Vielberth *et al. Cybersecurity* (2019) 2:23
https://doi.org/10.1186/s42400-019-0040-0

Cybersecurity

RESEARCH Open Access

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

Manfred Vielberth*, Florian Menges and Günther Pernul

Abstract

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

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

1 Introduction

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

During past years, humans or employees were generally seen as the weakest link in corporate IT security (Lineberry 2007). To mitigate the risk of humans for IT security, a lot of effort is put into awareness campaigns and training of employees (Mello 2017) to ensure that they receive a basic understanding of this topic. This also enables them to distinguish between "normal" events and events harming the organization. However, the ability to recognize malicious events is not harnessed to its full extent. Information about potential incidents might be hidden in the minds of humans and could be the missing link for attack detection or for forensic reconstruction of adverse events. Especially when it comes to nontechnical traces. Therefore, we argue that the connection of digital events with non-digital events observed by people is crucial to IT security.

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

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

2 Motivating example

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

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

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

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

3 Related work

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

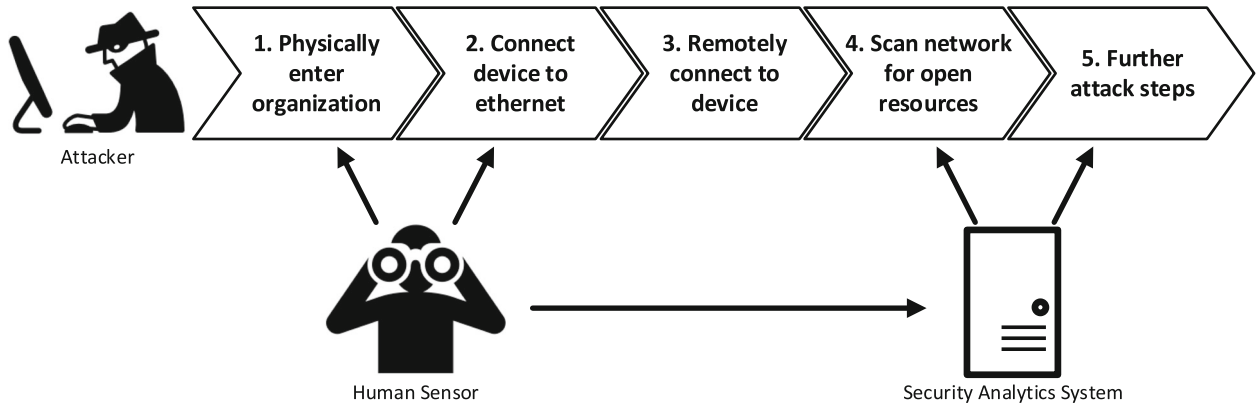

Fig. 1 Attack steps of DarkVishnya

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

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

To sum up the developments in this area, platforms for reporting potential malicious or unwanted emails were implemented at first. This was followed by the development of processes for human-to-human reporting and succeeded by more sophisticated approaches for detecting semantic social engineering attacks with the help of a human-as-a-security-sensor framework. However, to the best of our knowledge, there are no approaches that support reporting a wide range of possible attacks detectable by humans. Additionally, there are no concepts for integrating reported incidents into existing, and in many organizations already established, security systems (e.g. SIEM systems). Moreover, the participation of people with different knowledge in the field of cybersecurity, is currently neglected.

