted from a wide variety of different sources, SOCMINT tools use social media [22, p. 205] for the acquisition of CTI. Both, advanced OSINT and SOCMINT tools, may include information prioritization and credibility assessment functionalities.

Another category is related to the assessment of CySec *risks*, including cyber risk assessment and configuration management tools. Cyber risk assessment tools execute risk-assessment frameworks, provide automated solutions for infrastructure operators to determine the present state of their CySec efforts and, if required, identify specific areas for improvement [43]. A different type of tool with relevance for risk assessment has emerged to support configuration management [12]. These tools inform operators of misconfigured systems according to specific guidelines.

Finally, the category *others* comprises CySec tools, *e.g.*, packet sniffers, sandboxes, firewalls, deception tools, or penetration testing tools, which fit in none of these categories. While these tool types serve important functions within CySec operations, they are not described in detail in this section. Thus, they have also been omitted in Fig. 1.

3.3 Implications for Notions of Relevance

To answer RQ1, we searched for existing CySec tools and tool taxonomies, analyzed 16 academic publications to identify definitions and core functionalities of the tools, and finally derived a novel high-level categorization on this basis (*cf.* §3.2).

Since this is an ideal-typical categorization and the actual range of features of tools available on the market varies greatly, some tools can be clearly assigned to one tool type and category, *e.g.*, cyber risk assessment and data loss/leakage prevention systems, while others combine different functionalities and can thus be assigned to several tool types and categories at once. For instance, security information and event management (SIEM) tools that combine the features of security information management (SIM) and security event management (SEM) are now predominant [19], and some CTI systems provide OSINT and SOCMINT in addition to intelligence from commercial or closed sources. Likewise, some tools offer interfaces for integrating data from other tool types; *e.g.*, vulnerability management tools partly feature interoperability with vulnerability scanners as well as CTI platforms [17], and some SIEM systems also provide CTI [10]. Furthermore, the tool types may contribute to the different aspects of CSA; network, threat, and mission awareness [11]. While SIM, configuration management, endpoint (security) management, and vulnerability management tools, as well as vulnerability scanners, may primarily contribute to network awareness, tools of the intelligence and incident categories as well as SEM, data loss/leakage prevention systems, and endpoint detection and response tools along with vulnerability databases may particularly improve threat awareness. By contrast, cyber risk assessment tools may support the establishment of mission awareness.

The diverse operational scenarios of the tools within CySec, the broad spectrum of information types processed and generated, and the contribution of the tools to different aspects of CSA complicate the derivation of a comprehensive notion of relevance for CySec as a whole. Such an universally applicable notion would remain at a very abstract level and would be of limited use in identifying relevant sources and information for specific operational scenarios and tools. Instead, it seems reasonable to derive separate notions of security relevant information for each of the categories identified in this chapter (*cf.* §4).

4 DEDUCTION OF RELEVANCE NOTIONS

Relevance is a subjective information criteria, but necessary to mitigate or even prevent information overload. While information might be relevant for one user, it might be irrelevant for another [31, 60, 66]. Hence, different notions of relevance for the CySec domain based on the identified tool categories are necessary (*cf.* §3). Those categories focus varying information types and notions of relevant information.

To gain an overview over the information used in these categories, we analyze related work that propose methods for (automatic) analysis of information in the corresponding operational scenario. These results can be leveraged in subsequent research to guide automatic analysis methods to improve one's CySec. Hereby, we split the relevance notion into *context* and *precise* notions. An *context relevance notion* indicates the scope of information (*e.g.*, infrastructure), while *precise relevance notions* refer to information of a specific type (*e.g.*, blog posts or source code).