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

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

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

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

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

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

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

4.1 Integrating human sensors into SIEM

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

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

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

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

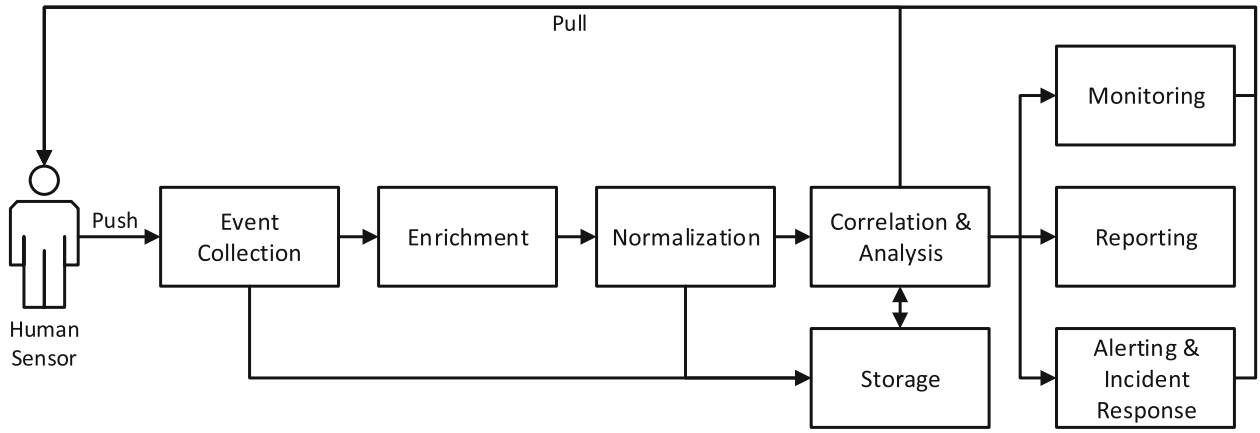

Fig. 2 Integrating human-as-a-security-sensor into the SIEM Data Flow

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

4.2 IHaaSS incident model and taxonomy

For being able to develop a format (Q2) and structure the acquisition of information (Q3), it is necessary to capture everything that human-security-sensors can perceive. Information security management standards and its associated resources provide a good basis by providing risk assessment guidelines. These consider and evaluate mostly future risks. However, in our approach we want to report past incidents, requiring to adjustment for some elements. Regarding the NIST Guide for Conducting Risk Assessments (Joint Task Force Transformation Initiative 2012) and Juliadotter and Choo (2015), the key risk factors are *Threat Sources*, *Threat Events*/*Vector*, *Targets*/*Vulnerabilities* and *Impact*. All four risk factors are observable or can at least be assessed by human sensors and thus, have to be dealt with. The resulting threat model can be seen in Fig. 3.

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

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

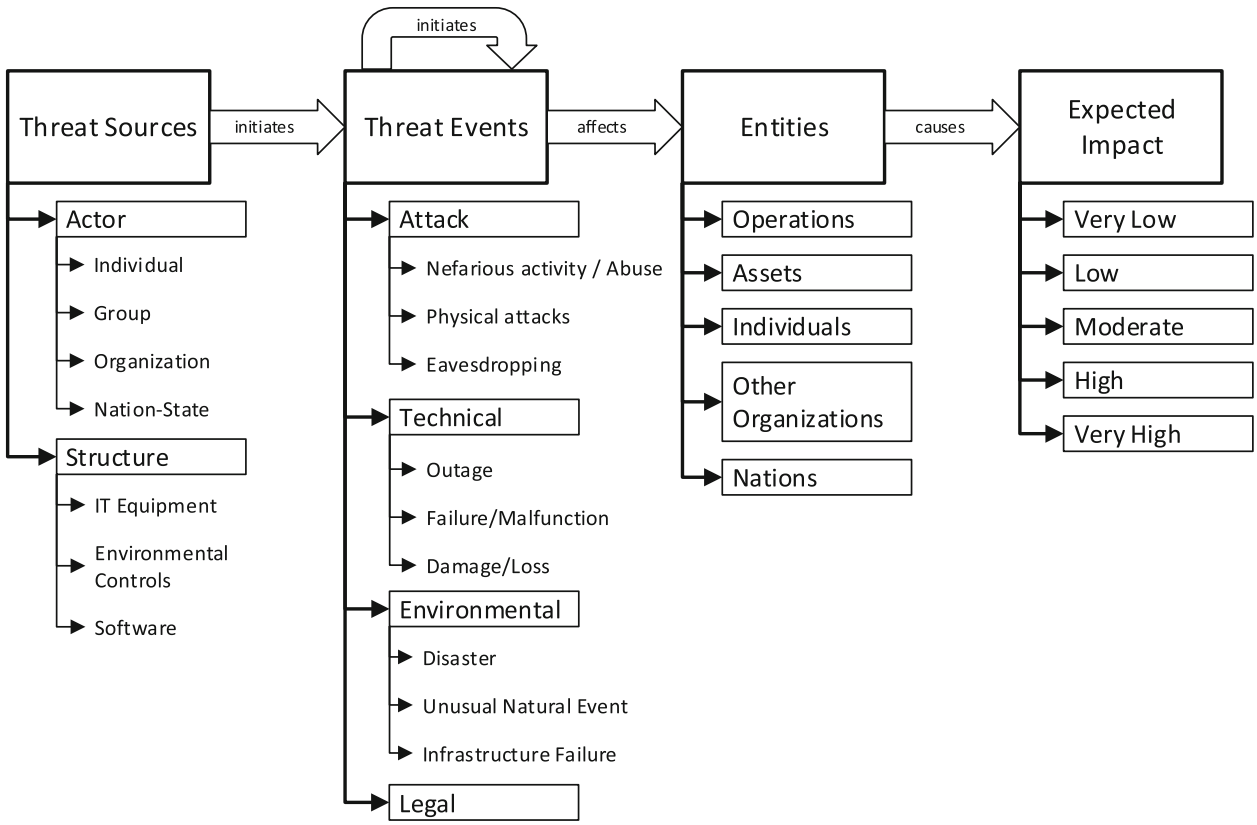

Fig. 3 IHaaSS Incident Model and Taxonomy

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

- **Threat Sources:** Threat sources are the starting point of the incident and can initiate subsequent threat events. This part of the taxonomy is based on the NIST Taxonomy of Threat Sources (Joint Task Force Transformation Initiative 2012). However, to avoid overlaps with subcategories of threat events, we narrow the scope. In the context of our paper, a threat source is an entity, which can decide and initiates events. Thus, the environment defined as a threat source by NIST is equal to a threat event in our taxonomy as we argue that environmental factors cannot take decisions. Additionally, environmental events might be initiated by sources and are therefore better classified as events (e.g. a fire can be set by a person). However, the risk model process remains unaffected, because events can initiate other events. As a result, environmental events can still trigger structural events such as outages.
- **Threat Events:** Threat Events are the processes actually causing harm to an organization and thus are the key component of an IHaaSS report. Our taxonomy of threat events is based on the ENISA Threat Taxonomy (Marinos 2016) with some changes in order to fit in the rest of our model. We have defined more general categories, which allow the distinction between intentional (Attack) and potentially unintentional (Technical, Environmental, Legal) events. This is especially important in order to form dependencies in Section 4.5. Furthermore, the environmental events are merged with environmental threat sources from the NIST Taxonomy of Threat Sources (Joint Task Force Transformation Initiative 2012).
- **Entities:** The identification of relevant assets (asset inventory) is much discussed in academic literature and by industry, due to its importance to risk management (Fenz et al. 2014). Our approach, however, is somewhat broader, which is why we talk about entities (e.g. other organizations may be affected, which are not necessarily an asset for the company). The entity taxonomy is taken from (Joint Task Force Transformation Initiative 2012), wherein it is called adverse impact.
- **Expected Impact:** The expectation of possible impacts is usually quite hard to classify for humans. Therefore, the human sensor commonly provides qualitative estimations, especially when the IT security knowledge is low. Nevertheless, this estimation can be very helpful for evaluating further actions and reactions. Very Low to Very High is a rating of the effect of the event as described by the NIST (Joint Task Force Transformation Initiative 2012). It ranges from "negligible" to "multiple severe or catastrophic effects".

4.3 Conceptualizing a CTI data structure for human sensor data

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

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

Table 1 CTI base model extensions

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

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

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

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

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

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

4.4 A structured representation for threat intelligence reported by humans

In the previous sections, we introduced connection points for integrating human knowledge into SIEM systems and a taxonomy that defines the information basis for the acquisition of threat information detected by humans. Subsequently, we introduced the theoretical foundation for a CTI data structure that is able to represent human sensor data. Based on these findings, we develop a CTI data format in this section that allows to capture information provided by human sensors and enables further technical processing according to research question Q2.

In developing the data format we pursue two main objectives. On the one hand, we aim to achieve a high compatibility to existing SIEM systems to allow a direct integration of additional information into the system. On the other hand, we aim to create a format that allows a complete representation of human sensor data. More specifically, the full scope of the taxonomy shown in Section 4.2 needs to be covered. In order to meet these requirements as completely as possible, we first select existing and well supported CTI data format standards as development basis in the following. Subsequently, we propose a specification of necessary extensions for the integration of human sensor data according to Section 4.3.