The *management* category, comprised of security event management-, security information management-, and security information and event management-tools, focuses on the local infrastructure. All

Table 1: Summary of context and precise information in each category. *Italic information relate to other categories.*

| Category | Context | Precise |
|---------------|---|---|
| Management | Infrastructure | Log files, used hard- and software-information (CPE) |
| Vulnerability | System weakness | Vulnerability databases, source code, CVSS, CPE, CWE, CAPEC, OVAL |
| Incident | Infrastructure, network, metadata | Log files, network traffic, payloads, SNORT, YARA |
| Treatment | Attacker behavior, metadata | Log files, network traffic, IoCs |
| Intelligence | Blogs, threat information, <i>vulnerabilities</i> | IoCs, CVSS, CWE, CVE, attacker behavior |
| Risk | <i>Management, vulnerability, intelligence</i> | CPE, CVE, CVSS |

information surrounding the infrastructure in scope is relevant, *e.g.*, log files, traffic information, and infrastructure information (soft-, hardware). Accordingly, Debnath et al. [13] propose a tool for anomaly detection in IT infrastructures, using unsupervised learning to detect patterns in application log files. Another system, proposed by Moh et al. [50], detects web attacks by applying a multi-stage log file analysis based on combination of pattern matching and supervised machine learning techniques.

The *vulnerability* category leverages different weakness information sources. First, information in vulnerability databases is primarily used [4, 14, 36]. Precise information that is used are Common Platform Enumeration (CPE), Common Weakness Enumeration (CWE), CVSS, Common Attack Pattern Enumerations and Classifications (CAPEC), and Open Vulnerability and Assessment Language (OVAL) [35]. Second, source code is another primary information source. You et al. [75] leverage source code, vulnerability descriptions, and issue descriptions to nail down vulnerabilities for automatic exploit creation.

The *incident* category uses network-based information. Jyothsna et al. [27] summarize recent studies in the field of anomaly based intrusion detection and explain why signature based anomaly detection is preferred for mainstream implementations. Signature based detection systems use either context information, *i.e.*, metadata retrieved during attacks, or content information, *i.e.*, payloads, to filter malicious content. Common standards combine these information, *e.g.*, SNORT or YARA. Ou et al. [54] implement a host-based intrusion detection systems that combines log file analysis technology with back-propagation techniques of neural networks to improve the efficiency and accuracy of intrusion detection systems. Kumar and Sangwan [38] introduce a signature based intrusion detection systems based on SNORT.

The *treatment* of threats use data at rest (inactive data), data in motion (any network data), or data in use (*e.g.*, RAM or cache data) [33]. One use case is the identification of the attack vector. These systems leverage information which are similar to that presented in the incident category, *e.g.*, network traffic, log files, or file hashes, domain names, and byte sequences (known as IoCs). A discussion of various possible attack scenarios shows that most state-of-the-art endpoint detection and response systems, which analyze system logs, fail to prevent the bulk of the attacks [28]. Alneyadi et al. [3] provide a survey on data loss/leakage prevention systems showing that such systems are based on firewalls, intrusion detection systems, and virtual private networks.

Intelligence tools use different sources containing varying types of information from the OSINT domain. Some systems [26, 42, 77]

crawl grey literature like websites, blogs, or social media to identify current threat information. They focus on IoCs and CVE information. Others use solely social media to identify intelligence information in order to obtain an overview of currently used vulnerabilities, *e.g.*, CVE, CWE, CPE information, and Tactics, Techniques, and Procedures (TTPs) [8, 9, 62]. There are different standards to publish IoCs, namely the STIX-, OpenIOC-, and MISP-standard [26, 42].

Risk assessment in CySec rests upon internal and external information. Lyu et al. [44] assess safety and security risks of cyber-physical systems. The assessment considers asset values, potential threats, and vulnerabilities. It is based on sector-precise standards [65, 72], the used infrastructure, and uses related management-, vulnerability-, and intelligence-information (*cf.* the previous paragraphs). Aksu et al. [2] introduce a CySec risk assessment methodology that is asset- as well as vulnerability-centric. Kure et al. [39] introduce a CySec risk assessment methodology that considers risk from a holistic stakeholder model perspective and integrates the cascading effects from interdependent cyber-physical system components. The information used for this approach are information about vulnerabilities, threats, and risks to an asset.

Tools unlisted in other categories have cyber threats (*cf.* vulnerability and intelligence) as a common baseline. All of them need information about vulnerabilities and threats (*e.g.*, CVSS-scores, signatures, and exploits) as well as information about the systems they protect (*e.g.*, system logs, operating system, and network ports).

Building on the categorization and findings of §3, we examined related work on automatic information analysis methods across different operational scenarios of CySec. Based on this, we elaborated context and precise notions of relevant information (*cf.* Tab. 1) as well as key information standards in the different CySec categories, thus answering RQ2. There are a few base categories, *e.g.*, management or vulnerability, whose information are used in others, *e.g.*, risk. In their current form, these notions remain vague due to the numerous categories in CySec, with the most precise information being the named standards. The information used in CySec is basically every stream of information in or connected to ones infrastructure (*e.g.*, system-, application-, and network-log files) as well as related external information and standards listed in Tab. 1.

5 TOWARDS AN AUTOMATED AND USABLE CYBERSECURITY DASHBOARD

In order to capture relevant information for CySec intelligence and vulnerability management in an mostly automatic and usable manner, we employ a human-centered process to iteratively design

Table 2: Overview of the derived user requirements.

| # | User Requirements |
|----|---|
| 1 | Manual efforts of data collection should be replaced by (semi-)automation. |
| 2 | Enable modularity to add new data sources and features in the later course of development. |
| 3 | Allow gathering of data from different sources and unify collected data for interoperability . |
| 4 | Offer ways to support data protection (e.g., anonymization and data sparsity). |
| 5 | Automatically detect and filter out redundant information across different sources. |
| 6 | Allow the visualization of important data to get an overview and accelerate decision making. |
| 7 | Facilitate information management for different users or organizational roles. |
| 8 | Allow customization of data sources, filters, features, and settings to fit individual needs. |
| 9 | Display only priority (relevant) information to prevent the overload of human capacities. |
| 10 | Evaluate information based on trustworthiness and provide data to the user for verification . |

and refine the Open Data Observatory, a CySec dashboard to visualize relevant CySec information across multiple channels. We shortly introduce the methodology of user evaluation, followed by the presentation of the dashboard.

5.1 Method

The requirements for our first iteration of design were identified through interviews with people working at or with CERTs in two rounds in 2019 and 2021. To acquire the necessary data, requests for semi-structured expert interviews were sent out in two rounds. After receiving the participants' acceptance and informed consent, each interview session lasted around 50 minutes using a web conferencing tool. In the first round of interviews (n=8), a stronger emphasis was put on organizational factors and collaborative practices. The interview guideline comprised nine open-ended questions organized in three parts: (i) an introduction of the interviewee and his/her organizational role, (ii) the deployment, organization, and work processes of the CERT, and (iii) the communication and cooperation between CERTs.

To get further insights into technology use by CERTs, we conducted a second round of interviews (n=7). In the second round, the perspective of some non-CERT organizations that share information with CERTs has also been included. The second interview guideline comprised technology-focused questions on the (i) interviewees' role and organization, (ii) reporting of cyber incidents, (iii) monitoring of cyber incident data (e.g., IoCs), (iv) analysis, prioritization, and verification of gathered evidence, as well as (v) communication of recommendations and warnings. We interviewed seven cyber incident managers, five team leaders, two information security officers, and one public safety answering point for CySec issues.

With the consent of the participants, all interview sessions were recorded and later transcribed. In order to account for the rich, qualitative interview data, we decided to conduct an inductive qualitative content analysis where categories emerge from the data analyzed [47]. Thus, the first and second authors independently employed open coding [69] to gather data into approximate categories reflecting the issues raised by respondents based on repeated readings of the data and its organization into similar statements. The authors then compared their categories and compiled a list of the most important user requirements identified during the coding

process. Based on this, all authors participated in a workshop to identify the requirements guiding the implementation of the Open Data Observatory.

The identified user requirements that serve as the foundation of our design and development process are displayed in Tab. 2. While the next subsection will briefly present details on the design of the CySec dashboard, a first evaluation of the CySec dashboard using scenario-based walkthroughs and corresponding design implications can be found in a different study [30].