Event collection modules within SIEM systems handle heterogeneous raw data from different log sources. This data is then translated into homogeneous indicator data structures to allow further processing. Literature provides different standards for the structured representation of indicators, such as CybOX[1] or openIoC[2]. These data structures are commonly referred to as Indicators of Compromise (IoC) as they depict a set of observations associated with a threat (Appala et al. 2015). These basic incident data can furthermore be enriched using intelligence- and attribution data, such as information about attackers, utilized attack patterns or attackers' objectives as shown by Burger et al. (2014). Together, they allow the representation of complex security incident information as shown in Section 4.3. Literature also offers different standards for representing enriched incidents information, such as STIX, IODEF, VERIS and X-ARF (Barnum 2014; Dandurand et al. 2015; Menges and Pernul 2018). In order to allow the representation of human delivered information, we chose the combination of the existing formats CybOX and STIX as development basis. Both formats are issued together by MITRE[3] and a combined usage is explicitly intended. Since these formats are most commonly applied to represent comprehensive threat intelligence information (Shackleford and SANS Institute 2015; Sauerwein et al. 2017), high compatibility to existing systems can be assumed. Moreover, they offer broader representation capabilities in their basic configuration than comparable formats as shown by Menges and Pernul (2018) and therefore, represent a solid foundation for the integration of human delivered information. Both CybOX and STIX are briefly introduced in the following and examined for necessary extensions to represent human delivered information afterwards.

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

[1] https://cyboxproject.github.io
[2] https://github.com/mandiant/OpenIOC_1.1
[3] https://www.mitre.org

entity for structuring the incident information. TTP (Tactics, Techniques and Procedures) and Course of Action to describe the Intelligence layer. Campaign, Threat Actor and Exploit Target describe the Attribution layer. Indicator, Observable serves as interface to the Indication layer that is essentially provided by CybOX. Moreover, numerous attributes for a detailed expression of these concepts are provided by the data model (Barnum 2014; Menges and Pernul 2018; Fransen et al. 2015).
After this short introduction of the data formats STIX and CybOX, we develop adjustments for these formats to represent human delivered incident information following the IHaaSS taxonomy (see Section 4.2) and CTI data structure (see Section 4.3) in the following. For this purpose, we first discuss the missing elements within the data formats based on the CTI basic data structure. On this basis, we propose the following changes to the formats to allow the integration of human sensors.

- **Intelligence:** Previously, it was shown that attack events can be mapped within the CTI base model, whereas other events are not available yet. Using the taxonomy, we are able to limit these additional events to the categories *Structural*, *Environmental* and *Legal*. In order to also support these events within the data format, we have defined the additional entities "Technical Event", "Environmental Event" and "Legal Event". All these entities are derived from the basic entity TTP, which describes tactics, techniques and procedures used in the course of an attack. An essential property of TTP objects is the structured representation of attack patterns. For this purpose, STIX uses the CAPEC Enumeration, a freely available data set of known attack patterns for the unambiguous description of specific attacks. In order to achieve a comparable functionality for the additionally defined events, we defined a corresponding vocabularies for structural, legal and environmental events. Each vocabulary offers predefined event definitions according to our taxonomy. In addition to the event definitions, the area of intelligence also offer possibilities for describing the expected impact of an incident. For this purpose it was previously shown that the base model only provides impact definitions that emerge from active attacks. Although this is basically also true for the data format, its data definitions do not explicitly restrict the representation of incident results to an underlying attack. As a result, no changes are necessary to enable the definition of specific event results.
- **Attribution:** It was shown that an integration of structural sources is necessary for addressing passive threats within the CTI base format. In addition, it was shown that the entities are limited to the expression of active attacks. The data format already provides elements such as Threat Actor, Exploit Target to represent active attacks and attackers, and Asset Vocabulary to define assets. To enable the integration of passive threats, we extend STIX with the definition of an additional entity "structural source" as intended in the CTI base format. Since this is an alternative threat source, the object is derived from the existing Threat Actor object and exists on the same level. This object is extended by an additional vocabulary "StructuralSourceTypeVocab" to be able to represent structural threat sources in a structured way. Since this extension of threat sources also extends the scope of attribution, we additionally defined an extension of the asset vocabulary. This makes it possible to define additional assets that can occur in connection with passive threats.
- **Indicator:** The indicator category is used to represent incident event information on a high level of detail, which are basically able cover the event information that may be delivered by humans. However, humans are usually not capable of delivering information on this level of detail and will rather provide unstructured data fragments. Consequently, such data fragments must be evaluated afterwards and the format must allow the unstructured data to be recorded at the time of acquisition. For this reason, we have also added an extension to the Observable object that allows to include unstructured data, which can later be translated into structured CybOX information.

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

Table 2 STIX extensions

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

work[4]. The repository includes XML-schema definitions for the STIX schema extension types and vocabularies that are developed with this work.

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

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

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

$$prohibits(Technical) = Legal \tag{2}$$

$$prohibits(Environmental) = Legal \tag{3}$$

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

Equation 1 defines that an attack cannot be initiated by an environmental or a legal event. The reason for this is that an attack requires action by a human being or at least some technical device and thus cannot be initiated by nonhuman events or sources. Furthermore, the cause of a technical security event cannot be a legal event (Eq. 2), *technical events* can only follow *physical events* or *sources*. The same applies to environmental events (Eq. 3). *Unusual natural events* (e.g. sunspots) cannot be caused by any other events or sources except *Environmental* ones as stated in Eq. 4, because they have a natural cause.

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

5 Evaluation

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

[4] http://tinyurl.com/y3h5k25t

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

5.1 Prototypical implementation

Our application prototype realizes the rendering of information delivered by human security sensors into the structured threat intelligence information. A working example of the IHaaSS prototype is available online[5]. The prototype pursues two different goals. On the one hand, it demonstrates the use of IHaaSS in a possible scenario for the structured acquisition of incident information to show the overall validity of our approach. On the other hand, it illustrates the value of information delivered by human security sensors and the combination possibilities with data from existing analytics processes. The application consists of two major components: First, a wizard component that allows the reception of incident information delivered by humans. Second, a server component that translates the acquired incident information into the structured format to be further processed afterwards. The frontend is implemented by using Angular[6] and Typescript[7]. Java EE in combination with a Glassfish[8] application server was used to implement the STIX conversion logic and the database access.

Figure 4 shows a screenshot of the first step in the wizard component. The wizard is divided into two components. In the first component, the information can be entered by the user. The second part (Captured elements) gives an overview of already declared incident elements so that the user can see what has been previously entered. The wizard is structured in four steps as specified by the taxonomy. At first, the threat sources can be reported. Thereby, an arbitrary number of sources can be added. For selecting a source, the user is presented a drop-down list containing the elements of the first layer of the taxonomy (Actor and Structure). When an element is selected, a second drop-down list with the elements of the next layer is displayed. This continues until there are no sub-elements left. The same selection mechanism is implemented for event types and entities in subsequent steps. Only for events a "triggered by" input field is added to specify the previously reported threat source or threat event that initiated the event. There the selectable events get filtered according to the constraints defined in chapter 4.5. In the fourth and final step, the estimated impact of the whole incident can be entered. Furthermore,the following additional information is requested:

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

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

5.2 Case study

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

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

Moreover, two threat events can be identified:

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

[5] http://tinyurl.com/yyqqlgg7
[6] https://angular.io/
[7] https://www.typescriptlang.org/
[8] https://javaee.github.io/glassfish/