5.2 Cybersecurity Dashboard

From an architectural point of view, the Open Data Observatory is a web application based on Vue.js as the overall framework, Bootstrap for responsive design, and Chart.js for data visualization. Besides some local filtering options, all other actions of the Open Data Observatory, such as searching for posts in open and social media or managing users, are forwarded to a backend called Open Data API. This API is implemented following the paradigm of a web-based and service-oriented architecture (SOA). It is a Java Tomcat application using the Jersey Framework for RESTful web services and the MongoDB database for document-oriented data management. Several libraries facilitate the automated and continuous real-time collection of data from open sources, such as NVD vulnerabilities, IoCs, and RSS feeds, or social media source APIs for Flickr, Reddit, Tumblr, Twitter, and YouTube.

The interface (cf. Fig. 2) comprises up to four feeds with security advisories, CVEs, IoCs, and social media data to fully cover the intelligence and vulnerability parts of our identified CySec category (cf. §3). The security advisories are embedded via RSS feeds provided by software and hardware vendors, and the API of the NVD database is used to populate the CVE feed with documented vulnerabilities. Furthermore, we decided to use the ThreatFox platform to obtain IoCs, and individual platform APIs to gather information from social media (e.g., Reddit or Twitter). For each feed, specific charts and a different set of available and characteristic information per entry is displayed (e.g., a textual description and the CVSS score for CVEs). The displayed data set can be selected in the upper left corner and is based on predefined individual parameter settings that are used to query the various sources. On demand, users can display or hide individual or all feed entries and pin important entries that are then highlighted and displayed in the black bottom pin menu for quick

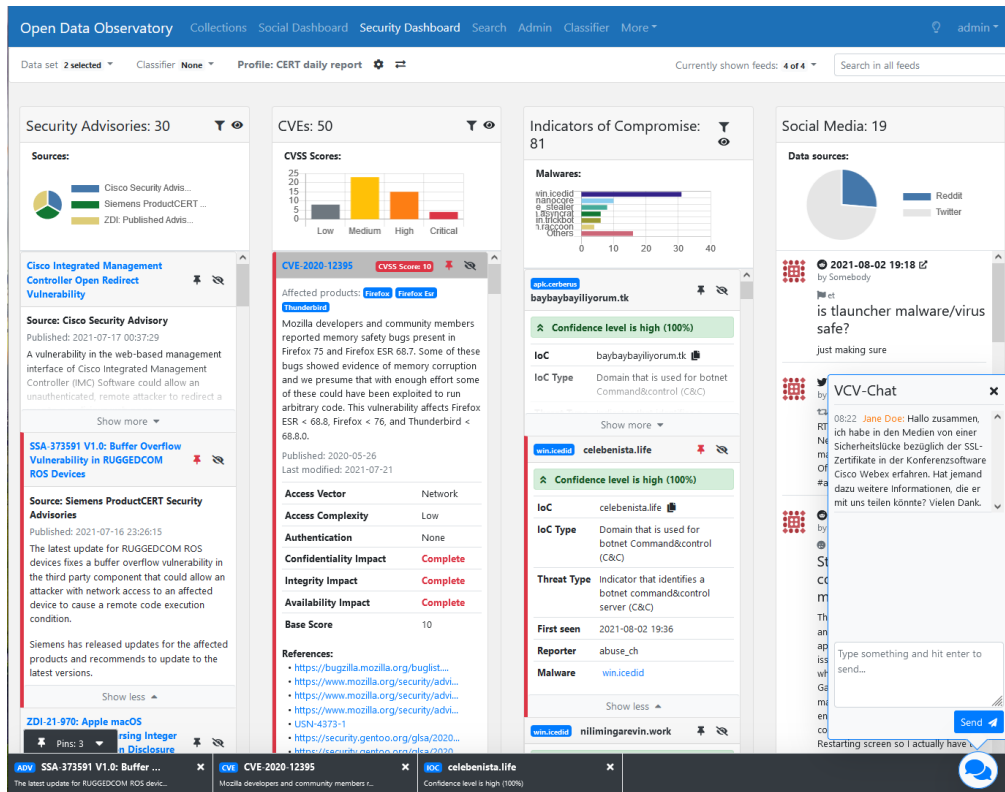


Figure 2: Interface of the Open Data Observatory, a cross-platform cybersecurity dashboard featuring security advisories, Common Vulnerabilities and Exposures, indicators of compromise, and social media feeds.