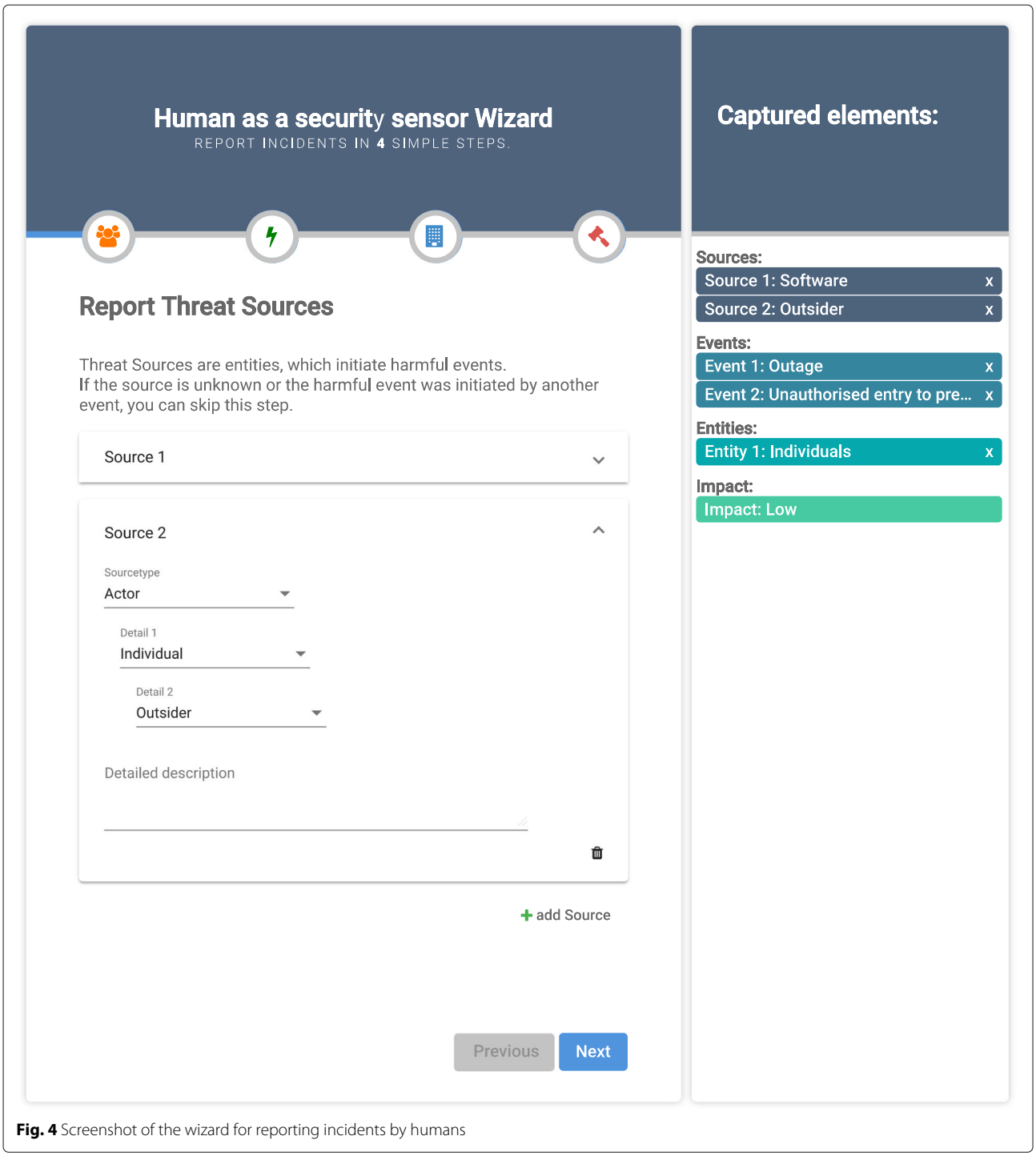

Fig. 4 Screenshot of the wizard for reporting incidents by humans

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

In addition, a network device was identified as a negatively affected entity. Thus, assets are selected from the wizard. Finally, the impact is estimated as low, since the employee may not be able to judge the whole extent of the incident. After the data is collected from the human sensor, it is translated into the corresponding STIX data objects by the server component as described in the following. The outsider (1) who falsely claimed to have legitimate access to the premises is translated

into a Threat Actor object. Its specific properties are mapped to the internal vocab "ThreatActorTypeVocab" that was extended within this work. The technique of gaining unauthorized access to the premises is translated into a TTP object and matching attack patterns from the CAPEC enumeration are mapped. The suspicious hardware device (2) attached to the internal network is then mapped to a structural source object and its specifics are mapped using the "StructuralSourceTypeVocab" created with this work. The action of planting a malicious device is described using a further TTP object and the corresponding CAPEC attack patterns analogous to the first TTP object. After creating these specific entities, the general descriptions of the incident as well as the time of the occurrence, affected assets, and the expected impact are recorded using an Incident object. All these objects are then finally wrapped using a Report object. The complete STIX report for this exemplary use-case is appended to this work as supplementary material. Moreover, it can be viewed under the past incidents overview section within the wizard prototype[9].

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

5.3 Discussion

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

[9] http://tinyurl.com/y5tsoxo3

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

6 Conclusion and future work

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

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

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

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

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

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

Funding
Not applicable.

Availability of data and materials
Source code - iHaaSS wizard
Project name: Client
Project home page: http://tinyurl.com/y6kjrb4q
Archived version: 1.0
Operating system(s): Platform independent
Programming language: HTML, Typescript/JavaScript
Other requirements: Apache Webserver or similar, NPM 6.2.0 or higher
License: GNU GPL v3
Source code - sTIX server
Project name: STIX Server
Project home page: http://tinyurl.com/y46hsvj8
Archived version: 1.0
Operating system(s): Platform independent
Programming language: Java EE 7
Other requirements: Glassfish version 4.1.1 or higher
License: GNU GPL v3
Additional sTIX schema files
Project name: STIX-Schema
Project home page: http://tinyurl.com/y2s3ba7k
Archived version: 1.0
Operating system(s): Platform independent
Programming language: xml-schema
License: BSD-3-Clause
Appended as supplementary material

Received: 24 April 2019 Accepted: 29 August 2019
Published online: 22 October 2019

References

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

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

Publisher's Note