access. In order to filter the displayed information, users can either click on the interactive charts (e.g., on the critical bar to show only critical vulnerabilities within the CVE feed) or open an advanced filtering menu with additional options (e.g., also filtering by the CVE ID or the affected product of a vulnerability). In the top-right corner, a full-text search field allows searching for keywords across the different feeds simultaneously. To render the collection and processing of data in the application transparent, each displayed information contains a link to the external source (e.g., the NVD for the vulnerabilities).

6 CONCLUSION

CySec is one major cornerstone to secure organizations and corporations, especially due to the increasing digitization. However, proper CySec has many different facets, depending on the specific operational scenarios and the applied tools, and requires to be up-to-date with information of relevance, e.g., on vulnerabilities and attack campaigns. This information is usually gathered from the web or domain specific sources and assists security experts in their task. Yet, there exists no study on what types of information is relevant for different tool categories in the field of CySec.

This paper answers which categories of CySec tools are used in research and practice (RQ1), by presenting six categories and their features in detail (cf. §3), and which security related relevance notions are found in these categories (RQ2), outlining information

contexts and specific information types (cf. §4). These results then inform the implementation of a security dashboard to guide CERT staff in their task of identifying current CySec information in OSINT sources, which mitigates the problem of information overload. We already conducted a first evaluation of the presented tool [30] and its findings will be used to refine the interface for a better (e.g., by the inclusion of machine learning classifiers for relevancy assessment [21]) and more usable detection of relevant CySec information. Our results present a solid baseline to identify new information of relevance, but should be further enhanced as part of a thorough study into each tool category.

ACKNOWLEDGMENTS

This research work has been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE, the German Federal Ministry for Education and Research (BMBF) in the project CYWARN (13N15407), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1119 (CROSS-ING) – 236615297.

REFERENCES

[1] Abuse.ch. 2022. ThreatFox: Share Indicators of Compromise. <https://threatfox.abuse.ch/>

- [43] Georgia Lykou, Argiro Anagnostopoulou, George Stergiopoulos, and Dimitris Gritzalis. 2018. Cybersecurity self-assessment tools: evaluating the importance for securing industrial control systems in critical infrastructures. In *International Conference on Critical Information Infrastructures Security*. Springer, 129–142. https://doi.org/10.1007/978-3-030-05849-4_10
- [44] Xiaorong Lyu, Yulong Ding, and Shuang-Hua Yang. 2019. Safety and security risk assessment in cyber-physical systems. *IET Cyber-Physical Systems: Theory & Applications* 4, 3 (2019), 221–232.
- [45] Lukáš Matta and Martin Husák. 2021. A Dashboard for Cyber Situational Awareness and Decision Support in Network Security Management. In *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. 716–717.
- [46] Vasileios Mavroicidis and Siri Bromander. 2017. Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence. In *2017 European Intelligence and Security Informatics Conference (EISIC)*, 91–98. <https://doi.org/10.1109/EISIC.2017.20>
- [47] Philipp Mayring. 2000. Qualitative Content Analysis. *Forum: Qualitative Social Research* 1, 2 (2000).
- [48] Mike McQuade. 2018. The Untold Story of NotPetya, the Most Devastating Cyberattack in History. *Wired* (2018), 1–6. <https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/>
- [49] Sudip Mittal, Prajit Kumar Das, Varish Mulwad, Anupam Joshi, and Tim Finin. 2016. CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 860–867. <https://doi.org/10.1109/ASONAM.2016.7752338>
- [50] Melody Moh, Santhosh Piniinti, Sindhusha Doddapaneni, and Teng-Sheng Moh. 2016. Detecting web attacks using multi-stage log analysis. In *2016 IEEE 6th international conference on advanced computing (IACC)*. IEEE, 733–738.
- [51] Raydel Montesino, Stefan Fenz, and Walter Baluja. 2012. SIEM-based framework for security controls automation. *Information Management & Computer Security* 20, 4 (2012), 248–263. <https://doi.org/10.1108/09685221211267639>
- [52] Vanamala Mounika, Xiaohong Yuan, and Kanishka Bandaru. 2019. Analyzing CVE Database Using Unsupervised Topic Modelling. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. 72–77. <https://doi.org/10.1109/CSCI49370.2019.00019>
- [53] Matunda Nyanchama. 2005. Enterprise Vulnerability Management and Its Role in Information Security Management. *Information Systems Security* 14, 3 (2005), 29–56.
- [54] Yang-jia Ou, Ying Lin, and Yan Zhang. 2010. The design and implementation of host-based intrusion detection system. In *2010 third international symposium on intelligent information technology and security informatics*. IEEE, 595–598.
- [55] Daniel Plohmann, Martin Clauss, Steffen Enders, and Elmar Padilla. 2017. Malpedia: a collaborative effort to inventorize the malware landscape. In *Proceedings of the Botconf*.
- [56] Pulsedive. 2022. Pulsedive: Threat Intelligence Made Easy. <https://pulsedive.com/>
- [57] Christian Reuter, Thomas Ludwig, Marc-André Kaufhold, and Volkmar Pipek. 2015. XHELP: Design of a Cross-Platform Social-Media Application to Support Volunteer Moderators in Disasters. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM Press, Seoul, Korea, 4093–4102. <https://doi.org/10.1145/2702123.2702171>
- [58] Thea Riebe, Marc-André Kaufhold, and Christian Reuter. 2021. The Impact of Organizational Structure and Technology Use on Collaborative Practices in Computer Emergency Response Teams: An Empirical Study. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 478:1–478:30. <https://doi.org/10.1145/3479865>
- [59] Ariel Rodriguez and Koji Okamura. 2019. Generating Real Time Cyber Situational Awareness Information Through Social Media Data Mining. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 2. 502–507. <https://doi.org/10.1109/COMPSAC.2019.10256>
- [60] Jan Philipp Rohweder, Gerhard Kasten, Dirk Malzahn, Andrea Piro, and Joachim Schmid. 2011. Informationsqualität – Definitionen, Dimensionen und Begriffe. In *Daten- und Informationsqualität: Auf dem Weg zur Information Excellence*. Knut Hildebrand, Marcus Gebauer, Holger Hinrichs, and Michael Mielke (Eds.). Vieweg+Teubner, Wiesbaden, 25–45. https://doi.org/10.1007/978-3-8348-9953-8_2
- [61] Robin Ruefle, Audrey Dorofee, David Mundie, Allen D. Householder, Michael Murray, and Samuel J. Perl. 2014. Computer Security Incident Response Team Development and Evolution. *IEEE Security & Privacy* 12, 5 (2014), 16–26. <https://doi.org/10.1109/MSP.2014.89>
- [62] Carl Sabottke, Octavian Suci, and Tudor Dumitras. 2015. Vulnerability disclosure in the age of social media: Exploiting twitter for predicting real-world exploits. *Proceedings of the 24th USENIX Security Symposium* (2015), 1041–1056.
- [63] Karen Scarfone and Peter Mell. 2010. Intrusion detection and prevention systems. In *Handbook of Information and Communication Security*, Peter Stavroulakis and Mark Stamp (Eds.). Springer, 177–192. https://doi.org/10.1007/978-3-642-04117-4_9
- [64] Karen Scarfone, Peter Mell, et al. 2007. Guide to intrusion detection and prevention systems (idps). *NIST special publication* 800, 2007 (2007), 94.
- [65] Christoph Schmittner, Zhendong Ma, and Peter Puschner. 2016. Limitation and Improvement of STPA-Sec for Safety and Security Co-analysis. In *Computer Safety, Reliability, and Security (Lecture Notes in Computer Science)*, Amund Skavhaug, Jérémie Guiochet, Erwin Schoitsch, and Friedemann Bitsch (Eds.). Springer International Publishing, Cham, 195–209. https://doi.org/10.1007/978-3-319-45480-1_16
- [66] G. Shankaranarayanan, Bala Iyer, and Donna B. Stoddard. 2012. Quality of Social Media Data and Implications of Social Media for Data Quality. *undefined* (2012). <https://www.semanticscholar.org/paper/Quality-of-Social-Media-Data-and-Implications-of-Shankaranarayanan-Iyer/4db13709cb106fb65be16bf85316c5806fc40241>
- [67] Florian Skopik, Timea Páhi, and Maria Leitner. 2018. *Cyber Situational Awareness in Public-Private-Partnerships*. Springer.
- [68] Stefan Stieglitz, Milad Mirbabaie, J. Fromm, and S. Melzer. 2018. The Adoption of Social Media Analytics for Crisis Management - Challenges and Opportunities. In *Proceedings of the 26th European Conference on Information Systems (ECIS)*. Portsmouth, UK.
- [69] Anselm L. Strauss and J Corbin. 1998. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications. <https://us.sagepub.com/en-us/nam/basics-of-qualitative-research/book235578>
- [70] Radwan Tahboub and Yousef Saleh. 2014. Data Leakage/Loss Prevention Systems (DLP). In *2014 World Congress on Computer Applications and Information Systems (WCCAIS)*. 1–6. <https://doi.org/10.1109/WCCAIS.2014.6916624>
- [71] Andrea Tundis, Samuel Ruppert, and Max Mühlhäuser. 2020. On the automated assessment of open-source cyber threat intelligence sources. In *International Conference on Computational Science*. Springer, 453–467. https://doi.org/10.1007/978-3-030-50417-5_34
- [72] Krishna K. Venkatasubramanian, Sidharth Nabar, Sandeep K. S. Gupta, and Radha Poovendran. 2012. Cyber Physical Security Solutions for Pervasive Health Monitoring Systems. <https://doi.org/10.4018/978-1-61350-123-8.ch007> ISBN: 9781613501238 Pages: 143-162 Publisher: IGI Global.
- [73] Cynthia Wagner, Alexandre Dulaunoy, Gérard Wagener, and Andras Iklody. 2016. MISP: The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform. In *Proceedings of the 2016 ACM Workshop on Information Sharing and Collaborative Security (Vienna, Austria) (WISCS '16)*. Association for Computing Machinery, New York, NY, USA, 49–56. <https://doi.org/10.1145/2994539.2994542>
- [74] Chris Williams, Scott Donaldson, and Stanley Siegel. 2020. *Building an Effective Security Program*. De Gruyter. <https://doi.org/10.1515/9781501506529>
- [75] Wei You, Peiyuan Zong, Kai Chen, XiaoFeng Wang, Xiaoqing Liao, Pan Bian, and Bin Liang. 2017. SemFuzz: Semantics-based Automatic Generation of Proof-of-Concept Exploits. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)*. Association for Computing Machinery, New York, NY, USA, 2139–2154. <https://doi.org/10.1145/3133956.3134085>
- [76] Himanshu Zade, Kushal Shah, Vaibhavi Rangarajan, Priyanka Kshirsagar, Muhammad Imran, and Kate Starbird. 2018. From Situational Awareness to Actionability: Towards Improving the Utility of Social Media Data for Crisis Response. In *Proceedings of the ACM on Human-Computer Interaction*.
- [77] Jun Zhao, Qiben Yan, Jianxin Li, Minglai Shao, Zuti He, and Bo Li. 2020. TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data. *Computers and Security* 95, May 2017 (2020), 1–27. <https://doi.org/10.1016/j.cose.2020.101867>